

HS²P：用于多模态遥感图像云和阴影去除的分层光谱与结构保持融合网络

摘要

光学遥感图像常常受到云和阴影的污染，导致数据缺失，这极大地阻碍了持续的地表观测任务。云和阴影的去除是光学遥感图像处理中最重要的任务之一。由于主动成像的特性使合成孔径雷达 (SAR) 能够穿透云层和其他气候条件，SAR 数据被广泛用于指导光学遥感图像的云和阴影去除。然而，SAR 数据受到斑点噪声的高度干扰，这会对从光学图像中提取的光谱特征产生伪影污染，并使得 SAR-光学融合在生成云和阴影去除结果时，难以保持高光谱保真度和合理的空间结构。为了克服上述缺点，提出了一种新颖的分层光谱和结构保持融合网络 (HS²P)，通过光学和 SAR 遥感图像的分层融合来恢复光学遥感图像中的云和阴影区域。HS²P 是一种深度层次结构，其中包含堆叠的残差组 (ResGroups)，这种结构逐步约束了重建过程。为了自适应选择更多具有信息量的特征进行融合，并减少对光学数据中的云和阴影带来的伪影特征或 SAR 中的斑点噪声特征的注意力，使用带有通道注意力机制的残差块 (RBCA)。此外，还提出了一种新颖的协同优化损失函数，在增强结构细节的同时保持光谱特征。在公开数据集 SEN12MS-CR 上进行的大量实验表明，复现的方法能够鲁棒地恢复各种云类型的光学遥感图像中的多种地面信息。

关键词：多模态融合；云去除；注意力机制；SAR

1 引言

持续监测地球表面对于理解世界非常重要。随着遥感技术的快速发展，光学遥感图像逐渐成为监测地球表面的主要手段。然而，光学遥感图像不可避免地会受到云层的污染，导致对地球表面的观测不连续。根据美国地质调查局 (USGS) 数据的分析，全球平均年云量约为 66%。而且，Landsat ETM+ 数据的统计显示，大约有 35% 的陆地面积被云层覆盖 [15]。云层的覆盖严重阻碍了光学遥感图像的广泛应用，对各种地球监测任务产生了巨大影响，这些任务涉及持续的监测。为了保证对地球表面的持续观测，光学遥感影像中的云和阴影去除已成为一个迫切需要解决的问题。

通常情况下，光学遥感影像中的云和阴影去除旨在利用互补信息重建被云污染的遥感影像中缺失的数据。根据辅助信息类型的差异，云和阴影去除方法可以划分为三大类：单图像重建方法、多时相融合方法和多模态融合方法。单图像重建方法通过使用剩余的空间部分或其它光谱的原始场景信息来填充缺失数据区域 [22, 31]。这些方法假设辅助的空间区域干净或光谱存在，通常无法重建大尺寸或厚实的非透明云层覆盖场景。多时相方法使用不同时期相同的场景来恢复缺失的地表信息 [19, 26]。多时相方法的局限性在于假设不同时期获取的数据

之间存在轻微的差异。然而，不能确保土地覆盖的时态稳定性，这导致多时相去云结果无法服务于细粒度监测或变化检测方法 [25]。在多模态融合方法的情况下，去云是由额外的数据源支持的 [7,8]。不同传感器具有不同的成像原理，它们捕获的多模态图像在描述场景时的重点也有很大差异。多模态融合方法可以通过融合不同模态图像中的互补信息来重建被遮挡区域。其中最受关注的话题之一是合成孔径雷达（SAR）数据与光学数据的融合。SAR 是一种全天候传感器，记录雷达的后向散射强度。由于具有强大的穿透性，SAR 能够在不考虑云层的情况下收集地面信息，从而提供互补的上下文和结构信息，充分补偿光学图像中的污染区域 [27]。因为，本文利用 SAR 数据对光学遥感影像进行云和阴影移除。

近年来，基于 SAR 数据与光学数据融合的云影去除方法表现出强大的性能。然而，由于后向散射信号的一致性处理，SAR 数据受到斑点噪声的高度影响 [16]，这会对光学图像中提取的光谱特征造成伪影污染，使得生成的云去除结果难以同时保持高光谱保真度和合理的空间结构。特别是小面积或薄云的光学图像含有大量未受污染的光谱信息，但由于斑点噪声的影响，这导致了去云结果中细节的模糊。尽管基于 SAR 与光学融合的去云和阴影移除方法在过去的几年中得到了改进，但大多数现有方法在利用 SAR 数据作为辅助输入信息时，直接将 SAR 数据和光学数据进行堆叠融合 [3,23]，忽略了斑点噪声带来的不良影响。为了减少光学图像中云和阴影产生的伪影对通道特征的注意力，同时自适应地强调更多有信息量的特征，HS²P 使用带有通道注意力机制的残差块（RBCA）构建用于云和阴影去除的深度网络。此外，一些基于 SAR-光学融合的方法通过元素损失重建与指定目标相似的光谱信息，而在云和阴影去除结果中忽略了几何结构信息 [17,23,32]。为了克服这一局限，HS²P 中使用一个协同优化损失函数，包含光谱保持损失和结构保持损失，网络能够重建更加丰富的光谱和结构信息。

考虑到上述因素，层次化光谱和结构保持融合网络（HS²P）[18] 能够基于光学数据与 SAR 数据的融合，重建云和阴影区域。HS²P 的架构设计为通过堆叠残差组（ResGroups）逐步约束重建，确保深层网络多个层面上的云去除结果质量。这种架构也有利于浅层特征的传递。为了减少云去除结果中的伪影，利用 RBCA 作为 ResGroups 的基本组件来引导网络自适应地选择更具信息量的通道特征进行融合。进一步地，开发了一种新颖的协同优化损失函数，在增强云去除结果中的结构细节的同时保留光谱特征。在公开的大型数据集 SEN12MS-CR 上进行了实验。复现结果表明，HS²P 方法能够在光学遥感影像中重建出多样性地面信息且具有更高的光谱保真度和更丰富的结构纹理。在定量评估和定性评估方面均优于当前最先进的云和阴影去除方法。

HS²P 的主要贡献可以概括如下：提出了一种名为 HS²P 的层次化光谱和结构保持融合网络，该网络逐步重建云和阴影区域；在 HS²P 中，使用了名为 RBCA 的残差块和通道注意力机制作为基本组件。内嵌的注意力模块引导网络更多地关注多模态影像中的信息特征；引入了一种协作优化损失，使 HS²P 能够学习更强大的光谱和结构特征表示，从而在去云结果中增强光谱保真度和突出结构特征。

2 相关工作

2.1 生成对抗学习驱动的云和阴影移除

随着深度学习的成熟,云和阴影移除的方法不断得到发展。生成对抗网络 (GANs) 在基于深度学习的方法中受到了极大的欢迎。GAN 由生成器和判别器组成。生成器的目标是产生判别器无法识别的图像,而判别器的目标是尽可能准确地辨别真实图像和生成图像。生成器-判别器的博弈使得 GAN 生成与相应目标相似的图像。Bermudez 等人提出了一种基于条件生成对抗网络 (cGAN) 的从配准 SAR 图像中映射无云光学图像的方法,该网络仅依赖 SAR 数据重建场景 [2]。然而,由于 SAR 数据在光谱方面缺乏信息,很难将 SAR 图像转换为精度较高的无云多光谱图像。为了解决上述问题,提出了一些通过协同利用 SAR 数据与光学数据的互补性来消除云和阴影的方法。Grohnfeldt 等人基于 Pix2Pix 开发了一种 SAR-Optical-cGAN,移除了合成的有云图像 [8]。高等研究者进一步考虑了 SAR 数据与光学数据之间的弱相关性,他们首先将 SAR 图像转换成具有强互补性的模拟光学图像,然后利用这些模拟光学图像、SAR 图像以及有云的光学图像生成无云图像 [7]。遵循这一思路,高等进一步平衡了全局损失、局部损失、感知损失和 GAN 损失 [1]。局部损失使得网络在重建缺失区域时投入更多注意力,而感知损失则使得结果在视觉感知上更佳。为了展示更加出色的云和阴影移除效果,提出了一种时空生成网络 (STGAN),该网络增加了额外的多时相信息作为输入 [24]。然而,它牺牲了时间分辨率,从而降低了持续监测的可能性。随后,Darbaghshahi 等人提出了两阶段 GANs,分别用于 SAR 到光学图像的转换和云去除,并通过使用扩张卷积来增加感受视野和防止信息丢失,改进了传统的 U-net 架构,这在对包含四个波段 (RGB 和 NIR) 的光学图像去云方面取得了进展 [3]。此外,循环 GANs 也被应用于减少对成对的有云和无云训练数据的依赖 [6]。然而,GANs 普遍存在的问题是,在训练过程中生成器和判别器难以达到理论上的完全均衡,导致生成虚假细节或意外的伪影。

2.2 残差学习驱动的云和阴影移除

深度残差网络 (ResNet) [9] 利用残差块 (ResBlocks) 作为基本组成部分。每一个 ResBlock 由多层组成,其输出是最后一层和输入的累加。通过这种方式,ResBlock 中的层被强制学习输入和输出之间的差异,这通常对应于噪声图像中的噪声干扰。残差学习还可以快速优化大型和深度网络,并保持稳定性能 [10]。许多视觉任务仅通过将普通的卷积神经网络 (CNNs) 替换为 ResNets [20] 就能进一步得到改善。因此,ResNet 常用于在云和阴影去除任务中重建受污染区域。李等人引入了一种深度残差对称连接网络 (RSC-Net),该网络被设计为一种对称架构,由多个残差卷积层和残差反卷积层组成 [17]。通过卷积层和反卷积层之间的对称连接,无云细节可以直接传递到顶层,从而减轻输入无云区域的损害。Meraner 等通过在 DSen2-CR [23] 中使用长跳跃连接实现了类似的想法。DSen2-CR 不仅处理了薄云问题,而且在重遮挡图像的云去除方面取得了优势。此外,需要使用大规模数据集来保证网络的泛化能力。为此,Meraner 等人发布了全球采样的 SEN12MS-CR 数据集,该数据集包含由有云、无云的 Sentinel-2 光学图像和 Sentinel-1 SAR 图像组成的三元组,促进了基于 SAR-光学融合的云去除研究。然而,DSen2-CR 模型无法有效表示非局部特征。为了解决这个问题,提出了一种多尺度深度残差网络 (MDRN),其中嵌入了多尺度卷积单元 [32],使得 MDRN 具有更大

的感受野，可以提取多尺度特征。何等人提出了一种可变形上下文特征金字塔 (DCFP) 模块，该模块将固定的过滤器感受野替换为基于云的形状和大小的自适应方式 [12]。然而，ResNet 在处理复杂场景时的云和阴影去除效果不佳 [30]，还需进一步改进。

2.3 注意力机制

使用注意力机制提高可以提取特征的代表性。注意力机制通过引导可用处理资源到最具信息性的输入部分，有助于图像信息的重建 [14]。Duanmu 等提出由几个密集空间注意力块 (DSAB) 组成的重建网络 [4]。DSAB 的基本组成部分是卷积块注意力模块 (CBAM)，它包含通道注意力模块和空间注意力模块。中间层的特征图通过两个连续的子模块在通道和空间维度上进行了不同程度地细化。周等进一步在多尺度注意力残差网络 (MSAR-Defognet) 中整合了通道注意力块和多尺度卷积块 [35]。此外，通道注意力机制被证明对于恢复薄云覆盖场景是有效的 [28]。为了更加关注云雾区域的恢复，徐等设计了一个注意力模块，该模块能够在他们基于注意力机制的生成对抗网络 (AMGAN-CR) [30] 中生成由云掩膜优化的注意力图。从图神经网络 (GNNs) 扩展而来，时空推理模块 (STeRe) 被提出，通过可微分的注意力机制构建长距离依赖，同时保持图中节点的空间信息，从而有效地追踪模糊或密集的目标对象 [13]。最近，He 等人提出了一种利用 Transformer 来捕捉多模态数据之间长距离依赖性的想法，并基于自注意力机制提出了一种注意力信息聚合机制，用于聚合异质信息 [11]。利用 Transformer 的优势，Xu 等人提出了一种合成孔径雷达 (SAR) 引导的全局上下文交互 (SGCI) 模块，该模块应用于基于全局-局部融合的云去除方法 (GLF-CR) 中。SGCI 模块通过合成孔径雷达特征，引导重建区域保持与无云区域一致的结构 [29]。

3 本文方法

3.1 本文方法概述

如图 1 所示，HS²P 中采用了一个数据融合模块，该模块包括一个串联层，后接一个卷积层和一个注意力模块，用于融合输入的成对合成孔径雷达图像 d_{SAR} 和光学图像 d_{OPT} 。在 HS²P 中， N 个堆叠的 ResGroup 构成了深层分层架构。在训练阶段，HS²P 中多层级的 ResGroup 在光谱项和结构项上都产生分层输出。在 ResGroup 的内部，RBCA 是基本组成部分。为了高质量地重建光谱特征和结构特征，提出了一种新的协同优化损失函数来优化 HS²P。HS²P 方法能够自适应地重建被遮挡的区域，而不依赖于准确的云和阴影检测结果。在云和阴影去除领域，一些方法利用云掩码将光学遥感图像中的有云区域和无云区域分开。然后，被掩膜的区域（有云区域）被认为是信息重建的空白区域。然而，不同厚度的云层覆盖的地面可见性各不相同。薄而半透明的云层仅遮蔽了光谱信息，仍然可以从这种情况下场景中提取丰富的特征，只有厚云完全遮挡了地面，导致几乎所有的地面信息丢失。因此，不将所有有云的区域都视为空白区域。

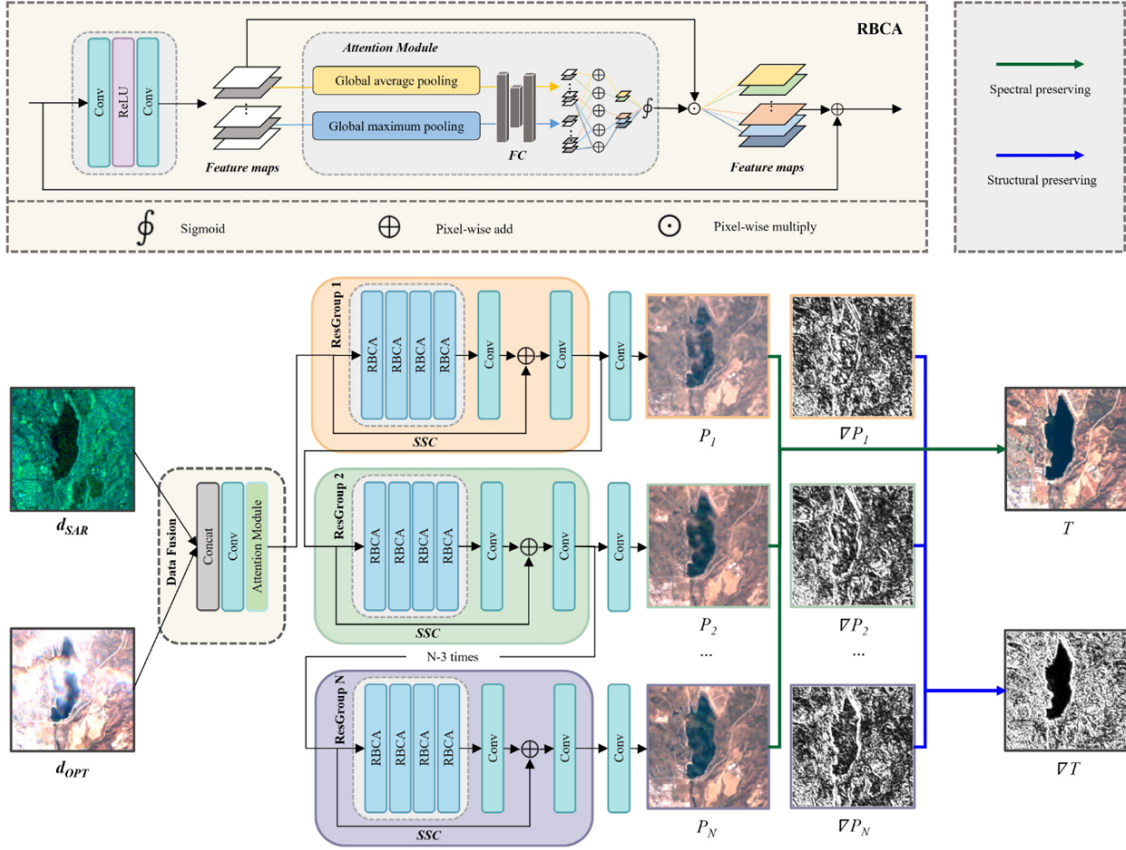


图 1. HS²P 的网络架构。最上面的矩形表示 RBCA。黑色箭头代表信息流，绿色箭头代表光谱保持功能，而蓝色箭头代表结构保持功能

3.2 深度分层架构

如图 1 所示，堆叠的 ResGroups 构成了 HS²P 的主干部分。在 N 个 ResGroups 之后，得到了 HS²P 的最终输出。ResGroups 的信息流可以表示为：

$$F_g = RG_g(F_{g-1}) = RG_g(RG_{g-1}(\dots RG_1(F_0))) \quad (1)$$

其中， RG_g 代表第 g 个 ResGroup，而 $g = 1, 2, \dots, N$ ， N 是 HS²P 中 ResGroups 的总数。第 g 个 ResGroup 生成的特征图表示为 $F_g \in \mathbb{R}^{f_c \times W \times H}$ ，在这里我们从表示中省略了批次维度。 f_c 代表 f_g 中的特征图数量， W 和 H 分别代表特征图的宽度和高度。 $F_0 \in \mathbb{R}^{f_c \times W \times H}$ 是数据融合模块的输出，该模块由公式 (2) 定义。

$$F_0 = DF(d_{SAR}, d_{OPT}) \quad (2)$$

其中 DF 表示数据融合模块的功能。 $d_{SAR} \in \mathbb{R}^{f_s \times W \times H}$ 和 $d_{OPT} \in \mathbb{R}^{f_o \times W \times H}$ 分别代表输入的 SAR 数据和输入的光学数据。请注意， $f_s = \tilde{2}$ 和 $f_o = \tilde{13}$ ，与原始数据波段一致。ResGroups 在光谱和结构层面上生成分层输出。受到 [33] 的启发，在 Res-Groups 中使用短跳连接 (SSC)，使其能够在粗糙级别上学习信息并稳定训练过程。然后，在每个 ResGroup 之后应用一个卷积层，将 F_g 的维度重新排列回其原始格式 ($\mathbb{R}^{f_c \times W \times H} \rightarrow \mathbb{R}^{f_o \times W \times H}$)。在训练过程中，对于每个输入 d_{SAR} 和 d_{OPT} ，会生成 N 个清晰图像 $P_n (1 \leq n \leq N)$ ，这些图像进一步用于与相应的真实、无云的光学图像 T 进行损失计算。辅助梯度信息在减轻模糊的几何结构 [21] 中起着重要

作用。因此，从生成的 N 个无云图像中得到了 N 个梯度图 $\nabla P_n (1 \leq n \leq N)$ ，通过计算图像中相邻元素之间的差异来表示文本特征。然后，我们使用这些梯度图与真实的梯度图 ∇T 进行损失计算，该梯度图是从相应的 T 中提取的。HS²P 的分层输出由公式 (3) 给出。

$$P(d) = \{\{P_1(d), P_2(d), \dots, P_N(d)\}, \{\nabla P_1(d), \nabla P_2(d), \dots, \nabla P_N(d)\}\} \quad (3)$$

其中 $d = [d_{SAR}, d_{OPT}]$ 代表输入。 P_1 到 P_n 是生成的光谱空间中的 N 个无云图像，而 ∇P_1 到 ∇P_n 是在结构空间中的 N 个梯度图。分层的输出被应用于约束网络中多个级别的云和阴影区域的重构，这使得恢复的信息被反复细化来避免明显的失真，同时提高生成的无云结果的保真度。

3.3 RBCA

胡等人提出了一种注意力机制，并证实了注意力机制可以使网络进行特征重标定，通过这种方式，网络可以学习使用全局信息来有选择地强调信息特征并抑制不太有用的特征 [14]。受此启发，在网络中使用了嵌入通道注意力机制的 ResBlocks，将其命名为 RBCA，来组成 ResGroup。在 RBCA 中，使用一个卷积层后跟一个 ReLU 层以及另一个卷积层来提取多模态特征。然后并行使用全局平均池化层和全局最大池化层将提取的特征压缩为通道描述符，从而平衡均值和极值标准。然后将通道描述符输入到两个全连接 (FC) 层，以灵活且非互斥的学习方式实现对通道依赖性的捕获。将 FC 生成的平均通道描述符和最大通道描述符相加，然后使用 sigmoid 层进一步生成每个通道的最终权重。然后，我们使用乘法根据计算出的权重调整特征图。RBCA 的最后一步是添加输入和调整后的特征图执行残差学习。RBCA 定义如下：

$$F_b(d) = F_{b-1}(d) + W_{b-1}(d) \times \text{Conv}(\text{ReLU}(\text{Conv}(F_{b-1}(d)))) \quad (4)$$

F_b 是第 b 个 RBCA 的输出，同时也是第 $b+1$ 个 RBCA 的输入。类似地， F_{b-1} 代表第 b 个 RBCA 的输入。在第 b 个 RBCA 中产生的逐通道权重表示为 W_{b-1} 。 Conv 和 ReLU 分别表示卷积层和 ReLU 层。通过嵌入通道注意力机制，RBCA 能够以逐通道的方式自适应调整多模态特征，并促进网络表示的质量。同样的通道注意力机制同样应用于数据融合模块中，这在图 1 中有所展示。在数据融合模块之后，多模态特征被提取并以判别性方式合并，而不是简单地按通道连接。

3.4 协同优化损失

一般来说，现有的云和阴影移除方法使用 L1 损失来进行信息重建，这种方法忽视了结构信息。受到之前工作 [34] 的启发，设计了一种协同优化损失，以保留光谱信息和结构信息，该损失由光谱保持损失和结构保持损失组成。自定义损失定义为：

$$\begin{aligned} \mathcal{L}_{S^2P} &= \sum_{n=1}^N \lambda_n (\mathcal{L}_{SP}^n + \alpha \mathcal{L}_{ST}^n) \\ &= \sum_{n=1}^N \lambda_n (\|P_n - T\|_1 + \alpha \|\nabla P_n - \nabla T\|_1) \end{aligned} \quad (5)$$

其中 \mathcal{L}_{SP}^n 和 \mathcal{L}_{ST}^n 分别代表第 $n(1 \leq n \leq N)$ 个 ResGroup 的光谱保持损失和结构保持损失。 P_n 和 ∇P_n 分别代表第 n 个 ResGroup 生成的无云图像和梯度图。 α 是一个正则化常数，用于调整两个损失的权重。 λ_n 是另一个正则化常数，它表示不同 ResGroups 的权重。

\mathcal{L}_1 损失可以对异常值提供鲁棒性。因此，用 \mathcal{L}_1 损失作为光谱保持损失 \mathcal{L}_{SP} 和结构保持损失 \mathcal{L}_{ST} 。 \mathcal{L}_{SP} 的基本思想是增强预测图像和目标图像在光谱项之间的相似性。梯度图表示梯度长度，考虑梯度强度，足以揭示给定图像中局部区域的清晰度。通过计算两个给定梯度图 ∇P_n 和 ∇T 之间的 \mathcal{L}_1 损失， \mathcal{L}_{ST} 可帮助模型从梯度空间中学习并捕捉结构依赖性，从而使预测的无云图像具有与目标相似的纹理。借助自定义损失，网络在光谱和结构方面都得到了优化，从而生成不仅外观精美而且轮廓清晰的云去除结果。此外，普遍认为，网络的较深层数伴随着更强的非线性表示。在这方面， λ_n 的值被设计为随着 n 的增加而增加，这意味着更深的 ResGroups 的输出被分配给更大的权重。在实践中，利用 sigmoid 函数为 λ_n 设计了一个单调递增函数。

4 复现细节

4.1 实验环境搭建

实验采用操作系统为 Ubuntu 20.04.6 LTS，搭载 CUDA 11.4 和 cuDNN 8.7。使用 PyTorch 深度学习框架，并通过 Anaconda 创建虚拟环境。主要依赖库版本：Python 版本：3.8.20，PyTorch：1.12.1，Torchvision：0.15.2，NumPy：1.24.2，Matplotlib：3.7.1。

4.2 数据集

为了充分展示 HS²P 方法的有效性，在 SEN12MS-CR 的公开大规模数据集上进行实验 [5]，该数据集包含由有云的 Sentinel-2 光学图像、无云的 Sentinel-2 光学图像和 Sentinel-1 SAR 图像组成的 triplet。公开发布的 SEN12MS-CR 包含 175 个非重叠的兴趣区域 (ROI)，每个区域被切割成多个 256×256 像素大小、步长为 128 像素的小块。总共有 122,218 个空间分辨率为 10 米的三元组。这些小块是在地球陆地表面和四个气象季节中采样的。根据 ROI 将所有小块分为三个子数据集，即训练集、验证集和测试集。确保每个子数据集都分布在四个气象季节中。在 SEN12MS-CR 中，选择了 Level-1C 大气顶层反射率产品作为 Sentinel-2 数据。对于 Sentinel-1 数据，选择了 Sentinel-1 IW Level-1 GRD 产品，其数值为转换到分贝 (dB) 规模的后向散射系数。为了减少由于建筑物变化或植被变化可能引起的时序差异，确保所有来自同一场景的三元组都在同一气象季节内获取。在实验中，利用了 Sentinel-1 SAR 数据的两个极化通道 (VV 和 VH)。为了充分利用光谱信息，使用了 Sentinel-2 光学数据中的 13 个波段 (B1、B2、B3、B4、B5、B6、B7、B8、B8a、B9、B10、B11 和 B12)。

4.3 定量评价指标

在实验中，使用了多种常见的指标来定量评估模型的性能。这些指标包括平均绝对误差 (MAE)、均方根误差 (RMSE)、峰值信噪比 (PSNR)、光谱角映射器 (SAM) [51] 和结构相似性 (SSIM)。MAE 和 RMSE 都是常用的元素级误差指标。不同的是，RMSE 首先计算误差的平方，这会放大较大的偏差。MAE 和 RMSE 的值越低，表示评估图像的精度越高。PSNR

是另一种用于评估恢复图像质量的元素级指标。PSNR 值与预测的质量成正比。SAM 是一种图像级度量，它将光谱视为高维向量，并通过计算向量之间的角度来量化两个给定图像之间的相似性。SAM 的值越低，目标和预测之间的相似性越高。SSIM 也是图像级的，旨在通过量化两幅图像在亮度、对比度和结构方面的差异来捕捉结构相似性。SSIM 值在 $[0,1]$ 范围内，与预测图像的结构质量呈正相关。对于多光谱图像，SSIM 是通过在每个波段的单独计算取平均值来计算的。

4.4 实验设置

所提出的模型在 NVIDIA GeForce GTX 1080 Ti 上训练了 30 个 epoch。在数据准备方面，使用排序改组、尺寸裁剪（ 128×128 像素）、随机旋转和翻转操作创建了 patch。输入光学数据被裁剪到 $[0,10,000]$ 范围内，而 SAR 数据的 VV 极化和 VH 极化的裁剪范围分别为 $[-25,0]$ 和 $[-32.5,0]$ 。另外，所有 Sentinel-2 数据的波段都被除以 2000 以保证数值稳定性。类似地，对 Sentinel-1 SAR 数据也应用了缩放操作，来匹配光学数据的分布。在训练过程中，每个训练周期使用验证集来评估训练模型的性能，并选择表现最好的模型。由于 PSNR（峰值信噪比）和 SSIM（结构相似性指数）是最广泛使用的图像客观评价指标，最佳性能的模型是通过在每个训练周期后使用它们来评估验证集来确定的。具体来说，验证集上 PS （即 $PS = PSNR + SSIM \times 10.0$ ）最高的模型将被选中进行测试。在测试阶段，预测结果与目标进行比较评估训练模型的性能。为了可视化去云结果，预测图像被乘以 2000 以恢复到原始范围 $[0,10,000]$ 。然后，通过拼接给定 Sentinel-2 图像的 B4、B3 和 B2 光谱来合成 RGB 图像。由于 Sentinel-1 SAR 数据采用极化成像模式，所以使用单一 VV 波段的灰度图像进行展示。

5 实验结果分析

5.1 训练过程

训练损失和验证损失是评估模型性能的重要指标，二者的关系可以反映模型的训练状况和泛化能力。训练损失是在训练集上计算的误差，通常会随着训练过程逐渐下降，而验证损失是在验证集上计算的误差，用于衡量模型对未见数据的性能。在训练 HS^2P 去云模型时，训练过程如图 2 所示。训练损失下降时，验证损失同步下降，表明 HS^2P 模型在学习并泛化良好，二者的趋势保持一致，并在验证损失最小时停止训练以防止过拟合。

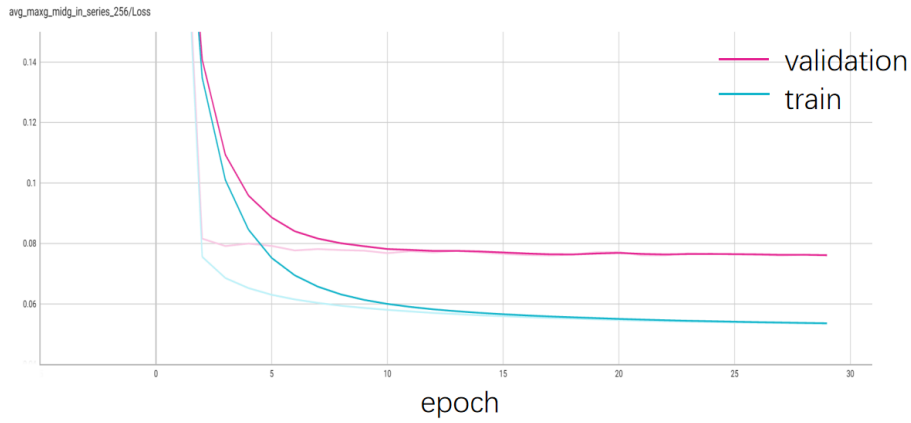


图 2. HS²P 训练过程中 loss 曲线

5.2 实验结果

HS²P 的定量结果如表 1 所示。MAE 自己跑的结果比原论文稍高，误差增加了约 0.0007，表明在绝对误差上稍逊色于原论文。RMSE 自己跑的结果 RMSE 略高，误差增加了 0.0007，与 MAE 的趋势一致，说明模型重现性稍有降低。PSNR 自己跑的结果略低于原论文，减少了约 0.1388，说明信噪比有所下降，可能暗示模型生成的图像质量稍差。SAM 自己跑的结果比原论文稍高，增加了约 0.22，说明光谱角误差有所上升。SSIM 自己跑的结果比原论文稍低，降低了约 0.0011，说明结构相似性稍逊。总体来看，自己跑的结果与原论文结果非常接近，但在各项指标上均略逊一筹。主要差异体现在 MAE、RMSE 和 SAM 上，表明模型在误差和光谱一致性上稍有下降；PSNR 和 SSIM 的差异较小，说明模型生成图像的整体质量变化不大。可能的原因包括：数据预处理方式存在微小差异；模型的训练过程或随机初始化的不同；实验环境（如硬件、软件版本）的差异。

HS²P 的云去除效果如图 3 所示，分别选取了山脉（第一行）、农田（第二行）、城镇（第三行）场景进行测试，从左到右分别是原始有云影像，有云影像对应的无云影像 GT，无云影像的梯度图，预测的无云影像，以及对应的 SAR 影像。可以看出 HS²P 的云去除结果跟原始无云图像接近，但仍存在一定程度的失真现象。

表 1. HS²P 定量结果

	MAE ↓	RMSE ↓	PSNR ↑	SAM ↓	SSIM ↑
原论文结果	0.0265	0.0376	29.3851	7.8649	0.8896
自己跑的结果	0.0272	0.0383	29.2463	8.0849	0.8885

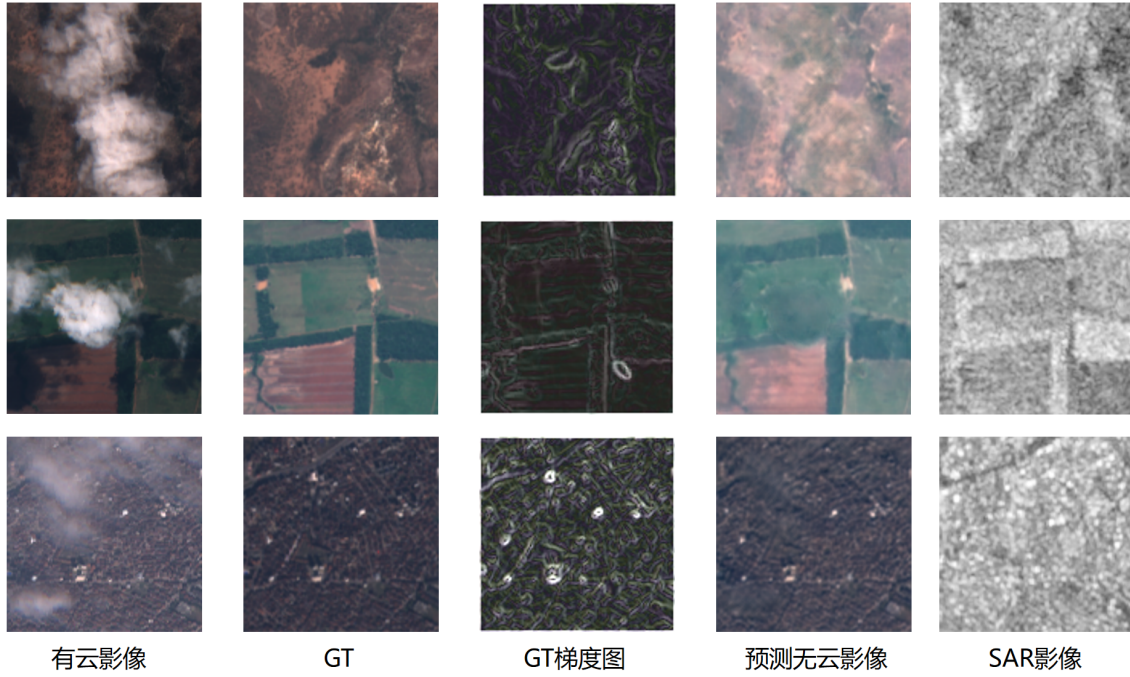


图 3. HS²P 云去除效果

6 总结与展望

光学遥感图像应用在多种领域中。然而，光学图像常常受到云层的污染。作为光学图像处理的基本步骤，云和阴影的移除为持续地面监测提供了数据支持。尽管 SAR 数据可以为光学遥感图像中的云和阴影移除提供互补的情境和结构信息，但它容易受到斑点噪声的影响，这使得 SAR-光学融合在生成高质量云移除结果方面处于不稳定状态。在 HS²P 方法中，提出了一种基于 Sentinel-2 光学数据和 Sentinel-1 SAR 数据融合的云和阴影移除方法。为了在不同网络层面上逐步限制重建，提出了一种深度层次化架构。此外，HS²P 中的 RBCA 嵌入了通道注意力机制，对更多信息特征的自适应选择进行融合，而不是平等对待多模态图像的通道特征。使用自定义协同优化损失来训练所提出的模型，使网络生成的去云结果不仅具有精细的外观，而且具有明确的轮廓。在 SEN12MS-CR 数据集上的实验结果表明，HS²P 能够处理各种类型的云层并重建多样的地表覆盖，实现了显著的改进。在未来的工作中，可以考虑尝试更多先进的深度网络架构，改善云和阴影移除的性能，并考虑融合更多遥感数据来进一步强化有云区域的重建。同时，也可以考虑利用 SAR 数据的极化信息和 SAR 成像系统的特性，更好地利用 SAR 数据作为互补数据源。

参考文献

- [1] Cloud removal with fusion of high resolution optical and sar images using generative adversarial networks. *Remote Sensing*, 12(1):191, 2020.
- [2] Jose D Bermudez, Patrick Nigri Happ, Dario Augusto Borges Oliveira, and Raul Queiroz Feitosa. Sar to optical image synthesis for cloud removal with generative adversarial net-

- works. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4:5–11, 2018.
- [3] Faramarz Naderi Darbaghshahi, Mohammad Reza Mohammadi, and Mohsen Soryani. Cloud removal in remote sensing images using generative adversarial networks and sar-to-optical image translation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–9, 2021.
 - [4] Chunjiang Duanmu and Junjie Zhu. The image super-resolution algorithm based on the dense space attention network. *IEEE Access*, 8:140599–140606, 2020.
 - [5] Patrick Ebel, Andrea Meraner, Michael Schmitt, and Xiao Xiang Zhu. Multisensor data fusion for cloud removal in global and all-season sentinel-2 imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 59(7):5866–5878, 2020.
 - [6] Patrick Ebel, Michael Schmitt, and Xiao Xiang Zhu. Cloud removal in unpaired sentinel-2 imagery using cycle-consistent gan and sar-optical data fusion. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 2065–2068. IEEE, 2020.
 - [7] Jianhao Gao, Hai Zhang, and Qiangqiang Yuan. Cloud removal with fusion of sar and optical images by deep learning. In *2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp)*, pages 1–3. IEEE, 2019.
 - [8] Claas Grohnfeldt, Michael Schmitt, and Xiaoxiang Zhu. A conditional generative adversarial network to fuse sar and multispectral optical data for cloud removal from sentinel-2 images. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 1726–1729. IEEE, 2018.
 - [9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
 - [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 630–645. Springer, 2016.
 - [11] Qibin He, Xian Sun, Wenhui Diao, Zhiyuan Yan, Dongshuo Yin, and Kun Fu. Retracted: Transformer-induced graph reasoning for multimodal semantic segmentation in remote sensing, 2022.
 - [12] Qibin He, Xian Sun, Zhiyuan Yan, and Kun Fu. Dabnet: Deformable contextual and boundary-weighted network for cloud detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16, 2021.

- [13] Qibin He, Xian Sun, Zhiyuan Yan, Beibei Li, and Kun Fu. Multi-object tracking in satellite videos with graph-based multitask modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022.
- [14] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.
- [15] Junchang Ju and David P Roy. The availability of cloud-free landsat etm+ data over the conterminous united states and globally. *Remote Sensing of Environment*, 112(3):1196–1211, 2008.
- [16] Samadhan C Kulkarni and Priti P Rege. Pixel level fusion techniques for sar and optical images: A review. *Information Fusion*, 59:13–29, 2020.
- [17] Wenbo Li, Ying Li, Di Chen, and Jonathan Cheung-Wai Chan. Thin cloud removal with residual symmetrical concatenation network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 153:137–150, 2019.
- [18] Yansheng Li, Fanyi Wei, Yongjun Zhang, Wei Chen, and Jiayi Ma. Hs2p: Hierarchical spectral and structure-preserving fusion network for multimodal remote sensing image cloud and shadow removal. *Information Fusion*, 94:215–228, 2023.
- [19] Jie Lin, Ting-Zhu Huang, Xi-Le Zhao, Meng Ding, Yong Chen, and Tai-Xiang Jiang. A blind cloud/shadow removal strategy for multi-temporal remote sensing images. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 4656–4659. IEEE, 2021.
- [20] Xiangyu Liu, Qingjie Liu, and Yunhong Wang. Remote sensing image fusion based on two-stream fusion network. *Information Fusion*, 55:1–15, 2020.
- [21] Cheng Ma, Yongming Rao, Yean Cheng, Ce Chen, Jiwen Lu, and Jie Zhou. Structure-preserving super resolution with gradient guidance. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7769–7778, 2020.
- [22] Fan Meng, Xiaomei Yang, Chenghu Zhou, and Zhi Li. A sparse dictionary learning-based adaptive patch inpainting method for thick clouds removal from high-spatial resolution remote sensing imagery. *Sensors*, 17(9):2130, 2017.
- [23] Andrea Meraner, Patrick Ebel, Xiao Xiang Zhu, and Michael 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.
- [24] Vishnu Sarukkai, Anirudh Jain, Burak Uzkent, and Stefano Ermon. Cloud removal from satellite images using spatiotemporal generator networks. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1796–1805, 2020.

- [25] Michael Schmitt, Lloyd H Hughes, Chunping Qiu, and Xiao Xiang Zhu. Aggregating cloud-free sentinel-2 images with google earth engine. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4:145–152, 2019.
- [26] Huanfeng Shen, Jingan Wu, Qing Cheng, Mahemujiang Aihemaiti, Chengyue Zhang, and Zhiwei Li. A spatiotemporal fusion based cloud removal method for remote sensing images with land cover changes. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(3):862–874, 2019.
- [27] Zhaobin Wang, Yikun Ma, and Yaonan Zhang. Review of pixel-level remote sensing image fusion based on deep learning. *Information Fusion*, 90:36–58, 2023.
- [28] Xue Wen, Zongxu Pan, Yuxin Hu, and Jiayin Liu. An effective network integrating residual learning and channel attention mechanism for thin cloud removal. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022.
- [29] Fang Xu, Yilei Shi, Patrick Ebel, Lei Yu, Gui-Song Xia, Wen Yang, and Xiao Xiang Zhu. Glf-cr: Sar-enhanced cloud removal with global–local fusion. *ISPRS Journal of Photogrammetry and Remote Sensing*, 192:268–278, 2022.
- [30] Meng Xu, Furong Deng, Sen Jia, Xiuping Jia, and Antonio J Plaza. Attention mechanism-based generative adversarial networks for cloud removal in landsat images. *Remote sensing of environment*, 271:112902, 2022.
- [31] Meng Xu, Xiuping Jia, Mark Pickering, and Sen Jia. Thin cloud removal from optical remote sensing images using the noise-adjusted principal components transform. *ISPRS Journal of Photogrammetry and Remote Sensing*, 149:215–225, 2019.
- [32] Qiaoqiao Yang, Guangxing Wang, Yaxuan Zhao, Xiaoyu Zhang, Guoshuai Dong, and Peng Ren. Multi-scale deep residual learning for cloud removal. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 4967–4970. IEEE, 2020.
- [33] Yulun Zhang, Kunkeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018.
- [34] Zhenzhou Zhang, Kun Gao, Junwei Wang, Lei Min, Shijing Ji, Chong Ni, and Dayu Chen. Gradient enhanced dual regression network: Perception-preserving super-resolution for multi-sensor remote sensing imagery. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2021.
- [35] Ying Zhou, Weipeng Jing, Jian Wang, Guangsheng Chen, Rafal Scherer, and Robertas Damaševičius. Msar-defognet: Lightweight cloud removal network for high resolution remote sensing images based on multi scale convolution. *IET Image Processing*, 16(3):659–668, 2022.