一、实验结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 原图像 | 插值重建 | 超分 | PSNR | SSIM |
| baby | baby | baby | 25.66161227 | 0.79573644 |
| bird | bird | bird | 20.83269869 | 0.67002514 |
| butterfly | butterfly | battlefy | 16.30550324 | 0.61234694 |
| head | head | head | 27.24681641 | 0.81015788 |
| woman | woman | woman | 19.82913088 | 0.71676738 |

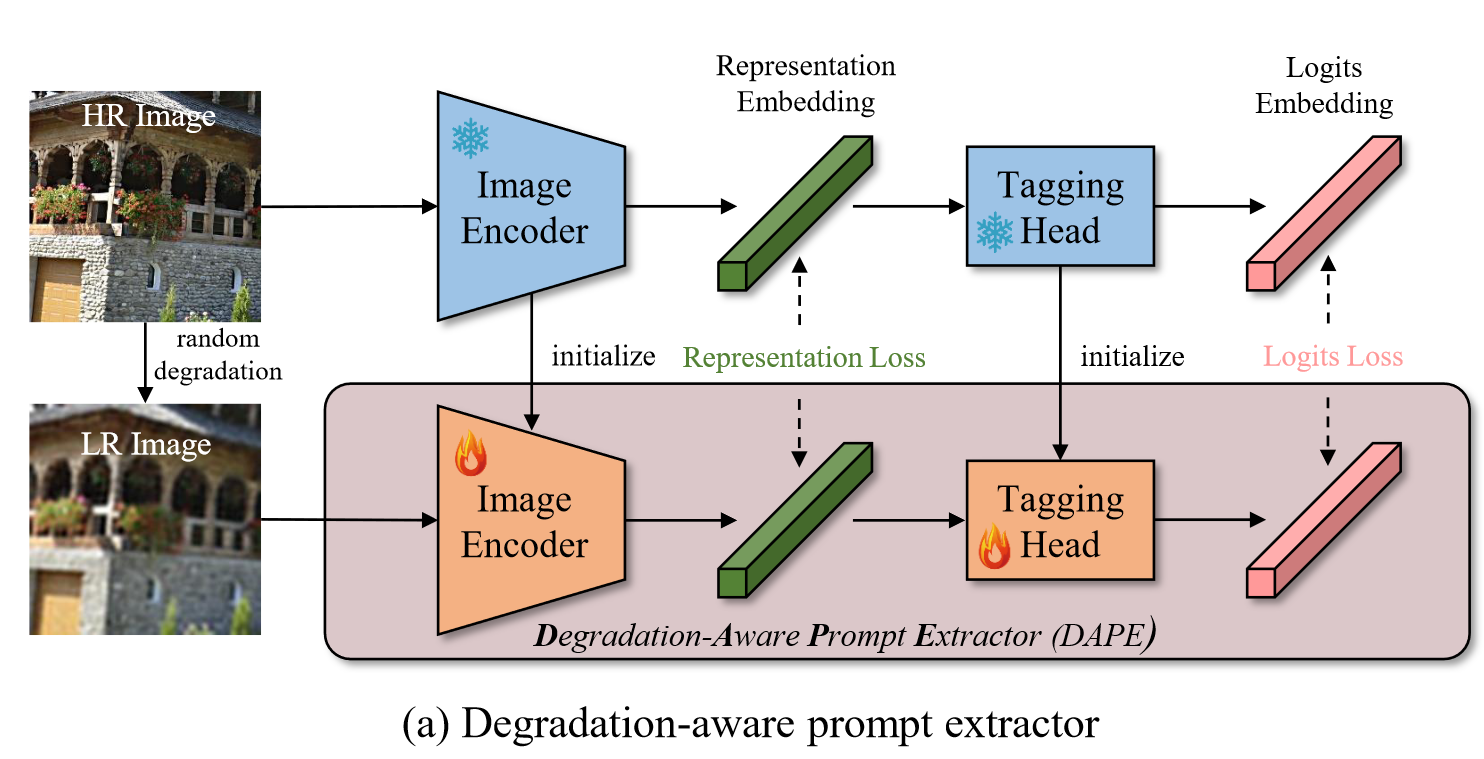
1. SeeSR方法介绍

SeeSR的模型训练分为两步，首先训练一个Degradation-aware prompt，这指的是模型在生成文本时能够感知到输入文本的质量降级或错误，并试图在输出中修正这些问题。期望通过使用原始标签模型，LR图像的特征表示和标签输出能够尽可能接近相应HR图像的特征表示和标签输出。

在第二阶段中，学习到的DAPE被复制到第二阶段，从输入 LR 图像中提取特征表示和标签（作为文本提示），这些特征表示和标签作为预训练 T2I 模型的控制信号，以生成视觉上令人愉悦的、语义上正确的 Real-ISR 结果。在推理过程中，只需要第二阶段来处理输入图像。

（1）DAPE

DAPE 是根据预训练的标签模型RAM进行微调的。如图所示，HR图像x通过冻结标签模型输出representation embedding和logits embedding作为锚点来监督DAPE的训练。 LR 图像 y 是通过对 x 应用随机降级获得的，并将它们输入到可训练图像编码器和标记头中。



（2）训练

使用ControlNet作为T2I模型的控制器。将预训练的Sable Diffusion中的Unet编码器作为可训练副本来初始化ControlNet。采用PASD中的交叉注意力机制来进行语义指导，表示交叉注意力被添加到Unet中并在文本交叉注意力模块之后。训练时冻结SD的参数，仅训练新添加的模块，包括了图像编码器，ControlNet以及RCA交叉注意力模块。

（3）推理

直接将LR latent嵌入到初始的随机高斯噪声中。这是为了避免使用SD预训练模型时其在训练阶段未将图片完全转换为随机高斯噪声，但在推理阶段以随机高斯噪声为起点导致的训练和推理过程中因噪声差异而存在的问题。

1. 优点

作者提出了 SeeSR，一种 Real-ISR 方法，利用语义提示来增强预训练 T2I 扩散模型的生成能力。通过探索不同风格的文本提示对生成结果的影响，发现图像标签可以极大地增强T2I模型的局部感知能力。然而，标签容易受到复杂图像退化的影响，并且受到手动设置阈值的影响。因此，作者提出了DAPE，它最大限度地减少图像退化对语义提示的影响，同时输出软和硬语义提示来指导图像超分辨率中的扩散过程。此外，为了解决扩散模型中训练测试不一致的不利影响，作者提出了一种简单而有效的 LRE 策略，该策略在扩散过程的起点嵌入 LR 潜伏，避免在平滑区域中产生伪影。正如我们广泛的实验所证明的那样，我们的工作朝着更好地利用生成先验来合成语义正确的 Real-ISR 图像迈出了一步。