

# 基于 MODIS 红外数据的云场景机器学习生成技术

王放<sup>1, 2</sup>, 戴聪明<sup>2, 3\*</sup>, 连文涛<sup>2, 3, 4</sup>, 陈舜平<sup>2, 3, 4</sup>, 张宇轩<sup>2, 3, 4</sup>, 魏合理<sup>2, 3</sup>

(1 安徽大学物质科学与信息技术研究院, 合肥 230601;

2 中国科学院合肥物质科学研究院安徽光学精密机械研究所, 中国科学院大气光学重点实验室, 合肥 230031;

3 先进激光技术安徽省实验室, 合肥 230037;

4 中国科学技术大学环境科学与光电技术学院, 合肥 230022)

**摘要:** 红外云场景模拟在光学工程的场景模拟中有重要的应用价值。为了快速、高质量地模拟云场景, 本文提出了一种基于 MODIS 红外数据的云场景机器学习生成技术: 使用 MODIS 红外数据对模型进行训练, 通过去噪扩散概率模型 (DDPM) 对随机数矩阵去噪产生低分辨率下不同通道的红外云场景图, 再使用微调后的超分辨模型 (Real-ESRGAN) 获得高分辨率下的云图像。将本文方法产生的云场景图与 MODIS 红外云图像数据和几种常见的云场景生成软件产生的云图像进行对比, 结果表明本文方法生成的云场景图更接近于真实的云场景, 在云场景模拟中有一定的应用价值。

**关键词:** MODIS、云场景生成、扩散模型、机器学习

**中图分类号:** TP391

**文献标识码** A

## Cloud scene machine learning generation technique based on MODIS infrared data

WANG Fang<sup>1,2</sup>, Lian Wentao<sup>2,3,4</sup>, Chen Shunping<sup>2,3,4</sup>, Zhang Yuxuan<sup>2,3,4</sup>, DAI Congming<sup>2,3\*</sup>, WEI Heli<sup>2,3</sup>,

(1 Institutes of Physical Science and Information Technology, Anhui University, Hefei 230601, China;

2 Key Laboratory of Environment Optics and Technology, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China;

3 Advanced Laser Technology Laboratory of Anhui Province, Hefei 230037, China;

4 School of Environmental Science and Optoelectronic Technology, University of Science and Technology of China, Hefei 230022, China)

**Abstract:** It is important to simulate the infrared cloud scene in optical engineering scene simulation. In order to

**基金项目:** 国家重点研发计划 (2019YFA0706004)

**作者简介:** 王放 (1997-), 男, 安徽宣城人, 硕士研究生, 主要从事计算机机器学习和大气光学方面的研究。E-mail: 936341603@qq.com

**导师简介:** 魏合理 (1965-), 男, 安徽宁国人, 博士, 研究员, 博士生导师, 主要从事大气辐射传输研究。Email: hlwei@aiofm.ac.cn

**收稿日期:** 年-月-日; **修改日期:** 年-月-日

**\*通信作者.** E-mail: cmdai@aiofm.ac.cn

simulate cloud scenes quickly with high quality , this paper presents a cloud scene machine learning generation technique based on MODIS infrared data: using random number matrix and the noise diffusion probability model (DDPM) to generate the infrared cloud scene image under different channels of low resolution, and then the cloud scene image of the high resolution is produced by the fine-tuning model (Real-ESRGAN). Comparing with MODIS data and several common cloud scene generation methods, the results show that the cloud scene image generated by this method is closer to the real cloud scene, and the model proposed in this paper has some application value in scene simulation.

**Key words:** MODIS, cloud image simulation, diffusion model, Machine learning

## 0 引 言

云是常见的自然现象,光学工程经常在有云场的环境下工作。天空背景下,不同波段下光谱的辐射情况十分复杂。在有云天空中,云雾会对光线辐射造成复杂的影响,研究有云天空辐射计算和图像仿真对气象科学和环境评估具有重要的意义,因此,需要进行云场景图像仿真。

通常情况下云图像生成主要分为三类方法:数值计算,分形算法和使用实际的云雾灰度纹理编织技术。数值模拟基于大气流体动力学的方法,将云看作自由大气中热力过程与动力过程的外部表现,通过求解Navier-Stokes方程展现云的动态效果<sup>[1]</sup>;分形方法通过控制算法的参数实现了一种符合客观事物的利用局部描述整体的方法,例如,20世纪90年代,美国美国空军菲利普斯实验室利用多尺度叠加算法生成特定天气背景下的云场景,即CSSM (Cloud Scene Simulation Model) <sup>[2][1]</sup>、杨春平提出的基于辐射传输理论和分形理论的云场景仿真<sup>[3]</sup>、成熟的商业动画软件Houdini提供的云雾仿真方法等;云雾灰度纹理编织技术<sup>[4]</sup>利用实测图像进行一系列的数据变换获得云模型,例如由美国光子协会开发的云空间建模软件CLDISM (Cloud Radiation Simulation) <sup>[1]</sup>。

数值计算方法通过数值模拟计算出具有物理意义的云的形成过程和云图像,但是这种计算计算量极大,很难做到快速地模拟出一幅云场景图像。分形方法虽然计算速度较快,但只能从形状上简单地近似云图形的纹理,细节上并不具有实际的物理意义。云雾灰度编制技术需要对云图像进行分类和叠加合成且在云图像纹理过渡处失真明显。由于数值计算方法计算代价大,速度较慢而后两种方法均不能产生真正符合物理规律的云图像,因此本文尝试利用基于卫星红外云图像,结合计算机图形学中通过随机数演化出物体合理分布的思想,通过机器学习,产生具有真正物理信息的云图像。

近年来,深度学习领域对于图像信息编解码技术日渐成熟,生成模型的兴起使得学习复杂规律成为可能。本文提出使用一种机器学习算法,对卫星测量的云图形数据集进行学习,获得了一个从随机数生成云场景的云图像生成模型。并对常见的两种计算机云图形仿真方法产生的云图像和本文提出的方法产生的云图像,在两种计算机图形学指标 (PSNR、SSIM)

和一种本文提出的深度神经网络分类判别器指标下分别进行对比。对比结果表明，本文提出的云场景仿真方法具有一定的应用价值。

## 1 数据来源

中分辨率成像光谱仪MODIS (Moderate Resolution Imaging Spectroradiometer) 是美国航空航天局NASA地球观测系统EOS (Earth Observation System) 项目开发的卫星传感器。MODIS卫星传感器能够提供地球表面的高分辨率观测数据，获得从可见光到长波红外36个波段的观测数据，涵盖了从大气到陆地和海洋的多个领域。其空间分辨率达到250-1000m<sup>[5]</sup>。

MODIS06产品包括云覆盖率、云底高度、云顶高度、云光学厚度等多个云参数。MODIS06产品中的云参数数据，提供了有关大气云层的关键信息，其空间分辨率为5km。云参数数据是从MODIS第29通道，第31-36通道的7个红外通道中反演得到的，这7个数据通道的波段范围是(8.400-8.700 $\mu\text{m}$ , 10.780-11.280 $\mu\text{m}$ , 11.770-12.270 $\mu\text{m}$ , 13.185-13.485 $\mu\text{m}$ , 13.485-13.785, 13.785-14.085 $\mu\text{m}$ , 14.085-14.385 $\mu\text{m}$ )。这7个通道用于云参数反演，能够反映云的特征。因此，本文选择这7个通道的辐射亮温数据作为机器学习训练的原始数据<sup>[6]</sup>。

本文利用MODIS06通道的辐射亮温数据，产生对应波段下的仿真云层。为了方便通过RGB图像显示，本文训练和生成涉及所有图像均为对应通道32-34的数据。图1是MODIS红外三通道（第32、33、34通道）合成并映射到RGB空间的图和该3个通道各自灰度的假彩色图<sup>[7]</sup>，从中可以明显地看出云的特征。

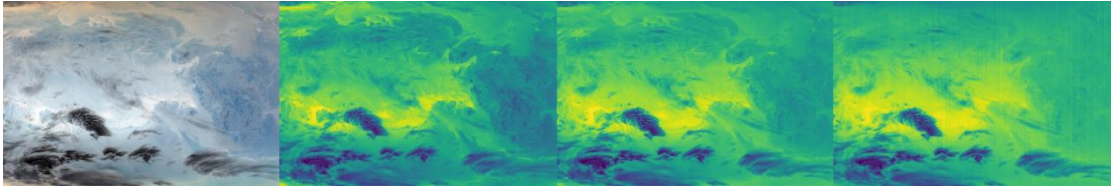


图 1 MODIS06 红外三通道数据合成 RGB 图与第 32、33、34 通道的假彩色图

Fig.1 Images of MODIS06 Data of channel 32-34 and their composite results

## 2 相关算法

### 2.1 去噪扩散概率模型

去噪扩散概率模型DDPM (Denoising Diffusion Probabilistic Model) 算法<sup>[8]</sup>中分为两个过程，前向的高斯扩散过程和反向去噪的逆过程，通过对前向扩散过程参数的学习，能够反向去噪去除增加的高斯噪声得到原始图像。高斯扩散过程是一种通过在原图像上增加高斯噪声直到原图像变为高斯分布图像的马尔科夫过程<sup>[8]</sup>。图2从右往左是对图加噪高斯扩散过程，从左往右是对扩散后得到的高斯分布的去噪过程。

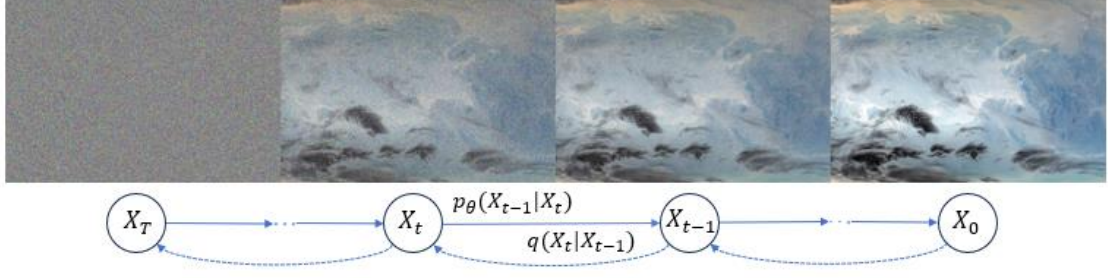


图2 高斯扩散和去噪过程

Fig.2 Gaussian Diffusion Processing and Denoising Processing

图2中， $q$ 代表扩散过程的状态， $p$ 代表去噪过程的状态，高斯扩散过程可以被描述为：

$$q(X_{1:T}|X_0) = \prod_{t=1}^T q(X_t|X_{t-1}) \quad (1)$$

将高斯扩散的结果描述为原始图像 $x_0$ 、方差相关的 $\alpha_t$ 和均值 $\mu_t$ 的函数：

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\alpha_t}x_0, (1 - \alpha_t)I) \quad (2)$$

在生成阶段，由于 $X_t$ 图像为已知均匀的、随机产生的高斯分布，反向去噪的过程也是马尔科夫过程，利用 $X_t$ 来生成 $X_{t-1}$ 的过程为：

$$X_{t-1} = X_t - k\epsilon_\theta(X_t, t) \quad (3)$$

$k$ 是一个与方差相关的参数，在状态 $\theta$ 下由DDPM网络预测得到的分布为 $\epsilon_\theta$ ， $X_T$ 仅与初始分布 $X_0$ 和加噪声的参数 $t$ （即加噪声的参数）相关。

DDPM的去噪过程实现了从随机产生的高斯分布中预测噪声并逐渐去除噪，刻画出原图像分布的过程。这一过程的逆过程相当于将一幅图像编码为多个采样参数 $t$ 对应噪声的序列。在DDPM深度学习网络架构的反向推理阶段，将会对状态量 $t$ 下得到的去噪结果进行学习和优化。通过调整网络结构使得网络能够推理加噪过程的采样参数 $t$ ，在反向去噪的过程中逐步去除参数 $t$ 对应的噪声就可以得到逐步逼近真实结果的图像。

在DDPM模型中，使用了注意力机制（attention）<sup>[9]</sup>得到了更大的全局视野、获得了更多像素间的互相关信息。但是，注意力机制的使用造成了更多的显存占用，在常规显卡上难以进行推理和训练<sup>[10]</sup>。所以，DDPM模型对应产生的图像尺寸较小。为了解决这一问题，在DDIM（Denoising Diffusion Implicit Model）模型<sup>[10]</sup>中，使用了VAE结构<sup>[11]</sup>。VAE模型是一个自编码器模型，通过对称的“编码器-解码器”结构可以将输入的数据变量编码为隐空间变量，这一过程类似于数据降维过程可以极大提升神经网络的数据抽象能力和处理速度。之后在隐空间内进行处理后，通过解码器将数据恢复到原先的数据格式。VAE的过程可以概括为以下两步：

$$h = g_{\theta_1}(X) = \sigma(w_1X + b_1); \hat{x} = g_{\theta_2}(h) = \sigma(w_2h + b_2) \quad (4)$$

其中， $h$ 是编码后的隐空间变量， $g_{\theta_1}$ 和 $g_{\theta_2}$ 为多层感知机组成的编码器结构和解码器结构。

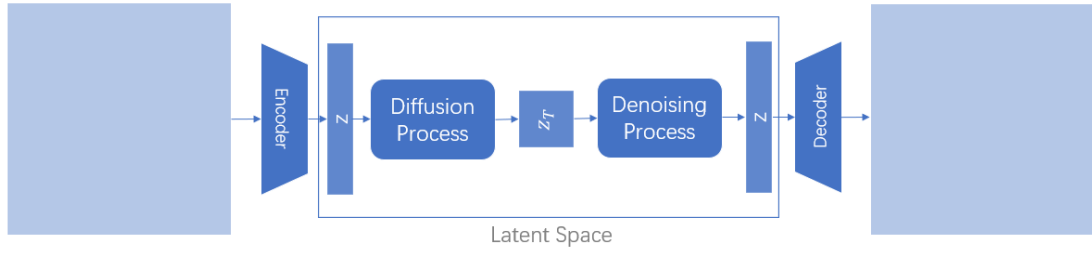


图 3 DDIM 模型结构

Fig.3 DDIM model construction

图3是DDIM模型结构。DDIM分为三个部分，VAE-encoder将原图像压缩降秩到低维度的潜空间 $h$ ，再对潜空间进行DDPM中的高斯扩散和去噪过程，最后通过VAE-decoder部分将潜空间中编码的信息提升到原本的维度 $\hat{x}$ 。此过程下，原本的正向扩散过程变为：

$$q(X_{1:T}|X_0) = q_{Encoder}(x_0|X_0) \prod_{t=1}^T q(x_t|x_{t-1}) \times q_{Decoder}(X_T|x_T) \quad (5)$$

最终，对已知随机产生的一个 $[3 \times H \times W]$ 的高斯分布矩阵，最终降噪产生的结果为一个趋近于真实分布的 $[3 \times H \times W]$ 分辨率图像。

## 2.2 超分辨率模型

盲超分辨率（Blind Super-resolution）旨在恢复遭受未知和复杂退化的低分辨率图像<sup>[13]</sup>。Real-ESRGAN<sup>[12]</sup>在ESRGAN网络<sup>[14]</sup>的基础上，吸取了BSRGAN的经验，认为对数据样本使用退化过程（degeneration，包括图像超分辨失真的几种情况：模糊、下采样、噪声、伪影和JPEG图压缩组成）可以让模型学习到如何在提高分辨率的同时去除图中的失真部分。在BSRGAN的基础上，Real-ESRGAN除了模型结构不同外，使用的训练数据全部来自合成数据，因此可以学习到几种失真情况的组合，即“高阶”退化情况<sup>[12]</sup>。

图4为Real-ESRGAN模型结构分为生成器和鉴别器两部分。Real-ESRGAN采用了对抗神经网络的网络设计，在训练过程中使用退化过程处理得到退化后的数据，输入到残差结构（RRDB block）和上采样结构形成的生成器中进行超分辨过程，得到的输出再次输入到鉴别器（U-net discriminator）中得到评价分数与初始退化之前的图像进行对比打分，并且通过分数反向调整生成器。此外，Real-ESRGAN还提出了使用U-net作为判别器来实现对更加复杂的“高阶”退化情况进行评判<sup>[12]</sup>，大大提升了网络训练的效率。

通过对这些退化数据的学习，Real-ESRGAN可以将一幅具有退化情况的图像进行超分辨得到一幅去除退化情况并且提升分辨率的图像。在本文提出的网络结构中，Real-ESRGAN部分代替DDIM中VAE结构的解码器部分，将DDPM结构运行得到的低分辨率图像通过超分辨得到近似MODIS06数据集中数据分辨率的云图像。

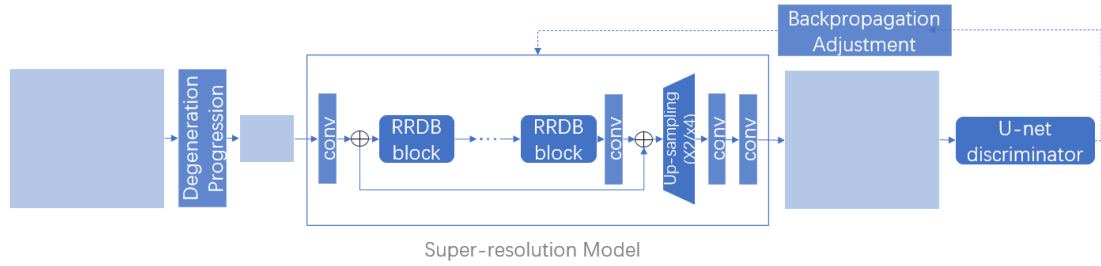


图 4 Real-ESRGAN 模型

Fig.4 Real-ESRGAN model construction

### 3 CLOUD 云场景生成算法

本文提出一种DDPM+微调Real-GAN的云图像生成技术，将之命名为：CLOUD: A Cloudy Image Overlay with Using Diffusion。在这一模型中，使用随机数矩阵生成云图像，同时对低分辨率云图像的超分辨处理，使之能够与MODIS卫星图像的分辨率相近。最终呈现出使用随机数产生的近似现实中MODIS卫星观测到的云图像。

对于大样本数量的分类，在多轮训练下模型性能会随着attention头数和卷积层CNN（convolutional neural network）的通道数增加而改进，但是对于小样本来说，神经网络分类和泛化的能力并不会随之增加，反而会在更多轮的训练下产生过拟合情况。如果使用训练过多种类样本的预训练模型，对小样本学习会产生很多的冗余信息，沉积在网络中。因此，如何取舍预训练和样本训练之间的影响、对模型进行数据驱动的训练十分重要。本文使用的数据集为基于MODIS06数据集筛选得到的3个红外通道的辐射亮温数据集。使用真实数据进行训练和微调，生成过程产生的结果会趋近于真实分布。

在DDIM模型中，VAE编解码时，损失函数优化的惩罚项会造成多重特征之间的缠结（entangling），这种缠结会使得新的分布趋近于高斯分布从而产生伪影和模糊<sup>[15]</sup>。当使用DDIM模型从零开始训练时，需要将模型的各个部分分开训练，且要尽量减少原始数据中的伪影和模糊。由于使用VAE仅是为了降低原本图像的维度将信息编码压缩到潜空间中进行处理。这种编解码信息的过程和图像超分辨（super-resolution）过程有相似之处。对此，本文提出一种将DDIM模型求解目标转换为DDPM+超分辨模型的思路。即在DDPM阶段，使用CNN或其他下采样方法降低图像的维度，采样其中的重要信息，之后在上采样阶段使用微调后的超分辨模型Real-GAN。

图5为CLOUD模型的主要结构，该模型主要分为DDPM部分和超分辨部分。DDPM部分为一个U-net网络<sup>[16]</sup>，每层都分为卷积层（convolution）和注意力层（attention）和下采样/上采样层（sampling），同时层间还有残差网络结构作为连接（concatenate）。超分辨部分为一个Real-ESRGAN网络，仅利用数据集微调了网络参数，结构见图4。其中，DDPM模块用于产生低分辨率云图像内容，通过对于MODIS 7个通道数据的学习调整模型各层的参数，并分



步骤逐渐刻画出云图像；微调后的Real-ESRGAN模块对低分辨率图像进行信息补全和细节优化，从而提升分辨率。

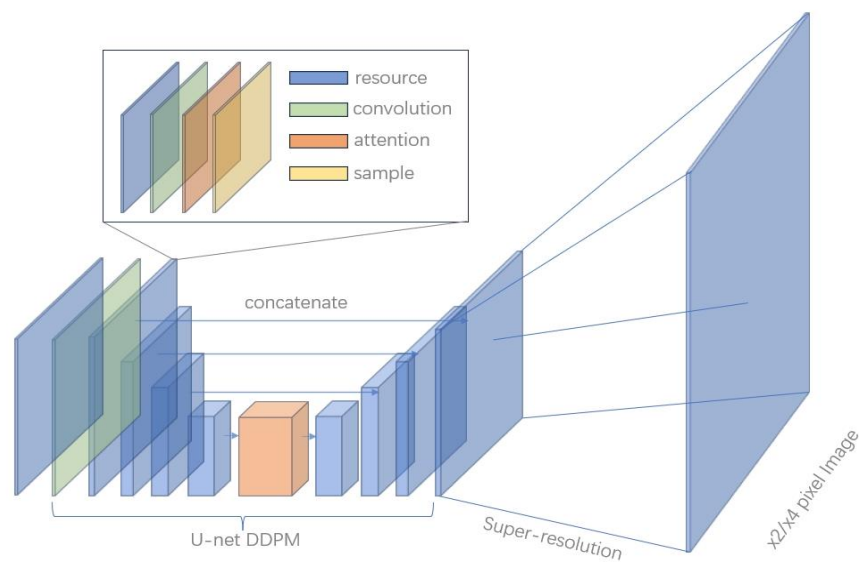


图 5 CLOUD 模型结构

Fig.5 CLOUD model construction

图6为使用此模型产生图像过程：①首先产生一个的高斯分布的随机数矩阵，②之后利用训练过程中DDPM模块的U-net网络学习到的扩散过程参数，对不同通道的随机数矩阵进行去噪（在实验部分，低分辨率指64×64矩阵元，高分辨率指256×256矩阵元）。③重复步骤②共T次获得的低分辨率图像矩阵；④最后通过使用MODIS06数据提前微调训练好的Real-GAN模型，产生×2/×4倍分辨率的图像。

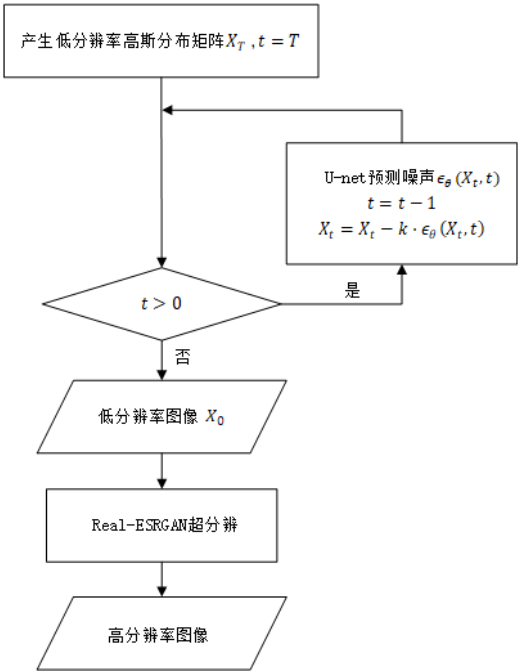


图 6 CLOUD 云图像生成流程

Fig.6 Cloud image generated procession

损失函数用来衡量计算结果和真实值之间差距，CLOUD模型的损失函数为：

$$Loss = \sum_{t=0}^T \frac{1}{2\sigma^2} [\|\epsilon - \epsilon_{\theta}(x_t, t)\|_2^2] + \lambda(-\zeta(x_r, x_f)) + \eta L_1 \quad (6)$$

即两部分损失值之和，DDPM阶段产生的损失和在超分辨率阶段产生的损失。其中， $\epsilon$ 为参数 $t$ 下的高斯分布， $\epsilon_{\theta}$ 为DDPM预测出需要去除的高斯分布噪声， $\zeta$ 为Real-ESRGAN在退化后图像 $x_f$ 和真实图像 $x_r$ 之间的PSNR差距， $L_1$ 为正则化参数， $\lambda$ 、 $\eta$ 为相关系数。通过对参数 $t$ 的优化和对超分辨率模型参数进行微调，可以优化损失值，从而达到优化模型性能的目的。

图7为使用DDIM和CLOUD分别训练后产生出的云图像。DDIM模型产生的图7(a)细节丰富，但是由于引入了VAE对图像进行分辨率升降，所以在图的边缘和本身透过率较低的云团部分具有颗粒感，处理效果不佳。而CLOUD模型产生的图7(b)由于使用Real-ESRGAN模型提升分辨率，在提升分辨率的同时对可能存在的图像退化进行了处理，所以不会产生图7(a)中的颗粒感和明显的物体边缘失真的问题。同时，CLOUD模型产生的云图像的颜色更加鲜艳，物体边缘过渡的部分更加清晰。

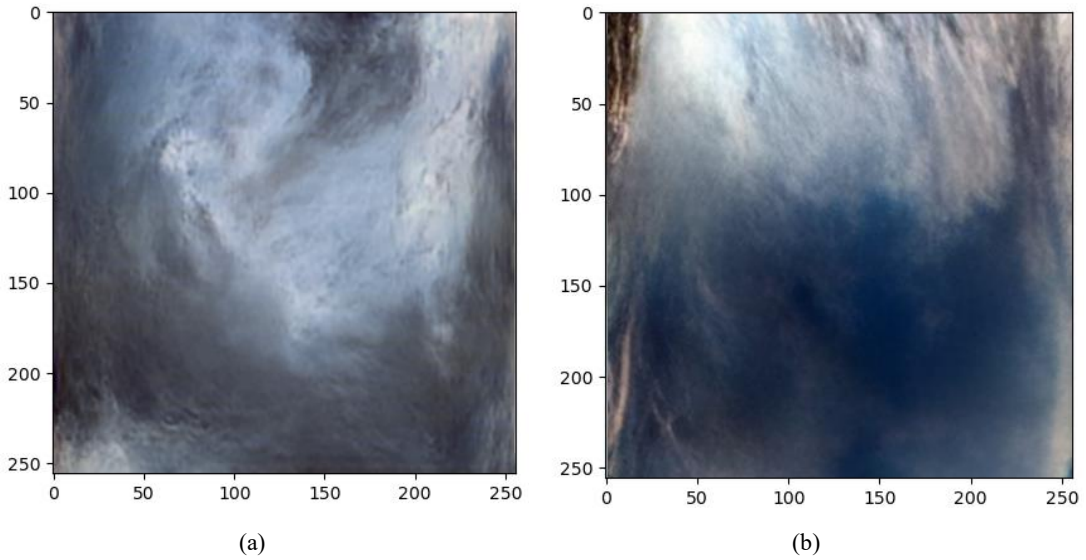


图 7 (a):DDIM 产生的云图像 (b):CLOUD 产生的云图像

Fig.7 Cloud image generated by CLOUD model

采用CLOUD模型的DDPM+微调Real-GAN方法有以下优势：①利用DDPM模型的生成能力，仅对1000张MODIS06样本数据进行学习就可以达到通过对随机数矩阵去噪产生近似真实云图像的目的；②将DDPM从零训练，使模型仅面向MODIS红外数据集，不会使大量的冗余信息沉积到网络中，相较于直接精调预训练DDIM的效果更好；③类似DDIM模型使用编解码信息的方式，让DDPM在不极大提升显存占用的情况下，产生了足够多的信息；④使用Real-GAN微调模型依然保持了大模型本身的泛化能力，可以较好地提取和丰富DDPM中产生的信息；⑤使用这种方法可以拆分模型进行训练，减小单次训练的显存压力同时使得微调变得简单，单机显存需求仅3Gb；⑥相较于直接使用DDPM计算高分辨率图像，使用这种方



法的训练成本更低。

图8为DDPM部分的训练误差折线图，收敛速度较快。需要注意的是收敛阶段模型的泛化能力不能仅由图像的一阶误差判定。

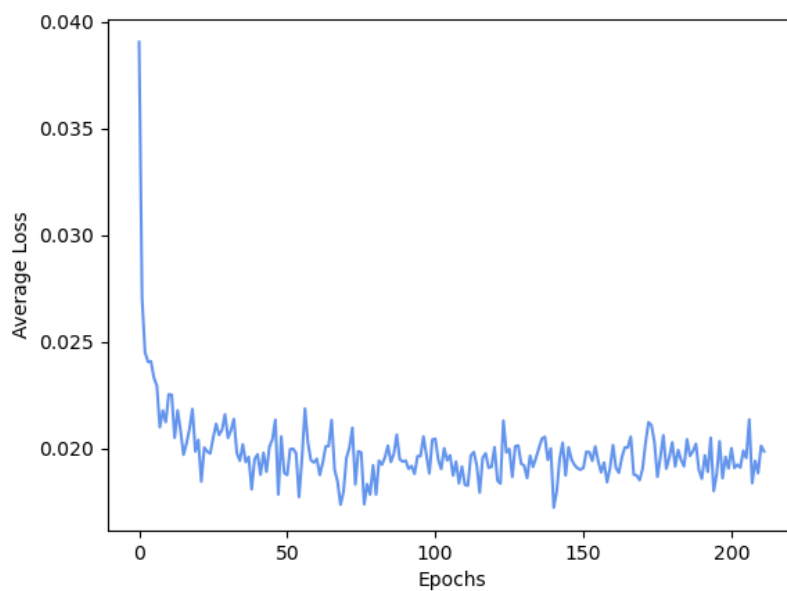
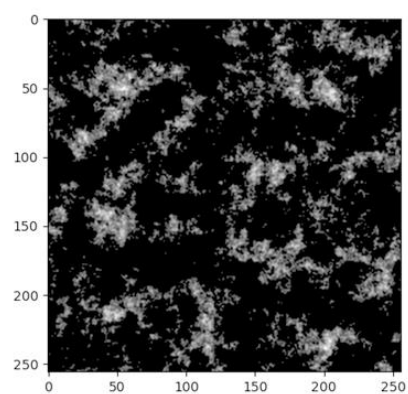


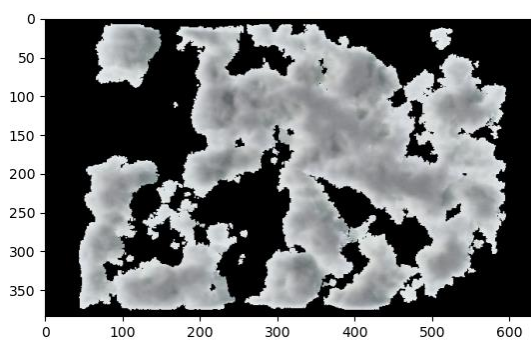
图 8 DDPM 部分训练的绝对误差折线图

Fig.8 DDPM model training metrics chart

### 3 实验和结果分析



(a)



(b)

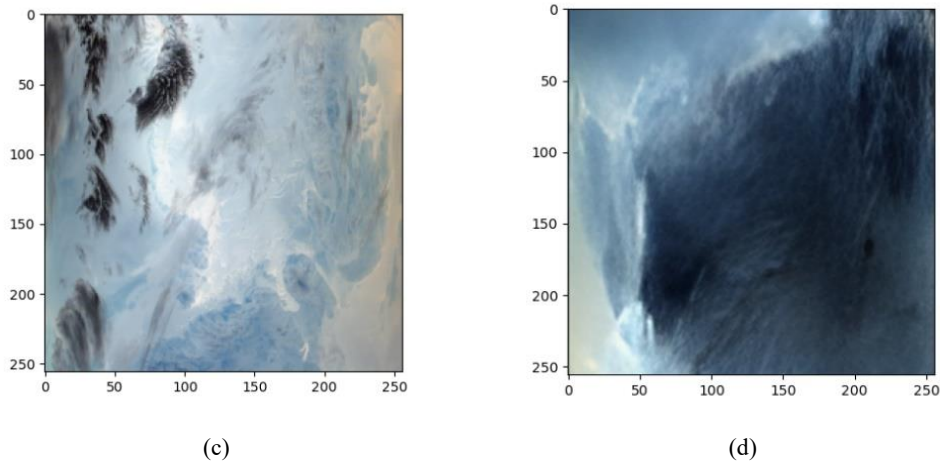


图9 利用 perlin 噪声分形产生的云图像 (a)、Houdini 软件产生的云图像(b)、MODIS06 云图像(c)、CLOUD 产生的云图像(d)

Fig.9 Cloud image: (a) Perlin Noise image; (b) Houdini business software; (c) MODIS06 data; (d) CLOUD model generation

图9的 (a) 和 (b) 为两种常见方法产生的云图像，图9 (a) 使用Perlin噪声<sup>[18]</sup>分形叠加产生，首先产生一个 $256 \times 256$ 随机数矩阵，之后分为 $8 \times 8$ 个格子，通过计算格子中每个点相对于四个角点的梯度并与格子中每个点计算点积，再对所有格子进行三次插值，对以上步骤重复2~4次后，通过数个通过以上方法得到的矩阵叠加和阈值筛选产生图像；图9 (b) 使用Houdini软件 ([www.sidefx.com/](http://www.sidefx.com/)) 提供的三维Perlin噪声，再对z轴方向增加环境光并拍摄产生图像。MODIS06三通道合成云图像和本文CLOUD产生的云图像为图9的 (c) 和 (d)。可以观察到图8中，图8(a)和图8(b)本质上是使用噪声叠加产生云图像，都和真实的MODIS红外三通道合成的RGB图8(c)有一定差别：这些图像虽然宏观上具有云的结构但是实际上每个像素点附近的分布关系都与真实的云分布不同，不具有云团和云雾的结构和过渡并且有噪声部分，不具有真实的物理意义。图9 (d) 则没有这些问题，更接近于图9 (c)。

CLOUD产生的数据 $data$ 为归一化后的辐射亮温，每个像素值的存储格式为32位浮点数。通过映射，能够将生成图像对应通道的灰度值 $data$ 映射为具有物理意义的辐射亮温值<sup>[19]</sup>。如果与MODIS卫星定标后通道 $i$ 的辐射亮温数据区间 $[T_{min}(i), T_{max}(i)]$ 对比，即可从 $[0,1]$ 区间映射得到通道 $i$ 的辐射亮温值 $Brightness\ Temperature(i)$ ：

$$Brightness\ Temperature(i) = data \times T_{max}(i) + (1 - data) \times T_{min}(i) \quad (7)$$

对于MODIS06数据，MODIS各个探测通道提供的原始数据 $DN(i)$ 可以映射到辐射亮温（单位：K），其校正系数 $k = 0.01$ 和校正值 $add\_offsets = -15000$ <sup>[19]</sup>：

$$Brightness\ Temperature(i) = k \times (DN(i) - add\_offsets) \quad (8)$$

表1为训练所使用的1120幅 MODIS06辐射亮温数据，在对应第29、31~36通道下的最大值和最小值。图10中右图为3个通道下DDPM模型产生的灰度图与对应合成的彩色图，左图为CLOUD产生的高分辨率下的云图像。

Channels	Minimum	Maximum
29	3018	17163
31	3148	17610
32	3111	17538
33	3512	14058
34	3803	12318
35	4259	11062
36	4599	8952

表 1 校正后训练数据集中 7 个通道辐射亮温 $DN$ 值的最大值、最小值

Table.1 The maximum and minimum values of radiation brightness temperature's Digital Number for different channels in the training dataset

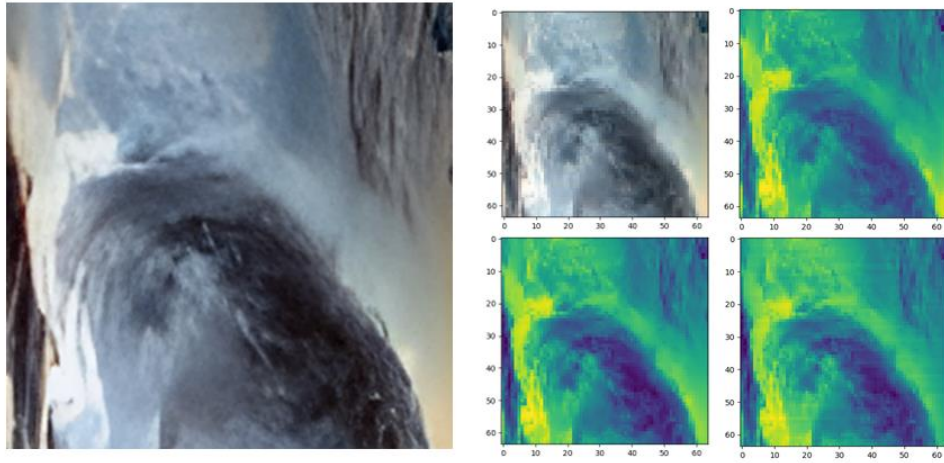


图 10 在不同通道上 CLOUD 的表现，左图：CLOUD 产生的云图像；右图：在 MODIS 32-34 通道下的表现  
Fig10. Cloud image generated by CLOUD at different channel, Left image: Cloud image generated by CLOUD;  
Right image: Performance at MODIS 32-34 channels

本文提出的方法产生的云图像具有与MODIS06相同的定标区间、相似的物理属性，还具有近似真正云图像的细微结构。

在实验的过程中发现，目前对于云图像的评价通常只针对单一云块的物理机制进行分析。使用的物理指标，包括：对云形状的定性描述、云的辐射亮温、光学厚度等反演出的云高、云粒子种类等物理信息。这些指标只判定一个像素点处云的状态，与图像识别和云分类问题中对于图像结构和像素点值的要求不同。因此，为了综合判断一副图像是否是云，除了人的主观判断以外，采用了以下几种判别标准，分别为两种结构化参数峰值信噪比（PSNR）、结构相似性（SSIM）<sup>[21]</sup>和神经网络判别器。

$$MSE(X, Y) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (X_{i,j} - Y_{i,j})^2 \quad (9)$$

$$PSNR = 10 \log_{10} \left( \frac{(2^n - 1)^2}{MSE} \right)$$

公式(9)为PSNR计算公式,其中 $H, W$ 为图高宽,  $X$ 为生成图像,  $Y$ 为MODIS图像。 $n$ 为灰度阶数通常取8。

$$SSIM = L \times C \times S \quad (10)$$

公式(10)为SSIM计算公式,  $L, C, S$ 分别为生成图像与MODIS图像之间的照明度、对比度、结构<sup>[21]</sup>, 这三个参数与两个图像矩阵的均值、方差和协方差有关。对于这两种指标参数, PSNR分数越高说明两幅图的相似程度越高; SSIM为亮度对比度和结构相似程度的组合, SSIM分数越接近于1说明两幅图的相似程度越高。表2中, 各方法的PSNR和SSIM指标值均为此种方法产生图与MODIS06图进行相应指标计算后的结果。

Name	PSNR	SSIM	Neural Discriminator
Fractal image	53.910	0.9539	1.0613
Houdini image	56.418	0.9810	1.0214
MODIS06	60.078	0.9913	0.9919
CLOUD	57.850	0.9853	0.9865

表 2 不同方法的指标结果

Table.2 Different metrics results of different method

由于是否为云图像难以以数学方法判断,所以本文提出了一种使用神经网络作为分类判断的方法。在神经网络作为判别器的情况下,首先利用神经网络训练一个判断是否为MODIS06数据集的网络,之后将生成的图像作为输入,最后输出判断是否为MODIS06数据的分数。这个打分网络使用一个VGG16网络作为判断的骨干网络,用1000张作为正例的MODIS06数据和2000张作为反例的其他航拍图片作为数据训练出的分类网络。该网络的输出值越接近于1表明输入数据与真实MODIS图像的相似程度越高,相反接近0则表示输入数据与真实MODIS图像无关。通过判别器计算的结果对比,使用分形噪声方法生成的图片输出分数均超过1,即图像具有云层的轮廓结构但是云层结构的边缘过渡因为噪声产生了失真情况。本文提出的CLOUD模型产生的图像的PSNR和SSIM分数更高,神经网络判别器打分更接近于MODIS06数据。

综合对比可以得出结论, CLOUD生成的云图像与真实的MODIS06云图像数据具有相似性;同时, CLOUD可以产生7个通道、不同波段下的更加贴近真实的云层分布。

## 4 结论

本文研究和提出了一种基于MODIS06数据和深度扩散神经网络产生云图像的方法。经

过一系列实验，确定了使用扩散模型和超分辨模型产生云图像的架构。通过模型对MODIS红外数据集进行学习，最终获得了一个从随机数矩阵生成云场景的云图像生成模型。这种方法对随机数矩阵进行逐步去噪产生低分辨率下的图像，之后通过超分辨模型得到高分辨率下的云场景图。

相较于传统的产生云图像的方法，这种新方法拥有更加接近真实云场景辐射亮温的像素分布，使产生的云图形具有一定的物理意义。本文提出的模型可以分别产生7个通道下具有差异性的灰度图像，文中为了方便通过RGB图像显示，本文训练和生成涉及所有图像均为对应通道32-34的数据。使用这种方法，可以对有云天空进行更精细、更真实的仿真，为对有云天空中的辐射传输等研究提供可靠的参考。

本文提出的产生低分辨率云图像的部分使用模型为DDPM，而输入去噪模型的图像矩阵来自随机数产生的高斯分布矩阵，所以难免与真实的云图像分布略有差异。想要解决这种差异问题，需要对去噪过程中使用交叉注意力机制（cross-attention）<sup>[22]</sup>注入实际云层的物理特征信息，使得图像保留更多的物理信息作为生成图像的辅助<sup>[23]</sup>。这就需要利用物理特征对云层进行描述和分类。能够更好地对云层进行分类，将会为图像注入相关的位置、形状、厚度、温度、云顶高度、粒子种类、附近地形等物理信息，做到更准确地对云层情况进行预测。可以在本文提出方法的基础上，使用物理参数控制，进一步地生成更加符合真实云场景的图像。

## 参考文献：

- [1] 邱航，杨珂，陈瑜等. 云的真实感模拟技术综述[J]. 计算机科学技术，2011，6:14-19.  
QIU Hang, YANG Ke, CHEN Yu, CHEN Lei-ting. Survey on Realistic Simulation of Cloud[J]. Computer Science. 2011,6. p.14-19.
- [2] 潘皓. 云背景红外辐射计算及图像仿真[D]. 西安电子科技大学，2016  
Pan Hao. Calculation and Image Simulation of Infrared Radiation of Cloud Background[D]. Xidian University. 2016
- [3] 杨春平，吴健等. 基于辐射传输理论和分形理论的云场景仿真[J]. 强激光与粒子束，2007，7：1085-1088.  
YANG Chun-ping, WU Jian, WEI Ling, LENG Jie. Cloud scene simulation based on radiative transfer theory and fractal theory[J]. High Power and Laser Particle Beams. 2007,7. p. 1085-1088.
- [4] HARRIS, Mark J., et al. Simulation of cloud dynamics on graphics hardware[C]. In: Proceedings of the ACM SIGGRAPH/EUROGRAPHICS conference on Graphics hardware. 2003. p. 92-101.

- [5] 曹亚楠, 魏合理等. 利用MODIS资料模拟计算水云大气红外辐射特性[J]. 光子学报, 2014, 6: 1001, 1-8.
- Cao Yanan, Wei Heli, etc. Simulation of Atmospheric Radiative Properties and IR Band under Water Clouds Based on MODIS Data[J]. ACTA PHOTONICA SINICA. 2014, 6: 1001, 1-8.
- [6] 李锡祥 麻金继 梁晓芳. 基于BP神经网络进行云相态识别方法的研究[J]. 大气与环境光学学报, 2010, 5(4): 299-304.
- LI Xi-Xiang, MA Jin-Ji, LIANG Xiao-Fang. Retrieving the Cloud Phase Based on BP Neural Network[J]. Journal of Atmospheric and Environmental Optics, 2010, 5(4): 299-304.
- [7] 程伟 杨世植 文奴 崔生成. 基于小波变换的高分辨率遥感图像复原算法实现[J]. 大气与环境光学学报, 2015, 10(5): 401-407.
- CHENG Wei, YANG Shi-Zhi, WEN Nu, CUI Sheng-Cheng. Implement of High Resolution Remote Sensing Image Restoration Algorithm Based on Wavelet Transform[J]. Journal of Atmospheric and Environmental Optics, 2015, 10(5): 401-407.
- [8] HO, Jonathan; JAIN, Ajay; ABBEEL, Pieter. Denoising diffusion probabilistic models[J]. Advances in neural information processing systems, 2020, 33: 6840-6851.
- [9] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need[J]. Advances in neural information processing systems, 30.
- [10] Rombach Robin, Blattmann Andreas, Lorenz Dominik, Esser Patrick, Ommer Björn. High-Resolution Image Synthesis with Latent Diffusion Models[C]. Computer Vision and Pattern Recognition 2022. arXiv:2112.10752.
- [11] BALDI, Pierre. Autoencoders, unsupervised learning, and deep architectures[C]. In: Proceedings of ICML workshop on unsupervised and transfer learning. JMLR Workshop and Conference Proceedings, 2012. p. 37-49.
- [12] Wang Xintao, Xie Liangbin, Dong Chao, Shan Ying. Real-esrgan: Training real-world blind super-resolution with pure synthetic data[C]. Proceedings of the IEEE/CVF international conference on computer vision. 2021: 1905-1914.
- [13] BIANCO, Simone, et al. On the use of deep learning for blind image quality assessment[J]. Signal, Image and Video Processing, 2018, 12: 355-362.
- [14] Zhang Kai, Liang Jingyun, Luc Van Gool, Radu Timofte. Designing a Practical Degradation Model for Deep Blind Image Super-Resolution[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021, pp. 4791-4800



- [15] Chen Ricky T. Q. and Li Xuechen and Grosse Roger and Duvenaud David. et al. Isolating sources of disentanglement in variational autoencoders[J]. Advances in neural information processing systems, 2018, 31.
- [16] OKTAY, Ozan, et al. Attention u-net: Learning where to look for the pancreas[C]. Conference on Medical Imaging with Deep Learning (MIDL 2018). arXiv preprint arXiv:1804.03999, 2018.
- [17] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation[C]. In Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.
- [18] Perlin. An Image Synthesizer[C]. Special Interest Group on Computer Graphics, 1985, 19:287-296
- [19] Lyapustin, A., Wang, Y., Korkin, S., & Huang, D. (2018). MODIS collection 6 MAIAC algorithm[J]. Atmospheric Measurement Techniques, 11(10), 5741-5765.
- [20] Platnick, Meyer, King, Wind, et. al, July 2018. Cloud Optical Properties - Collection 6 & 6.1 Users Guide[M]. <https://atmosphere-imager.gsfc.nasa.gov/products/cloud/documentation>
- [21] Hore, A., & Ziou, D. (2010, August). Image quality metrics: PSNR vs. SSIM[C]. In 2010 20th international conference on pattern recognition (pp. 2366-2369). IEEE.
- [22] Hertz, A., Mokady, R., Tenenbaum, J., Aberman, K., Pritch, Y., & Cohen-Or, D. (2022). Prompt-to-prompt image editing with cross attention control[J]. arXiv preprint arXiv:2208.01626.
- [23] Lee, K. H., Chen, X., Hua, G., Hu, H., & He, X. (2018). Stacked cross attention for image-text matching[C]. In Proceedings of the European conference on computer vision (ECCV) (pp. 201-216).