

基于扩散模型的超分辨率方法研究

卢栩瑶¹, 李四², 王五¹ ¹ 计算机学院, 西北大学, 西安 710127

² 通信工程学院, 北京大学, 北京 100871

{luxuyao, wangwu}@tsinghua.edu.cn, lisi@pku.edu.cn

日期: 2026 年 4 月 4 日 15:00

存在问题:

- **文章的整体结构和写作风格是参考的那篇论文? 在这里回答: 给出论文的名称, 并上传 overleaf 的《参考文献》目录**
回答信息:
1: 目前参考论文《Pixel-level and Semantic-level Adjustable Super-resolution:A Dual-LoRA Approach》, 也是我的 baseline。
2: 已经上传部分参考文献
- **论文摘要重新修改润色, 感觉不清晰! 着重突出: 1) 语言知识; 2) 低分辨率图像融合**
回答信息 (修改了哪些内容?):
1: 摘要修改, 开头修改, 先点明现有超分方法的缺陷, 再说自己的创新点; 将创新点改成 4 部分, 看起来更加清晰; 突出了 VAE 编码器中提取浅层、中层、深层多尺度特征和语义引导; 明确了指标和数据集, 实验结果。
2:
- **论文整体感觉逻辑混乱, 没有重点! 我的建议: 着重说明现有方法有什么问题? 我们采用了什么, 解决了什么问题? 当然也也可以按照参考论文的风格写**
回答信息 (修改了哪些内容?):
1: 现有方法的问题, 过度依赖单一尺度特征、细节表征不足、结构易失真; 我们提出一种基于 VAE 多尺度特征融合与跨模态对齐的高保真图像超分方法; 解决了尺度单一、细节不足、结构统一的问题。
2: 也将这部分同步修改在介绍和相关工作中。
- **论文整体文字还要好好修改! 我的建议: 所有问题修改润色, 整理思路**
回答信息 (修改了哪些内容?):
1: 论文重新阅读, 修改论文表述语言。
2: 这里打算重点参考 baseline, 然后修改技术部分语言。
- **论文图表都要修改! 1、图片的文件名称, 统一为 < 英文.PDF> 或者 < 英文.eps>; 2、论文图质量都有待提升, 感觉质量不好**
回答信息 (修改了哪些内容?):
1: 图片名称已经修改
2: 论文图片质量问题, 还在重新绘图

摘要—针对现有图像超分方法过度依赖单一尺度特征、细节表征不足、结构易失真、与真实场景退化不匹配等问题,本文提出一种基于 VAE 多尺度特征融合与跨模态对齐的高保真图像超分方法。首先,从预训练 VAE 编码器中提取浅层、中层、深层多尺度特征,弥补单一尺度特征信息缺失问题;其次,构建特征对齐模块,将退化图像特征映射到扩散模型隐空间,缓解域差异导致的结构漂移、细节模糊、伪影过多问题;同时引入跨模态语义引导,利用文本提示增强重建内容的语义一致性与真实感;最后,通过 ControlNet 结构约束,严格保持输入图像的结构与边缘信息,实现高质量超分。在 DIV2K、RealSR、RealSet 等标准数据集上的实验结果表明,本文方法在 PSNR、SSIM、LPIPS、FID 等指标上均优于现有主流方法,能够有效恢复清晰边缘、细腻纹理、真实细节,显著提升超分结果的保真度与视觉质量,为真实场景低质图像超分提供了一种有效解决方案。

Index Terms—深度学习; 超分辨率; 卷积神经网络; 注意力机制;

I. 引言

图像超分辨率任务作为计算机视觉领域的重要研究方向,旨在将低分辨率的图像转换为高分辨率版本,增强图像细节与清晰度,在众多领域中发挥着关键作用。传统方法如插值法、稀疏表示法依赖手工设计特征,难以捕捉复杂场景下的细节信息与语义关联,恢复效果易受噪声干扰 [?]; 深度学习技术的突破性进展为图像修复算法的革新提供了强大驱动力。深度学习方法中,卷积神经网络 (CNN) 凭借局部特征提取能力推动了超分性能提升,但存在长距离依赖挖掘不足、特征信息孤立等问题 [?]; Transformer 类方法虽能通过自注意力机制建模全局关联,却常忽视多尺度特征的互补性与跨模态信息的协同作用,而 Control-Net 驱动的生成式超分方法则缺乏对 LQ 图像多维度信息的充分利用,导致细节恢复准确性与语义一致性不足 [?]. 近年来,基于扩散模型的超分方法凭借强大的生成能力,取得显著进展。作为最先进的生成模型,将扩散模型应用于图像超分辨率任务,不仅需要恢复出看起来逼真的图像,而且还要求恢复后的图像与输入图像保持保真度和结构一致性。

现有方法仍存在以下关键问题: 特征表征单一: 多数方法仅使用 VAE 最后一层单一尺度特征,浅层细节、中层结构、深层语义信息利用不充分,导致细节丢失、纹理模糊; 域差异明显: 退化图像特征与扩散模型隐空间存在域不匹配,导致生成结果结构漂移、边缘失真、伪影严重; 语义引导缺失: 缺乏有效文本语义引导,重建内容语义不一致、真实感不足; 结构保持不足: 未对

输入图像结构、边缘、轮廓进行严格约束,导致超分结果结构变形、细节不自然。

为解决现有方法存在的信息融合不充分、特征关联挖掘不足等问题 [1], 本文围绕“多尺度特征充分利用、域差异有效缓解、语义精准引导、结构严格保持”这一核心目标,提出一种基于 VAE 多尺度特征融合与多源信息融合的高保真图像超分方法,将图像生成机制与超分辨率任务深度结合,实现低清图像细节的精准恢复与语义增强。实验结果表明,本文方法在峰值信噪比 (PSNR)、结构相似性指数 (SSIM) 等核心客观评价指标上均显著优于对比方法。消融实验进一步验证了所引入模块对模型性能的正向增益作用。本文的实验研究充分佐证了扩散模型在超分辨率任务中的技术优势,为复杂场景下的图像质量增强提供了可靠的技术范式与实验依据,对推动超分辨率技术在实际工程领域的落地应用具有重要理论与实践意义。

本文针对单图超分辨率算法的核心短板,提出多源信息融合创新方案,主要贡献如下: (1) 构建四模块协同端到端框架,首次融合 VAE 分层特征交互、UNet 约束与文本语义跨模态融合,解决信息碎片化问题。(2) 设计 VAE 多尺度特征交互模块,通过自适应对齐与 Transformer 关联挖掘,突破单一尺度信息孤立局限。(3) 创新跨特征交互控制模块与语义交叉注意力机制,提升超分结果结构保真度与语义一致性。(4) 构建特征对齐模块,有效缓解退化图像与扩散模型之间的域差异,减少结构失真。(5) 实现跨模态语义引导 + 结构约束的联合超分框架,显著提升超分结果的保真度与视觉质量。经实验验证,算法在 PSNR、SSIM 指标上优于主流模型,消融实验进一步验证了所引入模块对模型性能的正向增益作用,为复杂场景单图超分提供高效解决方案。

II. 相关工作

图像超分辨率作为计算机视觉领域的核心研究课题,单图像超分辨率旨在从遭受各种退化(如噪声、模糊、下采样等)的低分辨率图像(LQ)中精准恢复高清晰度、高保真度的高分辨率图像(HQ)。由于深度学习方法的发展,这一问题引起了广泛的关注,并取得了重大进展。然而,由于细节重建的固有模糊性,恢复高真实感和保真度的忠实 SR 图像仍是一个具有挑战性的问题。

大多数最先进的图像 SR 方法依赖于深度生成模型,其目的是学习 HQ 图像的分布,并从 LQ 输入中

恢复生动的纹理。生成对抗网络 (Generative Adversarial Networks, GANs) 作为其中的代表方法之一, 可以恢复具有丰富的锐利图像细节。然而, 基于 gan 的方法的训练过程是不稳定的, 生成的图像通常会出现感知上不愉快的伪影。最近, 扩散模型在各种图像生成、编辑、增强和翻译任务中表现出令人印象深刻的性能。特别是, 预训练的文本到图像 StableDiffusionModel 为具有挑战性的逼真图像超分辨率 (Real-ISR) 提供了潜在的解决方案。虽然 ldm 对 HQ 图像的结构和细节具有较强的先验性, 但它们在图像 SR 任务中面临挑战, 这不仅需要生动的生成细节, 还需要与 LQ 输入保持一致的结构。

为了解决这个问题, 有文章选择增强编码器以从 LQ 图像中提取退化鲁棒特征, 然后利用预训练的扩散模型来改进提取的特征。然而, 由于扩散模型是在 HQ 图像上进行预训练的, 因此提取的特征中的任何错误都可能被扩散模型误认为是图像结构, 从而误导扩散过程, 并对最终的恢复结果产生负面影响。因此, 仅仅依靠改进提取的特征来指导扩散过程, 在实现忠实恢复方面存在局限性。与上述方法不同的是, 我们微调扩散模型, 从退化的输入中识别有用的信息, 来提高图像的保真度。需要注意的是, 在基于 ldm 的 SR 方法中, 编码器和扩散模型扮演着不同但不可或缺的角色。单独优化这两个模块将限制 ldm 的能力, 阻碍精细尺度结构的恢复。因此, 本文联合优化编码器和扩散模型, 以帮助编码器与渐进扩散过程很好地对齐。通过这种方式, 编码器能够提供与扩散过程相吻合的有用潜在特征, 扩散模型可以进一步恢复这些特征, 以便更好地重建。

III. 网络设计

本研究围绕超分辨率任务的性能提升与细节保真展开, 创新地将图像生成技术引入超分辨率研究中, 通过多源特征提取、交互融合及注意力引导的条件注入, 构建了兼具细节恢复与语义一致性的超分辨率模型, 具体来说: 首先, 针对低质量图像 (LQ) 的特征挖掘不足问题, 采用 VAE 编码器提取 LQ 图像的前三层特征图, 通过特征拼接与自注意力机制挖掘各层特征间的内在关联, 来更加有效的向潜在的 LQ 特征中传递有用的信息, 有效避免了单一特征层信息孤立的问题, 提升了 LQ 特征的代表能力, 并确保它们与扩散过程很好地对齐。同时, 使用预训练的文本编码器从图像描述中提取的文本嵌入作为辅助信息, 挖掘图像底层特征与语义信

息的关联。最后, 采用注意力机制将 ControlNet 控制条件、LQ 融合特征及文本语义信息这三部分条件进行高效注入, 实现多源信息的深度融合与互补增强, 确保超分辨率重建结果既满足结构保真要求, 又具备良好的语义一致性, 有效提升了超分辨率任务的重建质量。

A. 多层融合 *multiple layer fusion*

给定低质 (LQ) 图像输入后, 本文首先采用预训练 VAE 编码器对其进行系统性的特征提取操作, 核心目标是获取能够全面表征 LQ 图像退化特性与结构细节的有效特征。针对现有方法普遍存在的特征提取不充分问题, 本文摒弃了传统单一依赖 VAE 编码器最后一层输出特征的思路, 重点挖掘编码器前三层中间特征的价值——现有方法所采用的编码器最后一层特征, 经过深度维度压缩后仅保留 8 个通道, 这种极端的通道压缩会导致 LQ 图像中包含的各类退化因素 (如模糊、噪声、压缩伪影、纹理失真等) 以及基础结构细节被大量丢弃, 无法为后续扩散生成过程提供充足的约束与支撑, 难以保证高清图像生成的准确性与保真度。为了获得足够的特征信息, 本文专门选取 VAE 编码器前三层输出的中间特征, 这些特征未经过过度的维度压缩, 保留了更丰富的原始信息, 随后通过自适应特征融合策略对这三层特征进行深度整合, 最终得到融合特征 f_1 。该融合特征 f_1 的通道数提升至 512 维, 相较于传统方法的 8 通道特征, 其包含了更全面的退化模式信息、图像纹理细节与结构先验知识, 浅层特征: 对应细节、纹理、边缘信息; 中层特征: 对应结构、形状、轮廓信息; 深层特征: 对应语义、全局、内容信息。通过多层特征互补, 解决单一尺度特征信息不足问题。将浅层、中层、深层特征通过自注意力模块进行融合: 对三层特征进行维度对齐、尺寸对齐; 通过自注意力计算特征之间相关性与权重; 输出融合后的多尺度增强特征。它有 512 个通道, 包含更多有用的信息, 能够为后续扩散模型的生成过程提供强有力的信息支撑, 有效规避了因特征信息不足导致的生成结果失真、细节缺失等问题, 为高质量的图像超分生成奠定坚实基础。为了将提取到的 LQ 多尺度融合特征 f_1 中所蕴含的丰富信息高效传递至扩散模型, 实现对扩散生成过程的精准控制, 确保生成的高清图像能够准确匹配 LQ 图像的原始结构与语义信息, 本文设计了一套针对性的特征交互与引导机制。具体而言, 首先对 LQ 融合特征 f_1 与扩散迭代过程中生成的噪声高清图 x_t^{HQ} 分别执行独立的卷积运算, 该卷

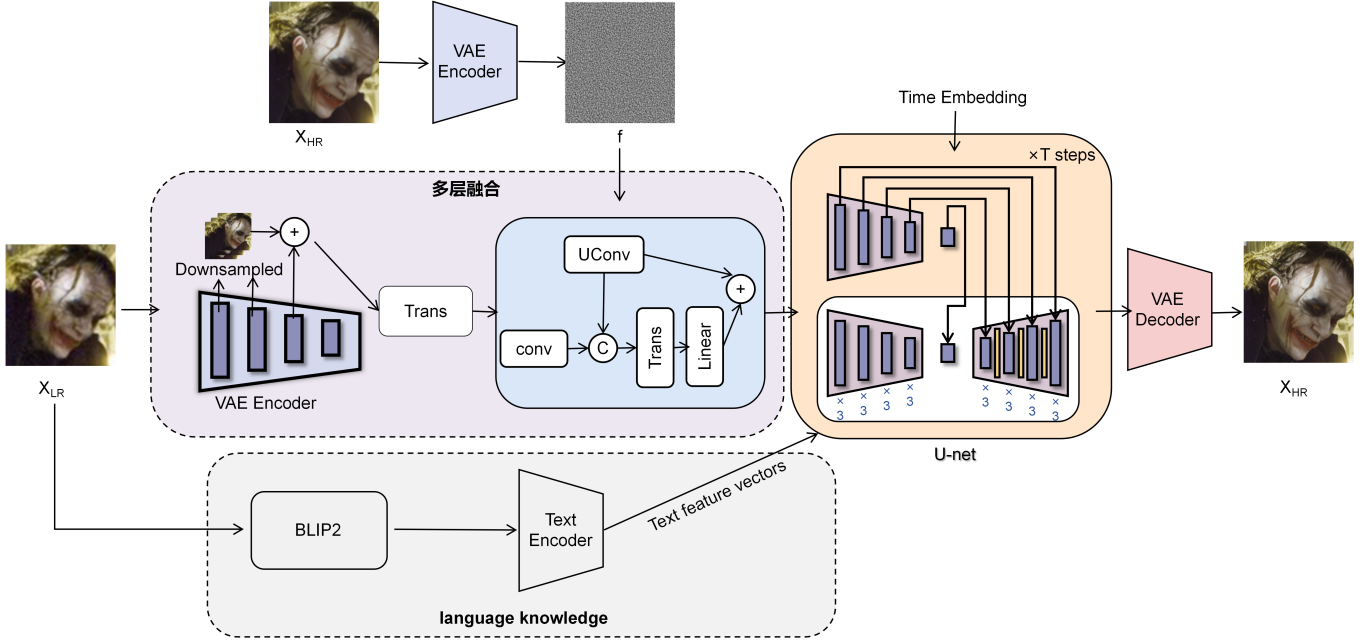


图 1: 本文提出的超分辨率网络结构

积操作采用轻量化卷积核设计，在不引入过多计算开销的前提下，实现两类特征在维度、空间分布上的精准对齐，消除特征之间的维度差异与空间错位，为后续特征交互奠定基础。经过卷积对齐后，将处理后的 LQ 融合特征 f_1 与噪声高清特征图 x_t^{HQ} 在通道维度进行拼接，得到初步融合特征 fc ，该特征初步整合了 LQ 图像的退化先验与扩散过程的噪声特征信息。随后，将初步融合特征 fc 输入至由两个 Transformer 块组成的堆栈结构中，利用 Transformer 块强大的自注意力与交叉注意力机制，充分挖掘 LQ 融合特征 f_1 与噪声高清图 x_t^{HQ} 之间的内在关联，促进两类特征的深度交互与信息互补，使 LQ 特征中蕴含的退化约束与结构信息能够高效渗透到扩散生成的每一个环节，最终输出经过深度交互与优化的特征 f_2 ，为扩散模型后续的去噪迭代与高清图生成提供精准、有效的引导，确保生成结果既能够还原清晰细节，又能够保持与原始 LQ 图像的结构一致性。

B. 文本语义嵌入 *language knowledge*

本文的方法是基于预训练的 SD 模型，其中文本用作输入，在 Real-ISR 等任务中，LQ 图像可用作输入在扩散过程中，因此本文还利用文本嵌入作为额外的辅

助信息，以帮助从扩散模型的潜在特征中提取有用的信息。这些文本嵌入是通过预训练的文本编码器从图像描述中提取的。利用交叉注意层来建立文本嵌入和潜在特征之间的交互，捕获 LDM 中固有的文本到图像的知识。总的来说，给定扩散步骤 t 的潜在 HQ 特征 x_t^{HQ} 和对齐特征 f_2 以及文本嵌入 c ，我们可以得到扩散步骤 $t-1$ 的潜在 HQ 特征为

$$x_{t-1}^{HQ} = \frac{1}{\sqrt{t}}(x_t^{HQ} - \frac{1-t}{\sqrt{1-t}} \hat{(f_2, c, t)}) + t z, \quad (1)$$

，使用预训练好的网络从 LQ 输入中提取图像语义信息，并使用 CLIP 编码器将文本信息转换为图像级特征，提供额外的语义信号来控制扩散过程。

使用文本提示 (Prompt) 作为语义引导：例如：“清晰、自然、高质量、细节丰富”；通过跨模态注意力将文本语义与图像特征结合；增强超分结果的语义一致性、真实感、自然度。

本文所提出的整体方法均基于预训练 Stable Diffusion (SD) 模型构建，在 Real-ISR 等图像超分任务中，低质 (LQ) 图像已被证实可作为有效输入参与扩散生成全过程，为进一步提升扩散模型的生成精度与语义一致性，本文额外引入文本信息作为辅助引导，借助文本语义的约束作用，帮助模型从扩散过程的潜在特

征中精准提取有用信息，进一步优化生成效果。具体而言，本文采用 BLIP2 模型对输入的 LQ 图像进行图像描述生成，BLIP2 作为性能优异的多模态生成模型，能够精准捕捉 LQ 图像的核心语义与结构特征，生成与图像内容高度匹配的文本描述，为后续文本编码提供高质量的原始文本素材。生成文本描述后，将其输入至预训练文本编码器中进行编码处理，转化为标准化的文本嵌入向量，同时结合 CLIP 编码器对文本信息进行进一步加工，将文本语义转换为与图像特征维度对齐的图像级特征，确保文本信息能够与扩散模型的潜在特征高效融合。随后，将得到的文本嵌入向量传递至 ControlNet，借助 ControlNet 的约束能力，实现文本语义信息与扩散生成过程的深度绑定，同时通过交叉注意层搭建文本嵌入与扩散模型潜在特征之间的交互桥梁，充分捕获潜在扩散模型 (LDM) 中固有的文本到图像的映射知识，让文本语义能够精准引导潜在特征的演化与优化。总的来说，在扩散过程的每一步 t 中，给定该步骤的潜在 HQ 特征 x_t^{HQ} 、前文得到的对齐交互特征 f_2 以及经过 BLIP2 生成、编码器处理后传递至 ControlNet 的文本嵌入 c ，我们即可通过以下公式计算得到扩散步骤 $t-1$ 的潜在 HQ 特征，完成扩散去噪的迭代过程：

$$x_{t-1}^{HQ} = \frac{1}{\sqrt{t}}(x_t^{HQ} - \frac{1-t}{\sqrt{1-t}} \hat{c}(f_2, c, t)) + \epsilon_t, \quad (2)$$

该公式清晰体现了文本嵌入 c 、对齐特征 f_2 与潜在 HQ 特征 x_t^{HQ} 之间的协同作用，其中文本嵌入所提供的额外语义信号，结合 UNet 的约束的作用，能够有效控制扩散生成的方向与细节，确保生成的高清图像不仅在结构上与 LQ 图像保持一致，更能在语义层面贴合文本描述的需求，进一步提升超分结果的质量与实用性。

IV. 实验结果与分析

在本节中，我们首先描述了所提出方法的数据集和实现细节。然后，我们使用公开可用的基准数据集来评估我们的方法与最先进的方法。

A. 数据集

采用 SDXL 的基本模型作为扩散模型，并使用其 VAE 编码器作为 LQ 编码器。使用 NVIDIA 4090 24G 来训练所提出的方法。在训练过程中，将图像裁剪成 512×512 像素的 patch，并将批大小设置为 256。以 5×10^{-5} 的初始学习率进行 6000 次迭代。然后，对整个

表 I: 在 DIV2K 数据集下不同方法的指标

Methods	PSNR↑	SSIM↑	LPIPS↓	FID↓
StableSR	22.3377	0.6162	0.4144	57.715
ResShift	24.69	0.6175	0.3374	36.01
DiffBIR	23.67	0.5653	0.3541	30.93
PASD	23.14	0.5489	0.3607	29.32
SinSR	24.43	0.6012	0.3262	35.45
OSDiff	23.72	0.6108	0.2941	26.32
SeeSR	21.7714	0.5943	0.3577	47.3327
baseline-PiSA	23.87	0.6058	0.2823	25.07
PassionSR	25.11	0.7097	0.2496	67.92
CCSR	24.17	0.6130	0.3152	36.08
OUR-N	24.8994	0.7018	0.2007	55.0423

网络进行微调，包括 LQ 编码器、校准模块和扩散模型，以端到端的方式进行 40,000 次迭代。在测试时，采样步长为 20 步。Python 版本为 3.10，CUDA 版本为 12.4，Pytorch 版本为 2.4.0。训练数据集。采用 DIV2K 的高分辨率图像训练数据集。并且采用 Real-ESRGAN 的退化管道合成 LQ-HQ 训练对象。使用 LLAVA 为每个图像生成文本描述。测试数据集，使用 DIV2K 和 RealSR 进行评估。评估指标。为了定量评估模型，因此采用了广泛使用的感知指标 PSNR、SSIM、LPIPS、FID 来进行比较。

B. 性能对比-定量比较

我们在 DIV2K 数据集和 REALSR 数据集上与最先进的图像 SR 方法进行了比较，如基于扩散的方法 StableSR、DiffBIR、PASD 等。我们使用 PSNR (峰值信噪比)、SSIM (结构相似性指数)、LPIPS (学习感知图像块相似度) 和 FID (弗雷歌距离) 来评估恢复图像的保本度和质量。

如表 I 和表 II 所示，其中红色表示最优结果，蓝色表示次优结果，↑ 表示该指标越高越好，↓ 表示该指标越低越好，数据来源见引用。

在多个评估指标上，在 2 个数据集上，我们的方法取得了不错的性能，均证明了我们方法的有效性。

C. 性能对比-定性比较

为了验证本章方法的有效性，我们在 DIV2K 数据集和 REALSR 数据集上对提出的方法和现有方法进行定性评估。对比方法包括基于扩散的方法 StableSR、

表 II: 在 RealSR 数据集下不同方法的指标

Methods	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FID \downarrow
StableSR	24.69	0.7052	0.3091	127.2
ResShift	26.31	0.7411	0.3489	142.81
DiffBIR	24.88	0.6673	0.3567	124.56
PASD	25.22	0.6809	0.3392	123.08
SinSR	26.3	0.7354	0.3212	137.05
OSDiff	25.15	0.7341	0.2921	123.5
SeeSR	26.5952	0.7689	0.2796	58.3206
baseline-PiSA	25.5	0.7417	0.2672	124.09
PassionSR	25.15	0.7196	0.4199	65.88
CCSR	25.86	0.7335	0.2941	126.12
OUR-N	26.4301	0.8135	0.2546	58.9979

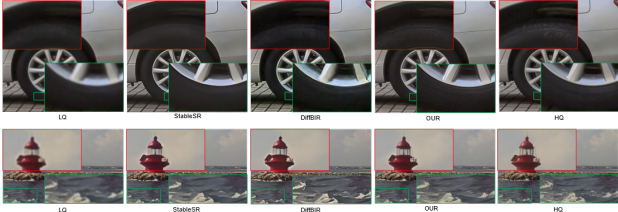


图 2: 超分结果可视化

DiffBIR、PASD 等。下图2展示了我们方法与现有方法在 DIV2K 数据集上的图像超分辨率可视化结果。

现有的基于扩散的方法并不能恢复令人满意的结果。直接将 LQ 输入注入扩散过程的方法会产生过平滑区域，因为冻结的生成先验很难与 LQ 输入很好地对齐，限制了强大的生成能力。尽管这些方法提出从 LQ 输入中提取退化鲁棒特征，但 LQ 特征中的任何错误都可能误导扩散过程，从而产生不准确的结果。相比之下，通过释放扩散先验并与编码器共同优化，我们的方法产生了更真实的图像，具有忠实的结构细节，更好的指标。

V. 消融实验

我们对所提出的网络进行了一系列消融研究，包括 LQ 特征提取融合重要性，高级信息的作用以及交叉注意力模块的作用。我们还研究了是否有必要使用 LQ 编码器的倒数第二层的视觉特征而不是最后一层的视觉特征。为此，我们进一步与 baseline 进行比较，该基线用 LQ 编码器最后一层的视觉特征替换 LQ 特征。我们的方法比 baseline 性能更好，在 DIV2K 数据集上将 LPIPS 降低了 0.0222。对比结果表明，从 VAE 编码器

表 III: 在 DIV2K 数据集下的消融实验

	A	B	C	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	FID \downarrow
A	×	√	√	24.2994	0.6668	0.2187	61.6423
B	√	×	√	22.7221	0.6135	0.2911	52.0012
C	√	√	×	21.6494	0.5543	0.3640	50.7819
OUR-N	√	√	√	24.8994	0.7018	0.2007	55.0423

的倒数第二层提取的特征可以有效地捕获更多有用的信息。LQ 特征提取融合重要性。我们通过排除 LQ 特征的融合来评估，即简单地将 LQ 提取的倒数第二层特征输入到 UNet 中。如表所示，指标明显降低。这验证了 LQ 特征提取融合在结构中的重要性。文本嵌入高级信息的作用。文本嵌入高级信息对模型网络非常有用。我们只需将它们替换为空文本提示符来评估它们的效果。从表 4 中较差的 FID 和 LPIPS 指数中得到了验证。放弃文本高级信息会导致过度平滑的结果。我们的完整模型同时利用了文本的高级信息，并在干净平滑和详细肮脏的输出之间取得了很好的平衡 (表中的较佳 FID 评分)。

VI. 结论

本文提出了一种有效的图像 SR 方法，进行扩散先验，更好地利用了 LDM 的强大表示能力。本文提出一种基于 VAE 多尺度特征融合与跨模态对齐的高保真图像超分方法，通过多尺度特征充分利用、特征空间对齐、跨模态语义引导、UNet 结构严格约束，有效解决现有超分方法细节表征不足、结构易失真、伪影过多、真实感不足等问题。这种统一的优化策略使编码器能够提供与扩散过程相吻合的有用特征，而潜在扩散模型可以进一步恢复这些特征。得益于它们之间的相互作用，整个网络可以有效地区分退化效应和固有结构信息，从而实现高质量的恢复。实验结果表明，本文方法在定量指标、定性效果、视觉质量上均优于现有主流方法，能够实现清晰、自然、高保真、高质量的低质图像超分，具有良好的实用性与扩展性，为真实场景低质图像超分提供了一种有效、可靠、高质量的解决方案。扩散先验在从退化的输入中探索有用信息和恢复忠实结果方面更有效。在结构保真度和视觉质量方面较优于最先进的方法，证明了本文提出方法的有效性。

参考文献

- [1] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention

u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.