

# 基于扩散模型的高分辨率星系图像生成\*

赵瑞妮<sup>1,2†</sup> 王形锋<sup>1,2</sup> 吕涛<sup>1,2</sup>

(1 中国科学院西安光学精密机械研究所 西安 710119)

(2 中国科学院空间精密测量重点实验室 西安 710119)

**摘要** 在大型巡天项目的数据积累和信息技术双快速发展的背景下,使用智能化方法对海量星系图像进行自动化分类是理解星系形成、演化及宇宙环境的基石.星系智能化分类模型的精度与训练样本规模及样本均衡性存在显著相关性,采用传统的数据增强方法对复杂星系图像的增强效果有限.基于生成式人工智能在图像生成领域的突破性进展,提出一种基于扩散模型的星系图像生成算法,生成的星系图像具有真实的形态学特征及纹理细节.使用增广数据集训练得到的分类模型其精确率、召回率及F1分数均优于原始数据集下学习的模型,3项指标最大可分别提升4.01%、4.3%和5.06%.利用生成的高保真清晰图像使星系形态分类模型的精度得到显著提升,为未来大型巡天项目的星系形态分类研究提供了新的思路.通过对生成式人工智能模型的持续探索,将为天文学研究提供强大工具,帮助人类更深入地洞察宇宙奥秘.

**关键词** 星系: 普通, 方法: 数据分析, 技术: 图像处理, 技术: 扩散模型, 技术: 生成式模型, 技术: 数据增广  
**中图分类号**: P152; **文献标识码**: A

## 1 引言

星系形态结构作为其形成历史、动力学演化及所处环境的直接记录,是解开星系乃至宇宙演化之谜的关键钥匙.从哈勃序列的经典分类到如今基于大型巡天的多维度定量描述,对星系形态的精确刻画始终是天文物理学研究的核心前沿<sup>[1]</sup>.随着斯隆数字巡天(Sloan Digital Sky Survey, SDSS)<sup>[2]</sup>、泛星计划(Panoramic Survey Telescope & Rapid Response System, Pan-STARRS)<sup>[3]</sup>及即将全面运行的薇拉鲁宾天文台(Legacy Survey of Space and Time, LSST)<sup>[4]</sup>等项目的推进,我们正步入一个数据洪流的时代. LSST预计每晚将产生超过20 TB的观测数据,识别并分类数以亿计的星系.为应对处理海量星系数据洪流,传统依赖人工

视觉检查的分类方法已失效,更因其固有的主观性强和一致性差等缺陷,严重制约了星系演化研究的速度.因此,开发高效、客观、可复现的星系形态自动化分类模型,已成为从海量星系数据中深化对星系形成和演化机制理解的必然要求和核心路径.

星系形态学分类模型已取得了一定的发展<sup>[5]</sup>,但其性能深受数据规模及样本均衡性的制约.当数据量较少及各类别样本不均衡时,模型易因信息不足而陷入“欠拟合”,难以捕捉到特征与标签间的复杂关联,导致精度受限;随着数据量的增加,模型能够学习到更丰富的样本多样性和更鲁棒的特征表示,从而有效提升模型精度并增强对未知数据的泛化能力.

2025-09-28收到原稿, 2025-12-15收到修改稿

\*国家自然科学基金项目(12103081)资助

† zhaoruini@opt.ac.cn

现有的数据增广方法基于不同的生成方式被分为单数据变形、多样本融合及数据分布学习3类. 基于数据变形的的方法侧重于对单个样本进行一定的变换操作以获取新的样本数据. 对应用于深度学习任务的单数据变形方法, 包括图像的仿射变换、缩放、裁剪<sup>[6]</sup>、镜像<sup>[7-8]</sup>等操作, 均取得了一定的成效. 单数据变形类方法除了基于上述几何变换的方法, 还涉及到基于色域变换<sup>[9]</sup>、基于清晰度变换<sup>[10]</sup>、噪声注入<sup>[11]</sup>及局部擦除<sup>[12]</sup>等方式.

对于多数据混合的数据增广方式, 通常将多个样本的特征信息进行融合, 获取更多的信息生成新的数据. 例如SamplePairing通过平均两张图像像素并随机分配其中一个标签生成新的训练样本<sup>[13]</sup>; MixUp通过线性插值构造虚拟训练样本, 不仅混合两张输入图像, 还混合它们对应的标签, 鼓励模型行为线性化<sup>[14]</sup>; SMOTE (Synthetic Minority Over-sampling Technique)通过在特征空间内合成新的少数类样本平衡数据集<sup>[15]</sup>. 但基于单数据变形和多样本融合的增广方法能够学习到的特征信息较少, 局限于选择的样本本身, 无法关注整个数据集的分布.

基于分布学习的数据增广方法可以通过学习数据的潜在概率分布, 生成新的、与原始数据相似但又不完全相同的数据. 变分自编码器通过编码-解码架构学习数据的隐变量分布, 并从中采样生成新样本<sup>[16]</sup>. 自回归模型通过精确建模像素间的序列关系实现高质量的图像生成<sup>[17]</sup>. 对抗生成网络通过生成器和判别器之间的极小极大博弈, 使两者在不断学习过程中, 分别提升自身生成能力及鉴别能力<sup>[18-19]</sup>.

综上所述, 上述基于单数据变形、多样本融合及数据分布学习等方法, 均能够为缓解特定类别星系图像数据稀缺、改善深度学习模型的泛化能力提供有益的解决方案. 但传统增广方法多依赖于低层次图像, 所生成的数据多样性有限, 且难以模拟宇宙天体复杂多变的形态学特征. 基于生成对抗网络(Generative Adversarial Network, GAN)<sup>[20]</sup>的方法虽然能够生成较为逼真的图像, 但其训练过程造成了其训练不稳定, 易出现模式崩溃问题. 鉴于扩散模型在高质量、高多样性图像

生成方面的潜力, 本文提出一种基于扩散模型的星系图像生成网络, 其生成图像作为原始数据集的增广, 在多个分类模型上进行了性能验证.

## 2 基于扩散模型的星系图像生成

### 2.1 扩散模型原理

去噪扩散模型包括前向扩散及反向去噪两个阶段, 其核心思想是通过给清晰图像样本逐渐添加噪声, 使原本清晰的图像变为完全随机、没有任何规律的噪声, 模型将在该过程中观察学习清晰图像是如何逐步被噪声破坏. 在进行图像生成时, 将利用学习到的知识, 逐步将噪声还原成类似源域的图像. 在反向去噪过程中通过添加少量随机噪声, 使采样过程具有随机性, 可用于确保生成模型的多样性.

如图1所示,  $x_0$ 为原始图像, 其满足如下分布  $x_0 \sim q(x_0)$ . 前向扩散指正向加噪过程, 对于  $t \in [1, T]$ 时刻, 在图像上逐步加入标准正态分布噪声  $\epsilon \sim N(0, 1)$ .

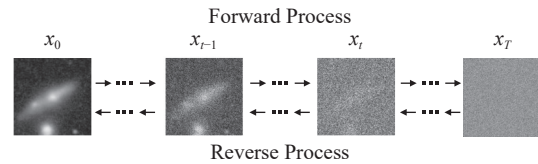


图 1 去噪扩散模型原理图

Fig. 1 Schematic of the denoising diffusion principle

$x_t$ 和 $x_{t-1}$ 满足如下关系式

$$x_t = \sqrt{1 - \beta_t}x_{t-1} + \sqrt{\beta_t}\epsilon. \quad (1)$$

令 $\alpha_t = 1 - \beta_t$ ,  $\beta_t$ 为固定常数, 且 $0 < \beta_t < 1$ , 其随 $t$ 的增加而增加, 则上式变为

$$x_t = \sqrt{\alpha_t}x_{t-1} + \sqrt{1 - \alpha_t}\epsilon. \quad (2)$$

同时又有

$$x_{t-1} = \sqrt{\alpha_{t-1}}x_{t-2} + \sqrt{1 - \alpha_{t-1}}\epsilon, \quad (3)$$

$$x_t = \sqrt{\alpha_t}(\sqrt{\alpha_{t-1}}x_{t-2} + \sqrt{1 - \alpha_{t-1}}\epsilon) + \sqrt{1 - \alpha_t}\epsilon. \quad (4)$$

因此

$$x_t = \sqrt{\alpha_t \alpha_{t-1}} x_{t-2} + \sqrt{\alpha_t} \sqrt{1 - \alpha_{t-1}} \epsilon + \sqrt{1 - \alpha_t} \epsilon. \quad (5)$$

依据正态分布的叠加性, 公式(5)末尾两项分别服从如下分布:

$$\sqrt{\alpha_t} \sqrt{1 - \alpha_{t-1}} \epsilon \sim N[0, \alpha_t(1 - \alpha_{t-1})], \quad (6)$$

$$\sqrt{1 - \alpha_t} \epsilon \sim N(0, 1 - \alpha_t), \quad (7)$$

则有如下分布

$$\sqrt{\alpha_t} \sqrt{1 - \alpha_{t-1}} \epsilon + \sqrt{1 - \alpha_t} \epsilon \sim N(0, \alpha_t - \alpha_t \alpha_{t-1} + 1 - \alpha_t) = N(0, 1 - \alpha_t \alpha_{t-1}). \quad (8)$$

于是

$$x_t = \sqrt{\alpha_t \alpha_{t-1}} x_{t-2} + \sqrt{1 - \alpha_t \alpha_{t-1}} \epsilon. \quad (9)$$

根据数学归纳法, 可进一步得到

$$x_t = \sqrt{\alpha_t \alpha_{t-1} \cdots \alpha_1} x_0 + \sqrt{1 - \alpha_t \alpha_{t-1} \cdots \alpha_1} \epsilon. \quad (10)$$

令  $\bar{\alpha} = \alpha_t \alpha_{t-1} \cdots \alpha_1$ , 则公式可进一步化简为

$$x_t = \sqrt{\bar{\alpha}} x_0 + \sqrt{1 - \bar{\alpha}} \epsilon. \quad (11)$$

因此任意时刻的加噪图像都可由初始时刻的图像直接得到, 无需进行多次迭代, 这为模型学习不同时间步下的对应噪声提供可能. 由于  $\beta_t$  一直变大时,  $\alpha_t$  一直在变小, 因此当  $t \rightarrow T$  时,  $\bar{\alpha}_T \rightarrow 0$ , 则  $x_T \rightarrow \epsilon$ . 所以经过多步加噪, 最终将清晰图像变成高斯噪声.

在反向去噪时, 利用当前时刻图像  $x_t$  预测前一时刻图像  $x_{t-1}$ , 意味着去除一部分噪声, 还原到上一时刻的图像, 该过程可用贝叶斯公式计算后验概率, 如下式所示:

$$P(x_{t-1}|x_t) = \frac{P(x_{t-1}x_t)}{P(x_t)} = \frac{P(x_t|x_{t-1})P(x_{t-1})}{P(x_t)}. \quad (12)$$

式中,  $P(x_{t-1}|x_t)$  为在  $x_t$  发生的前提下  $x_{t-1}$  发生的概率,  $P(x_{t-1})$  为  $x_{t-1}$  发生的概率,  $P(x_{t-1}x_t)$  为  $x_t$  与  $x_{t-1}$  同时发生的概率. 在已知原图  $x_0$  时, 上式可写为

$$P(x_{t-1}|x_t, x_0) = \frac{P(x_t|x_{t-1}, x_0)P(x_{t-1}|x_0)}{P(x_t|x_0)}. \quad (13)$$

将公式(2)和(11)代入上式可得到

$$P(x_{t-1}|x_t, x_0) = \frac{N(\sqrt{\alpha_t} x_{t-1}, 1 - \alpha_t) N(\sqrt{\bar{\alpha}_{t-1}} x_0, 1 - \bar{\alpha}_{t-1})}{N(\sqrt{\bar{\alpha}_t} x_0, 1 - \bar{\alpha}_t)}. \quad (14)$$

所有时刻的分布均为正态分布, 则公式(7)  $P(x_{t-1}|x_t, x_0) \sim \exp\left(-\frac{x^2 + \mu^2 - 2x\mu}{2\sigma^2}\right)$ , 其中  $\mu$  和  $\sigma$  分别是均值和方差. 则由当前时刻分布  $x_t$  获得上一时刻分布  $x_{t-1}$  的均值和方差分别如下式所示

$$\mu = \frac{\sqrt{\bar{\alpha}_{t-1}}(1 - \alpha_t)}{1 - \bar{\alpha}_t} x_0 + \frac{(1 - \bar{\alpha}_{t-1})\sqrt{\alpha_t}}{1 - \bar{\alpha}_t} x_t, \quad (15)$$

$$\sigma^2 = \frac{(1 - \alpha_t)(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t}. \quad (16)$$

使用公式(11)替换公式(15)中的  $x_0$ , 则  $\mu$  如下式所示

$$\mu = \frac{1}{\sqrt{\alpha_t}} \left( x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon \right), \quad (17)$$

即

$$P(x_{t-1}|x_t) = N \left[ \frac{1}{\sqrt{\alpha_t}} \left( x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon \right), \frac{(1 - \alpha_t)(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} \right]. \quad (18)$$

反向去噪过程中, 从当前时刻分布  $x_t$  获得上一时刻分布  $x_{t-1}$  时所需噪声  $\epsilon$  可通过神经网络学习得到, 其具体介绍见2.2节.

## 2.2 噪声预测网络架构

基于扩散模型的星系图像生成算法框架如图2所示, 该噪声预测网络基于编码器-解码器架构, 核心为UNet架构, 将指定时间步下的带噪图像和对应时间步输入该UNet架构, 对时间步进行编码, 将标量时间步  $t$  转换为丰富的、高维的向量表示. 带噪图像经过初始特征提取后, 利用级联的下采样编码器捕获图像的上下文信息, 在该过程中, 将编码后的时间步与图像编码特征进行自适应融合, 该时间条件信息用于指导反向去噪过程. 自适应融合后特征依次通过级联的上采样解码器, 并与编码器对应层的特征图进行拼接, 以恢复星系图像的空间细节.

对时间步  $timestep$  的正弦位置编码(Sinusoidal position encoding)将离散时间步转换为高维连续

的向量表示,用于学习时间步间的相对关系.噪声预测网络基于UNet架构,不同时间步下的带噪图像及编码后的时间步共同作为噪声预测网络的输

入,输出预测噪声.训练完成的噪声预测网络,噪声在 $T$ 个反向时间步中被预测并逐渐移除,将高斯噪声还原为原始分布.

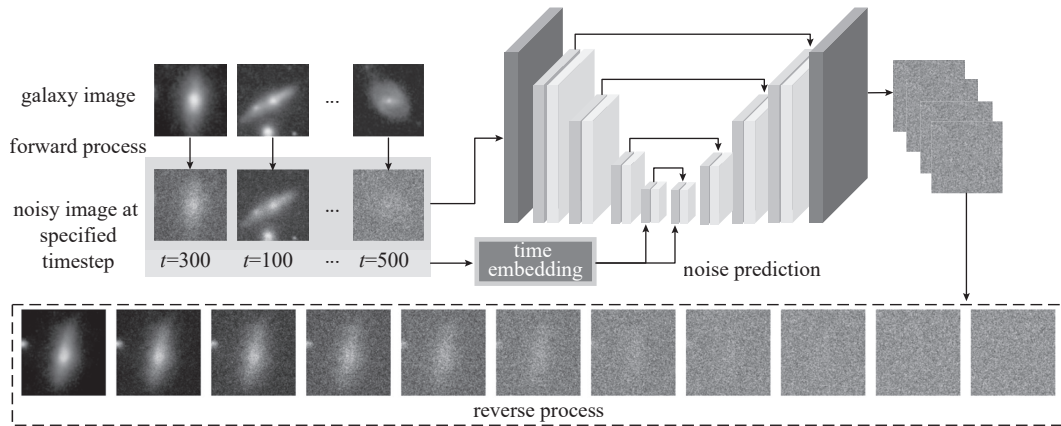


图 2 基于扩散模型的星系图像生成算法框架

Fig. 2 Framework for a diffusion model-based galaxy image generation algorithm

本文对于时间步的编码使用正弦位置编码,利用自然对数相关的递减频率参数,用于覆盖更大的数值范围,同时提供平滑且富有表现力的编码.高频分量能够提供相邻时间步之间的区分度,包含足够的信息确定任何一个时间步.低频分量能够保证变化平滑性、保证编码在低频部分是逐渐变化的,帮助网络理解相邻时间步间的关系.对不同频率作用下的时间步,分别使用正弦及余弦函数进行编码,为每个时间步生成一个唯一的、可区分的编码,避免单一周期性函数可能引发的不同位置编码相同的问题.同时能够捕获每个频率的完整相位信息,增加特征表示的鲁棒性及稳定性.

在扩散模型的前向传播过程中,时间步信息作为一个动态的、与内容无关的“调制信号”,对图像特征图的每个通道进行全局缩放,模型通过该方式学习不同噪声水平下,如何增强或抑制各通道的响应.具体而言,在噪声严重的早期时间步,调制信号引导网络关注恢复图像的整体结构和轮廓,在噪声较少的后期时间步,调制信号引导网络更加专注于生成细节和纹理.整个自适应过程,调制参数由时间步通过网络动态生成,使得模型能够精准地根据输入的噪声级别调整其去噪能力,

最终实现依据特征噪声水平控制去噪行为.

### 3 训练样本选取及处理

本文使用来源于SDSS的星系动物园2 (Galaxy Zoo 2, GZ2)数据集进行训练,共61578张样本图片.该数据集通过一个复杂的决策树<sup>[21]</sup>提供了每个星系从整体形状(圆、椭圆、螺旋等)到精细结构(棒、旋臂数量等)的多层次、概率性形态信息.  $f_{\text{smooth}}$ 、 $f_{\text{completelyround}}$ 、 $f_{\text{inbetween}}$ 、 $f_{\text{cigar-shaped}}$ 、 $f_{\text{features/disk}}$ 、 $f_{\text{edge-on,yes}}$ 与 $f_{\text{spiral,yes}}$ 分别表示一张星系图片被分类为平滑星系、圆形星系、介于圆形和雪茄状星系之间、雪茄状星系、平滑且有盘状结构、侧向星系与漩涡星系的概率.

本文研究的星系分为圆形(Completely round smooth)星系、中间(In-between smooth)星系、雪茄(Cigar-shaped smooth)星系、侧向(Edge-on)星系、漩涡(Spiral)星系5种<sup>[5, 22]</sup>,选择标准如表1所示.表中 $T0_i$ 表示GZ2分类树中的第 $i$ 项分类任务<sup>[21]</sup>,以漩涡星系为例,需同时满足3项条件:一张星系图片被分类为平滑且有盘状结构的概率 $\geq 0.430$ 、非侧向星系的概率 $\geq 0.715$ 、漩涡星系的概率 $\geq 0.619$ .依据该数据样本选取规则,从61578个样本

中选取的共5类样本数量如表2所示, 每个类别选择100个样本作为测试样本.

表 1 数据样本选取规则

Table 1 Data sample selection criteria

| Galaxy class            | Tasks | Threshold                              | Number of clean samples |
|-------------------------|-------|--|-------------------------|
| Completely round smooth | T01   | $f_{\text{smooth}} \geq 0.469$         | 8434                    |
|                         | T07   | $f_{\text{completelyround}} \geq 0.50$ |                         |
| In-between smooth       | T01   | $f_{\text{inbetween}} \geq 0.469$      | 8069                    |
|                         | T07   | $f_{\text{smooth}} \geq 0.50$          |                         |
| Cigar-shaped smooth     | T01   | $f_{\text{smooth}} \geq 0.469$         | 578                     |
|                         | T07   | $f_{\text{cigar-shaped}} \geq 0.450$   |                         |
| Edge-on                 | T01   | $f_{\text{features/disk}} \geq 0.430$  | 3903                    |
|                         | T02   | $f_{\text{edge-on,yes}} \geq 0.602$    |                         |
|                         | T01   | $f_{\text{features/disk}} \geq 0.430$  |                         |
| Spiral                  | T02   | $f_{\text{edge-on,no}} \geq 0.715$     | 7806                    |
|                         | T04   | $f_{\text{spiral,yes}} \geq 0.619$     |                         |

表 2 星系数据构成及其规模

Table 2 Composition and size of the galaxy dataset

| Galaxy class            | Original datasets |           | Augmented datasets |           |
|-------------------------|-------------------|-----------|--------------------|-----------|
|                         | Training Sets     | Test Sets | Training Sets      | Test Sets |
| Completely round smooth | 8334              | 100       | 8334               | 100       |
| In-between smooth       | 7969              | 100       | 7969               | 100       |
| Cigar-shaped smooth     | 479               | 100       | 5642               | 100       |
| Edge-on                 | 3803              | 100       | 3803               | 100       |
| Spiral                  | 7706              | 100       | 7706               | 100       |

为了直观地展示使用本文所提出扩散模型进行星系图像生成的作用原理, 并验证其从随机噪声中逐步生成具有合理星系特征图像的能力, 图4展示了5类星系在共计1000步反向去噪过程中每100步的模型生成能力. 从上到下依次为雪茄星系、侧向星系、中间星系、圆形星系及漩涡星系. 从左至右展示了从纯高斯噪声开始( $t = 1000$ ), 模型如何通过一系列连续迭代去噪步骤, 逐步注入

原始图像尺寸为 $424 \times 424$ , 对所有图像进行中心裁剪获得 $224 \times 224$ 的图像, 能够最大化保留主体星系特征的同时减少计算量. 该网络使用的损失函数如下式所示.

$$L(\theta) = \|\epsilon - \epsilon_\theta\|_2^2. \quad (19)$$

式中,  $\theta$ 为噪声预测网络的网络参数,  $\epsilon_\theta$ 为噪声预测网络在网络参数为 $\theta$ 时预测出的噪声.

## 4 实验结果与分析

### 4.1 星系生成结果可视化

本部分分别展示使用本文提出的数据生成模型基于真实的星系图像所生成的圆形星系、中间星系、雪茄星系、侧向星系、漩涡星系与真实星系的对比, 结果分别如图3 (a)–(e)所示. 从图3可见模型在生成不同类别星系时的样本多样性, 因此所生成的样本不仅有效补充了原始数据集, 同时在纹理细节及形态上展现了较大的相似性, 极大地保留了各星系类别的特征, 诸如旋臂、雪茄形状、圆盘形状等细节.

其学习到的星系形状、颜色、分布等先验知识, 到最终生成高保真图像( $t = 0$ )之间多个采样时间步的输出.

每一行均从初始随机噪声, 逐步显现星系形态轮廓与合理的亮度分布, 在此基础上, 再次移除噪声并细化星系形态细节, 包括圆形星系、中间星系的平滑梯度, 漩涡星系的旋臂缠绕度等特征进一步增强. 在迭代末期, 模型生成出具有高度真

实感且细节丰富的星系图像, 其纹理、形状、星光分布、背景噪声等均与真实观测数据十分接近.

该迭代可视化过程不仅能够揭示生成模型重构复杂星系图像的深层能力, 同时验证了本文所提出基于扩散模型的增广数据集并非简单的噪声

叠加或模式复制, 而是基于学习到的各类别数据分布进行的符合观测数据统计分布的创造性生成, 从而为本文后续利用增广数据集提升分类模型性能的有效性和合理性提供了关键依据.

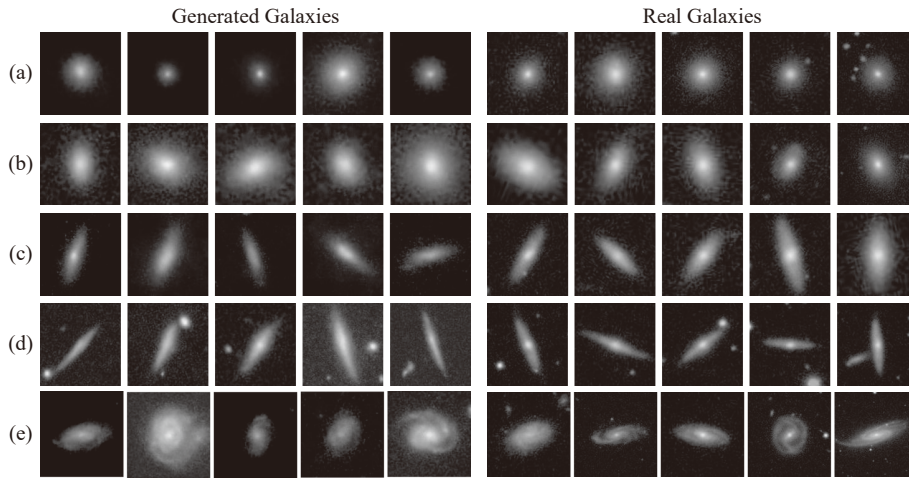


图 3 生成星系与真实星系对比

Fig. 3 Comparison between generated galaxies and real galaxies

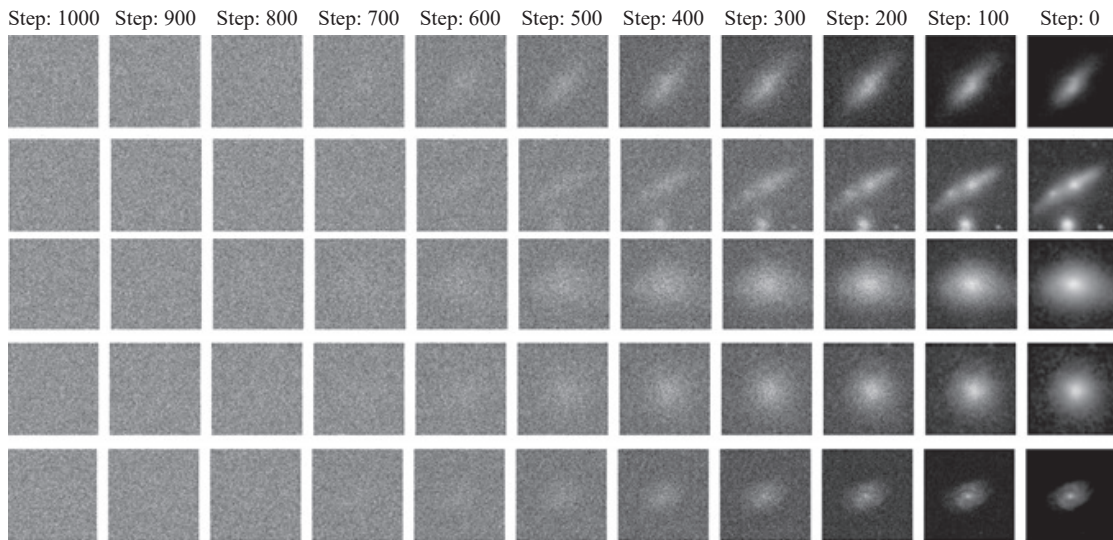


图 4 星系反向去噪过程可视化

Fig. 4 Visualization of the galactic reverse denoising process

鉴于GAN及其变体在图像生成细节清晰度及视觉保真度方面的早期突破, 其仍被广泛视为生成式任务中重要的基准模型. 针对该星系图像生

成任务, 本文使用生成对抗网络框架下的代表性模型—自注意力生成对抗网络 (self-attention GAN, SAGAN)进行对比分析, SAGAN通过引入

自注意力机制, 能够显著提升GAN在建模全局依赖关系上的能力. 基于SAGAN的星系生成结果如图5所示, 从上到下与图3的星系类别一致((a)-(e)分别表示圆形星系、中间星系、雪茄星系、侧向星系、漩涡星系), SAGAN能够勾勒出不同星系的整体轮廓与基本分类特征, 但其生成的部分图像存在明显的棋盘伪影、星系的形态学特征表达不够充分, 例如对于漩涡星系, 其旋臂细节模糊.

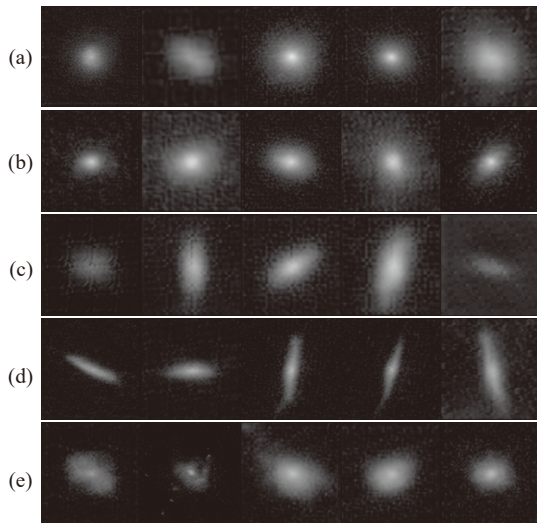


图 5 基于SAGAN的星系生成结果

Fig. 5 Galaxy generation results based on SAGAN

## 4.2 星系生成结果应用

为评估生成式增广数据对分类模型能力的提升效果, 并验证其有效性与合理性, 本文使用两种具有代表性的先进深度学习分类模型EfficientNetV2<sup>[23]</sup>和SwinTransformer<sup>[24]</sup>进行验证. EfficientNetV2分类模型在追求高精度的同时, 极大地优化了模型的效率和训练速度, 其复合缩放策略(Compound Scaling)与融合深度可分离卷积模块(Fused-Mobile Inverted Bottleneck Convolution, Fused-MBConv)使其对数据集的局部纹理与模式特征极为敏感, 因此EfficientNetV2的性能提升将

直接证明生成数据在细化局部特征方面的有效性. 同时SwinTransformer这一基于移位窗口的层次化视觉Transformer模型, 旨在通过其强大的全局上下文建模能力和长距离依赖关系捕捉能力, 使其性能高度依赖于数据集中整体结构与语义关联的丰富性与真实性. 若增广数据集能够同样提升SwinTransformer模型的精度, 则能够说明本文所生成的增广数据并非简单的纹理复制, 而是具有与真实星系观测相符的、连贯的全局形态学特征, 从而被具有全局视野的模型学习.

本文使用精确率(Precision)、召回率(Recall)和F1分数(F1-Score) 3项指标分别评价原始数据集和增广数据集对分类模型精度的影响. 精确率表示模型预测为正的样本中, 实际为正的比例; 召回率表示实际为正的样本, 被模型成功预测为正的比例; F1分数是精确率和召回率的调和平均数, 用于综合两者的表现, 该值越接近于1表示模型在精确率和召回率之间的平衡越好.

表3给出了3种不同参数量的EfficientNetV2模型对5种星系图像进行分类的定量评价结果, 表中加粗数值是为突出利用不同模型在两种数据集上训练后的较优性能. 在增广数据集上训练得到的3个分类模型, 大部分指标优于在原始数据集上得到的模型. 表4分别给出了使用不同参数量的EfficientNetV2模型和SwinTransformer模型对5类别星系图像分类总体精度评价的结果, 使用增广数据集训练得到的分类模型其所有精度指标均优于原始数据集下学习的模型. 在EfficientNetV2模型中, EfficientNetV2S模型的精度最高, Precision、Recall和F1分数相对于原始数据集下学习到的模型分别提升了4.01%、4.30%和5.06%; 在SwinTransformer模型中, SwinTransformerBase模型的精度最高, Precision、Recall和F1分数相对于原始数据集下学习到的模型分别提升了2.56%、3.75%和4.61%.

表 3 原始数据集与增广数据集分类模型不同类别精度验证

Table 3 Validation of per-class accuracy for the classification model on the original and augmented datasets

| Classification models | Galaxy class            | Original datasets |             |              | Augmented datasets |             |              |
|-----------------------|-------------------------|-------------------|-------------|--------------|--------------------|-------------|--------------|
|                       |                         | Precision         | Recall      | F1-Score     | Precision          | Recall      | F1-Score     |
| EfficientNetV2S       | Completely round smooth | <b>0.887</b>      | 0.95        | 0.917        | 0.873              | <b>0.97</b> | <b>0.919</b> |
|                       | In-between smooth       | 0.841             | <b>0.9</b>  | 0.869        | <b>0.926</b>       | 0.88        | <b>0.902</b> |
|                       | Cigar-shaped smooth     | 0.864             | 0.51        | 0.641        | <b>0.875</b>       | <b>0.7</b>  | <b>0.777</b> |
|                       | Edge-on                 | 0.707             | 0.92        | 0.8          | <b>0.764</b>       | <b>0.94</b> | <b>0.843</b> |
|                       | Spiral                  | 0.938             | <b>0.91</b> | <b>0.923</b> | <b>0.967</b>       | 0.88        | 0.921        |
| EfficientNetV2M       | Completely round smooth | 0.893             | 0.92        | 0.906        | <b>0.921</b>       | <b>0.94</b> | <b>0.930</b> |
|                       | In-between smooth       | 0.892             | 0.91        | 0.9          | <b>0.894</b>       | <b>0.93</b> | <b>0.911</b> |
|                       | Cigar-shaped smooth     | 0.904             | 0.38        | 0.535        | <b>0.959</b>       | <b>0.47</b> | <b>0.63</b>  |
|                       | Edge-on                 | 0.621             | <b>0.97</b> | 0.757        | <b>0.685</b>       | 0.96        | <b>0.8</b>   |
|                       | Spiral                  | <b>0.938</b>      | 0.91        | <b>0.923</b> | 0.895              | <b>0.94</b> | 0.917        |
| EfficientNetV2L       | Completely round smooth | 0.932             | <b>0.96</b> | <b>0.945</b> | <b>0.957</b>       | 0.91        | 0.933        |
|                       | In-between smooth       | 0.882             | <b>0.98</b> | <b>0.928</b> | <b>0.887</b>       | 0.95        | 0.917        |
|                       | Cigar-shaped smooth     | <b>0.957</b>      | 0.45        | 0.612        | 0.948              | <b>0.55</b> | <b>0.696</b> |
|                       | Edge-on                 | 0.653             | <b>0.98</b> | 0.784        | <b>0.697</b>       | 0.97        | <b>0.811</b> |
|                       | Spiral                  | <b>0.977</b>      | 0.87        | 0.92         | 0.95               | <b>0.96</b> | <b>0.955</b> |

表 4 原始数据集与增广数据集分类模型精度验证

Table 4 Validation of the classification model's overall accuracy on the original and augmented datasets

| Classification Models | Original Datasets |        |          | Augmented Datasets |              |              |
|-----------------------|-------------------|--------|----------|--------------------|--------------|--------------|
|                       | Precision         | Recall | F1-Score | Precision          | Recall       | F1-Score     |
| SwinTransformerTiny   | 0.862             | 0.844  | 0.833    | <b>0.867</b>       | <b>0.848</b> | <b>0.841</b> |
| SwinTransformerBase   | 0.876             | 0.854  | 0.846    | <b>0.898</b>       | <b>0.886</b> | <b>0.885</b> |
| SwinTransformerLarge  | 0.881             | 0.874  | 0.872    | <b>0.886</b>       | <b>0.876</b> | <b>0.873</b> |
| EfficientNetV2S       | 0.847             | 0.838  | 0.830    | <b>0.881</b>       | <b>0.874</b> | <b>0.872</b> |
| EfficientNetV2M       | 0.855             | 0.818  | 0.804    | <b>0.871</b>       | <b>0.848</b> | <b>0.838</b> |
| EfficientNetV2L       | 0.88              | 0.848  | 0.838    | <b>0.888</b>       | <b>0.868</b> | <b>0.862</b> |

## 5 结论与展望

本文成功构建了一个基于扩散模型的星系图像生成框架, 并系统性地验证了生成数据在天文数据增广中的有效性与应用潜力. 结果表明, 本文

所提出的扩散模型能够生成高度真实、细节丰富的星系图像. 通过对5个类别星系的生成结果与真实观测图像进行可视化对比, 展示了生成星系图像在星系形态结构、星光分布、旋臂特征及背景

噪声等细节上均具有出色的保真度. 同时给出了反向去噪过程的逐步可视化, 揭示了模型从随机噪声中逐步生成具有合理星系特征高保真图像的过程, 证明了生成过程的合理性和多样性.

本文将生成图像作为增广数据集, 能够明显提升星系形态学分类模型性能. 使用增广数据集训练所得分类模型的精确率、召回率及F1分数均优于原始数据集下学习的模型, 3项指标最大可分别提升4.01%、4.3%和5.06%. 这一结果表明本文所生成的增广数据集并非简单的噪声叠加, 而是有效补充了原始数据集的分布, 提供了更加多样化且具有判别性的特征样本.

本文所提方法为解决天文学中因观测数据有限、类别不平衡所带来的挑战提供了一个新颖、高效的解决方案. 未来可探索更先进的扩散模型变体, 以期在降低计算成本的同时, 生成更高分辨率、更多样化的星系图像. 通过对生成式人工智能模型的持续探索, 将为天文学研究提供强大工具, 帮助人类更深入地洞察宇宙奥秘.

### 参考文献

- [1] 中国学科及前沿领域发展战略研究(2021—2035)项目组. 中国天文学2035发展战略. 北京: 科学出版社, 2023: 1-78
- [2] Pan Z R, Qiu B, Liu C X, et al. *RAA*, 2024, 24: 045020
- [3] Lin Y F, Yu P C, Huang J C, et al. *A&A*, 2017, 600: A28
- [4] Buchanan J J, Schneider M D, Armstrong R E, et al. *ApJ*, 2022, 924: 94
- [5] 孔希阳, 窦江培. *天文学报*, 2024, 65: 14
- [6] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. International Conference on Learning Representations, Banff, Canada, April 14-16, 2014
- [7] He K, Zhang X, Ren S, et al. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2016: 770
- [8] Huang G, Liu Z, van der Maaten L, et al. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2017: 2261
- [9] Krizhevsky A, Sutskever I, Hinton G E. *Communication of ACM*, 2017, 60: 84
- [10] Shorten C, Khoshgoftaar T M. *Journal of Big Data*, 2019, 6: 60
- [11] Ma D G, Tang P, Zhao L J. *IGRSL*, 2019, 16: 1046
- [12] Zhong Z, Zheng L, Kang G L, et al. Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto, California USA: AAAI Press, 2020, 34: 13001
- [13] Inoue H. arXiv, 2018: 1801.02929
- [14] Zhang H, Cisse M, Dauphin Y N, et al. Mixup: Beyond Empirical Risk Minimization. International Conference on Learning Representations, Vancouver Canada, Apr 30 – May 3, 2018
- [15] DeVries T, Taylor G W. Dataset Augmentation in Feature Space. International Conference on Learning Representations, Toulon France, April 24-26, 2017
- [16] 刘牧云, 卞春江, 陈红珍. *计算机工程与应用*, 2024, 60: 244
- [17] Chen X, Mishra N, Rohaninejad M, et al. Proceedings of the 35th International Conference on Machine Learning. New York: Curran Associates, Inc, 2018, 80: 1364
- [18] 王礼, 李琳, 徐树振, 等. *激光与光电子学进展*, 2026, 63: 0217002
- [19] 郭全民, 张义康, 梁晨霄, 等. *激光与光电子学进展*, 2026, 63: 0637002
- [20] Zhang H, Goodfellow I, Metaxas D, et al. Proceedings of the 36th International Conference on Machine Learning. Online: PMLR, 2019, 97: 7354
- [21] Willett K W, Lintott C J, Bamford S P, et al. *MNRAS*, 2013, 435: 2835
- [22] 戴加明, 佟继周. *天文学进展*, 2018, 36: 384
- [23] Tan M, Le Q. Proceedings of the 38th International Conference on Machine Learning. Online: PMLR, 2021, 139: 10096
- [24] Liu Z, Lin Y, Cao Y, et al. 2021 IEEE/CVF International Conference On Computer Vision (ICCV 2021). New York: IEEE, 2021: 9992

# High-Resolution Galaxy Image Generation Based on Diffusion Models

ZHAO Rui-ni<sup>1,2</sup>    WANG Xing-feng<sup>1,2</sup>    LV Tao<sup>1,2</sup>

*(1 Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an 710119)*

*(2 Key Laboratory of Space Precision Measurement Technology, Chinese Academy of Sciences, Xi'an 710119)*

**ABSTRACT** Against the backdrop of rapid advancements in both data accumulation within large-scale sky survey projects and information technology, the automated classification of massive galaxy images using intelligent methods has become a cornerstone for understanding galaxy formation, evolution, and the cosmic environment. The accuracy of galaxy classification models is significantly correlated with the scale and balance of training samples, yet traditional data augmentation methods exhibit limited effectiveness in enhancing complex galaxy images. The study proposes a diffusion model-based algorithm for generating galaxy images, which produces galaxy images with realistic morphological features and textural details. The classification model trained using the augmented dataset demonstrates superior precision, recall, and F1-score compared to the model trained on the original dataset, with maximum improvements of 4.01%, 4.3%, and 5.06% in the three metrics respectively. These high-fidelity, clear generated images significantly improve the accuracy of galaxy morphological classification models. This process provides novel insights for future research on galaxy morphology classification in large-scale sky survey initiatives. Continued exploration of generative artificial intelligence models will provide powerful tools for astronomical research, enabling humanity to gain deeper insights into the mysteries of the universe.

**Key words** galaxies: general, methods: data analysis, techniques: image processing, techniques: diffusion model, techniques: generative model, techniques: data augmentation