

基于 transformer 的 diffusion 超分辨率模型

李孟霖

December 6, 2025

Introduction

Introduction

研究背景

主流超分辨模型

EDSR(CNN)

SwinIR(transformer)

StableSR(Diffusion)

Diffusion 模型在生成任务中往往具有比较好的高频细节生成能力, 本研究中考考虑使用 Diffusion 方法构造超分辨模型

UNet

DiT

Transformer 模型相比于 UNet 具有更好的全局性、更强的灵活性和更低的计算开销，本研究考虑使用 Transformer 作为去噪模块

去噪方式

epsilon-prediction

x-prediction

v-prediction

Introduction

相关工作

何恺明团队 CFG 条件生成模型 (JiT), Back to Basics: Let Denoising Generative Models Denoise

低维流形假设

以 vit 为去噪模块 backbone, 模型做 x-prediction, x-prediction 的预测值变换为 v-prediction 构造损失函数

Methods

主体与 JiT 相同，删去 CFGtoken 拼接，尝试按下述两种方式将低清图信息引入

低清图插入方式

- 将低清图编码为 token, 和带噪声图像拼接成一个序列
- 将低清图和带噪声图像按通道拼接

Experiments

Experiments

按 token 拼接

psnr ssim

效果图

Experiments

按通道拼接

psnr ssim

效果图

Experiments

基线对比

谢谢大家！