

USDR

状态：CVPR2023

单位：北京大学

文章链接：<http://arxiv.org/abs/2303.00354>

Github 链接：<https://github.com/wyhuai/DDNM>

目录

摘要	1
不足之处	2
背景	2
扩散模型	2
DDNM	2
RND	3
基于扩散模型的图像修复	3
贡献	3
现存问题	3
贡献	3
实现	4
MSR	4
HiR	4
运用的通用性	4
实验	5

摘要

预训练的扩散模型可以被轻易的使用在不同的下游任务之中。但之前的工作总是基于某一特定的尺寸来讨论效果。而对于大尺寸的图像，则需要分块进行处理，但这一方案既没有考虑到整体图像的语义，也没有考虑到邻近图像的一致性，因此，本文提出了一种特殊的分块方案，能够使得扩散模型具有处理任意尺寸图像的能力。

不足之处

1. 实质上加大了计算开销
2. 基于预训练模型（不觉得是个问题）
3. 无法很好的在隐空间上兼容
4. 需要反向的模糊核（即把图像信息恢复的线性核）

背景

当前，预训练的扩散模型可以用于多种领域，比如图像修正（IR）。但大多数工作都只在特定分辨率上进行研究，而实际使用中，却可能会有各种分辨率的输入。如果输入与模型所适应的尺寸差距过大，就可能会出现 OOD 问题。

扩散模型

本文的扩散模型原理较为简洁，可以作为参考。

DDNM

一种奇怪的算法，大致意思是，对于一个由原图 x 转化而来的 y ：

$$y = Ax$$

我们想根据 y ，求解出一个 \hat{x} ，使其尽可能接近真实的 x ，实际上要满足：

$$\text{一致性: } A\hat{x} \sim y \quad \text{真实性: } \hat{x} \sim x$$

而为了实现这两个要求， \hat{x} 可以表示为：

$$\hat{x} = A^\dagger y + (I - A^\dagger A)x_r$$

其中， A^\dagger 为变化矩阵 A 的伪逆矩阵，而 $A^\dagger y$ 则意味着通过 y 中剩余的信息去倒推回原始的图像。但这并不足够，因为变化的过程中会损失一部分信息， $(I - A^\dagger A)$ 就是这部分损失的信息，通过预测的 x_r 进行补充。

而在扩散模型领域，此处的 x_r 即为根据 x_t 预测出的原图 $x_{0|t}$ 。

具体的伪代码如下：

DDNM 算法

- 1: $x_T \sim N(0,1)$
 - 2: **for** $t = T, \dots, 1$ **do**
 - 3: $\epsilon_t = Z_\theta(x_t, t)$
-

```
4:      $x_{0|t} = \frac{1}{a_t}(x_t - \theta_t \epsilon_t)$ 
5:      $\hat{x}_{0|t} = A^\dagger y + (I - A^\dagger A)x_{0|t}$ 
6:      $x_{t-1}$  使用  $\hat{x}_{0|t}$  进行计算
7: return  $x_0$ 
```

可以看出，DDNM 的实质就是向普通的 DDPM 反向过程中，不断加入了 y 的信息，来引导最后的生成结果。而具体的引导方向和强度由 A^\dagger 决定，也就是由实际处理问题的种类决定。

RND

RND (Range-Null space Decomposition) 给反向问题定义了信息恢复的上界。也就是说，存在可恢复的部分+不可恢复的部分。而神经网络的先验知识被广泛用于恢复（实际上是生成）不可恢复的部分。

基于扩散模型的图像修复

即使用预训练扩散模型中的先验知识，去对图像进行恢复。扩散模型的作用即为预测不可恢复的部分。

贡献

现存问题

当下的预训练扩散模型，由于是在特定大小的数据上进行训练的，其保守 OOD 问题（超出生成范围）困扰。一种解决方案是使用不同大小的数据进行训练，使模型本身具有处理任意尺寸图像的能力。其问题在于：

1. 解决不够彻底，如果输入的尺寸是完全没训练过的，效果依然很差。
2. 仍然对尺寸有限制，比如被 32 整除。
3. 大尺寸图像会带来极大的算力需求。

另一种解决方案便是把超出范围的图像切块，并针对每一块图像进行独立生成，最后合并成最终结果。但这种方案有两大问题：

1. 没有考虑与周边图像的相似度。
2. 无法获得全图的语义信息。

贡献

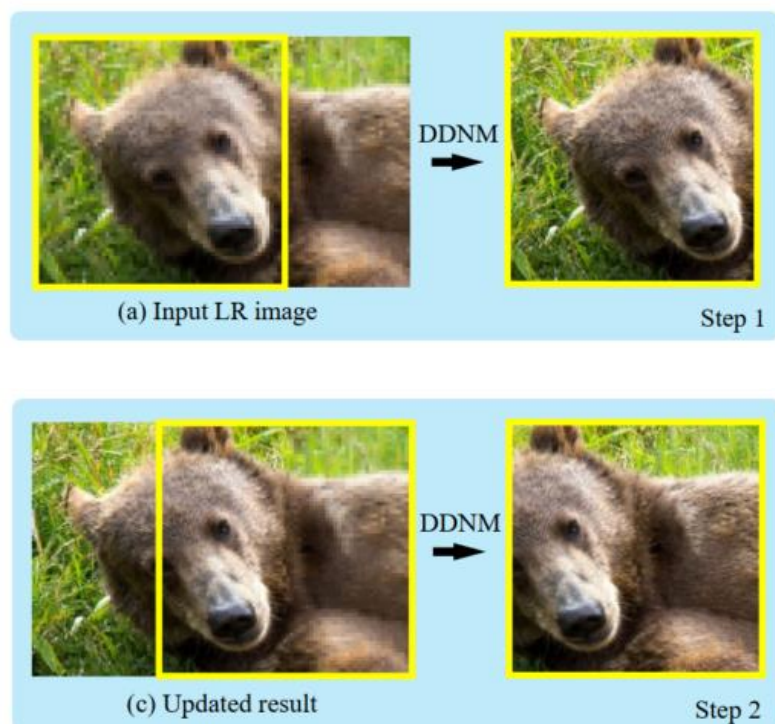
1. 我们提出了 MSR，一种留出重叠部分以保证相邻块相似度的方案。

2. 我们提出了 HiR，一种保证每个分块都具有全局信息的方案。
3. 我们提供了使用 HiR 和 MSR 的基本架构，可以用于各种不同的下游任务。

实现

MSR

主要的实现方案就是通过留出一些重叠的部分来保证相邻块之间的相似程度。



HiR

MSR 通过一种简单的方式保证了相邻图像之间的相似程度，并很好的拟合了边界。但对于一个较大的图像而言，缺乏全图的整体信息依然会带来一定的影响。为了保证这种全局语义信息，HiR 方案被提出。

其整体流程分为两步：

1. 首先把原图缩小，缩放到每一个分块都足以有足够语义信息的程度（即每一个分块都能包含较为完整的物体，而非过于琐碎的部分）。对这个缩小的图像进行 MSR，得到缩小的恢复图像。
2. 再对原图中的每个部分开始分块处理，同时，使用上一步得到的小恢复图中相对应的部分作为指导，以保证全具语义信息的完整。

运用的通用性

MSR+HiR 的组合，可以十分契合的用于超分辨率，图像修复，色彩调节等种种图像翻译任务之中。

实验

本文只有定性对比实验，较为简单。

