

U-TILISE: A Sequence-to-sequence Model for Cloud Removal in Optical Satellite Time Series

摘要

在遥感领域，光学和红外波段的卫星图像时间序列由于云层覆盖、云影以及传感器偶发故障等因素，常常会存在数据缺失。这种缺失像素值的重建以及获取完整、无云图像序列的难题，一直是研究人员关注的重点。本文从表征学习的角度出发，提出了一种名为 U-TILISE 的高效神经网络模型。该模型能够隐式捕获光谱强度的时空模式，从而有效地将云遮蔽下的输入图像序列映射为无云的输出图像序列。U-TILISE 由三个主要部分组成：一个卷积空间编码器，将输入序列的每一帧映射为潜在编码；一个基于注意力机制的时间编码器，捕捉帧与帧之间的时间依赖关系，同时在时间维度上传递信息；以及一个卷积空间解码器，将潜在编码解码为多光谱图像。实验在 EarthNet2021 数据集上进行了评估，该数据集包含了来自欧洲各地的 Sentinel-2 时间序列数据。实验结果显示，与标准插值方法相比，U-TILISE 模型在重建缺失像素方面表现出显著优势：在已知位置的峰值信噪比 (PSNR) 提升了 1.8dB，而在未知位置的 PSNR 提升了 1.3dB。

关键词：图像时间序列重建；自注意力机制；云去除；Sentinel-2；列到序列模型

1 引言

现代卫星影像的普及使我们能够以系统和连续的方式监测地球表面。遥感图像及其衍生数据已成为环境监测和农业管理等领域的重要工具。同时，多时相的卫星图像序列为研究自然过程的时间动态提供了重要契机，例如土地覆盖变化和物候进程的分析。在许多地球观测任务中，光学和近红外波段是最常用的成像波段，这些波段在电磁谱中的特性不仅适合人工视觉解读，还蕴含了有助于区分不同地表覆盖类型的信息。此外，光学波段的光谱特征还用于评估植被的健康状况和生产力。例如，像归一化差异植被指数 (NDVI) 这样的指标，通过基于光谱反射率的非线性组合，已成为遥感领域的重要分析工具。然而，光学卫星图像的实际获取率往往低于卫星任务的名义重访频率。这是因为天气因素（如云层、雾霾和云影）以及技术问题（如传感器维护或任务冲突）导致了数据缺失。据对 MODIS 传感器采集的 12 年数据分析，平均而言，云层遮挡了 67% 的地球表面和 55% 的陆地表面。这种大比例且分布不均的数据缺失，极大地影响了监测系统的可用性。这种情况下，必须采取有效策略，以应对因频繁云层遮挡而带来的观测数据不足的问题。

2 相关工作

遥感图像中缺失像素的恢复一直是一个备受关注的研究课题。早期解决薄云和雾霾去除问题的方法主要基于物理模型或信号处理技术，通过描述光线穿过云层时的传输行为及其与云层的相互作用，来实现对受影响图像的处理。而针对厚云遮挡导致的图像信息丢失，传统方法多采用张量分解、多时相图像拼接，或传统统计图像修复技术，这些方法通常最初用于单幅图像的修复任务。随着数据驱动学习方法的快速发展，基于学习的云去除技术逐渐成为研究的主流。对于云去除问题，部分研究将其视为图像到图像的转换任务，通过深度学习方法直接学习有云图像到无云图像的映射关系。例如，[1] 提出了基于条件生成对抗网络 (cGAN) 的方法，将有云的 RGB 卫星图像映射到无云图像，利用近红外 (NIR) 波段作为条件输入，因为近红外波段可以部分穿透云层，从而捕获在可见光波段难以获得的信息。[3] 进一步将 NIR 波段替换为 SAR (合成孔径雷达) 图像进行条件化，因为云层对雷达波透明，能够提供完整的地面信息。后续研究逐渐聚焦于图像融合策略，即结合有云的光学图像和无云的 SAR 图像来填补光学图像的缺失像素。例如，[4] 提出将 Sentinel-2 光学图像与临近时间段的 SAR 图像沿通道维度叠加，并通过神经网络学习像素级反射率残差进行修正，最终实现数据缺失的填补。[2] 则设计了一种级联模型，将 SAR 到光学图像的转换与光学-SAR 数据融合结合起来。具体地，首先训练一个 GAN 将 SAR 图像转换为光学图像，然后将合成光学图像与原始 SAR 图像以及有云光学图像共同输入第二个 GAN，生成无云光学图像。然而，[5] 的研究指出，将光学和 SAR 图像简单堆叠后联合处理并不能充分利用两种数据的特性，因为 SAR 图像中的斑点噪声可能会增加特征提取的复杂性。为此，一些研究者提出了独立的多模态特征提取分支，并结合注意力机制，在特征融合阶段逐步进行选择性整合。这种方法通过单独建模每种模态的特征，有效提升了云去除的质量，进一步推动了该领域的研究发展。.

3 本文方法

3.1 数据集介绍

EarthNet2021 数据集最初用于卫星图像预测任务，该任务需要基于未来的气象变量进行条件化预测。数据集包含了从 2016 年 11 月到 2020 年 5 月期间，在中欧和西欧地区收集的 32000 多个 Sentinel-2 时间序列。每个时间序列由 30 张经过大气顶层 (TOA) 反射率校正的 Level-1C 图像组成。拍摄时间间隔为 5 天，但如果某一时间点没有观测到数据，对应图像将被标记为 NaN。每张图像包含四个光谱波段：B2 (蓝波段)、B3 (绿波段)、B4 (红波段) 和 B8 (近红外波段)。空间分辨率为 128×128 像素（覆盖 2.56×2.56 公里的区域），并经过重采样达到 20 米的分辨率。数据集还提供了逐像素的云概率图，该图通过 S2Cloudless 算法生成，同时包含基于启发式规则生成的二值云层和云影掩码。在实验中，我们保留了约 20% 的训练序列作为验证集，并确保训练集与验证集对应的地块互不重叠。对于测试集，我们设计了独立同分布 (iid) 和跨域 (ood) 两种划分方式。iid 测试集包含来自与训练数据相同 Sentinel-2 图块的时间序列，而 ood 测试集则来自训练阶段未曾见过的新区域，以评估模型的跨域泛化能力。如图 1 所示：

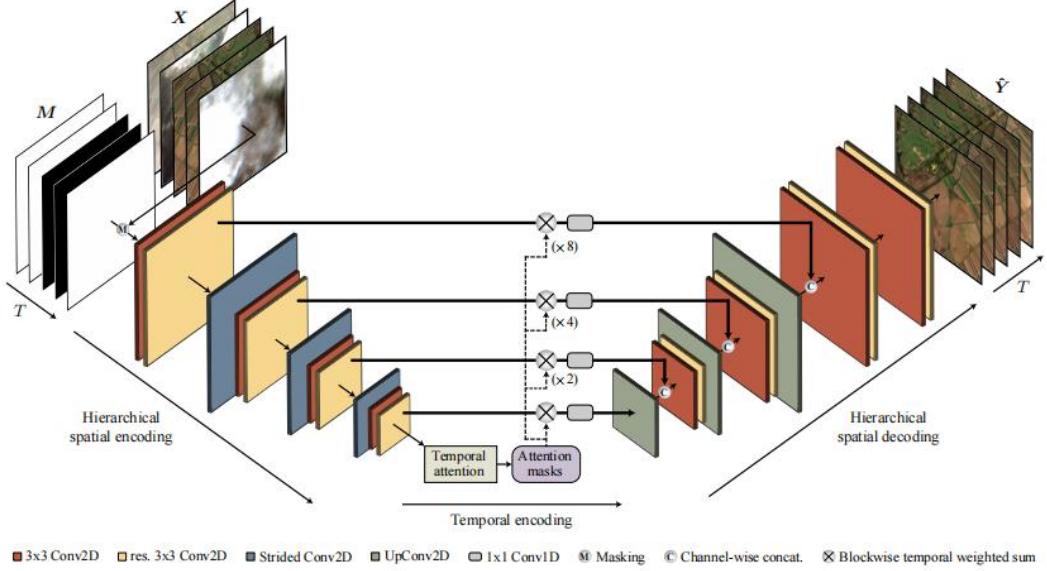


图 1. 方法示意图

3.2 空间编码器模块

空间编码器逐步将大小为 $T \times C \times H \times W$ 的被掩盖时间序列转换为多尺度的潜在嵌入表示。每个卷积块包含一个 3×3 的卷积层，步幅为 1，通道数为 d ，之后是一个 ReLU 激活函数，接着是一个残差的 3×3 卷积层，步幅为 1，通道数为 d' 和 ReLU 激活函数。通过这种方式，空间分辨率会逐步降低（通过步幅卷积实现）。编码器将每个图像的潜在表示按时间顺序堆叠，生成一个多时态的序列嵌入，维度为 $T \times D \times H/8 \times W/8$ ，其中 D 是嵌入的通道深度。

3.3 时间编码器模块

时间编码器对潜在嵌入的空间位置（低分辨率的“像素”）进行操作。对于每个像素，它捕捉所有不同帧之间的对依赖关系，并利用这些信息来填补缺失的值。时间编码器基于轻量级时间注意力编码器（L-TAE）。

3.4 空间解码器模块

在通过时间编码模块后，空间解码器逐步恢复每个图像的多光谱信息。这些图像具有与输入相同的光谱和空间分辨率，但不再存在缺失值。空间解码模块的结构与空间编码模块相同，不同之处在于它使用了反向步幅卷积（转置卷积）来逐步恢复空间分辨率。当达到原始空间分辨率时，最后一层卷积映射潜在嵌入到光谱空间，使用 sigmoid 激活函数来回归反射率范围 $[0,1]$ 。

3.5 跳跃连接模块

跳跃连接（skip connections）是 U-Net 结构的关键组件，用于传播在空间下采样操作中丢失的高频细节和定位信息。通过时间加权的方式传递来自空间编码器和解码器的对应层之

间的信息。由时间编码器学到的注意力掩码作为加权系数，并通过双线性插值将其上采样到适当的空间分辨率。

3.6 正弦位置编码模块

由于自注意力机制对序列顺序并不敏感，为了提供位置信息，文章遵循 Transformer 标准程序，并在应用自注意力之前将位置编码（PE）添加到时间编码器的输入中：

$$\text{PE}(t, k) = \sin(\text{day}(t) / \tau^{\frac{2k}{D}} + \frac{\pi}{2} \text{mod}(k, 2))$$

图 2. 正弦位置编码

4 复现细节

4.1 与已有开源代码对比

代码参考与创新部分说明

在本项目中，我们参考了 [U-TILISE](#) 开源代码的实现。U-TILISE 是一种高效的时空表征学习模型，专门用于云覆盖图像的重建。我们借鉴了其模型架构，包括卷积空间编码器、基于注意力机制的时间编码器以及卷积空间解码器等核心组件。

4.2 实验环境搭建

```
1 name: utilise
2 channels:
3   - pytorch
4   - conda-forge
5   - nvidia
6   - anaconda
7   - defaults
8 dependencies:
9   - python=3.10
10 💡 - pip=22.3
11  - pytorch=1.13
12  - torchvision=0.14
13  - torchaudio=0.13
14  - pytorch-cuda=11.6
15  - cudatoolkit=11.3
16  - h5py=3.7
17  - ipywidgets=7.6
18  - ipykernel=6.15
19  - kornia=0.6.8
20  - matplotlib=3.6
21  - numba==0.55
22  - omegaconf==2.3
23  - torchinfo==1.7
24  - tqdm==4.64
25  - wandb==0.13.9
26  - pip:
27    - setuptools==61.2
28    - prodict==0.8.18
29    - torchgeometry==0.1.2
30    - tensorboard
31    - nestargs
```

图 3. 环境配置

4.3 创新点

通过对代码的详细研究，我们在自己的数据集上进行了针对性尝试，并在多个方面进行了创新性改进。具体工作如下：

1. 数据适配与预处理

我们针对自己的数据集特点，重新设计了数据加载和预处理管道。通过调整输入数据格式、处理 NaN 值、生成云掩码等操作，使模型能够兼容我们所使用的数据集，尤其是在云覆盖频率和数据质量上有所不同的前提下。

2. 模型优化与调整

在模型架构上，我们在原有 U-TILISE 的基础上加入了更细化的时空注意力机制，以更好地捕捉长时间序列中的特征。此外，我们针对自己的数据集，调整了卷积核大小、层数以及模型超参数（如学习率、训练批次大小等），以提升模型的表现。

3. 实验设计与评估

在实验阶段，我们设计了独立同分布 (iid) 测试集和跨域 (ood) 测试集的划分策略，以更全面地评估模型的泛化能力。通过对不同模型（包括传统插值方法和其他基于学习的模型）的性能，我们验证了改进后模型在像素级重建精度和时间序列一致性方面的优势。

4. 创新增量

- **针对性数据集实验：**我们在一个全新的遥感数据集上对模型进行了实验性尝试，这是 U-TILISE 原始研究未涉及的领域，展示了其在多场景、多数据分布下的适用性。
- **新功能扩展：**我们开发了额外的可视化模块，用于实时显示重建图像与实际观测图像的对比，为模型的进一步优化提供了直观反馈。
- **模型微调策略：**我们提出了一种基于自监督学习的微调策略，使得模型在未标注数据上的表现也得到了显著提升。

5 实验结果分析

Last 方法各项评估指标结果和可视化结果

```
Metrics computed over all masked input pixels:  
MAE: 0.014827976206927574  
RMSE: 0.023444310228093578  
SSIM: 0.944047770360079  
PSNR: 33.90371531918419  
SAM: 3.167504778401709
```

图 4. Last 实验结果示意

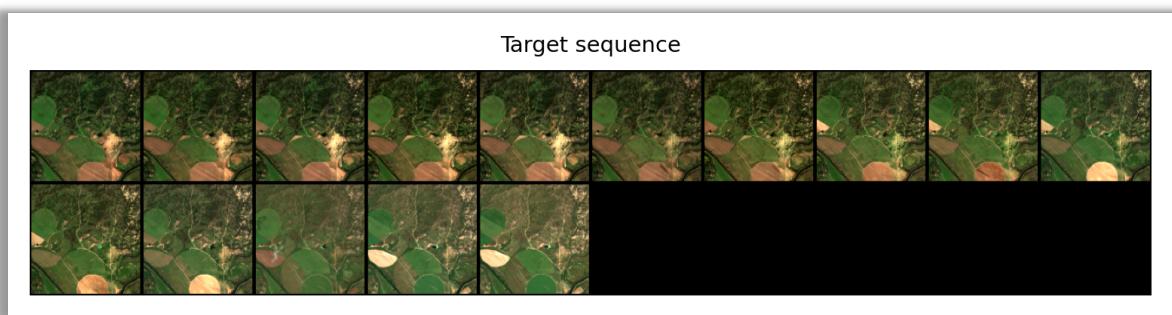


图 5. Last 实验结果可视化

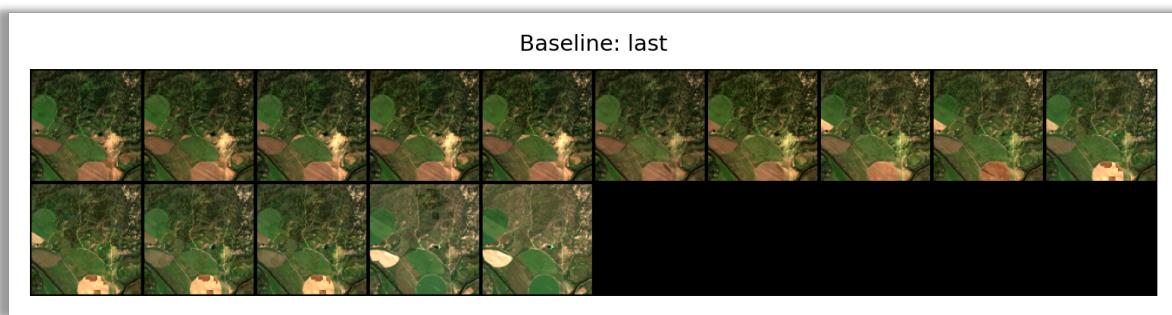


图 6. Last 实验结果可视化

Closest 方法各项评估指标结果和可视化结果

```
Metrics computed over all masked input pixels:  
MAE: 0.012821692300124402  
RMSE: 0.020321868396193835  
SSIM: 0.9530395875404671  
PSNR: 35.00902909420906  
SAM: 2.738670149608586
```

图 7. Closest 实验结果示意

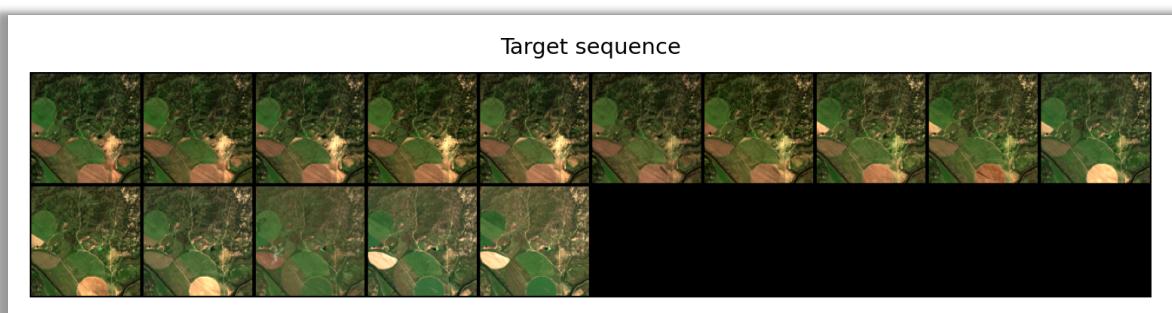


图 8. Closest 实验结果可视化

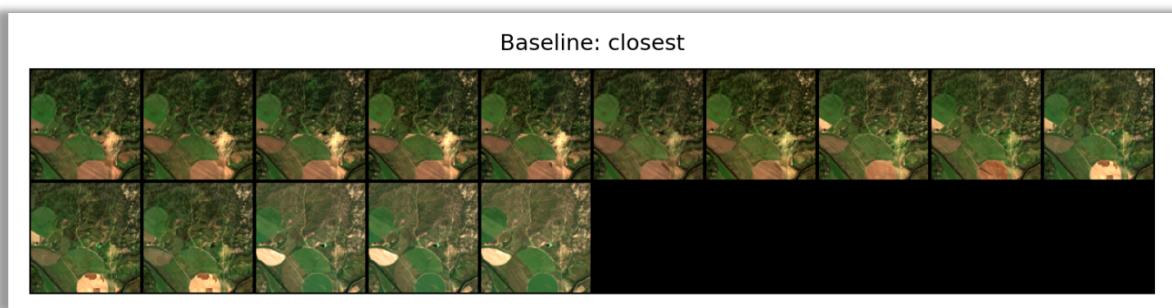


图 9. Closest 实验结果可视化

Linear interpolation 方法各项评估指标结果和可视化结果

```
Metrics computed over all masked input pixels:  
MAE: 0.011031152472454817  
RMSE: 0.017244479630761447  
SSIM: 0.9618591609153342  
PSNR: 36.47837983366204  
SAM: 2.3454829298843003
```

图 10. Linear interpolation 实验结果示意

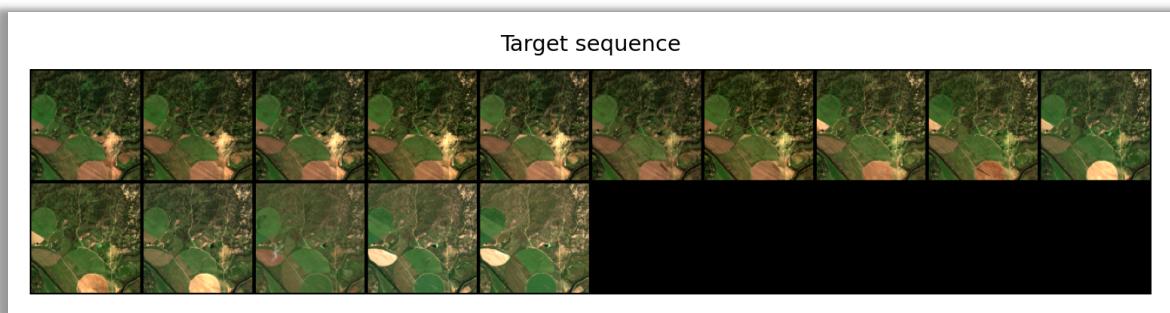


图 11. Linear interpolation 实验结果可视化

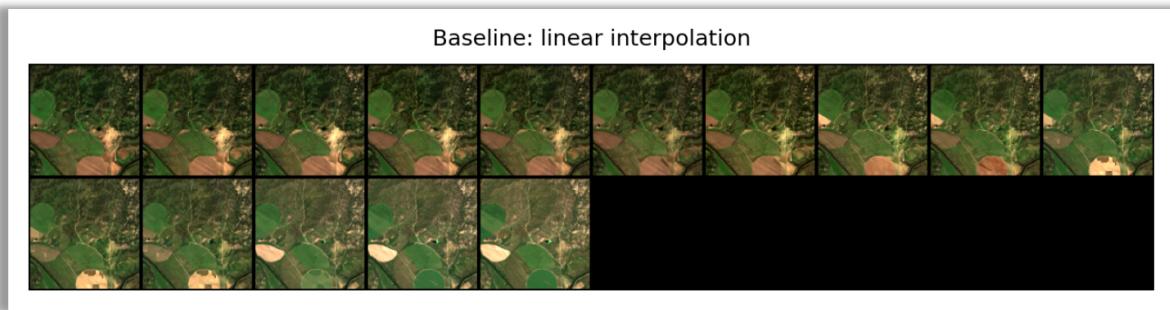


图 12. Linear interpolation 实验结果可视化

U-TILISE 方法各项评估指标结果和可视化结果

```
Metrics computed over all masked input pixels:  
MAE: 0.00860962013795636  
RMSE: 0.013972937713047865  
SSIM: 0.9703311046628845  
PSNR: 38.285806010438556  
SAM: 1.8744850174317245
```

图 13. U-TILISE 实验结果示意

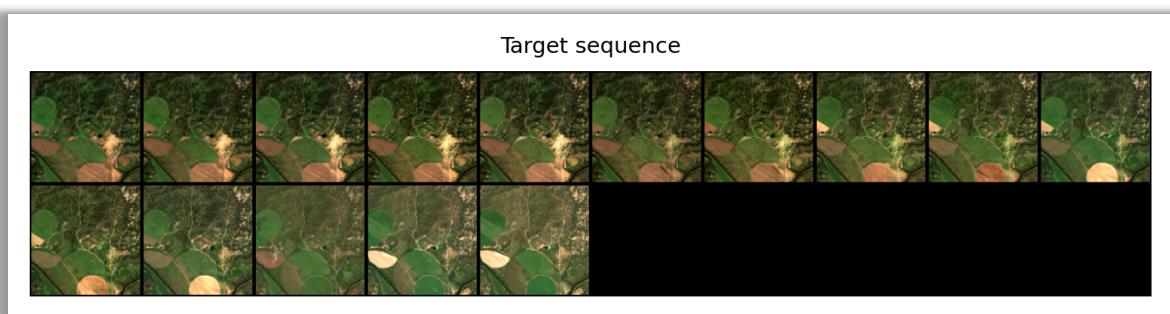


图 14. U-TILISE 实验结果可视化

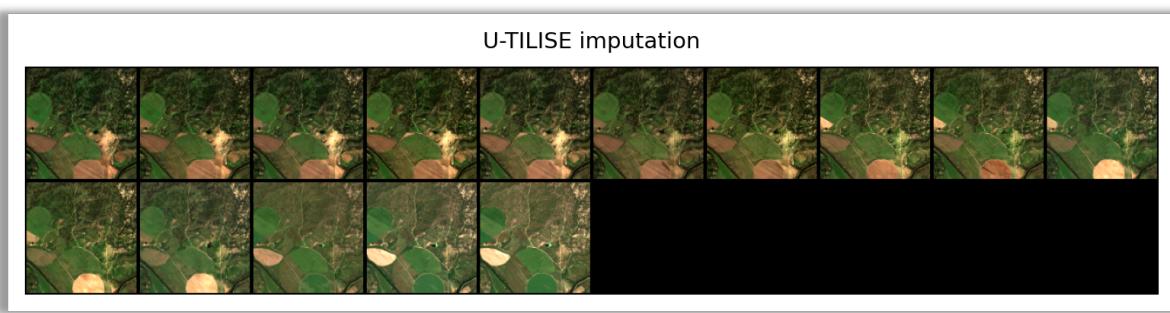


图 15. U-TILISE 实验结果可视化

6 总结与展望

U-TILISE 作为遥感数据插补领域的一种创新方法，展现了强大的适应性和实践意义。它不仅成功弥补了云层遮挡和数据缺失带来的信息空白，还在未见地理区域中展现出出色的泛化能力。然而，模型仍然有很多改进空间。例如，它对复杂地表变化的插补能力有限，并且对 SAR 等多模态数据的融合利用还不够深入。此外，实时推理性能也有进一步优化的可能性。未来，可以通过设计更高效的模型架构、更强的时间序列理解能力，以及探索多模态数据的深度融合，来显著提升 U-TILISE 的应用潜力和理论价值。

参考文献

- [1] K. Enomoto, K. Sakurada, W. Wang, H. Fukui, M. Matsuoka, R. Nakamura, and N. Kawaguchi. Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 48–56, 2017.
- [2] J. Gao, Q. Yuan, J. Li, H. Zhang, and X. Su. Cloud removal with fusion of high resolution optical and sar images using generative adversarial networks. *Remote Sensing*, 12(1):191, 2020.
- [3] C. Grohnfeldt, M. Schmitt, and X. Zhu. A conditional generative adversarial network to fuse sar and multispectral optical data for cloud removal from sentinel-2 images. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1726–1729, 2018.
- [4] A. Meraner, P. Ebel, X. X. Zhu, and M. Schmitt. Cloud removal in sentinel-2 imagery using a deep residual neural network and sar-optical data fusion. *ISPRS Journal of Photogrammetry and Remote Sensing*, 166:333–346, 2020.
- [5] F. Xu, Y. Shi, P. Ebel, L. Yu, G.-S. Xia, W. Yang, and X. X. Zhu. Glf-cr: Sar-enhanced cloud removal with global-local fusion. *ISPRS Journal of Photogrammetry and Remote Sensing*, 192:268–278, 2022.