

# 复杂背景下提出的改进卷积神经网络云检测方法

高琳<sup>1\*</sup>, 盖晨曦<sup>1</sup>, 芦偲俊<sup>2</sup>, 田育星<sup>1</sup>

<sup>1</sup>沈阳理工大学信息科学与工程学院, 辽宁 沈阳

<sup>2</sup>中冶沈勘工程技术有限公司, 辽宁 沈阳

收稿日期: 2024年2月16日; 录用日期: 2024年3月18日; 发布日期: 2024年3月26日

## 摘要

对遥感影像云检测面临的下垫面复杂、厚云与雪的光谱特征相似导致同谱异物等复杂问题导致的传统的单类地物提取方法在云提取上效果不佳, 提出了一种复杂背景下的改进卷积神经网络的遥感云检测方法 MSANet。通过在浅层加入膨胀卷积以扩大首次感知野范围, 同时关注整体结构特征传递, 在解码部分加入包含多头“软”注意力的空间信息建模模块, 增强了网络对全局信息的感知能力。使用 ZY-3、38-cloud、GF1-WHU 数据集上进行验证实验, 实验结果表明该方法在复杂背景下云雪混淆的遥感影像上表现优秀, 模型能够更好地应对复杂背景下的云检测任务, 有效的提高了模型的精度。

## 关键词

遥感, 深度学习, 编码解码结构, 云提取

# An Improved Convolutional Neural Network Cloud Detection Method Proposed in Complex Background

Lin Gao<sup>1\*</sup>, Chenxi Gai<sup>1</sup>, Sijun Lu<sup>2</sup>, Yuxing Tian<sup>1</sup>

<sup>1</sup>School of Mechanical Engineering, Shenyang Ligong University, Shenyang Liaoning

<sup>2</sup>MCC Shenyang Prospecting Engineering Technology Co., Ltd., Shenyang Liaoning

Received: Feb. 16<sup>th</sup>, 2024; accepted: Mar. 18<sup>th</sup>, 2024; published: Mar. 26<sup>th</sup>, 2024

\*第一作者。

文章引用: 高琳, 盖晨曦, 芦偲俊, 田育星. 复杂背景下提出的改进卷积神经网络云检测方法[J]. 计算机科学与应用, 2024, 14(3): 66-77. DOI: 10.12677/csa.2024.143059

## Abstract

In view of the complex underlying surface of remote sensing image cloud detection and the similar spectral characteristics of thick clouds and snow, which lead to the same spectrum for foreign bodies, the traditional single-class ground object extraction method is not effective in cloud extraction. An improved convolutional neural network remote sensing cloud detection method MSANet under complex background is proposed. By adding expansive convolution in the shallow layer to expand the range of the first perception field, and paying attention to the overall structural feature transmission, the spatial information modeling module containing multiple “soft” attention is added to the decoding part, which enhances the perception ability of the network to the global information. Verification experiments were carried out on ZY-3, 38-cloud and GF1\_WHU data sets. The experimental results show that the proposed method performs well on remote sensing images with cloud snow confusion under complex background, and the model can better cope with cloud detection tasks under complex background, effectively improving the accuracy of the model.

## Keywords

Remote Sensing, Deep Learning, Encoding and Decoding Structure, Cloud Extraction

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

遥感云检测的目标是准确地识别遥感图像中的云层区域，将云层从图像中去除或标记出来，从而提高遥感图像的可用性。而在高纬度或者高海拔地区的遥感图像往往存在云与雪共存情况，由于两者的光线反射率和光学特征是非常相似的，导致的难以区分。如何解决在包含云雪的复杂背景高分辨率遥感影像云检测面临着存在边缘难以检测、以及下垫面复杂、厚云与雪的光谱特征相似导致同谱异物等复杂问题，仍是遥感数据处理过程中对云层的检测和提取是进一步分析和解译任务需要解决的首要问题之一。

目前深度学习在云检测中已经有了广泛的应用[1] [2] [3]。U-Net 是一种在许多图像分割任务中具有卓越性能的经典图像分割方法，尤其是医学图像分割任务。许多研究发现，基于 U-Net 架构的模型在卫星遥感图像分割中表现出优异的性能。例如基 U-Net 的遥感图像云检测算法 CloudU-Net [4]，CloudU-Net 使用空洞卷积代替了传统的卷积层来增加过滤器的感受野，通过批量归一化提高训练速度，防止过度拟合。尽管该算法在小物体密集的情况下有很好的表现，但它没有考虑到云的多样性，因为云的特征是不确定的。在医学图像分类、目标检测等领域，注意力机制是一种非常有效的方法[5]，可以为目标分配更多的处理资源，并且出现了大量将 U-Net 与注意力机制结合的算法。文献[6]提出 Cloud-AttU 模型采用对称的编码器-解码器结构，通过跳接操作实现高层特征和低层特征的融合，使输出结果包含更丰富的多尺度信息。这种对称的网络结构简洁稳定，显著增强了图像分割的效果。注意机制起源于人类的视觉认知科学[7]。在阅读文本或看物体时，人类往往会更加注意有关目标的详细信息，而抑制其他

无用的信息[8]。

上述文献并未有效的解决边缘难以检测、下垫面复杂、厚云与雪的光谱特征相似等问题[9] [10]。本文提出一个加入空间信息建模的编码解码网络 MSANet, MSANet 结合空间信息建模和卷积神经网络,并在编码部分加入适应云特性的膨胀卷积,弥补网络对全局信息的感知能力,进一步提升了在复杂背景下的云检测识别能力。这种方法使得模型能够更准确地捕捉云层的边界,从而提高了模型的精度。

## 2. MSANet 网络结构

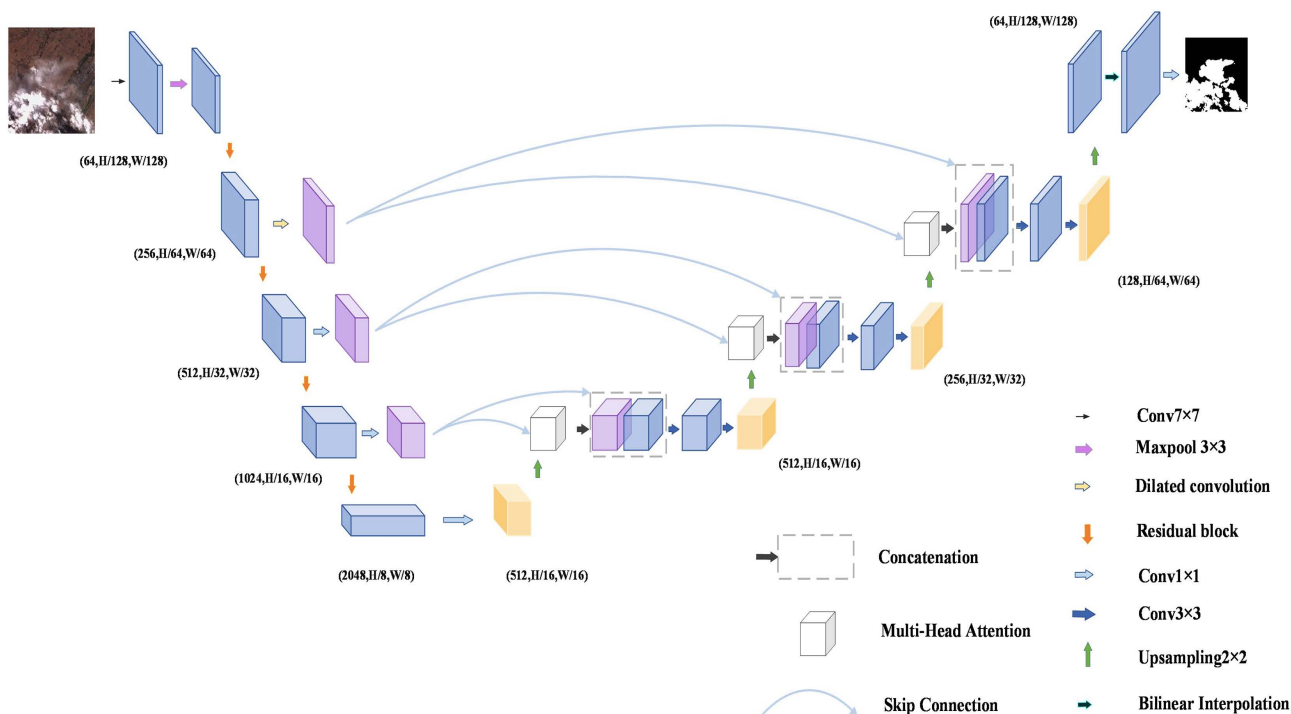


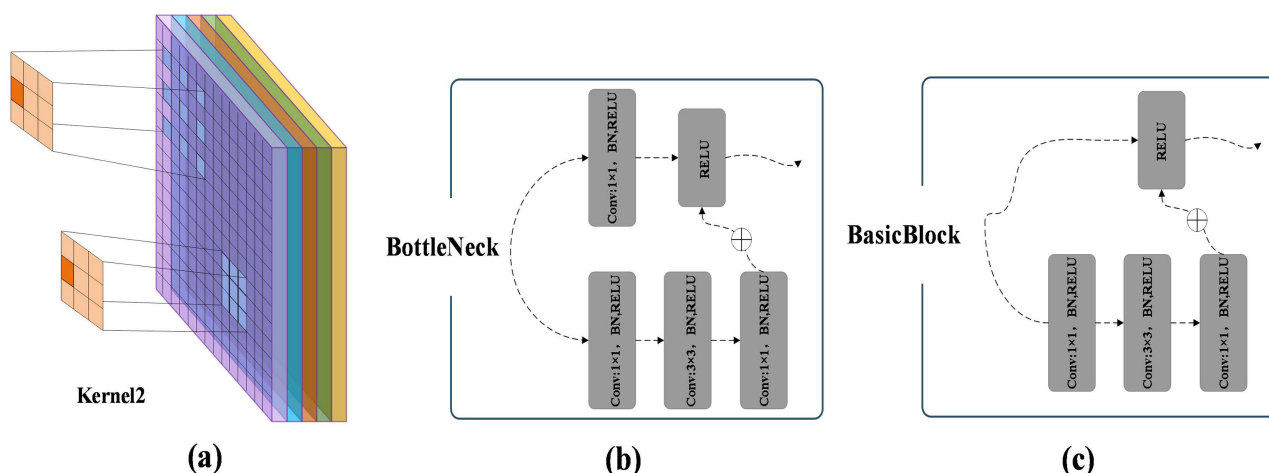
Figure 1. MSANet structure diagram

图 1. MSANet 结构图

首先提出注意力机制和卷积神经网络的融合, MSANet 结构主要包括特征提取模块和云语义预测模块。在 U-Net 的基础上, MSANet 的 backbone 使用了 ResNet50, 其扩张路径和收缩路径也存在对应关系。对于感知, 特征提取部分通过 ResNet50 的残差模块降低模型时间成本, 更好的学习复杂的特征表示以提高模型的表征能力。在将特征提取部分通过跳跃连接输入到云语义预测模块之前通过卷积改变形状以适应预测模块, 并且在第一个卷积使用了膨胀卷积, 以此来扩大感受野, 保留空间分辨率。云语义预测模块中加入了多头注意力模块。如图 1 所示, 多头注意力模块由这些白色块组成。每个多头注意力模块都基于串联特征, 空间信息建模模块接收来自编码器和解码器部分的两个输入, 并对其输出和上采样结果进行拼接和卷积操作, 以保证网络的深度。在复杂的图像分割任务中取得良好的性能。注意力机制认为网络中不同层次的特征具有不同的重要性, 通过为重要特征分配更大的权重, 它可以通知后续层更多地关注这些重要信息并抑制不重要的信息。这种改进的方法使模型能够更准确地捕捉云的空间位置信息和边界细节, 从而提高模型的精度。通过上采样和注意力模块的多次操作, 网络能够逐步恢复空间细节并进行精细的边界分从而提高云边界的准确性和清晰度。

## 2.1. 深度特征提取模块

具有“编码”意义的深度特征提取模块兼有层次深、参数量小、语义信息保留完整的功能。MSANet 选用 ResNet50 作为模型的 backbone, ResNet50 的基本组成单元就是残差块, 其中包括多个 stage, stage 0 包括一系列的卷积、BN、ReLU、MaxPooling 得到输出, 而 stage1~4 包括多个 bottleneck 和 basicblock 模块, 如图 2(b)~(c)所示。两个模块通过引入跨层的快捷连接帮助解决深度卷积神经网络都存在的梯度消失问题, 使网络更容易训练和优化。其中 basicblock 的  $\text{expansion} = 1$ , 因此输出特征图的通道数与输入相同, bottleneck 在 basicblock 多了一个右侧的卷积层, 并且  $\text{expansion} = 4$ , 也就是说此时输入与输出通道数不同。为了跳跃连接后与云语义预测模块形状相匹配, 在 ResNet50 中间部分添加了  $1 \times$  卷积以此来改变通道匹配形状。为了更好的接收云的特征, 将感受野扩大并且要保持输入图像的空间分辨率, 将第一个卷积替换为  $\text{dilation} = 2$  的膨胀卷积, 如图 2(a)所示, kernel1 为  $\text{dilation} = 2$  的膨胀卷积, 而 kernel2 是普通卷积。通过一系列的残差块和卷积层, 通道数逐渐增大, 高宽逐渐变小, 最后得出一个形状为  $[2048, 8, 8]$  的初步有效特征层。



**Figure 2.** (a) Illustration of the fusion of dilated convolution and deep feature extraction module; (b) Illustration of the bottleneck module in ResNet50; (c) Illustration of the basicblock module in ResNet50.

**图 2.** (a) 膨胀卷积与深度特征提取模块融合的图示; (b) ResNet50 中 bottleneck 模块的图示; (c) ResNet50 中 basicblock 模块的图示。

## 2.2. 云语义预测模块

该模型引入了具有“解码”功能的云语义预测模块, 并结合了空间信息建模模块, 如图 3 所示。空间信息建模模块中的每个注意力头都是一个“软”注意力模块, 由多种卷积和最终的注意力分数组成。这种注意力模块允许模型根据输入的不同特征自适应地分配注意力。与传统的注意力模块不同, 软注意力模块能够接受具有空间维度的输入表示, 如图像、特征图或其他类型的数值数据。来自跳跃连接的  $x_1$  输入和上一层生成的  $x_2$  输入都经过  $1 \times 1$  卷积处理, 使它们具有相同数量的通道数。由于  $x_2$  来自于  $x_1$  的下一层, 其尺寸是  $x_1$  的一半, 因此对  $x_1$  进行了下采样操作。然后, 它们被相加, 并通过 ReLU 激活函数, 经过另一个  $1 \times 1$  的卷积和 sigmoid 激活函数处理, 得到一个介于 0 到 1 之间的重要性分数。这个分数分配给特征图的每个部分, 然后用这个注意力图乘以跳跃连接的  $x_1$  输入, 生成该注意力块的最终输出。将同样的 4 个或 8 个软注意力模块的输出结果进行 concat, 然后对不同的注意力头进行平均池化, 得到最终输出。

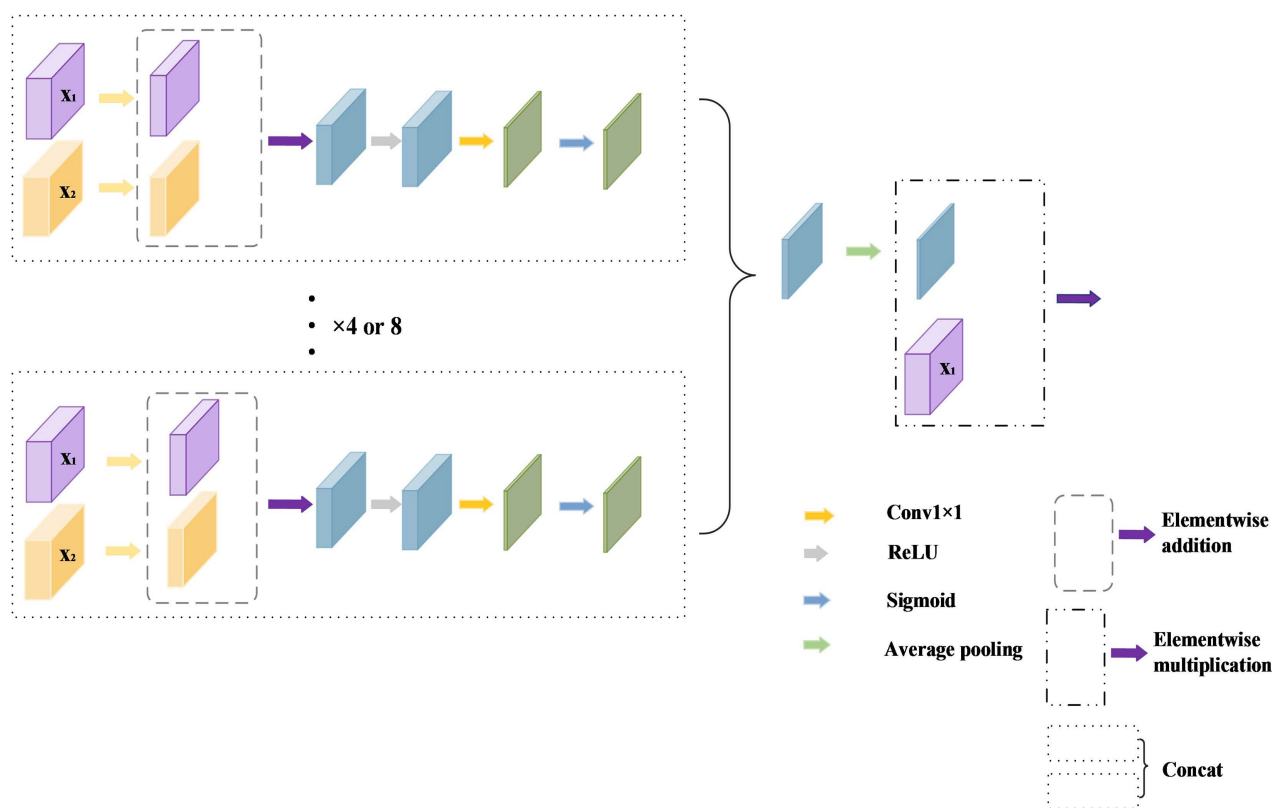


Figure 3. Workflow of spatial information modeling module

图 3. 空间信息建模模块的工作流程

MSANet 采用了一个 U 字形的架构。语义云预测模块是为了组建一个更精确的输出。该模块的结构与深度特征提取模块的结构一一对应。利用这五个初步的有效特征层与上一层得到的输出进行拼接 (concat) 然后进行特征融合, 特征融合的方式就是对特征层进行上采样并且进行堆叠。

$$x = \text{merge}(x\_1, x\_2) \quad (1)$$

在公式(1)中,  $x\_1$  是来自深度特征提取模块跳跃连接的输出,  $x\_2$  是来自上层卷积输出的特征。经过第五个卷积池化块后, 获得一个大小为 [16, 16, 512] 的特征层, 随后通过上采样操作。最终, 利用  $1 \times 1$  卷积进行通道调整, 并通过跳跃连接将其输出的  $x\_1$  与第五个卷积池化块上采样之后的  $x\_2$  作为多头注意力模块的两个输入, 再将多头注意力模块的输出与  $x\_1$  拼接 (concat) 到一起得到一个 [32, 32, 256] 的特征层, 以此类推, 最后通过双线性插值的方法将特征层形状还原回输入图像大小, 通过应用  $1 \times 1$  卷积进行通道调整, 将最终特征层的通道数调整为类别数 (num\_classes)。

### 3. 实验与分析

在训练阶段, 我们使用 AdamW 优化器对  $256 \times 256$  的图片进行训练, 参数为: 学习率 = 0.001,  $\text{beta1} = 0.9$ ,  $\text{beta2} = 0.999$  和  $\text{epsilon} = 10\text{e}-8$ 。在我们的网络顶部, 我们设置了 Sigmoid 激活函数来整理结果。损失函数采用 Dice 损失函数。训练优化指标包括 OA, R, P, F1, IoU。训练阶段的批量大小被固定为 16。我们的网络花了大约 200 个 epochs 来收敛。

如上所述, 使用 pytorch 框架来建立和训练我们的模型。所有的训练都是以 16 个批次的规模进行的, 并使用验证集来评估训练过程中的表现。具体来说, 数据被分为以下几组: 训练(70%), 验证(25%), 测



试(5%)。主要使用的软件包包括 python 3.7、CUDA 11.6、cudnn 8.5、pytorch1.12 等。实验损失图如图 4。

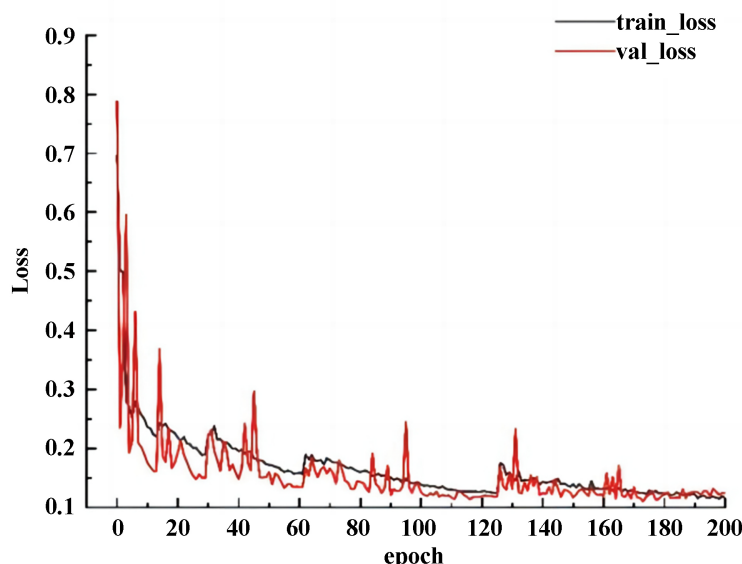


Figure 4. Loss diagram of ZY-3 dataset on MSANet

图 4. ZY-3 数据集在 MSANet 上的损失图

为了验证 MSANet 算法的有效性,在 ZY-3 卫星影像数据集上做了全面实验。我们将所提出的方法与具有相同 CNN 架构的其他方法和基于 Transformer 进行了比较。如图 6 所示,实验包括基于 Transformer 的算法、基于传统 CNN 的算法以及本章所提出的实验进行结果比较。其中 TransUNet 融合了 U-Net 与 Transformer,Transformer 模块使 TransUNet 能够捕捉图像的全局依赖关系,有助于更好地理解图像的上下文信息,这在一些需要全局信息的任务中很重要。U-Net 的结构允许 TransUNet 保留并利用图像的细节信息,特别是在分割任务中,对目标的准确定位和边界的清晰识别非常重要。通过计算不同位置的权重来捕捉全局依赖关系。TransUNet 通过梯度下降等优化算法来更新模型参数,使损失函数最小化,从而提高图像分割的准确性。虽然 TransUNet 在小尺寸或低分辨率图像上表现较好,但需要大量数据参与训练,且对局部细节信息的提取能力相对较弱。SwinUNet 引入了窗口注意力机制和 Swin Block 块,窗口注意力机制包括对输入序列进行划分、局部区域内的注意力计算、全局位置的整合等步骤。Swin Block 作为基本构建块包括位移层和窗口注意力层,其中传统的自注意力机制在计算注意力权重时需要对整个输入序列进行全局操作,这对于大尺度图像来说计算复杂度很高。Swin Transformer 通过将大尺度图像输入分割为多个固定大小的图像块(称为窗口),在每个窗口级别上计算自注意力权重,降低计算复杂度。这种混合结构使得 SwinUNet 在各种图像处理任务中更具通用性,既能处理全局关系,又能保留局部细节,适用于不同尺寸和分辨率的图像。但 SwinUNet 在处理高分辨率图像或大规模数据集时,可能会遇到资源限制的问题。

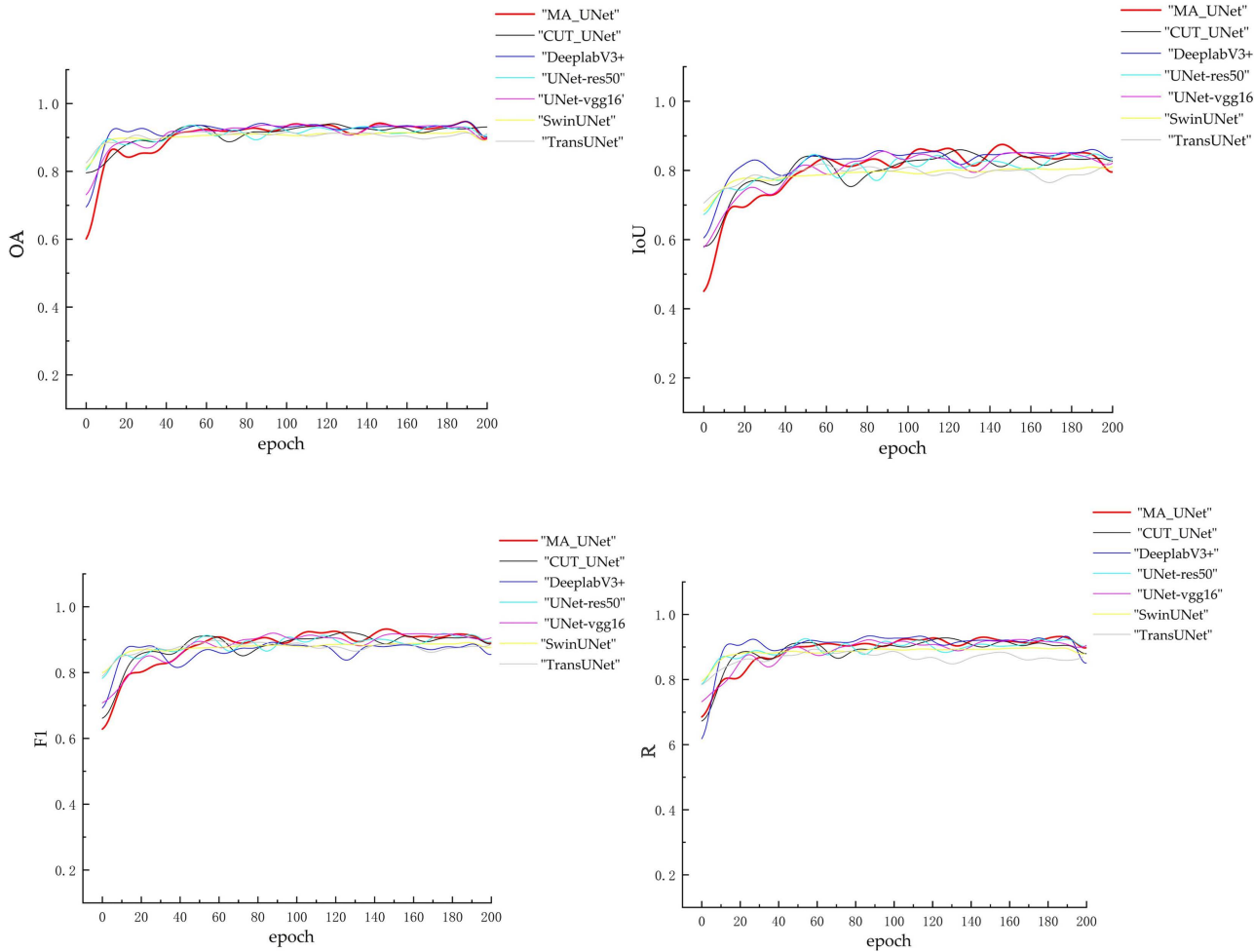
### 3.1. 结果分析

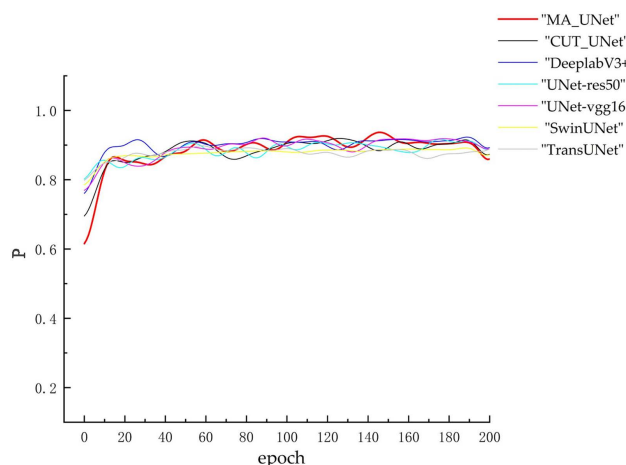
为了与传统的基于卷积神经网络的图像分割模型进行比较,引入了 U-Net 与各种 backbone 的结合网络和 DeeplabV3+进行对比实验。U-Net 是一种经典的卷积神经网络图像分割模型,由于网络结构的局部感受野和上采样操作的限制,U-Net 往往在细微边界的分割中表现不佳,包括对于全局上下文信息重要性较高的任务,U-Net 性能受限。而 DeeplabV3+是一种基于金字塔池化的卷积神经网络图像分割

模型,该模型通过引入空洞卷积和解码器模块来改进传统的 DeepLabV3+,但是对于小尺寸的目标物体,其检测和分割准确性可能会有所下降。CUT\_UNet 是以 VGG16 为 backbone 的 UNet 网络,并删减了 backbone 的下采样卷积层数,以探究卷积层数与精度的关系。MSANet 即本文提出的主要网络。表 1 中 MSANet 的多头注意力模块数量为 8。在 2.2 中消融实验中会进一步探究多头注意力模块数量与模型精度的关系。

**Table 1.** Comparison table of indicators between MSANet and various models  
**表 1.** MSANet 与多种模型的指标比较表

| 方法          | OA            | IoU           | P             | R             | F1            |
|-------------|---------------|---------------|---------------|---------------|---------------|
| TransUNet   | 0.9653        | 0.8988        | 0.9528        | 0.9409        | 0.9464        |
| SwinUNet    | 0.9189        | 0.8147        | 0.8908        | 0.9011        | 0.8941        |
| U-Net-vgg16 | 0.9704        | 0.9285        | 0.9595        | 0.9665        | 0.9628        |
| U-Net-Res50 | 0.9725        | 0.9439        | 0.9721        | 0.9703        | 0.9711        |
| DeeplabV3+  | 0.9714        | 0.9218        | 0.9589        | 0.9559        | 0.9574        |
| CUT_UNet    | 0.9746        | 0.9503        | 0.9749        | 0.9720        | 0.9734        |
| MSANet      | <b>0.9857</b> | <b>0.9510</b> | <b>0.9761</b> | <b>0.9737</b> | <b>0.9748</b> |





**Figure 5.** (a) OA indicator chart; (b) IoU indicator chart; (c) P indicator chart; (d) R indicator chart; (e) F1 indicator chart.

**图 5.** (a) OA 指标图; (b) IoU 指标图; (c) P 指标图; (d) R 指标图; (e) F1 指标图。

此表格中, IoU 和 OA 取所有轮数中的最优值, P、R 和 F1 则是取了 IoU 最优值所在的轮次的值(以下 MSANet 均指 MSANet-8 head)。图 5 即为表格中五项评价指标在 200epoch 中的走势图。从表格上可以看出, 可以看出在资源三号卫星影像数据集中 MSANet-8 head 表现突出, OA 比基于 Transformer 的 TransUNet 高 2.03%, 比基于 CNN 的 U-Net-Res50 高 1.32%。IOU 比 TransUNet 高 5.32%, 比 U-Net-vgg16 高 2.35%。从定性和定量分析的结果可以看出, 所提出的方法具有良好的性能, 并且在云分布和云含量差异较大的场景中也具有良好的检测效果。其中使用了 ResNet50 的预训练权重进行迁移学习, 通过迁移学习技术对模型进行改进, 使模型快速适应遥感影像的云端识别。

为了增加方法的普适性, 我们对遥感影像与卷积神经网络的关系进行了进一步的讨论, 数据集分别使用中分辨率的 GF1\_WHU 数据集和低分辨率的 38-cloud 数据集。为了进行实验比较的一致性, 对数据集中的所有图片统一的尺寸调整为  $256 \times 256$  大小, 所用超参数相同。

**Table 2.** Comparison table of indicators of MSANet in three datasets

**表 2.** MSANet 在三个数据集的指标比较表

| 数据集      | OA     | IoU    | P      | R      | F1     |
|----------|--------|--------|--------|--------|--------|
| ZY-3     | 0.9857 | 0.9510 | 0.9761 | 0.9737 | 0.9748 |
| 38-cloud | 0.9686 | 0.9375 | 0.9652 | 0.9708 | 0.9677 |
| GF1WhU   | 0.9920 | 0.9777 | 0.9878 | 0.9897 | 0.9887 |

如表 2 所示, 可以看出 MSANet 在 ZY-3 数据集上表现最好, 虽然 GF1\_WHU 数值较高但是相对于之前所做的对比实验并没有一个好的提升。

**Table 3.** Metric comparison table of MSANet on the same cloud dataset

**表 3.** MSANet 在同一云数据集上的指标比较表

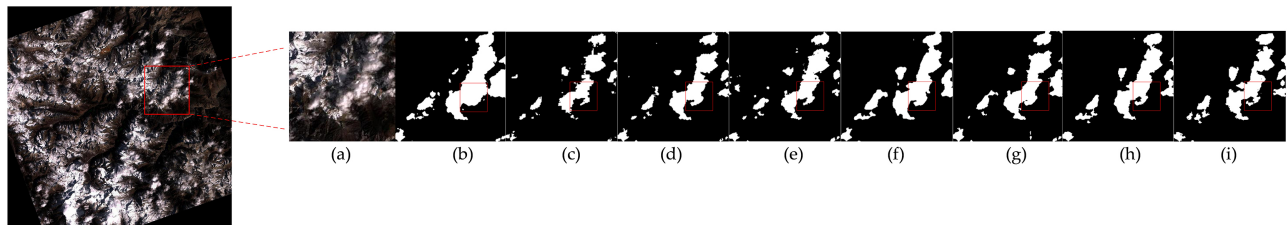
| 图片   | 方法          | OA     | IoU    | P      | R      | F1     |
|------|-------------|--------|--------|--------|--------|--------|
| 图片 1 | TransUNet   | 0.9171 | 0.9033 | 0.6792 | 0.9024 | 0.7751 |
|      | SwinUNet    | 0.9306 | 0.9217 | 0.8159 | 0.7256 | 0.7681 |
|      | U-Net-vgg16 | 0.9508 | 0.8159 | 0.8867 | 0.7905 | 0.8358 |



续表

|      |             |               |               |               |               |               |
|------|-------------|---------------|---------------|---------------|---------------|---------------|
| 图片 2 | U-Net-Res50 | 0.9221        | 0.9117        | 0.8532        | 0.6711        | 0.7513        |
|      | DeeplabV3+  | 0.9504        | 0.9400        | 0.7993        | <b>0.9172</b> | 0.8542        |
|      | CUT_UNet    | 0.9387        | 0.9297        | 0.8026        | 0.8129        | 0.8077        |
|      | MSANet      | <b>0.9641</b> | <b>0.9580</b> | <b>0.9210</b> | 0.8617        | <b>0.8904</b> |
|      | TransUNet   | 0.9246        | 0.8968        | 0.8637        | 0.8914        | 0.8773        |
|      | SwinUNet    | 0.9319        | 0.9088        | 0.9321        | 0.8357        | 0.8812        |
|      | U-Net-vgg16 | 0.9401        | 0.9197        | 0.9575        | 0.8392        | 0.8944        |
|      | U-Net-Res50 | 0.9254        | 0.9013        | 0.9383        | 0.8063        | 0.8673        |
|      | DeeplabV3+  | 0.9505        | 0.9322        | 0.9392        | 0.8944        | 0.9162        |
|      | CUT_UNet    | 0.9487        | 0.9297        | 0.9328        | 0.8949        | 0.9135        |
|      | MSANet      | <b>0.9594</b> | <b>0.9441</b> | <b>0.9575</b> | <b>0.9061</b> | <b>0.9311</b> |
|      | TransUNet   | 0.8675        | 0.7517        | 0.9068        | 0.8147        | 0.8582        |
| 图片 3 | SwinUNet    | 0.8880        | 0.7763        | 0.9223        | 0.8307        | 0.8741        |
|      | U-Net-vgg16 | 0.8764        | 0.7574        | 0.9267        | 0.8057        | 0.8619        |
|      | U-Net-Res50 | 0.8884        | 0.7752        | 0.8842        | 0.9170        | 0.9003        |
|      | DeeplabV3+  | 0.8995        | 0.7925        | 0.8851        | <b>0.9390</b> | 0.9113        |
|      | CUT_UNet    | 0.7070        | 0.7851        | 0.9128        | 0.8488        | 0.8796        |
|      | MSANet      | <b>0.9238</b> | <b>0.8446</b> | <b>0.9334</b> | 0.9274        | <b>0.9304</b> |
|      | TransUNet   | 0.8675        | 0.7517        | 0.9068        | 0.8147        | 0.8582        |

除了上表中的数值结果，对于复杂背景下云覆盖的情况下，从资源三号数据集中选出三张具有不同特点的遥感云覆盖影像进行分析，这些图像具有不同程度云层覆盖和背景，与所提出方法的比较结果如图 7~9 所示。图中从左到右依次是原始图像、TransUNet、SwinUNet、UNet\_vgg16、UNet\_res50、DeeplabV3+、CUT\_UNet、MSANet 和 Ground Truth。从视觉结果来看，总体上来看，所提出的 MAT\_UNet 方法更接近于 ground truth，其中黑色代表背景像素，白色代表云像素。

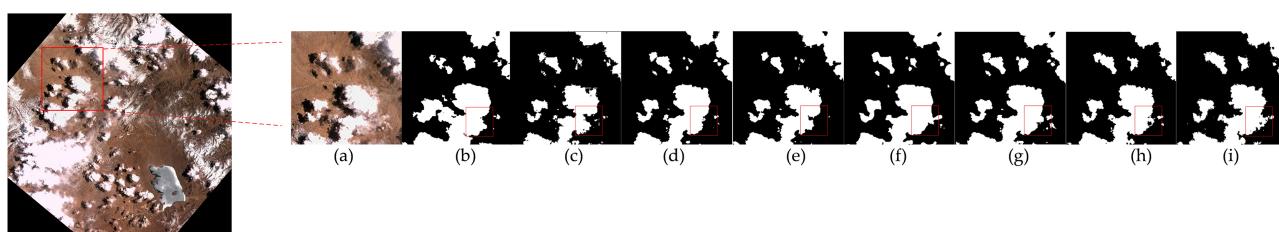


**Figure 6.** Image1 experimental results. (a) Original image; (b) TransUNet; (c) SwinUNet; (d) U-Net-vgg16; (e) U-Net-Res50; (f) DeeplabV3+; (g) CUT\_UNet; (h) MSANet; (i) Ground Truth

**图 6.** 图片 1 的实验结果. (a) Original image; (b) TransUNet; (c) SwinUNet; (d) U-Net-vgg16; (e) U-Net-Res50; (f) DeeplabV3+; (g) CUT\_UNet; (h) MSANet; (i) 真实标签

在图 6 中可以看出，云层覆盖了一个雪山。而云层和雪之间的反射是极其相似的，遥感影像中对云层检测就比较容易受到这种高反射率地物例如雪或冰因素的干扰。图 6(h)显示了 MSANet 的结果，MSANet 与图 6(b)~(c)的基于 Transformer 算法、图 6(d)~(g)中基于 CNN 的算法相比，图中红框标注的区域中，MSANet 可以清楚的将云层从含雪的复杂背景中区分开来。由此看来，MSANet 具有优异的云检

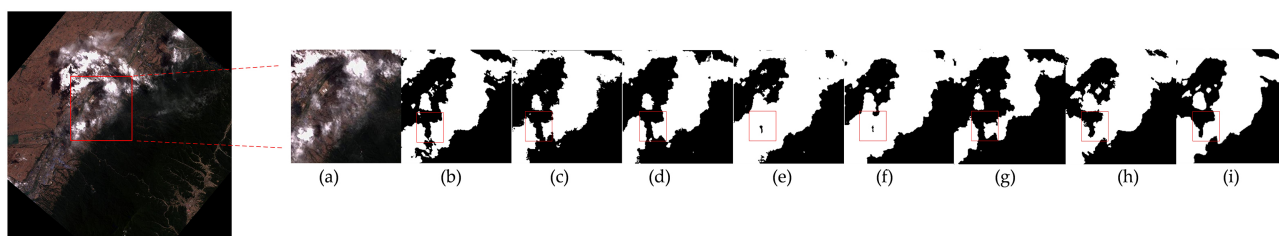
测能力,能够抵抗复杂环境中雪因素的干扰。从表 3 中图 1 的数据指标看出,MSANet 的 OA 比 TransUNet 高 4.7%,比 U-Net-vgg16 高 1.33%;MSANet 的 F1 比以 Swin Transformer 为 backbone 的 SwinUNet 高 12.23%,比 U-Net-vgg16 高 5.46%。尽管 MSANet 对云层的边界和细节更加关注,但由于其重点放在细节上,导致了 Recall 指标表现较差。



**Figure 7.** Image2 experimental results. (a) Original image; (b) TransUNet; (c) SwinUNet; (d) U-Net-vgg16; (e) U-Net-Res50; (f) DeeplabV3+; (g) CUT\_UNet; (h) MSANet; (i) Ground truth

**图 7.** 图片 2 的实验结果(a) Original image; (b) TransUNet; (c) SwinUNet; (d) U-Net-vgg16; (e) U-Net-Res50; (f) DeeplabV3+; (g) CUT\_UNet; (h) MSANet; (i)真实标签

图 7 显示了一个厚云的图像,整个图像的下垫面包括一块水域、大面积的土地和山坡。这些使云层检测任务变得困难。从图 7 中可以看出,MSANet 检测云层细节的过程要比其他两种方法更加有效。在图 7(e)中,U-Net-vgg16 的结果对于云检测来说是比较合理的。而 MSANet 的结果比 U-Net-vgg16 好,但那些邻域像素的薄云段却很模糊。这有两个原因,首先,云图像的分辨率很高,包含各种物体。云的分割与自然图像的分割不同,该任务是一个像素到像素的二元分类问题,它更注重保留高频信息。然而,我们提出的网络通过结合多头注意力模块和卷积神经网络的优势获取类间距离信息,加强对重要特征的控制,从而使得模型具有更好的空间适应性。这种整合的方法为云层边界的精细化划分提供了一种有效的解决方案,为云检测任务带来了显著的改进和提升。图 7 中的红框所标注区域中,可以明显看出 MSANet 所展现的优势。如表 3 中指标所示,MSANet 的 OA 比 SwinUNet 高 2.75%,比 U-Net-vgg16 高 1.93%;MSANet 的 IoU 比基于 Transformer 的 SwinUNet 高 3.53%,比基于 CNN 的 DeeplabV3+高 1.19%。



**Figure 8.** Image3 experimental results. (a) Original image; (b) TransUNet; (c) SwinUNet; (d) U-Net-vgg16; (e) U-Net-Res50; (f) DeeplabV3+; (g) CUT\_UNet; (h) MSANet; (i) Ground truth

**图 8.** 图片 3 实验结果。(a) Original image, (b) TransUNet; (c) SwinUNet; (d) U-Net-vgg16; (e) U-Net-Res50; (f) DeeplabV3+; (g) CUT\_UNet; (h) MSANet; (i) 真实标签

在图 8 中,有雾霾、厚薄不一的云层覆盖了一半的图像。相比之下,U-Net-Res50 和 DeeplabV3+在区分薄云和无云区域方面表现不佳。我们利用图 8 中图片 3 部分来评估 MSANet 在面对各种边界薄云和厚云时的性能。如图 8(c)所示,可以看出 DeeplabV3+无法处理细节信息,导致在细云的区分上存在困难。根据表 3 中图片 3 的指标表明,MSANet 的 OA 比基于 Transformer 高 3.58%,比 DeeplabV3+提升 2.43%;MSANet 的 F1 比 SwinUNet 提升 5.63%,比 U-Net-Res50 高 3.01%。实验结果表明,不论复杂背景还是薄厚不均的云层,我们的方法都能获得优异的结果,各项指标的结果证明了所提出的遥感云检测检测方法

的可行性和有效性。

3.2. 消融实验

为了探索所提出的 MSANet 的有效性，分别从 backbone 的不同、注意力模块数量的不同两个方面探究 backbone 的不同于注意力模块数量是否对整体模型精度产生影响。而 MSANet-s 是在 MSANet 的基础上(backbone 为 ResNet50)删除语义云预测模块中连续的卷积块，以探究 MSANet 的网络结构复杂度是否合适。从表 4 中可以看出，backbone 采用 ResNet50 并且多头注意力模块数量设置为 8 效果最好。

**Table 4.** Accuracy evaluation of different network structure complexities and different backbones  
**表 4.** 不同网络结构复杂度与不同 backbone 的精度评估

| 主干网络     | 注意力头数量 | OA            | IoU           | P             | R             | F1            |
|----------|--------|---------------|---------------|---------------|---------------|---------------|
| vgg16    | 4      | 0.9699        | 0.9317        | 0.9679        | 0.9612        | 0.9645        |
|          | 8      | 0.9684        | 0.9267        | 0.9643        | 0.9592        | 0.9617        |
| ResNet50 | 4      | 0.9667        | 0.9283        | 0.9693        | 0.9569        | 0.9626        |
|          | 8      | <b>0.9857</b> | <b>0.9510</b> | <b>0.9761</b> | <b>0.9737</b> | <b>0.9748</b> |
| MSANet-s | 4      | 0.9643        | 0.9255        | 0.9625        | 0.9600        | 0.9612        |
|          | 8      | 0.9775        | 0.9423        | 0.9699        | 0.9707        | 0.9702        |

4. 结论

提出了一种融合包含多头“软”注意力机制的空间信息建模模块的编码解码卷积神经网络的方法 MSANet。MSANet 是一个端到端的结构，主要由深度特征提取和云语义预测模块组成。实验证明，MSANet 选取了恰当的注意力头数，且 backbone 采取高效的 ResNet，对于 ZY-3 数据集这种高分辨率复杂背景遥感图像，效果相比基于 transformer 和传统卷积神经网络相比效果更好。

基金项目

项目类别名称(编号)(例如：国家自然科学基金项目(12345678))。

参考文献

[1] Li, X., Zheng, H., Han, C., *et al.* (2021) SFRS-Net: A Cloud-Detection Method Based on Deep Convolutional Neural Networks for Gf-1 Remote-Sensing Images. *Remote Sensing*, **13**, Article 2910. <https://doi.org/10.3390/rs13152910>

[2] 刘广进, 王光辉, 毕卫华, 等. 基于 DenseNet 与注意力机制的遥感影像云检测算法[J]. 自然资源遥感, 2022, 34(2): 88-96.

[3] 张家强, 李潇雁, 李丽圆, 等. 基于深度残差全卷积 网络的 Landsat8 遥感影像云检测方法[J]. 激光与光电子学进展, 2020, 57(10): 364-371.

[4] Shi, C., Zhou, Y., Qiu, B., *et al.* (2021) CloudU-Net: A Deep Convolutional Neural Network Architecture for Daytime and Nighttime Cloud Images' Segmentation. *IEEE Geoscience and Remote Sensing Letters*, **18**, 1688-1692. <https://doi.org/10.1109/LGRS.2020.3009227>

[5] Ghaffarian, S. and Valente, J. (2021) Effect of Attention Mechanism in Deep Learning-Based Remote Sensing Image Processing: A Systematic Literature Review. *Remote Sensing*, **13**, Article 2965. <https://doi.org/10.3390/rs13152965>

[6] Guo, Y., Cao, X., Liu, B., *et al.* (2020) Cloud Detection for Satellite Imagery Using Attention Based U-Net Convolutional Neural network. *Symmetry*, **12**, Article 1056. <https://doi.org/10.3390/sym12061056>

[7] Van, P., Olbrich, S. and Arns, M. (2018) Predicting Sex from Brain Rhythms with Deep Learning. *Scientific Reports*, **8**, Article No. 3069. <https://doi.org/10.1038/s41598-018-21495-7>

[8] Yang, J., Guo, J., Yue, H., *et al.* (2019) CDnet: CNN-Based Cloud Detection for Remote Sensing Imagery. *IEEE*

- 
- Transactions on Geoscience and Remote Sensing*, **57**, 6195-6211. <https://doi.org/10.1109/TGRS.2019.2904868>
- [9] Singh, R., Biswas, M. and Pal, M. (2022) Cloud Detection Using Sentinel 2 Imageries: A Comparison of XGBoost, RF, SVM, and CNN Algorithms. *Geocarto International*, **38**, 1-32. <https://doi.org/10.1080/10106049.2022.2146211>
- [10] Qiu, S., Zhu, Z. and He, B. (2019) Fmask 4.0: Improved Cloud and Cloud Shadow Detection in Landsats 4-8 and Sentinel-2 Imagery. *Remote Sensing of Environment*, **231**, Article ID: 111205. <https://doi.org/10.1016/j.rse.2019.05.024>