

文章编号:1001-5078(2024)11-1784-07

· 图像与信号处理 ·

## 基于多尺度注意力的 MODIS 云检测算法

张煜辉<sup>1,2</sup>, 边志强<sup>1,3</sup>, 魏倩茹<sup>4</sup>

(1. 上海卫星工程研究所, 上海 201109; 2. 西北工业大学 国家卓越工程师学院, 陕西 西安 710129;

3. 南京航空航天大学 航天学院, 江苏 南京 211106; 4. 西北工业大学 软件学院, 陕西 西安 710129)

**摘要:**云检测算法的研究可被应用于灾害预测、气象研究等领域, 本课题研究的内容是 MODIS(中分辨率光谱成像仪) 图像的云检测算法, 通过使用深度学习的语义分割算法来实现 MODIS 数据的云检测效果。本文结合 U-Net、注意力机制、多尺度网络, 设计了一种新型的深度学习模型, 该模型能够精确地检测图像中的云区域和非云区域。在实验环节, 本文介绍说明了使用的数据集以及所选取的包括近红外的数据波段等, 模型对于云检测的精确率和召回率分别为 88.58% 和 94.80%。结果表明本文设计的深度学习模型在 MODIS 图像云检测方面具有良好的性能。

**关键词:**云检测; MODIS; 深度学习; 语义分割; 注意力机制; 多尺度网络

**中图分类号:** P407.8; TP75 **文献标识码:** A **DOI:** 10.3969/j.issn.1001-5078.2024.11.020

## A cloud detection method for MODIS based on multiscale attention

ZHANG Yu-hui<sup>1,2</sup>, BIAN Zhi-qiang<sup>1,3</sup>, WEI Qian-ru<sup>4</sup>

(1. Shanghai Institute of Satellite Engineering, Shanghai 201109, China;

2. National Elite Institute of Engineering, Northwestern Polytechnical University, Xi'an 710129, China;

3. College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;

4. School of Software, Northwestern Polytechnical University, Xi'an 710129, China)

**Abstract:** The investigation into cloud detection algorithms holds significant potential for applications in disaster prediction, meteorological research, and beyond. The focus of this research endeavor lies in the development of a cloud detection algorithm tailored for MODIS imagery, leveraging the power of deep learning's semantic segmentation techniques to enhance the accuracy of cloud detection from MODIS data. This study introduces a novel deep learning model, which integrates the strengths of U-Net, block self-attention mechanisms, and multi-scale network modules, to achieve a more precise differentiation between cloud and non-cloud regions in remote sensing images. Building upon the robust foundation of the U-Net architecture, our model incorporates attention modules and multi-scale network elements. These enhancements are specifically designed to bolster the model's capability in identifying subtle features of cumulus humilis and fractocumulus clouds, addressing the limitations of traditional cloud detection algorithms in detecting thinner cloud layers. The attention mechanism employed in this work harmoniously combines block self-attention and multi-scale channel attention. The former enhances the model's sensitivity to global contextual information, thereby mitigating the challenge of poor detection in thin cloud layers. The latter, by extracting channel-wise relevant features, complements the detection of smaller cloud formations that might otherwise be overlooked. In the experimental phase,

**基金项目:**国家重点研发计划项目(No. 2022YFB3902905); 中国航天科技集团公司第八研究院产学研合作基金项目(No. SAST2023-026)资助。

**作者简介:**张煜辉, 男, 硕士研究生, 主要研究方向为计算机视觉技术。E-mail: zhangyh0814@foxmail.com

**通讯作者:**边志强, 男, 研究员, 主要从事卫星总体和姿轨控方面的研究。E-mail: 15925802118@163.com

**收稿日期:** 2024-08-23

we meticulously detail the dataset utilized, including near-infrared spectral bands among other carefully selected data channels. The evaluation results showcase the model's remarkable performance, with precision and recall rates of 88.58% and 94.80% respectively for cloud detection. These findings conclusively demonstrate the effectiveness of our designed deep learning model in accurately detecting clouds from MODIS imagery, underscoring its promising applications in advancing the field of remote sensing and related meteorological endeavors.

**Keywords:** cloud detection; MODIS; deep learning; semantic segmentation; attention mechanism; multiscale network

## 1 引言

本研究的内容是遥感图像的云检测算法。在气候变化方面,云检测可以帮助我们更好地了解云的性质和分布,以及它们对地球辐射平衡的影响。通过对云的观测和分析,可以更好地了解云的形成和演变过程,从而更好地预测和控制大气环境<sup>[1-2]</sup>。

对于浓积云层,相关云检测算法能够轻松地提取到有关于云层的特征信息,取得较为理想的云检测效果,但对于淡积云、碎积云的检测效果却仍仅能够完成小概率的识别,无法稳定地将遥感图像中的云层信息进行提取并分割。

传统的云检测算法主要被分为两种,包括基于光谱阈值的云检测算法、基于统计学的云检测算法。基于光谱阈值的云检测算法主要根据反射率、温度等物理属性获得各波段的阈值,从而达到云检测的效果,对于某些特定的传感器获得的遥感数据简单且高效,但缺乏普适性。而基于统计学的云检测算法通过提取遥感数据的物理特征送入到模型中进行训练,该种方法的性能取决于手工选取的特征的质量,且由于遥感环境的多样性,在实际使用场景中受到限制。付华联等,许赞等均提出了基于随机森林的遥感图像云检测算法,前者针对风云2号G星云检测产品创建数据集,选取5个波段中样本点的邻域灰度值、均值、方差作为训练特征<sup>[3]</sup>,后者通过自行搜集云雪雾数据创建数据集,选取灰度均值、灰度标准差、平均梯度、信息熵、对比度、逆差矩、自相关性、分数维等八种数据作为训练特征<sup>[4]</sup>。

近些年来,许多基于深度学习的算法被提出,并被应用于云检测领域<sup>[5-6]</sup>。目前已提出的基于CNN的强监督云检测算法大部分是基于自编码器的算法,即编码器-解码器(Encoder-Decoder)结构<sup>[7]</sup>。尽管编码器提取得到的高层特性具有良好的语义信息,可以对图像中的大部分像素进行粗粒度的预测,但是连续的下采样操作会导致图像的高频信息丢失,进而影响云检测性能。为解决该问题,

一些方法参考如U-Net的手段,来将底层特征和更多的高频信息引入到解码器中<sup>[8]</sup>。除此之外,Wang等人通过引入注意力机制来增强跳跃连接中的底层语义信息<sup>[9]</sup>。但是,Wang等人分析指出,当属性卷积的膨胀率超过一定程度时,会对模型的性能产生不利影响<sup>[10]</sup>。针对多通道扫描成像辐射仪(AGRI)获取的遥感信息,高军等人提出了使用反向传播神经网络的云检测算法,利用基于归一化动云指数的动态阈值云检测算法建立数据集,再使用神经网络拟合云层和各波段物理量之间的数学关系<sup>[11]</sup>。

但以上基于光谱阈值的算法、基于统计学的算法和深度学习算法均不能稳定的对输入图像中淡积云、碎积云进行检测。为解决常规云检测算法对于该种云层检测效果一般的问题,本文分别采用了注意力机制<sup>[12]</sup>、多尺度网络<sup>[13]</sup>两种模块,其中注意力机制用于改善模型对淡积云、碎积云的特征提取能力,多尺度网络则用于从不同粒度的特征图提取信息,增强模型对云层的检测效果。本文所提出的云检测算法,采用了U-Net<sup>[14]</sup>为基础构建深度学习模型,随后介绍了本次研究所使用的数据集MODIS,并将该数据集应用于模型训练,得到基于MODIS数据集训练的模型权重,最后对模型进行客观的评价分析。

## 2 基本原理

### 2.1 模型总体架构

如图1所示,即为本研究提出的基于语义分割的遥感图像云检测模型的总体框架。在本研究中,输入图像为MODIS中10个波段的灰度图像,通过在通道尺度的级联操作,将10个波段的图像融合成一个特征图(Feature Map)。操作示意如下:

$$\text{input} = \text{concat}(x_1, x_2, x_3, \dots, x_{10})$$

其中,  $x_1, x_2, x_3, \dots, x_{10}$  分别对应为MODIS中10个波段的灰度图像,  $\text{concat}(\cdot)$  表示级联操作。

然后将input送入之后的通道注意力模块中,在该模块中通过一种非线性方法来聚合来自多个卷积核的信息,以融合来自于不同通道的特征信息。操作示意如下:

$$MSCA = CA(\text{input})$$

其中,  $CA(\cdot)$  表示对传入的 Feature Map 进行通道注意力的提取操作; MSCA 为通道注意力模块的输出, 可用于后续输入; input 为前述级联操作的输出信息。

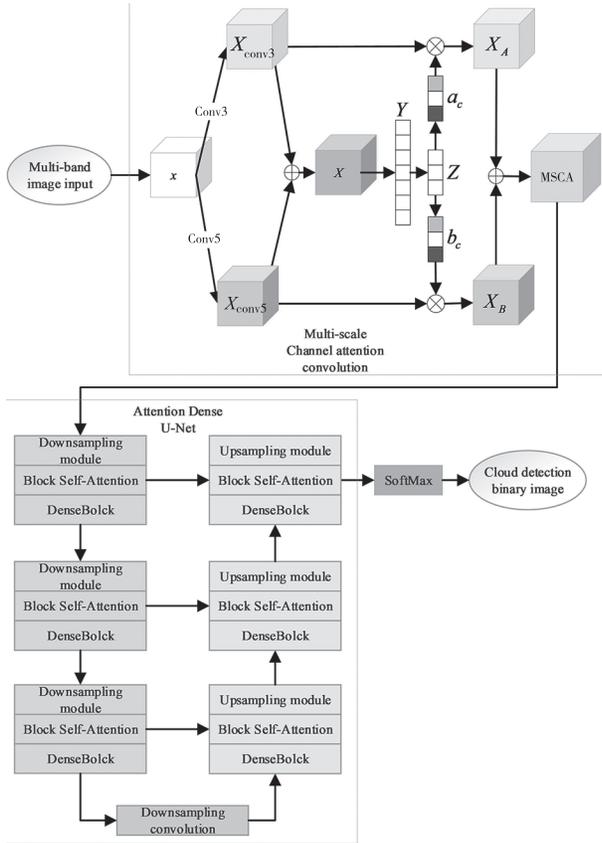


图 1 基于语义分割的遥感图像云检测模型总体框架  
Fig. 1 The overall framework of the cloud detection model for remote sensing images based on semantic segmentation

在 Attention Dense U-Net 模块中, 该模块的左右两侧分别为 Encoder 和 Decoder, 两者能分别对输入的特征图下采样和上采样, 提取不同粒度的特征信息, 再通过相同层级的跳跃连接来实现不同粒度信息的融合, 最终输出  $2 \times H \times W$  的矩阵,  $H, W$  为原始输入图像的尺寸, 分别为高度和宽度。操作示意如下:

$$Z = \text{Decoder}(\text{Encoder}(MSCA))$$

其中  $\text{Encoder}(\cdot), \text{Decoder}(\cdot)$  分别表示 Attention Dense U-Net 中对输入的 MSCA 的编码和解码操作;  $Z$  为该模块的输出。

在模型的最后, Attention Dense U-Net 输出将被送入到 SoftMax 中, 对各点像素值归一化为  $0 \sim 1$ , 用于表示对云区域的预测概率。

### 2.2 多尺度注意力机制

如图 2 所示, 为多尺度的通道注意力卷积<sup>[15]</sup>。

在输入特征图后, 参考类似 Inception 的分支计算结构<sup>[16]</sup>, 减小模型的计算参数量, 使用两个  $3 \times 3$  卷积来代替  $5 \times 5$  卷积, 该操作能够提升模型的计算效率。其操作示意如下:

$$X_{\text{conv}3} = \text{Conv}3(x)$$

$$X_{\text{conv}5} = \text{Conv}3(\text{Conv}3(x))$$

其中,  $x$  为输入的特征图;  $X_{\text{conv}3}, X_{\text{conv}5}$  为  $3 \times 3$  卷积和  $5 \times 5$  卷积的输出;  $\text{Conv}3(\cdot)$  为  $3 \times 3$  卷积操作。

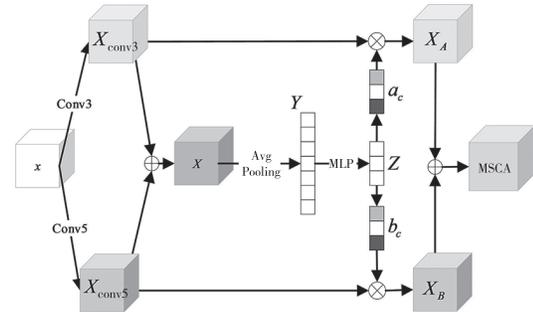


图 2 多尺度通道注意力卷积

Fig. 2 Multi-scale channel attention convolution

分支卷积后得到的  $X_{\text{conv}3}, X_{\text{conv}5}$  携带有不同尺度的信息, 通过对不同分支特征图的矩阵元素加法融合各种层次的语义特征信息。对于来自得到两个分支卷积得到的特征图, 依次通过平均池化和多层感知机, 生成多个尺度的通道注意力操作结果, 其中全局平均池化操作可以生成用于统计通道信息的全局特征信息, 多层感知机类似 SENet<sup>[17]</sup> 的通道注意力结构, 用于创建一个通道维度紧凑的特征, 保存原始输入特征图中的重要信息, 为本文所提出的模型提供更为精确和自适应的特征搜索空间, 其结构为 2 个串行的全连接层。操作示意如下:

$$X = X_{\text{conv}3} + X_{\text{conv}5}$$

$$Y = \text{AvgPooling}(X)$$

$$Z = \text{MLP}(Y)$$

其中,  $X_{\text{conv}3}, X_{\text{conv}5}$  为  $3 \times 3$  卷积和  $5 \times 5$  卷积的输出;  $X, Y, Z$  分别对应于矩阵元素加法、全局平均池化层、多层感知机的输出;  $\text{AvgPooling}(\cdot)$  为全局平均池化层操作;  $\text{MLP}(\cdot)$  为多层感知机操作。

为了能够得到跨通道关注信息, 本模块使用两个权重矩阵  $A, B$  来对初始得到分支卷积的输出  $X_{\text{conv}3}, X_{\text{conv}5}$  进行加权计算操作, 最终进行求和得到多尺度通道注意力卷积模块的输出信息。其操作示意如下:

$$a_c = \frac{e^{A \cdot Z}}{e^{A \cdot Z} + e^{B \cdot Z}}$$

$$b_c = \frac{e^{B_c Z}}{e^{A_c Z} + e^{B_c Z}}$$

$$MSCA_c = a_c \cdot X_{\text{conv}3} + b_c \cdot X_{\text{conv}5}$$

$$MSCA = [MSCA_1, MSCA_2, \dots, MSCA_n]$$

其中,  $a, b$  为使用 SoftMax 对  $Z$  计算得到软关注信息向量;  $a_c, b_c$  分别表示  $a, b$  的第  $c$  个元素, 且  $a_c + b_c = 1$ ;  $Z$  为上述多层感知机的输出;  $A, B$  生成的权重矩阵,  $A_c, B_c$  分别代表  $A, B$  的第  $c$  行元素;  $MSCA_c$  为使用多尺度通道注意力卷积生成第  $c$  行的特征映射,  $[MSCA_1, MSCA_2, \dots, MSCA_n]$  表示将生成的各行特征映射以通道维度进行组合, 计算得到最终的多尺度通道注意力卷积的输出 MSCA。

### 2.3 Attention Dense U-Net

如图 3(a) 所示, 为 Attention Dense U-Net 单元层的下采样模块。对于输入到下采样模块的特征图, 模型将会操作其通过两个相似的  $3 \times 3$  卷积 - 批归一化-ReLU 激活的操作, 完成对特征图的下采样。操作示意如下:

$$\text{down}_{x_2} = \text{ReLU}(\text{BatchNorm}(\text{Conv}(\text{down}_{x_1})))$$

$$\text{down}_{x_3} = \text{ReLU}(\text{BatchNorm}(\text{Conv}(\text{down}_{x_2})))$$

其中,  $\text{down}_{x_1}, \text{down}_{x_2}, \text{down}_{x_3}$  分别对应于下采样的输入、中间值、输出;  $\text{ReLU}(\cdot)$  表示激活函数;  $\text{BatchNorm}(\cdot)$  为批归一化操作;  $\text{Conv}(\cdot)$  为  $3 \times 3$  卷积操作。

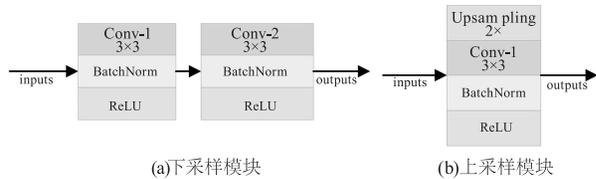


图 3 Attention Dense U-Net 单元层采样模块

Fig. 3 The sampling module of the Attention Dense U-Net unit layer

如图 3(b) 所示, 为 Attention Dense U-Net 单元层的上采样模块。对于输入到上采样模块的特征图, 模型将会操作其通过一个 2 倍上采样 -  $3 \times 3$  卷积 - 批归一化 - ReLU 激活的操作, 完成对特征图的下采样。操作示意如下:

$$\text{up}_{x_2} = \text{ReLU}(\text{BatchNorm}(\text{Conv}(\text{up\_sample}(\text{up}_{x_1}))))$$

其中,  $\text{up}_{x_1}, \text{up}_{x_2}$  分别对应于下采样的输入和输出;  $\text{up\_sample}(\cdot)$  表示对特征图的 2 倍上采样操作。

如图 4 所示, 为 Attention Dense U-Net 单元层块自注意力模块, 该模块以自注意力模块为基础<sup>[12]</sup>, 在注意力计算前进行分块操作, 便于模块从局部中

提取信息, 有利于模型在小尺寸的特征图中抽取相关云层特征。对于输入到 Attention Dense U-Net 任意单元层的特征图, 都先将其通过块自注意力机制, 提取到对于云检测效果有价值的特征信息, 然后将其送入到密集连接网络<sup>[18]</sup>中, 通过密集的跳跃连接来实现特征在通道尺度上的特征重用, 该种连接方式能够让本模型在参数量和计算成本更少的情况下增强模型性能。Attention Dense U-Net 单元层自注意力总体模块操作示意如下:

$$x_{i+1} = \text{Dense}(\text{SA}(x_i))$$

其中,  $x_i$  表示第  $i$  层的输入或输出;  $\text{SA}(\cdot)$  表示自注意力单元操作;  $\text{Dense}(\cdot)$  表示密集连接操作。

而对于密集连接模块, 其操作示意如下:

$$\text{Dense}(x_i) = H_l([x_0, x_1, \dots, x_{l-1}])$$

其中,  $[x_0, x_1, \dots, x_{l-1}]$  表示将各层级的特征图以通道维度进行级联,  $H(\cdot)$  表示一系列的卷积、批归一化、ReLU 激活、Pooling 池化操作。

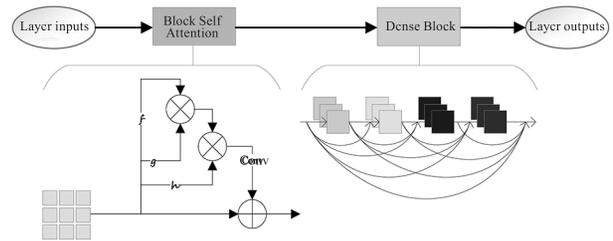


图 4 Attention Dense U-Net 单元层块自注意力总体模块

Fig. 4 The overall self-attention module of the Attention Dense U-Net unit layer

综上所述, 通道注意力机制能够帮助模型选择性地关注那些对任务最重要的特征通道, 抑制关联度低的特征。这有助于减少模型的计算复杂度, 并提高其性能; 多尺度设计使得模型能够在不同的尺度上捕捉信息, 进而实现不同尺度特征的融合。这种多尺度特征融合有助于模型更好地处理复杂场景和识别不同大小的目标。

## 3 实验设置与分析讨论

### 3.1 数据集

本研究使用的数据集来源于 MODIS, 拥有 36 个光谱波段, 覆盖从  $0.405 \sim 14.385 \mu\text{m}$  的宽光谱范围。本文中使用的 MODIS 波段为 10 个, 分别为 1, 3, 4, 18, 20, 23, 28, 29, 31 和 32, 包含有训练集用例 17895 例, 验证集用例 1200 例, 测试集用例 2250 例, 其具体参数如表 1 所示。

### 3.2 实验平台及训练超参数

本研究实验平台参数如表 2 所示。

表1 MODIS数据集参数  
Tab.1 MODSI dataset paremeters

MODSI band	MODIS wave length/ $\mu\text{m}$	usage
1	0.62 ~ 0.67	Chlorophyll uptake by vegetation
3	0.459 ~ 0.479	Soil vegetation difference
4	0.545 ~ 0.565	Green vegetation difference
18	0.931 ~ 0.941	Cloud, atmospheric properties
20	3.66 ~ 3.84	Ocean surface temperature
23	4.02 ~ 4.08	Clouds, surface temperatures
28	7.175 ~ 7.475	Mid-tropospheric humidity
29	8.4 ~ 8.7	Surface temperature
31	10.78 ~ 11.28	Clouds, surface temperature
32	11.77 ~ 12.27	Cloud height, surface temperature

表2 实验平台  
Tab.2 Experimental platform

OS	CPU	Memory	GPU
Windows 10	Intel i9 12900KF	DDR5 4800 MHz 32 GB	NVIDIA GeForce RTX 3090

训练超参数如表3所示。

表3 训练超参数  
Tab.3 Training hyperparameter

Hyperparameter	Value
learning rate	0.0001
batch size	8
dataloader	2
image size	512
loss	CE loss
learning rate scheduler	poly
epoch	100
momentum	0.9
weight decay	$5.00 \times 10^{-4}$

表3中CEloss表示为交叉熵损失,表达式如下:

$$L = \frac{1}{N} \sum_i L_i = \frac{1}{N} \sum_i - [y_i \cdot \exp(p_i) + (1 - y_i) \cdot \exp(1 - p_i)]$$

对于每个类别预测得到的概率为  $p$  和  $1 - p$ ;  $\log(\cdot)$  底数为  $e$ ,  $y_i$  表示样本  $i$  的标签信息。

Poly 学习率更新器计算表达式如下:

$$lr = base\_lr \times \left(1 - \frac{epoch}{max\_epoch}\right)^{power}$$

其中,  $base\_lr$  表示最开始设置的初始学习率;  $epoch$

表示当前模型的迭代次数;  $max\_epoch$  表示设置的最大模型迭代次数;  $power$  为权重参数。

### 3.3 评价参数

本研究对于云检测模型的检测效果,将采用以下指标进行客观的量化评价。

其中  $TP, FP, TN, FN$  分别为被模型预测为正类的正样本、被模型预测为负类的正样本、被模型预测为负类的负样本、被模型预测为正类的负样本。

$$JaccardIndex = \frac{TP}{TP + FN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$OverallAccuracy = \frac{TP + TN}{TP + TN + FN + FP}$$

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}}$$

### 3.4 消融实验

如表4所示为本研究提出的多尺度注意力 U-Net 模型 (MSA-UNet) 的消融实验的结果。

本研究的基础模型为语义分割领域的经典模型 U-Net。在将其应用于遥感图像的云检测后,对其模型结构进行微调,以使得模型所需要的输入能够满足云检测任务的目标,并删减了其中一层单元层,以减小模型的计算量,其实验结果如上所示。消融实验结果表明,本文在 U-Net 基础上加入的多尺度注意力结构和块自注意力结构能够提升模型云检测性能。

### 3.5 对比实验

如表5所示为本研究进行的对比实验结果。本文所提出的 MSA-UNet 模型在除 Precision 指标外的其他评价指标上,均取得了最佳效果。且相比起 DeepLap 在 Precision 取得的精度,本文所提出的 MSA-UNet 模型与其差距仅为 0.09%。以上结果表明 MSA-UNet 相比以上其他语义分割算法,其在云检测方面的性能更优。

### 3.6 热力图

如图5所示为 MSA-UNet 中经过单层块自注意力、多层块自注意力的的中间热力图输出值,以及所对应的真值标签。

表 4 消融实验结果(单位:%)

Tab.4 Ablation results(unit:%)

Model	JaccardIndex	Precision	Recall	F1-score	OverallAccuracy	MIoU
U-Net	85.27	87.70	90.26	88.96	91.37	86.53
Attention-UNet	85.68	88.35	91.66	89.97	92.34	86.82
MS-UNet	85.79	88.16	92.45	90.25	92.76	86.94
MSA-UNet	86.06	88.58	94.80	91.58	93.23	87.22

表 5 对比实验结果(单位:%)

Tab.5 Contrast results(unit:%)

Model	JaccardIndex	Precision	Recall	F1-score	OverallAccuracy	MIoU
U-Net	85.27	87.70	90.26	88.96	91.37	86.53
FCN	83.20	87.28	86.70	86.99	90.28	84.35
DeepLap	84.70	88.69	92.79	90.69	92.84	87.08
MSA-UNet	86.06	88.58	94.80	91.58	93.23	87.22

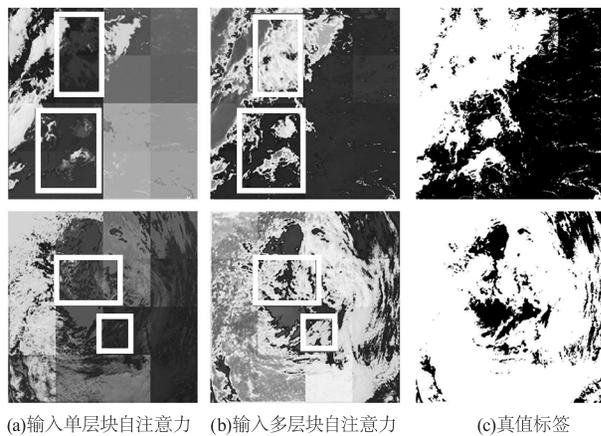


图 5 热力图

Fig.5 Heat map

如图中方框标注区域,所输入的待云检测图像在经过单个块自注意力层后提取到的云层注意力信息已经开始初步作用于模型最终的输出结果,但其仅大致局限于输出厚积云层的注意力信息,并未提取到淡积云、碎积云层的特征;但在经过多个块自注意力层后得到的热力图输出显示,该模块在得到厚积云层信息的基础上,提取到了云层中厚薄信息,相比单层块自注意力中局限的厚积云特征,多层块自注意力能够更精确地分辨输入图像中淡积云和碎积云的语义信息,并使用注意力模块对该种类型的语义信息进行注意力分数的提升,最终作用于模型的输出,兼顾多种类型的云层。

### 3.7 云检测效果图

如图 6 所示为本文所提出的 MSA-UNet 模型与 UNet 的云检测效果对比图,左侧两列展示输入图

像,为方便观察仅展示 1 波段和 20 波段,波长分别为  $0.62 \sim 0.67 \mu\text{m}$  和  $3.66 \sim 3.84 \mu\text{m}$ ;第三列为标签真值数据,第四列为 MSA-UNet 的推理值,第五列为 UNet 的推理值。模型的推理值能够在像素层级上精确地完成云检测的任务目标。

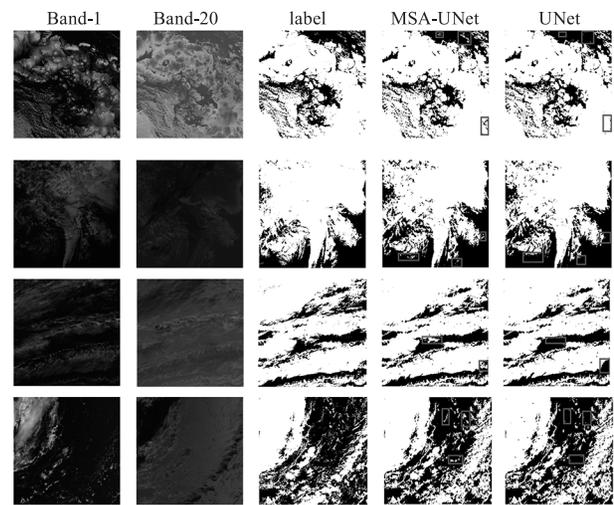


图 6 MSA-UNet 模型云检测效果图

Fig.6 MSA-UNet model cloud detection samples

## 4 结论

本文基于语义分割领域的经典模型 U-Net,通过引入自注意力模块和多尺度通道注意力模块,对遥感图像云检测算法进行了深入的研究。在综合考虑算法在复杂场景下的云检测性能后,本文结合自注意力模块和多尺度网络模块,提出了一种多尺度的注意力机制神经网络,命名为 MSA-UNet。这种模型在继承 U-Net 优秀结构的基础上,通过引入注意力机制和多尺度特征提取能力,提升了云检测的准

准确性和鲁棒性。与基础模型 U-Net 相比,MSA-UNet 在各项量化评价指标上均展现出更大的优势,尤其是在处理复杂场景时表现更为出色。因此,MSA-UNet 模型为遥感图像云检测任务提供了一种高效且可靠的解决方案,对于提升遥感图像处理的自动化程度和准确性具有重要意义。

综上所述,本文通过引入自注意力模块和多尺度网络模块,成功地对 U-Net 模型进行了改进,并提出了一种多尺度的注意力机制神经网络 MSA-UNet。该模型在云检测任务中表现出色,为相关领域的研究和应用提供了新的思路和方法。在将来的研究中,将着眼于提升云检测算法的通用性能,而不仅限于 MODIS 数据,达到一次训练重复使用的效果。

### 参考文献:

- [1] Sun Yunzhu, Jiang Guangwei, Li Yunduan, et al. Overview and application prospect of Gaofen-5 satellite [J]. *Space Return and Remote Sensing*, 2018, 39(3): 1-13. (in Chinese)  
孙允珠, 蒋光伟, 李云端, 等. “高分五号”卫星概况及应用前景展望[J]. *航天返回与遥感*, 2018, 39(3): 1-13.
- [2] Sun Yunzhu, Jiang Guangwei, Li Yun, et al. Hyperspectral observation satellite and the application prospect of [J]. *Journal of Shanghai Aerospace*, 2017, (3): 1-13. (in Chinese)  
孙允珠, 蒋光伟, 李云端, 等. 高光谱观测卫星及应用前景[J]. *上海航天*, 2017, 34(3): 1-13.
- [3] Fu Hualian, Feng Jie, Li Jun, et al. Based on random forest FY-2 g cloud detection method [J]. *Bulletin of Surveying and Mapping*, 2019, (3): 61-66. (in Chinese)  
付华联, 冯杰, 李军, 等. 基于随机森林的 FY-2G 云检测方法[J]. *测绘通报*, 2019, (3): 61-66.
- [4] Xu Yun, XU Aiwen. Cloud, snow and fog classification detection of remote sensing image based on random forest [J]. *Remote Sensing for Land and Resources*, 2021, 33(1): 96-101. (in Chinese)  
许赞, 许艾文. 基于随机森林的遥感影像云雪雾分类检测[J]. *国土资源遥感*, 2021, 33(1): 96-101.
- [5] Le Goff M, Tourneret J Y, Wendt H, et al. Deep learning for cloud detection [C]//8th International Conference of Pattern Recognition Systems (ICPRS 2017), IET, 2017: 1-6.
- [6] Shi M, Xie F, Zi Y, et al. Cloud detection of remote sensing images by deep learning [C]//2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE, 2016: 701-704.
- [7] Kingma D P, Welling M. An introduction to variational autoencoders [J]. *Foundations and Trends® in Machine Learning*, 2019, 12(4): 307-392.
- [8] Yan Z, Yan M, Sun H, et al. Cloud and cloud shadow detection using multilevel feature fused segmentation network [J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, 15(10): 1600-1604.
- [9] Wang F, Jiang M, Qian C, et al. Residual attention network for image classification [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 3156-3164.
- [10] Wang P, Chen P, Yuan Y, et al. Understanding convolution for semantic segmentation [C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018: 1451-1460.
- [11] Gao Jun, Wang Kai, Tian Xiaoyu et al. Cloud detection algorithm of Fengyun-4 remote sensing image based on BP neural network [J]. *Journal of Infrared and Millimeter Wave*, 2018, 37(4): 477-485. (in Chinese)  
高军, 王恺, 田晓宇, 等. 基于 BP 神经网络的风云四号遥感图像云检测算法 [J]. *红外与毫米波学报*, 2018, 37(4): 477-485.
- [12] Shaw P, Uszkoreit J, Vaswani A. Self-attention with relative position representations [J]. *arXiv preprint arXiv: 1803.02155*, 2018.
- [13] Fan T, Wang G, Li Y, et al. Ma-net: A multi-scale attention network for liver and tumor segmentation [J]. *IEEE Access*, 2020, 8: 179656-179665.
- [14] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, part III 18. Springer International Publishing, 2015: 234-241.
- [15] Wu W, Zhang Y, Wang D, et al. SK-Net: Deep learning on point cloud via end-to-end discovery of spatial keypoints [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(4): 6422-6429.
- [16] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. , 2015: 1-9.
- [17] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [18] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.