

---

# CuMoLoS-MAE: 一种用于 遥感数据重构的掩码自编码器

---

Anurup Naskar<sup>1</sup> Nathanael Zhixin Wong<sup>1</sup> Sara Shamekh<sup>1</sup>

<sup>1</sup> Courant Institute of Mathematical Sciences, New York University  
{an4462,nw2648,ss18284}@nyu.edu

## Abstract

来自多普勒雷达、雷达和辐射计等遥感仪器的精确大气廓线经常受到低信噪比 (SNR) 门限、距离折叠和虚假不连续性的影响。传统的间隙填补方法会模糊细尺度结构，而深度模型缺乏置信度估计。我们提出了**铜钼洛斯-MAE**，一个 **铜riculum-Guided Monte Carlo Stochastic Ensemble Masked Autoencoder**，旨在 (i) 恢复细尺度特征如上升气流和下降气流核心、切变线和小旋涡，(ii) 学习大气场的数据驱动先验知识，以及 (iii) 量化像素级别的不确定性。在训练过程中，CuMoLoS-MAE 使用了掩码比率课程，迫使 ViT 解码器从逐渐稀疏的上下文中进行重建。在推理时，我们通过蒙特卡洛方法对随机掩码实现进行近似后验预测，多次评估 MAE 并聚合输出以获得后验预测均值重建  $\bar{X}$  以及一个精细解析的逐像素不确定性图  $\sigma_X$ 。除了高保真度的重建外，这一新颖的基于深度学习的工作流程还支持增强对流诊断、实时数据同化，并改进长期气候再分析。

## 1 介绍

理解气候变化对极端事件的影响需要从多普勒雷达、雷达或辐射计等遥感仪器中获得准确且连续的测量数据。然而，这些数据通常会受到缺失或损坏返回值的影响，在可靠地用于下游应用之前必须填补这些空缺。经典的填隙程序（如滑动窗口平均滤波器 [1]）会使关键的小尺度特征（如剪切线和上升气流核心）变得模糊。近年来，深度学习方法（例如变分自动编码器 [2]）的进步可以恢复更清晰的结构，但当这些方法用于下游同化和警报框架时，它们无法提供衡量重建不确定性的信息。

为了解决这些缺点，我们提出了**铜钼洛斯-MAE**，一种 **铜riculum-Guided Monte Carlo Stochastic Ensemble Masked Autoencoder** 的新方法，该方法包含三个核心机制：(1) 课程掩码 [3] 以稳定训练并鼓励从更稀疏的上下文中进行重建，(2) 基于掩码自编码器的小块 [4] 以捕捉精细结构和中尺度动态，以及 (3) 蒙特卡洛 [5] 集合以生成最终重建图像的每像素不确定性图。在本论文中，我们使用了大气辐射测量 (ARM) 提供的南部大平原 (SGP) 的多普勒激光雷达测量数据 [6] 作为案例研究，并展示了我们对垂直速度场的重建及其对应的不确定性量化。

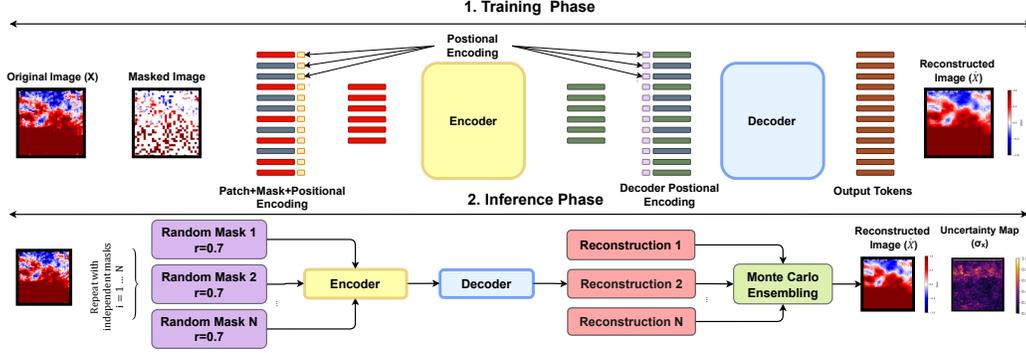


图 1: 从多普勒激光雷达的时间-高度数组中, 我们形成  $64 \times 64$  图像。在训练过程中, 我们随机隐藏一部分 patch tokens (掩码比例从 0.5 增加到 0.7), 添加位置编码, 将可见的 tokens 通过 ViT 编码器传递, 并使用轻量级解码器重构整个场。损失仅计算在隐藏像素上。在测试时, 对于每个未见过的 patch, 我们绘制 50 个独立的随机掩码并运行相同的流程以生成多个重建结果。然后我们将这些重建结果取平均值以获得单次最优去噪估计, 并使用集合的逐像素差异作为不确定性图。

## 2 数据

我们使用了来自 ARM SGP 站点的多普勒雷达垂直速度数据, 这些数据以 NetCDF 格式存储为时间-高度数组, 包含 320 个范围门, 间隔为 30 米, 每个范围门代表一个固定垂直层的平均返回值。重点关注最低的 64 个范围门 (1.92 公里), 我们提取了非重叠的  $64 \times 64$  块作为训练和评估的输入。模型在 2011 年 6 月 1 日至 9 日的数据上进行训练, 并在未见过的一天, 即 2011 年 6 月 15 日的数据上进行评估。

## 3 方法论

### 3.1 CuMoLoS-预测架构

受掩码自编码和随机集成方法启发, 用于不确定性量化, CuMoLoS-MAE 框架包括两个阶段, 如图 1 所示:

- **微小块化的 MAE 与课程安排:** 每个  $64 \times 64$  多普勒速度切片被分割成  $2 \times 2$  微块。一个 12 层的 ViT [7] 编码器处理可见标记, 而一个 4 层解码器重构该场。在训练过程中, 掩码比例  $r$  在前 5 个周期从 50% 开始, 然后通过余弦插值到第 30 个周期的 70% 并保持不变; 优化使用隐藏像素的掩码 - 均方误差。

- **蒙特卡洛集成:** 在推理时, 每次前向传递都会生成一个新的随机掩码, 并重复掩码、编码和解码过程  $N = 50$  次。然后聚合预测集合以计算平均值和逐像素标准差:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^N \hat{X}^{(i)}, \quad \sigma_X = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{X}^{(i)} - \bar{X})^2} \quad (1)$$

这产生了高保真去噪重建  $\bar{X}$  和逐像素不确定性估计  $\sigma_X$ 。

### 3.2 训练程序

数据通过应用信噪比滤波器 (强度  $\geq 0.005$ ) 进行预处理, 并将有效速度限制在范围  $[-5, 5] \text{ m s}^{-1}$  内。从生成的场中, 提取非-重叠的  $64 \times 64$  补丁。使用单个 NVIDIA A100 GPU

在 AdamW 优化器 [8] (基础学习率  $1.5 \times 10^{-4} \cdot \frac{32}{256}$ , 权重衰减 0.05) 上进行训练, 批处理大小为 32, 共 500 个周期。采用与掩码比率课程周期对齐的余弦学习率计划。

### 3.3 不确定性地图

在推理过程中, 每张图像抽取  $N = 50$  个独立的随机掩码, 每个被遮罩的输入通过解码器传递, 得到的集成通过公式 (1) 聚合以获得后验均值  $\bar{X}$  和像素级标准差  $\sigma_X$ 。N 的值选择用于平衡集成稳定性和计算成本。

## 4 结果

在保留的 1028 个测试集上, CuMoLoS-MAE 超过了  $8 \times 8$  均值滤波器、Noise2Void U-Net [9]、Noise2Void 去噪 CNN (DnCNN) [10], 以及卷积变分自编码器 (CVAE [11]) (表 1)。它还达到了最高的低频谱保真度 (93.75%)。光谱完整性通过以下方式比较去噪输出  $P_{\text{den}}(f)$  和原始信号  $P_{\text{raw}}(f)$  的时间功率谱密度:

$$\varepsilon_{\log}(f) = \frac{\log_{10} P_{\text{den}}(f) - \log_{10} P_{\text{raw}}(f)}{\log_{10} P_{\text{raw}}(f)}. \quad (2)$$

保真度是具有  $|\varepsilon_{\log}(f)| \leq 0.5$  的 bin 分数  $f \leq 0.01$  Hz。CuMoLoS-MAE 在低频 bin 的 93.75% 中满足这一条件, 保留了风暴尺度的能量。Gate 0 是主要的例外情况, 这可能是由于它位于大气边界层底部的位置, 如图 2 所示。DnCNN (Noise2Void) 基线尽管重建效果较差且光谱保真度较低, 但提供了最佳 FID (0.12), 这可能是因为在低方差区域中存在感知相似性。

表 1: 重建性能和低频谱保真度在 1028 个保留的雷达补丁上。向上箭头 ( $\uparrow$ ) 表示越高越好; 向下箭头 ( $\downarrow$ ) 表示越低越好。

方法	峰值信噪比 (分贝)	SSIM	均方误差	FID	频谱保真度
8CE8 MeanFilter	23.41	0.4950	0.5186	5.13	91.67%
CVAE	26.70	0.4190	0.4036	3.28	80.21%
DnCNN (Noise2Void)	23.09	0.6466	0.6232	<b>0.12</b>	36.46%
U-Net (Noise2Void)	27.70	0.7016	0.2581	0.44	49.48%
<b>CuMoLoS-MAE (我们的)</b>	<b>29.45</b>	<b>0.7857</b>	<b>0.1854</b>	1.87	<b>93.75%</b>

### 4.1 不确定性量化与气候效用

通过计算像素级皮尔逊相关性来评估不确定性估计的可靠性, 该相关性是蒙特卡洛标准差图  $\sigma_X$  与绝对重建误差之间的。

在 1028 张测试图像中观察到了强烈的对齐现象, 每个块的相关性的平均值为  $r = 0.961 \pm 0.037$ , 全局相关性为  $r = 0.961$ , 斯皮尔曼等级相关性为  $\rho = 0.926$ 。如图 2 所示, 这些高相关性表明  $\sigma_X$  紧密追踪真实的重建误差, 即该模型可靠地预测其输出在哪一部分可能准确, 在哪一部分不准确。这些像素被  $\sigma_X$  分成十个区间, 并计算每个区间的平均绝对误差 (MAE)。MAE 在  $\sigma$  十分位数上单调上升, 从 0.02845 到 0.99939 (一个  $35.1 \times$  差距), 并且按  $\sigma$  排序的前 1%, 5%, 10% 和 20% 的像素分别捕获了总  $|\text{error}|$  的 10.1%, 30.6%, 43.4% 和 59.4%, 这证实了  $\sigma_X$  在下游同化和预警系统中的错误分类中的实用性。

表 2: 重建质量和不同时间窗口大小下的频谱保真度。上箭头 ( ) 表示数值越高越好; 下箭头 ( ) 表示数值越低越好。

窗口 (时间 $\oplus$ 范围门限)	峰值信噪比 (分贝)	SSIM	均方误差	FID	光谱保真度
$64 \times 64$	29.45	0.7857	0.1854	1.87	93.75%
$128 \times 64$	30.11	0.7697	0.2253	3.73	87.50%
$256 \times 64$	28.55	0.6103	0.3205	5.50	38.02%

## 4.2 时间上下文对重建和频谱保真度的影响

从表 2 可以看出, 随着窗口从  $64 \times 64$  增加到  $128 \times 64$  再到  $256 \times 64$ , 重构质量和低频谱保真度下降, 这表明主要处于局部去噪模式, 在这种模式下,  $64 \times 64$  提供了足够的上下文。在  $128 \times 64$  处 PSNR 的轻微上升源自一些近似常数、超低误差区间被对数平均值过重加权 (MSE 几乎不变), 这对任何有意义的推断都不具有统计显著性。在  $256 \times 64$  处, 标记数量和遮罩区域增加但没有额外容量, 导致各指标上的错误增加。

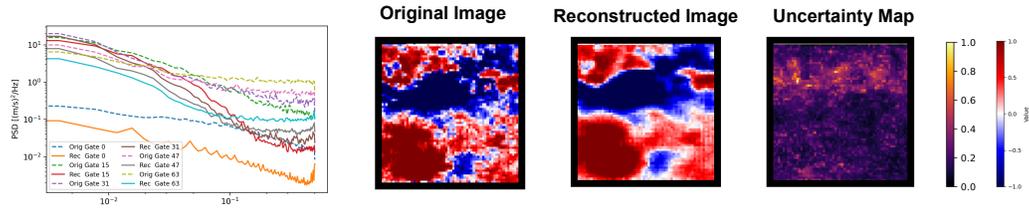


图 2: 单样本的可视化诊断。左侧: 选定范围门处的对数-PSD——重建与原始数据匹配 (小间隙) 除了在边界层底部附近的 Gate 0。左中: 原始多普勒雷达垂直速度。右中: 铜模洛斯-MAE 重建。右侧: 每像素不确定性地图  $\sigma_X$ 。

## 4.3 掩码比率课程对训练效率的影响

课程掩码加速了收敛, 同时保持感知质量。重建损失在第 198 个周期时降至低于 0.20 (固定情况下为 224; 快约 13 分钟), 并在第 286 个周期时达到 0.189 (固定情况下为 333)。如表 3 所示, 指标是可比较的, 因此我们得出结论, 课程掩码通过  $\sim 10\%$  的效率提高了训练效果, 尽管固定掩码的最终分数略有更好。

表 3: 掩码比率课程对重建、FID 和频谱保真度 (1028 个补丁) 的影响。上箭头 ( ) 表示数值越高越好; 下箭头 ( ) 表示数值越低越好。

配置	峰值信噪比 (分贝)	均方误差	SSIM	FID	光谱保真度
Without curriculum	29.45	0.1854	0.7857	1.87	93.75%
With mask-ratio curriculum	28.90	0.2106	0.7868	1.88	93.23%

## 5 结论与未来工作

我们介绍了 CuMoLoS-MAE, 这是一种基于课程指导的蒙特卡罗集成掩码自编码器, 它提供了最先进的重建效果以及逐像素不确定性。通过恢复因噪声或掩码丢失的细尺度特征并标记

低置信度区域，CuMoLoS-MAE 提升了大气相干结构的检测能力。这种功能对于地球观测和气候建模具有价值，因为更准确的重建场可以改进（1）我们对极端事件及其在全球变暖下的演变理解，（2）数值天气预报系统中的数据同化，以及（3）大气科学中的过程级分析。不确定性地图进一步通过根据置信度加权观测值来支持重构。

具体来说，我们计划使用该模型创建垂直剖面数据集以研究对流从浅层到深层的过渡。重建去噪后的垂直剖面及其相关不确定性将有助于（a）上升气流的检测，（b）更好地表征其宽度和动能，并且（c）改进对转换时机和可能性的预测。我们将评估该模型在 ARM 提供的激光雷达系统上的通用性（例如班克黑德国家森林的新数据集），并评估其实时部署以进行操作同化的情况。我们还计划将训练从天扩展到月和年，以便捕捉气候变异，支持长期再分析，并测试在变化的环境条件下系统的稳健性。

## 参考文献

- [1] Lesti G, Spiegel S. A Sliding Window Filter for Time Series Streams. In: IOTSTREAMING@PKDD/ECML; 2017. .
- [2] Pinheiro Cinelli L, Araújo Marins M, Barros da Silva EA, Lima Netto S. Variational autoencoder. In: Variational methods for machine learning with applications to deep networks. Springer; 2021. p. 111-49.
- [3] Madan N, Ristea NC, Nasrollahi K, Moeslund TB, Ionescu RT. Cl-mae: Curriculum-learned masked autoencoders. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision; 2024. p. 2492-502.
- [4] He K, Chen X, Xie S, Li Y, Dollár P, Girshick R. Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2022. p. 16000-9.
- [5] Harrison RL. Introduction to monte carlo simulation. In: AIP conference proceedings. vol. 1204; 2010. p. 17.
- [6] Sisterson D, Peppler R, Cress T, Lamb P, Turner D. The ARM southern great plains (SGP) site. Meteorological Monographs. 2016;57:6-1.
- [7] Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, et al. A survey on vision transformer. IEEE transactions on pattern analysis and machine intelligence. 2022;45(1):87-110.
- [8] Llugsi R, El Yacoubi S, Fontaine A, Lupera P. Comparison between Adam, AdaMax and Adam W optimizers to implement a Weather Forecast based on Neural Networks for the Andean city of Quito. In: 2021 IEEE Fifth Ecuador Technical Chapters Meeting (ETCM). IEEE; 2021. p. 1-6.
- [9] Song TA, Yang F, Dutta J. Noise2Void: unsupervised denoising of PET images. Physics in Medicine & Biology. 2021;66(21):214002.
- [10] Zhao Y, Li Y, Dong X, Yang B. Low-frequency noise suppression method based on improved DnCNN in desert seismic data. IEEE Geoscience and Remote Sensing Letters. 2018;16(5):811-5.
- [11] Bao J, Chen D, Wen F, Li H, Hua G. CVAE-GAN: fine-grained image generation through asymmetric training. In: Proceedings of the IEEE international conference on computer vision; 2017. p. 2745-54.