

Cloud removal in Sentinel-2 imagery using a deep residual neural network and SAR-optical data fusion

张瑞铎

摘要

光学遥感图像是许多地球观测活动的核心。卫星数据的定期、一致和全球范围的性质在许多应用中得到了利用，如耕地监测、气候变化评估、土地覆盖和土地利用分类以及灾害评估。然而，有一个主要问题严重影响了地表观测的时间和空间可用性，即云层覆盖。从光学图像中去除云层的任务自几十年来一直是研究的主题。卫星遥感的大数据时代的到来为使用强大的数据驱动的深度学习方法来解决这个问题提供了新的可能性。本文设计了一个深度残差神经网络架构，用于从多光谱的 Sentinel-2 图像中去除云层。SAR-光学数据融合被用来利用两个成像系统的协同特性来指导图像重建。此外，还提出了一种新的云层适应性损失，以最大限度地保留原始信息。该网络在一个全球采样的数据集上进行了训练和测试，该数据集包括真实的有云和无云图像。所提出的设置允许通过重建底层陆地表面结构的光学表示来消除甚至是光学上的厚云层。

关键词：去除云层；光学图像；SAR-光学图像；数据融合；深度学习；残差网络

1 引言

虽然近年来卫星观测的质量和数量大幅增加，但从第一次观测到今天，光学领域的遥感仍然存在一个共同的问题。云层。由于厚厚的云层在所有光学频段上都是不透明的，它的存在完全破坏了反射信号并阻碍了对其下表面的观察。这在空间和时间领域都造成了相当大的数据差距。对于需要一致的时间序列的应用，如农业监测，或者必须在特定的时间观察特定的场景，例如灾害监测。云层是一个严重的障碍。考虑到地球表面每天经历的云量，云层的问题变得更加明显。每天都有大量的云层。中分辨率成像分光仪 (MODIS) 仪器 12 年来的观测分析。卫星 Terra 和 Aqua 的数据显示，地球表面平均有 67% 被云层覆盖。(King 等人, 2013)。在陆地表面，云层比例平均为 55%，具有明显的季节性特征。考虑到这些云层遮挡百分比的重要性，很明显，一个成功的云层去除算法会大大增加有用数据的可用性。自地球观测活动开始以来，从卫星图像中检测和去除云层的任务已经被解决了，而且今天仍然是一个积极研究的领域。在这项工作中，我们提出了一个深度学习模型，能够从 Sentinel-2 图像中去除云层。网络设计和额外的 Sentinel-1 SAR 数据的整合使其对广泛的云层覆盖条件具有鲁棒性。该模型是在包含全球范围内获得的场景的大型数据集上训练的，确保其在任何土地覆盖类型上的普遍适用性。

2 相关工作

遥感数据中缺失信息的重建是一个长期研究的问题。在 Shen 等人 (2015)¹ 的文章中，全面回顾了传统技术。在过去的几十年里，人们提出了许多方法来完成去除光学图像中的云层这一特定任务。遵循传统方法的方法可以分为三大类，即多光谱、多时空和画中画技术。许多方法是这些类别的混合组

合。多光谱方法适用于在雾霾和薄卷云的情况下，光学信号并没有被完全阻挡，而是经历了部分波长的变化。在这种情况下，光学信号并没有被完全阻挡，而是经历了部分与波长有关的吸收和反射。在这种情况下，表面信息是部分存在的，并且可以被恢复，例如使用数学方法（Xu 等人，2019²。Hu 等人，2015³）或物理模型（Xu 等人，2016⁴；Lv 等人，2016⁵）。多光谱方法的优点是可以利用原始场景的信息，而不需要额外的数据，但却仅限于丝状、半透明的云层。多时相方法通过整合参考图像的信息来恢复多云的场景在晴朗的天空条件下获得的信息（Lin 等人，2013⁶；Li 等人，2015⁷。Ramoino 等人，2017⁸；Ji 等人，2018⁹）。为此，也可以使用多时空词典学习技术（Li 等人，2014¹⁰）。多时态数据也可能来自不同卫星上的不同传感器（Li 等人，2019¹¹）。多时空方法是最受欢迎的，因为他们用真实的无云观测数据来替代被破坏的像素。然而，在重建具有快速变化的表面条件的场景时，会出现问题。（例如，由于物候事件）因为要重建的场景和参考采集的场景之间存在着时间差。绘画方法通过利用同一云层影响的图像的清晰部分的表面信息来填补损坏的区域（Meng 等人，2017¹²）。这种直接涂抹方法不需要额外的图像，但只有在小的云层中才能取得良好的效果。为了缓解这个问题，选择最合适的类似像素进行克隆的过程通常由辅助数据来指导，如多时空（Cheng 等人，2014¹³）或 SAR 图像（Eckardt 等人，2013¹⁴）。这类方法能提供良好的结果，但由于需要多时空或多传感器的额外数据，其复杂性也随之增加。在采用传统方法去除云层的同时，使用深度学习的数据驱动方法最近越来越受到关注。传统算法中出现的许多问题都有可能通过深度神经网络的端到端学习来解决（DNN）。例如，云的检测和分割作为一个初步步骤往往是不需要的。因为它可以由网络隐含地学习。在多传感器数据融合的情况下，不同传感器领域之间的转换也可以被学习。此外，DNNs 可以被训练来应对任何类型的云和残留的大气条件。第一篇论文利用 DNNs 的潜力来恢复遥感图像中缺失信息的潜力，发表在 Zhang 等人（2018¹⁵）。该方法使用一个空间-时间-光谱卷积神经网络（CNN）来恢复 Landsat TM 数据中的数据缺口。在云层的情况下，同一场景的额外多时空图像被用来支持重建。最近的论文集中在使用现代 CNN 架构，称为条件生成对抗网络（cGAN）（Mirza 和 Osindero，2014¹⁶）。在 Enomoto 等人（2017¹⁶）中，训练了一个 cGAN，以从使用近红外图像作为辅助数据的 Worldview-2 RGB 图像中移除模拟云，而在 Grohnfeldt 等人（2018¹⁷）的 cGAN 使用 SAR 数据作为附加信息，从 Sentinel-2 图像中移除模拟云。一个进化的演变，称为 Cycle-GAN，可用于避免训练时需要成对的无云图像（Singh 和 Komodakis，2018¹⁸）。一个生成无云图像的不同方法是使用 cGANs 直接从 SAR 转换到光学（Bermudez et al, 2018¹⁹；Bermudez 等人，2019²⁰；He 和 Yokoya，2018²¹；Fuentes Reyes 等人，2019 年²²）。除了强大的生成能力外，cGANs 还可能遭受训练和预测的不稳定性，当输入不良的数据（大量云覆盖），正如一些参考的研究和 Mescheder 等人的报告所指出的那样。基于这些经验，本文提出的工作开发了一个模型架构，该架构对输入数据中存在的大型和光学厚的云层具有鲁棒性。除了概念上的考虑，对大型数据集的需求也是深度学习去云的一个突出问题。

3 数据集介绍

本工作中提出和使用的数据集，称为 SEN12MS-CR。SEN12MS-CR 包含 169 个不重叠的区域 (ROI)，在四个季节对所有大陆进行采样。场景位置是从两个均匀分布中随机抽取的，即一个是所有陆地上的，一个是城市地区的。这引入了对城市景观的偏差，因为城市景观通常是遥感研究的重点，包含更复杂

的模式。ROI 的平均大小约为 5200×4000 px，这对应于 52×40 公里的地面覆盖，每个 ROI 都是由一个正射、地理参照的多云和无云 Sentinel2 图像以及相应的 Sentinel-1 图像的三联体。所有这三张图像都是在同一气象季节获得的，以限制地表变化。为了评估光学图像的云层覆盖情况，使用了 Schmitt 等人描述的云层探测器。无云的 Sentinel-2 图像的云层覆盖率阈值为 10%，而多云图像是在 20% 和 70% 的云层覆盖率范围内。Sentinel-2 其数值范围为 $[0, 10,000]$ 。所有 13 个原始波段都包括在内。Sentinel-1 的数据来自 Level-1 GRD 在 IW 模式下获得的产品，有两个偏振通道（VV 和 VH）。这些数值是已经转化为分贝尺度的 0 背散射系数。为了使图像适应 CNN 的要求，ROI 被切割为 256×256 px 的小斑块，跨度为 128px。因此，相邻斑块之间的重叠量为 50%，以最大限度地增加可从图像中提取的斑块数量，同时还能确保可接受的独立性。对生成的斑块进行自动和手动检查，以消除镶嵌假象和其他损坏的区域。最终经过质量控制的 SEN12MS-CR 数据集包括 157521 斑块-三联体。图1显示了该数据集三联体的例子。

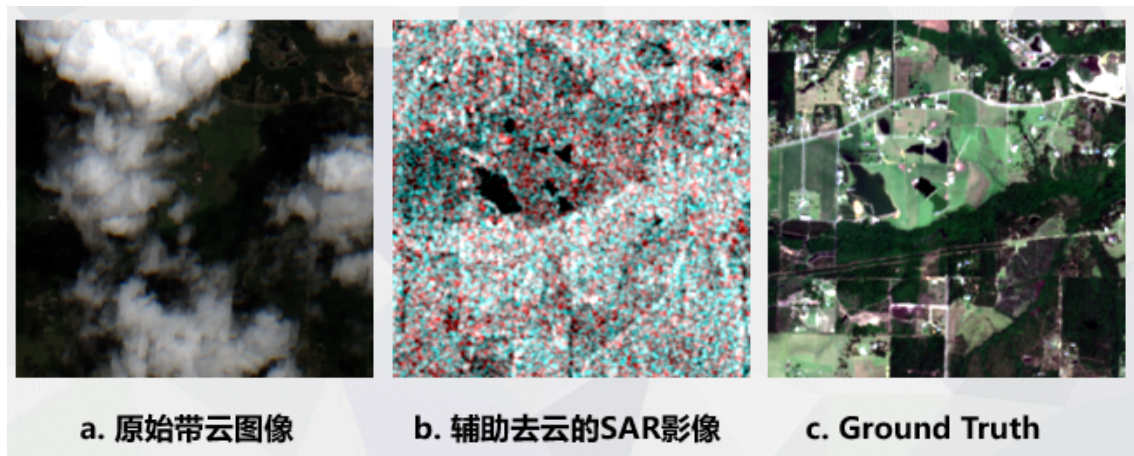


图 1: 数据组织形式

4 本文方法

4.1 DSen2-CR 模型

所提出的模型，称为 DSen2-CR，是基于 Lanaras 等人（2018）提出的超分辨率 Deep Sentinel-2 (DSen-2) ResNet，它本身是由最先进的单图像超分辨率 EDSR 网络（Lim 等人，2017）衍生出来的。与超分辨率类似，去除云层可以被看作是一项图像重建任务，其中缺失的空间和光谱信息必须被整合以恢复完整的信息内容。为了指导厚重的、光学上无法穿透且没有地面信息的云层下的重建过程，DSen2-CR 利用了 SAR 图像作为一种先验形式。为此，同一场景的 Sentinel-1 图像作为额外的输入被引入到网络中。该图像的 SAR 通道与输入的光学图像其他通道叠加起来。高度非线性的 SAR 到光学的转换，以及云层检测和处理，都是在网络中隐含地学习和执行的。训练是在一个端到端的设置中完成的。同一场景的无云图像被提交给网络作为损失计算的目标。图2显示了 DSen2-CR 模型的图示和使用的残差块设计。在下文中，将进一步介绍该网络的特性和特点。

- 长跳过连接。正如 Lanaras 等人（2018）最初提出的那样，一个加法捷径将输入的多云图像穿梭到最终输出前的加法层。这基本上意味着，整个网络正在学习预测一个残差图，其中包含对输入多云图像的每个像素的修正。在晴空万里的输入或丝状云的情况下，预测的修正将是轻微的或不存在的。相反，对于具有明亮外观的厚厚的云层，修正值会更大。

- 残差块。网络的主要部分是由几个残余单元依次叠加。网络中单元 B 的具体数量是一个超参数，它定义了网络的深度。残余单元各包含四个层和一个用于残余连接的附加层。四个跳过的层是一个带有后续 ReLU 激活的二维卷积层、第二个二维卷积层和最后的剩余缩放层（见下一点）。在第一个卷积层之后只使用了一个 ReLU 激活但在第二个卷积层之后没有，因为该网络应该预测的是纠正，既可以是正的也可以是负的。对于两个卷积层，都使用了 3×3 的核，遵循了一般的社区的普遍趋势是在更深的模型中使用较小的内核（Lanaras 等人，2018）。输出特征维度 F，即不同过滤器的数量。不同的过滤器，对所有单元都是固定的，是一个超参数。一个像素的跨度和零填充始终被使用，以便在整个网络中保持数据的空间尺寸。考虑到自己的实验和 Lanaras 等人报告的经验，在表示能力和计算复杂性之间进行折衷。Lanaras 等人（2018）、Lim 等人（2017）中报告的经验。残余单元与 $F = 256$ 特征的剩余单元被选为 DSen2-CR 架构的基线。

- 残差缩放。残差缩放层是一个自定义层，它将其输入与一个恒定的标量相乘。是 Szegedy 等人（2017）首次提出，这种激活缩放的效果是在不引入额外参数的情况下稳定训练。例如在批量归一化层中。本工作中选择 0.1 作为缩放常数。

- 额外的卷积。在网络的开始，一个连接层将输入的光学层和 SAR 层垂直堆叠，以实现联合处理。以便进行联合处理。在这之后，一个 3×3 的卷积层与 RELU 激活层被引入，在数据通过残差层之前处理联合处理。在最后一个残差单元之后，最后一个 3×3 卷积层将光谱尺寸恢复到与光学图像的波段数相匹配，然后到达残差添加层。

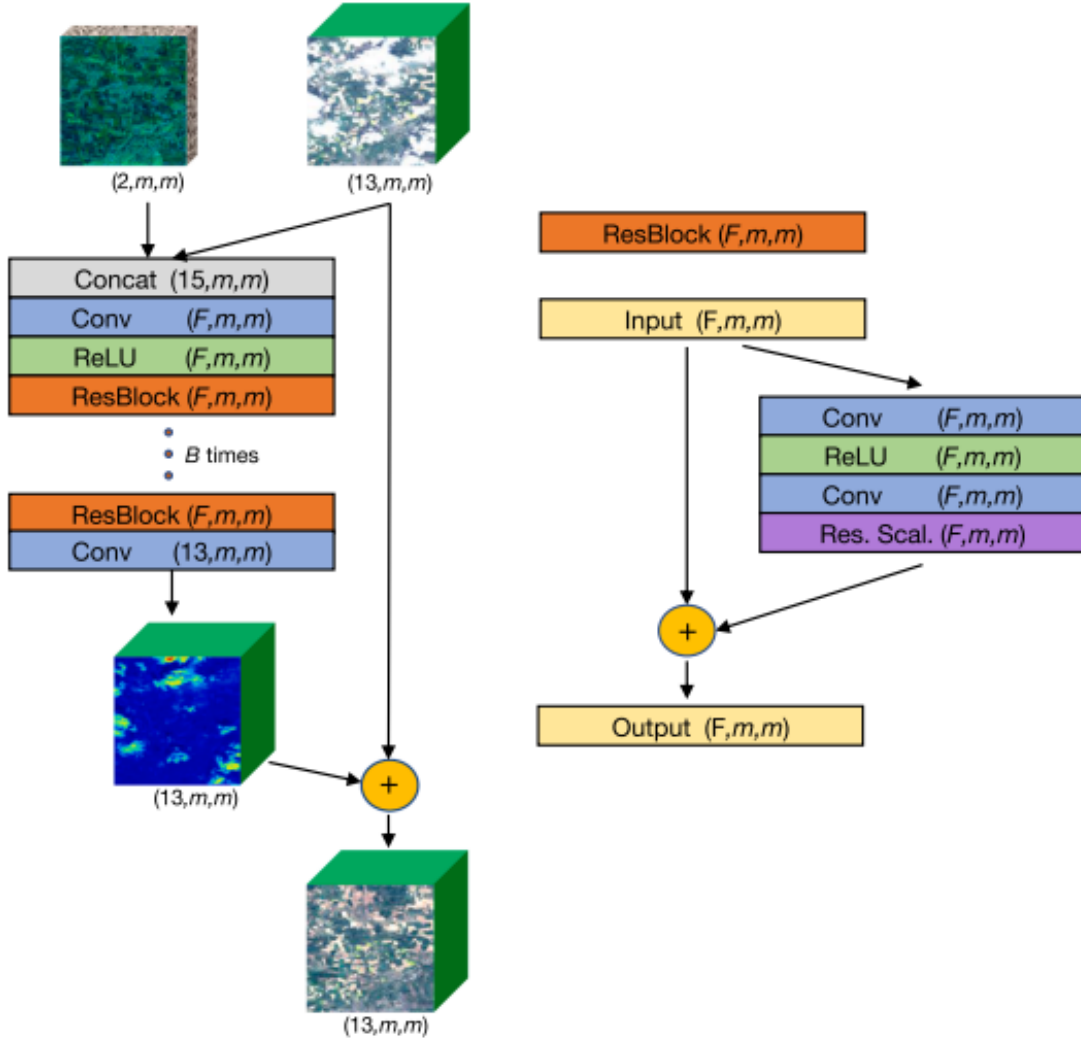


图 2: DS Sen2-CR 网络模型

4.2 云自适应正则化损失函数

如数据集部分所述，输入的多云图像和目标无云光学图像是在同一季节不同日子获得的。虽然时间上差异是有限的，但图像之间的地表状况变化可经常被观测到。由于除云算法的目标是在不修改清晰部分的情况下恢复云层下的地面信息，因此在输出中保留输入图像中尽可能多的信息是非常重要的。为了尽量减少目标图像中地面变化的影响，在这项工作中开发了一个自定义的训练损失。根据 Lanaras 等人（2018）的建议，L1 指标（平均绝对误差）被用作基本误差函数，因为它对大的偏差和 Sentinel-2 数据的高动态范围不敏感。将预测的输出图像定义为 P ，将无云目标图像定义为 T 。经典的目标损失 T 是基于预测和目标之间的简单 L1 距离，可以表述为公式1所示

$$L_t = \frac{\|P - T\|_1}{N_{tot}} \quad (1)$$

N_{tot} 是光学图像所有通道的像素总数。对这种普通的 L1 损失的优化是简单而直接的，但它有一个缺点：由于是在地面条件变化的多时相数据上训练的，网络会被诱导学习、预测和应用不需要的表面变化。为了减少这些假象，我们开发了一种新的损失原理。这个想法是将二元云和云影掩码（CSM）纳入损失计算，并利用这些信息引导学习过程，使输入信息得到最大限度的保留。这种定制的损失，我们称之为云自适应正则化损失（CARL），表述为公式2所示

$$L_{CARL} = \frac{\overbrace{\|CSM \odot (P - T) + (1 - CSM) \odot (P - I)\|_1}^{\text{cloud-adaptive part}}}{N_{tot}} + \lambda \frac{\overbrace{\|P - T\|_1}^{\text{target reg. part}}}{N_{tot}} \quad (2)$$

P , T , I , 分别表示预测的、目标和输入的光学图像。CSM 掩码具有与图像相同的空间尺寸，云层和阴影像素的像素值为 1，未被破坏的像素为 0。标有 CSM 和图像差异的乘法是逐元的，并应用于所有通道。在云适应部分，针对输入图像的阴天或阴影像素计算平均绝对误差损失，而针对晴天像素计算的是输入图像本身。如此，网络可优化预测以匹配输入的无云部分，并且只在需要时使用多时空信息，即用于云和阴影的重建。然而，当只用这个云适应性的部分进行训练时，据观察，由于对遮罩的学习过于精确，该网络在预测的图像中引入了假象。为了避免这种影响，在损失函数中加入了一个额外的目标正则化项，其形式是预测和目标之间的典型平均绝对误差损失（相当于公式（1）中的 L_T ）。这个额外的损失诱导网络学习生成仍然具有与目标图像相似的自然、平滑外观的图像。正则化因子 λ ，在公式（2）中对这个目标正则化项进行缩放，是一个能有效平衡输入信息保留和预测神经器的超参数。经过广泛的调整， $\lambda=1$ 的值被发现提供了最佳的权衡。对于训练期间需要的 CSM 掩码实现，采用了 Schmitt 等人（2019a）（云检测）和 Zhai 等人（2018）（云影检测）中提出的方法的组合。图3显示了创建掩膜的不同处理步骤的流程图。云层二值化的阈值 $TCL=0.2$ 是在视觉评估后选择的。云层检测的阈值是用以下参数计算的 $T_{CSI} = 3/4$ 和 $T_{WBI} = 5/6$ 。

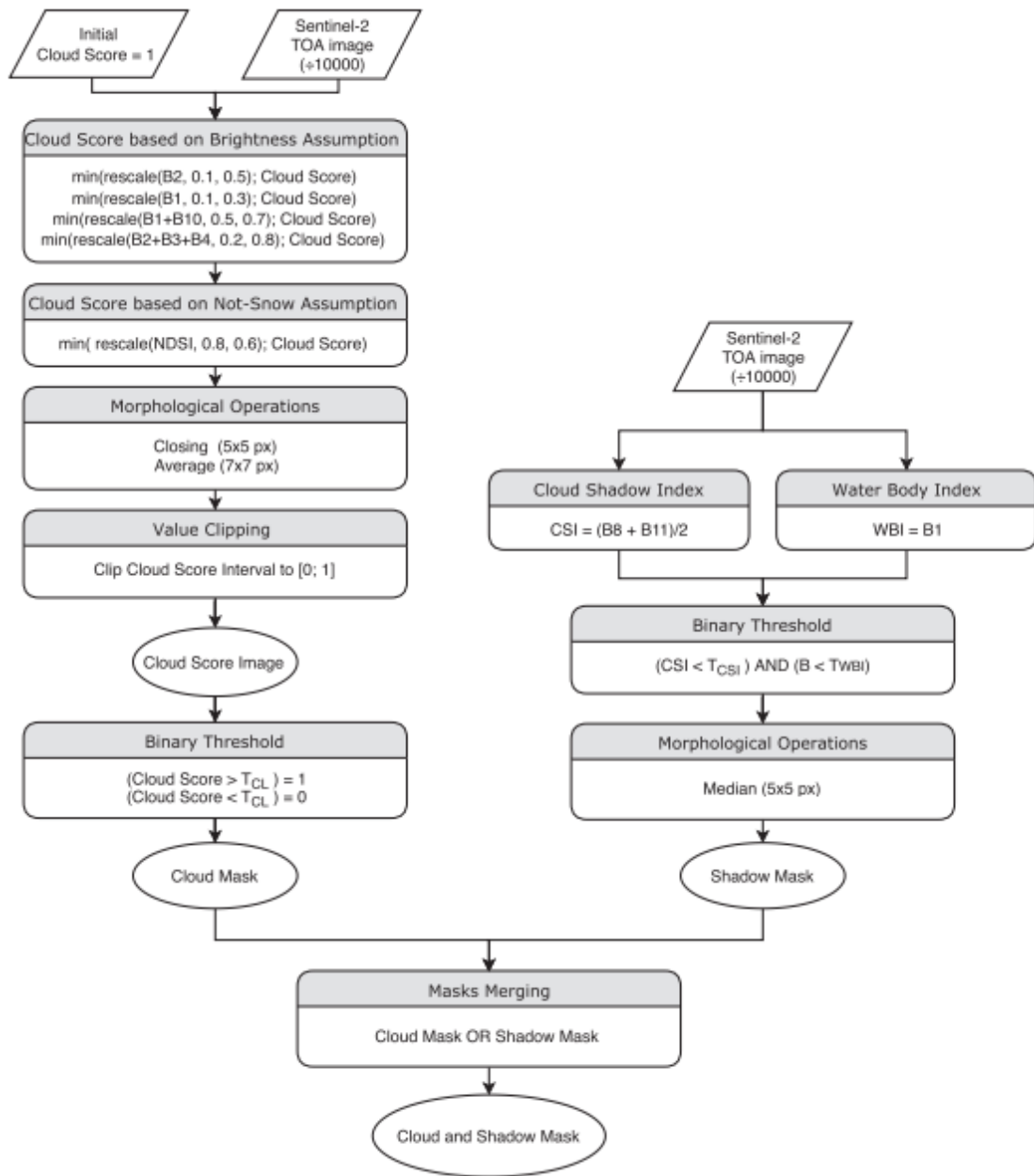


图 3: 云影掩码的生成步骤

5 复现细节

5.1 与已有开源代码对比

本文复现了作者所提出的基于深度残差网络的 Dsen2-CR 模型，并且对网络的输入端进行了修改。原始网络的输入端为多云光学图像与 SAR 图像，本实验验证了作者所提出模型的有效性。同时，在输入端取消了 SAR 影像以检测其在去除云层方面的有效性

本文参考了他人的相关代码，链接如下：

<https://github.com/ameraner/dsen2-cr>

https://github.com/Phoenix-Shen/DSEN2_CR_PYTORCH

5.2 实验环境搭建

表 1: 本地实验环境

操作系统	Windows10 专业版 21H2
CPU	11th Gen Intel(R) Core(TM) i7-11700 @ 2.50GHz
机带 RAM	40GB
存储器容量	1000GB

表 2: 服务器端配置

CPU	12 × Xeon Gold 6271
GPU	NVIDIA Tesla P100-16GB
显存	16GB
内存	48GB
硬盘	200GB

5.3 实验过程

由于原始数据集过大，故从原始的数据集中随机挑选出 9141 个 ROI 区域对应的影像，每一个 ROI 为一幅三联图像，故总图像数量为 27,423 张。划分训练集、验证集、测试集比例为 7：2：1，即训练集包含 6512 个 ROI 区域，验证集包含 1629 个 ROI 区域，测试集包含 1000 个 ROI 区域。训练数据范围在 [0,10000] 之间，每个波段除以缩放系数 2000，保证训练的稳定性。SAR 图像取 [VV、VH] 通道，范围在 [-25,0] 与 [-32.5,0]。移动数值到正数域并缩放到 [0,2] 范围内。

训练参数设置采用原论文中的推荐数值，不使用 batchsize 进行训练，学习率参数设置为 7e-5. 由于机器算力原因，训练的迭代次数设置为 50 轮。由训练完成后 loss 值与在验证集上的表现判断，模型已正常收敛，将该模型定义为模型一。

此外，本人修改了网络的结构，作为对照组。该组取消了模型的辅助 SAR 影像输入，在输入端只保留原始的多云图像，其余训练设置参数均保持不变，该模型定义为模型二。

6 实验结果分析

模型评估主要由定性与定量两方面展开。定性方面主要是根据目视解译的效果来检测模型是否完成了去除云层的工作，定量方面主要根据以下三种评估指标进行精度的评定。

- 峰值信噪比 PSNR。衡量最大值信号和背景噪音之间的差异。定义公式如下

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

(3)

- 结构相似指标 SSIM。衡量两张图片的相似程度。与 PSNR 衡量绝对误差不同，SSIM 是感知模型，即更符合人眼的直观感受。
- 光谱角制图 SAM。度量光谱间的相似性，夹角越小，两光谱越相似，属于同类地物的可能性越大

模型一在测试集上的定性表现如图4所示。由图中可知，网络能较好的检测到云层，并且也能正确的去除云层。虽然距离 Ground Truth 还有一定差距，但是图像也基本是可用状态，而原始图像基本是不可用的。

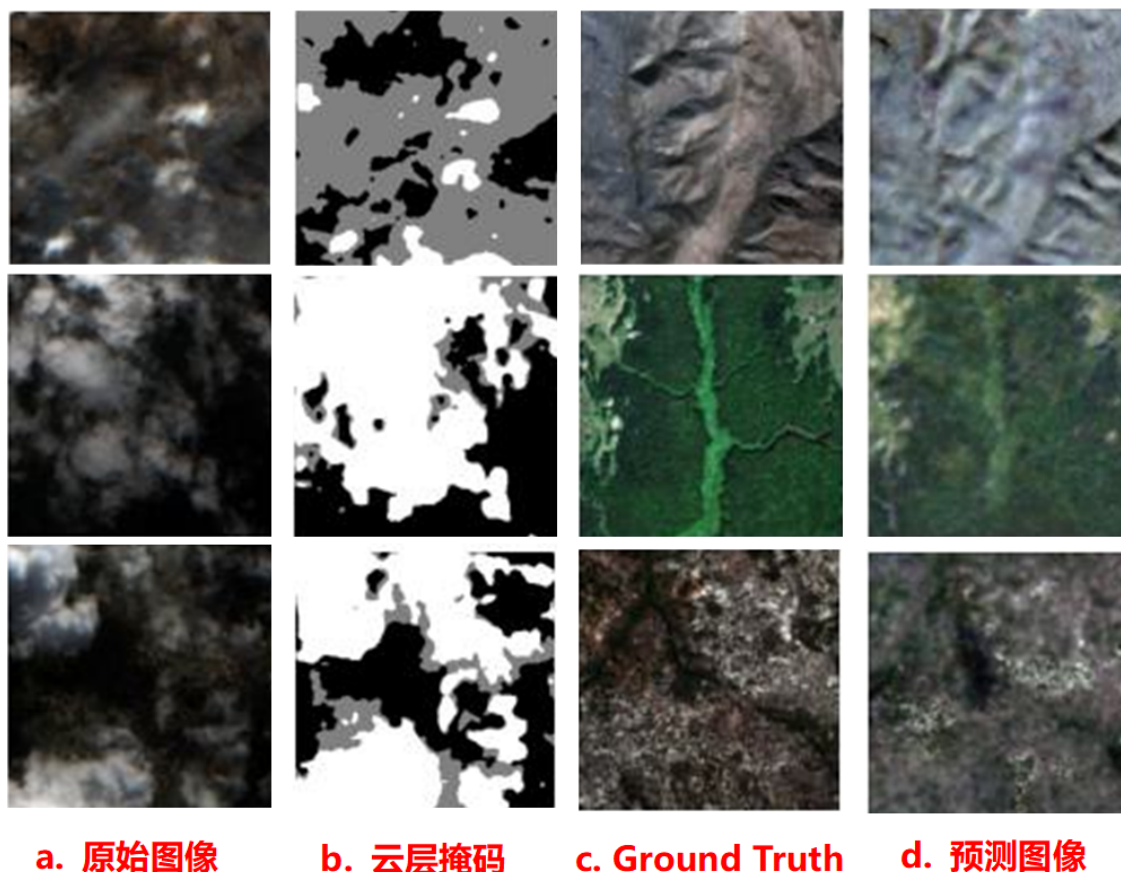


图 4: 模型一去云表现

定量分析部分。将模型一、模型二在测试集上的表现与论文原指标相比较。评估结果如下表所示，模型 1 有 SAR 图像输入，模型 2 无 SAR 图像输入。由于本人使用的数据集是从原数据集中随机选择的，因此部分指标稍微高于论文中的标准。但是可发现模型 1 的指标明显好于模型 2，即 SAR 图像输入确实对去除云层有一定的帮助

表 3: 模型结果对比

	PSNR	SSIM	SAM
论文指标	28.7	0.87	8.15
模型一	29.3	0.83	8.83
模型二	28.6	0.79	10.53

7 总结与展望

在本文中，介绍了一个用于清除单时态 Sentinel-2 卫星图像中云层的深度残差神经网络。该方法的主要特点有两个方面。一方面，作者在除云过程中加入了数据融合策略，以提供关于基于 Sentinel-1 SAR 图像的目标场景的表面特征的进一步信息。另一方面，作者提出了一个云适应性损失，以规避受云影响和无云训练图像不能同时获得的问题。

本文复现了作者所提出的相关工作，并验证了 SAR 图像辅助去云的有效性。从目视效果看，网络训练效果较好，基本能够完成去云任务，同时也能较好的分离云影与地表信息。同时由测试集的结果表明，对比无 SAR 图像输入的网络，有 SAR 图像辅助去云的效果更好。未来考虑在更大规模的数据集上进行训练与验证，同时修改模型的架构，尝试不同的网络模型，对比去云训练的效果。