

DOI:10.19651/j.cnki.emt.2210535

基于多尺度特征融合与混合注意力的云检测算法^{*}

胡 威 管雪元 付 珩

(南京理工大学瞬态物理国家重点实验室 南京 210096)

摘要: 针对传统云检测方法对特殊场景识别效果较差而造成的边缘信息丢失和薄云、碎云误判等问题,提出了一种基于多尺度特征融合与混合注意力的高精度云检测 MSHA-DeepLab 算法。首先,在原始 DeepLabV3+ 算法的基础上引入注意力模块,提高重要特征权重,增强网络对局部特征的感受能力。其次,使用深度可分离卷积提取不同尺度的语义信息,减少网络参数量。最后,进行逐级上采样和特征融合,减少特征信息丢失。选择多种方法与改进算法对比,使用不同场景、不同波段组合的数据集进行测试。结果表明,改进后算法的精确率达到了 86.376 9%,召回率达到了 85.895 9%,特异性达到了 96.915 6%,交并比达到了 82.846 7%,精确度达到了 94.600 8%,相比原始算法和其他方法有明显提高。验证了提出算法能在不同条件下实现高精度的云检测。

关键词: 遥感;云检测;深度学习;机器视觉;注意力机制

中图分类号: TP751 **文献标识码:** A **国家标准学科分类代码:** 520.20

Cloud detection algorithm based on multi-scale feature fusion and hybrid attention

Hu Wei Guan Xueyuan Fu Heng

(State Key Laboratory of Transient Physics, Nanjing University of Science and Technology, Nanjing 210096, China)

Abstract: The traditional cloud detection methods are less effective in recognizing special scenes, which cause problems such as edge information loss and thin and broken clouds misjudgment. In this study, MSHA-DeepLab algorithm based on multi-scale feature fusion and hybrid attention is proposed for high-precision cloud detection. First, the attention module is introduced based on the original algorithm, which to increase the weight of important features and improve the sensibility of local features. Second, depthwise separable convolutions are used to extract the multiscale semantic information and reduce the amount of network parameters. Finally, continuous up-sampling and feature fusion are performed to reduce the loss of feature information. After testing and comparing the datasets with different scenes and different band combinations using various methods and the improved algorithm, it can be seen that the precision of the algorithm reaches 86.376 9%, the recall reaches 85.895 9%, the sepecificity reaches 96.915 6%, the IoU reaches 82.846 7%, the accuracy reaches 94.600 8%, which is a significant improvement compared with the original algorithm and other methods. It is verified that the proposed algorithm can achieve high accuracy cloud detection under different conditions.

Keywords: remote sensing; cloud detection; deep learning; machine vision; attention mechanism

0 引 言

地球上空约有 67% 的区域被云覆盖^[1],云的存在导致光学遥感图像有大量信息被污染,对气候研究和地面目标提取造成重要影响^[2-3]。而卫星的存储资源宝贵,为提升图像数据利用率,进行高准确率的云检测成为遥感图像应用的必要步骤^[4]。云检测的主要方式是对卫星拍摄的遥感图像进行像素级分割,将云和背景的像素二分类,而遥感图像

应用于不同场景,并且包含大量纹理信息,因此对于云检测方法的泛化性有较高要求,能够在多个场景下完成云检测,且减少边缘信息丢失和薄云、碎云误判等问题,成为云检测方法所面临的主要困难。

目前,常用的云检测方法主要分为 3 种,分别为光谱阈值法^[5]、纹理特征法^[6]、基于机器学习的检测方法^[7]。光谱阈值法是根据地物和云在不同波段呈现出的光谱特性差异,将待测图像与设定的单阈值或多阈值对比实现云检测。

收稿日期:2022-07-01

^{*} 基金项目:国家自然科学基金(U2031138)项目资助

文献[8]中提出的 Fmask 算法针对不同传感器设置特定阈值完成云检测,文献[9]中则使用自适应阈值法中的最大类间方差法(Otsu)对 MERSI 影像云检测。阈值法运算简单,但容易将高亮区域误判为云,泛化性不够。纹理特征法则通过分析云和背景的纹理特征差异实现云检测,文献[10]中使用灰度共生矩阵完成了对 IKONOS 拍摄图像的云检测。但该方法不能判断无云图像。随着对检测精度和泛化性要求的不断增强,机器学习算法开始应用于云检测研究。文献[11]中提出了一种基于 K-means 聚类与多光谱阈值结合的云检测方法。但该方法首先要找到合适的特征有效空间,并且依赖于图片的分辨率。

近些年,深度学习在遥感图像处理方面有了更多的应用。相比于传统机器学习算法,卷积神经网络(convolutional neural network,CNN)可以自主提取图像特征,建立像素级的分割模型,对每个像素进行分类^[12]。针对遥感云图,文献[13]中提出了一种基于残差网络 ResNet34^[14]的改进 U-Net^[15]算法,完成对 Landsat8 图像的云检测,但算法中较多的上采样过程容易造成局部特征信息丢失,边界分割效果较差,因此需要更有效的解码模块。在此基础上,文献[16]中提出了一种采用编码器解码器结构的 DeepLab 结构,该算法广泛应用于遥感图像当中。文献[17]中使用 DeepLabV3+算法和 Fmask 算法分别对 L8Biome 数据集进行云检测,结果证明 DeepLabV3+算法具备更好的泛化性。但从分割结果来看,该算法对局部分布的云识别效果不好,云图的轮廓分割不够细致。针对高分辨率遥感图像纹理信息复杂,包含特征丰富等特点,引入注意力机制^[18-19]成为图像重要特征增强的主要手段。而遥感图像的局部特征丰富,因此对于模型的感受野有了

更高要求。引入深度可分离卷积(depthwise separable convolutions,DSC)^[20]能够使得网络在参数量不变的情况下获得更大的感受野。

针对以上方法的不足,本文提出了一种基于多尺度特征融合与混合注意力的 MSHA-DeepLab 网络。该网络保留传统 DeepLabV3+算法中编码层与解码层结构,使用骨干提取网络 ResNet-50 提取不同尺寸的低级特征。针对较大倍数的上采样过程而导致图像信息丢失,像素特征不连续等问题,本文算法采用逐级上采样的方式进行特征融合,充分利用图像的语义信息。而原始算法的 ASPP 模块中参数量过大,模型收敛速度较慢,且针对于高分辨遥感图像的感受能力不足。因此将模块中的卷积替换为深度可分离卷积,并且改变空洞率,增大网络的感受野,使得模型局部特征提取能力增强,参数量减少。在解码层引入混合注意力模块,提取更多云图的边界特征和语义信息,从而更好的完成云检测。

1 方 法

1.1 DeepLabV3+算法

DeepLabV3+是有效的深度学习语义分割算法。网络分为编码层和解码层两部分,在编码层中的骨干提取模块为 ResNet-50,该模块共包含 4 个 Block,在 Block_1 中完成对输入图像浅层特征的提取,经过 4 个 Block 后,得到图像的深层特征,深层特征进入 ASPP 模块,通过不同空洞率的卷积获取不同尺度的图像特征并进行特征融合。在解码层,经过 4 倍的上采样后,再和浅层特征进行特征融合,最后使用卷积核大小为 3×3 的卷积与 4 倍上采样操作得到与原图大小相同的分割结果。DeepLabV3+网络结构如图 1 所示。

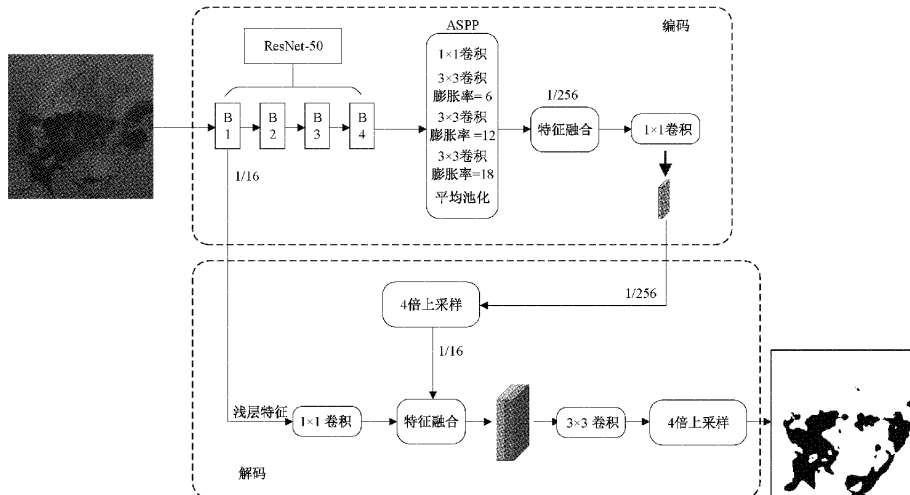


图 1 DeepLabV3+网络结构

使用 DeepLabV3+网络云检测存在很多问题。ASPP 模块中使用空洞率为 18 的卷积并不能遍历数量较多的像素,虽然增大了感受野,但会造成图像局部特征信息提取不足,并且增大计算量。在解码层,4 倍上采样后的深层特

征与低级特征直接融合,会导致重要语义信息丢失,破坏了像素的连续性。网络只提取了原图大小 1/16 的低级特征,而骨干提取网络的中间层包含丰富的空间细节信息和不同通道分别映射的语义信息,解码层并没有利用,特征

信息被忽略。

1.2 MSHA-DeepLab 算法

针对 DeepLabV3+ 算法存在的问题,主要从如下 4 个方面对网络进行改进。

1)将 ASPP 模块中空洞率为 6、12、18 的卷积操作,替换为空洞率为 1、3、5、9、13 的深度可分离卷积,由于云图局部特征丰富,使用更多不同空洞率的卷积操作有利于提取不同尺度的局部特征信息。相比标准卷积,深度可分离卷积对每个通道进行卷积操作,输出不同通道分离的特征属性。在训练过程中可以减少网络的参数量,降低计算压力。在保证精度的前提下,增强网络对局部特征的提取能力。深度可分离卷积的结构如图 2 所示。

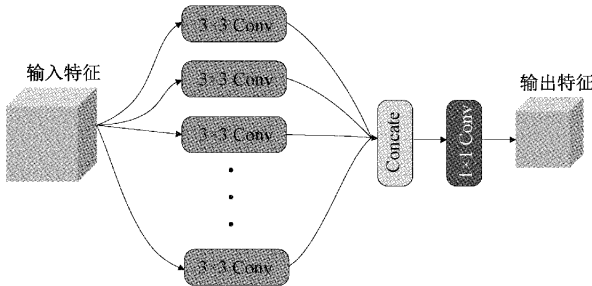


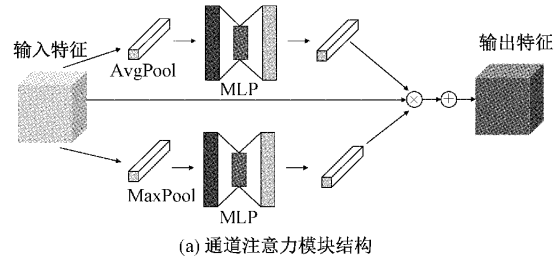
图 2 深度可分离卷积结构

2)将解码层 4 倍上采样后与低级特征融合的操作,替换为连续的 2 倍上采样,再与骨干提取网络中不同尺寸的低级特征进行逐级融合。

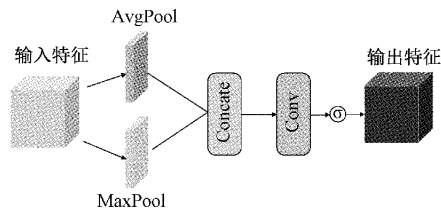
3)在解码层引入了通道注意力模块(channel attention module,CAM)。利用通道注意力模块对包含重要语义信息的通道分配较高权重,对不包含重要特征的通道分配较低权重。其具体过程为,浅层特征与深层特征先进行拼接,然后分别进入全局平均池化层和全局最大池化层得到

两个特征,分别使用 1×1 的卷积调整通道数后,获取通道间的依赖关系,再利用 ReLU 激活函数保证输出为正,最后使用 Sigmoid 函数生成权重,与输入特征图相乘获取重要特征突出的最终特征图。通道注意力模块结构如图 3(a)所示。

4)从浅层特征中可以得到云图的位置、边界等信息,因此在低级特征中引入空间注意力模块(SAM),能够有效的过滤掉干扰背景,专注于云图的局部特征,提升云图边界分割的效果。首先提取列均值和列最大值,将二者融合后,使用 3×3 的卷积进行特征细化,通过 Sigmoid 函数后与原始特征相乘,最终得到边界特征增强的低级特征。空间注意力模块结构如图 3(b)所示。



(a) 通道注意力模块结构



(b) 空间注意力模块结构

图 3 通道注意力模块和空间注意力模块

修改后得到 MSHA-DeepLab 网络,其结构如图 4 所示。

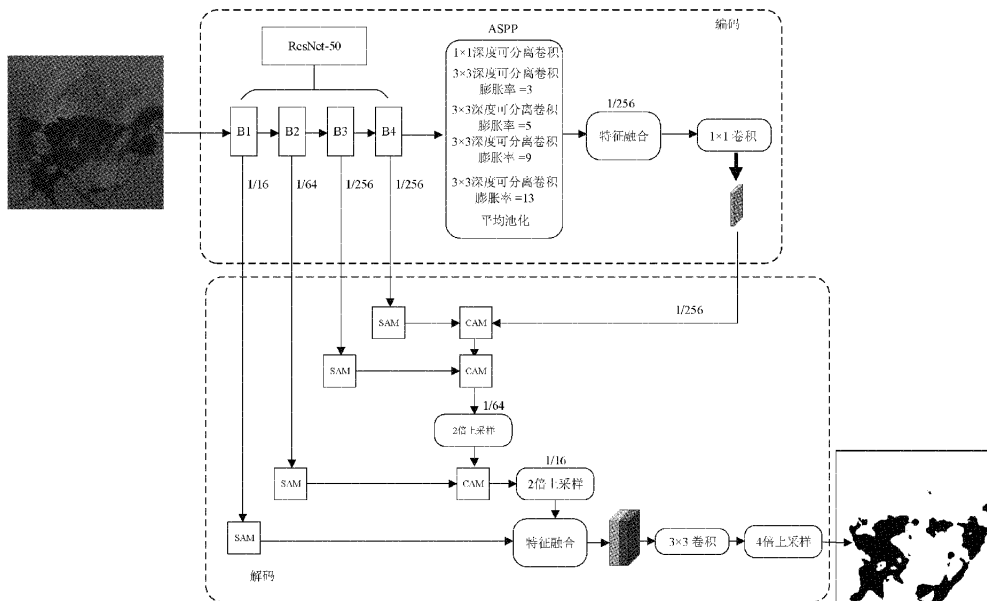


图 4 MSHA-DeepLab 网络结构

2 实验与评价指标

2.1 数据集

本文使用 38-Cloud 云检测数据集^[21]作为原始数据集,该数据集包含 38 个由 Landsat 8 拍摄的不同场景,主要选自北美,场景覆盖类型包括冰雪、森林、草地、城市、灌木、湿地、裸地、海洋以及手动提取的地面实况,每类场景中包含不同云量的图像,通过人工目视对云像素进行标注,最终得到精度很高的解译数据。为满足训练需要,并结合 GPU 算力,对这些场景的完整图片进行切片处理,生成多个分辨率为 384×384 的小块图片,用于对模型的训练和测试。最终生成的数据集,有 8 400 张图片用于训练,9 201 张图片用于测试,每张图片有 4 个对应的光谱通道,1 张伪彩色图,1 张人为标注的标签图,如图 5 所示。图 5(a)、(b)、(c)、(d)分别为蓝色、绿色、红外、红色波段图像,图 5(e)和(f)分别为伪彩色图像和标签。本实验涉及波段的光谱范围如表 1 所示。

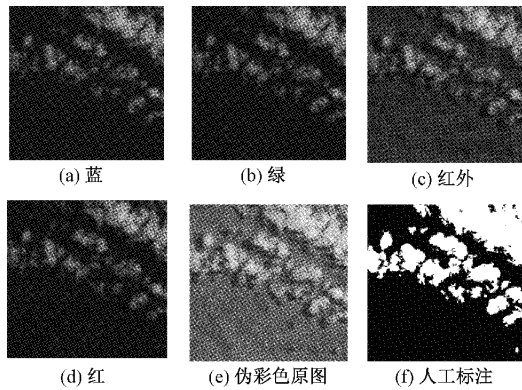


图 5 数据集示例

表 1 数据集波段的波长范围

光谱区域	波长范围/nm
红	630~680
绿	525~600
蓝	450~515
红外	845~885

为从多个方面验证网络的泛化性和稳定性,需要对切片后的数据集划分,最终得到 3 种不同类型的数据集。

1)原始训练集划分为多个不同场景的训练集,并整理与其对应的测试集。本文选择了雪山、河流、海洋、冰川、山地、农田、城市作为场景。

2)对原始数据集进行不同的光谱波段组合。本文构建了 4 种波段组:红-蓝、红-绿、红-红外、蓝-绿。

3)对原始训练集进行随机采样后作为训练集,使用原始测试集进行测试。

2.2 实验环境

实验所用的操作系统为 ubuntu,CPU 为 AMD Ryzen 7

5800H,内存 32 GB,GPU 为英伟达 GeForce RTX 3060,用来对网络结构和训练过程加速。编程平台为 Python3.9,网络结构搭建和测试使用 Pytorch 1.11.2。batch size 设置为 16,epoch 设置为 140,使用二分类交叉熵为损失函数,Sigmoid 将损失重构至 0~1,防止因对数运算产生无穷小,导致网络无法收敛,最后对该批次的样本损失求均值,式(1)为二分类交叉熵公式。

$$\left\{ \begin{aligned} L_i &= -(y_i \times \ln(z_i) + (1 - y_i) \times \ln(1 - z_i)) \\ \text{loss}(z, y) &= \frac{\sum_{i=1}^N L_i}{N} \end{aligned} \right\} \quad (1)$$

式中:z 为模型预测样本是正例的概率,y 为样本的标签,N 为样本个数。

2.3 MSHA-DeepLab 训练步骤

场景相同的训练集训练 MSHA-DeepLab 网络,可以得到 7 个云检测模型权重,对应了各自场景的测试集,以此来评估在特殊场景下不同算法的云检测正确率。用不同波段组合后的训练集进行训练,得到 4 个云检测模型权重,用来判断不同波段对云检测结果的影响,测试模型的稳定性。对数据集随机采样后训练,以测试网络的泛化性。实验还设计了 4 种其他方法与 MSHA-DeepLab 算法的云检测准确率进行对比,实验具体流程如图 6 所示。

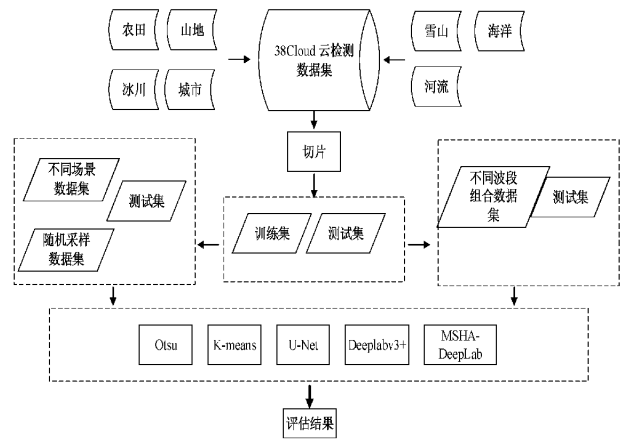


图 6 实验流程

1)使用最大类间方差法(Otsu)进行云检测。

2)将不同训练集切片,得到分辨率大小为 4×4 的小图,选取均值、方差、梯度、熵值 4 个特征值,用 K-means 聚类算法得到云和地面的特征点,随后计算测试图片的最小欧氏距离得到云检测结果。

3)采用与 DeeplabV3+ 网络相同的训练和测试方法,对采用残差网络 ResNet34 的 U-Net 模型的进行训练,得到云检测结果。

4)使用原始的 DeeplabV3+ 架构,进行训练和测试,将所得结果与 MSHA-DeepLab 算法对比。

2.4 评价指标

将云检测后分辨率为 384×384 的小图拼接,还原成与

原图分辨率相等的图像作为最终分割图。使用查准率 (Precision), 召回率 (Recall), 特异性 (Specificity), 交并比 (Intersection over Union, IoU), 精确度 (Accuracy), 作为分割效果的评价指标。其公式如下:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$Specificity = \frac{TN}{FP + TN} \quad (4)$$

$$IoU = \frac{TP}{TP + FN + FP} \quad (5)$$

$$Accuracy = \frac{TP + FN}{TP + TN + FN + FP} \quad (6)$$

式中: TP 为网络预测为云且预测正确的样本, TN 为网络预测为背景且预测正确的样本, FP 为网络预测为云但预测错误的样本, FN 为网络预测为背景但预测错误的样本。

3 结果与分析

3.1 训练过程

训练过程以随机采样数据集为例, 经过 140 轮的训练后, 可以发现模型逐级收敛, 随着迭代次数的增加, 模型逐渐达到最优参数。训练损失曲线如图 7 所示。

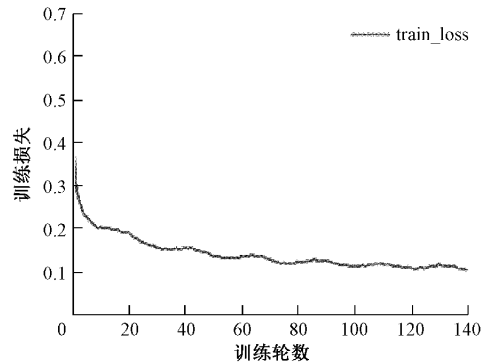


图 7 随机采用数据集的训练损失曲线

3.2 不同场景数据集的云检测结果

不同场景数据集在各种方法下的测试结果示例如图 8 所示。分割结果和人工标注对比可以看出, 针对背景和云图对比度较低或无云的特殊场景, 传统算法容易出现误判, 在冰川场景和河流场景中的识别结果完全劣于深度学习算法。从雪山场景和河流场景的分割结果中不难看出, U-Net 网络对云图特征的感知能力有限, 分割结果明显劣于 DeepLabV3+ 网络。以冰川场景为例, MSHA-DeepLab 网络能够有效识别无云场景。在河流场景中, DeepLabV3+ 网络出现了明显的边缘误判, 而 MSHA-DeepLab 网络对零散分布的云图上识别精确率更高, 在云图和背景的边缘处, MSHA-DeepLab 网络分割轮廓也更清晰。

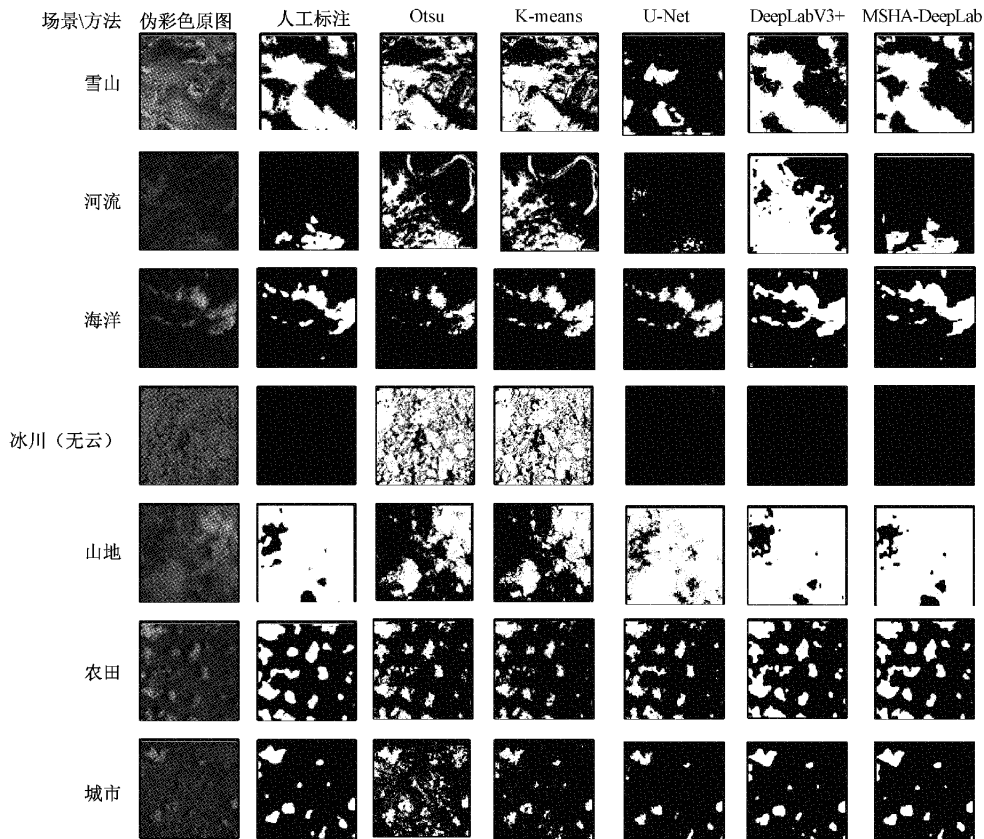


图 8 不同场景数据集的测试结果示例

3.3 不同波段组合数据集的云检测结果

图 9 为不同波段组合数据集在各种方法下的云检测示例。从分割结果看出,传统算法对不同波段组合后的图像识别效果较差。U-Net 的云检测性能受波段组合影响

很大,分割效果不好。DeepLabV3+网络改进前后云检测效果都很好,明显优于其他方法。整体来说,DeepLabV3+算法性能不易波段改变的影响。MSHA-DeepLab 网络对零散分布的云图具有更好的识别效果。

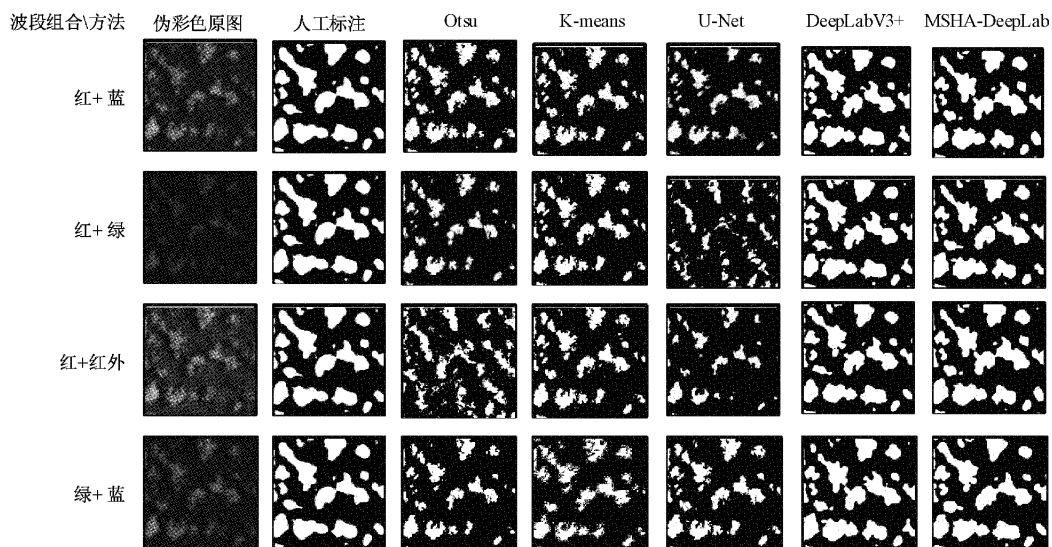
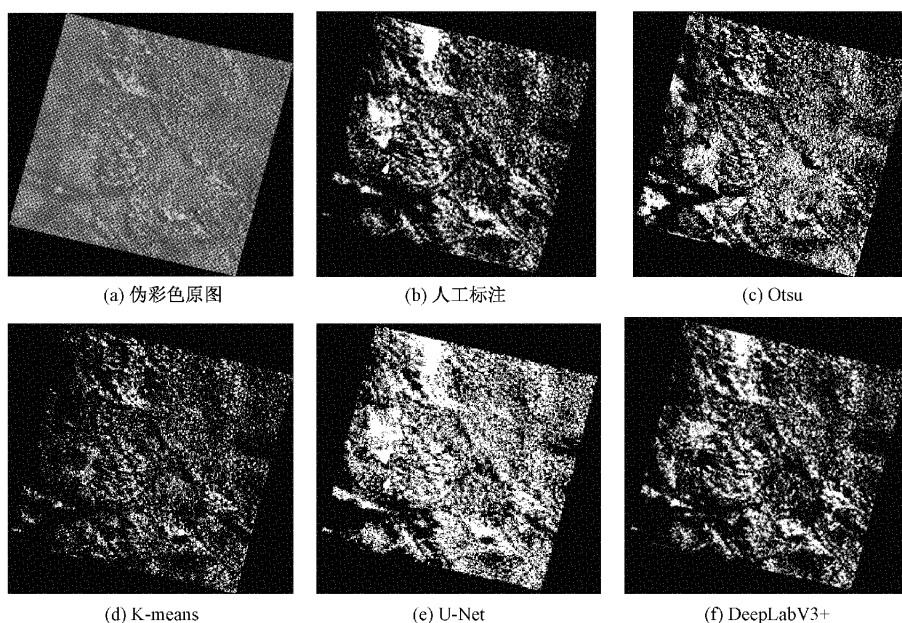


图 9 不同场景数据集的测试结果示例

3.4 不同波段组合数据集的云检测结果

图 10 为随机采样数据集所得切片云检测结果重新组合后的完整分割结果示例,图片分辨率为 $7\ 621 \times 7\ 791$ 。其中二值图片图 10(b)~(g)为分割结果图,彩色图片图 10(h)~(l)为分割误差图。彩色图片中图注从左到右依次为预测正确的正类、预测正确的负类、预测错误的正类、预测错误的负类。从分割结果可以看出,Otsu 算法和 K-means 算法在背景和云的像素亮度接近时,对于云图的

识别效果不好,无法完成云检测。U-Net 网络可以粗略地识别云图和背景分布的轮廓,但对薄云和边界容易发生误判,整体检测结果并不理想。而 DeepLabV3+ 的云检测效果明显好于其他方法,但与 MSHA-DeepLab 算法对比可以看出,MSHA-DeepLab 算法识别精确度更高,分割更加接近人为标注,尤其对于局部分布的云图识别效果明显提升。对于薄云和背景亮度接近的部分,MSHA-DeepLab 的分割效果也更好。



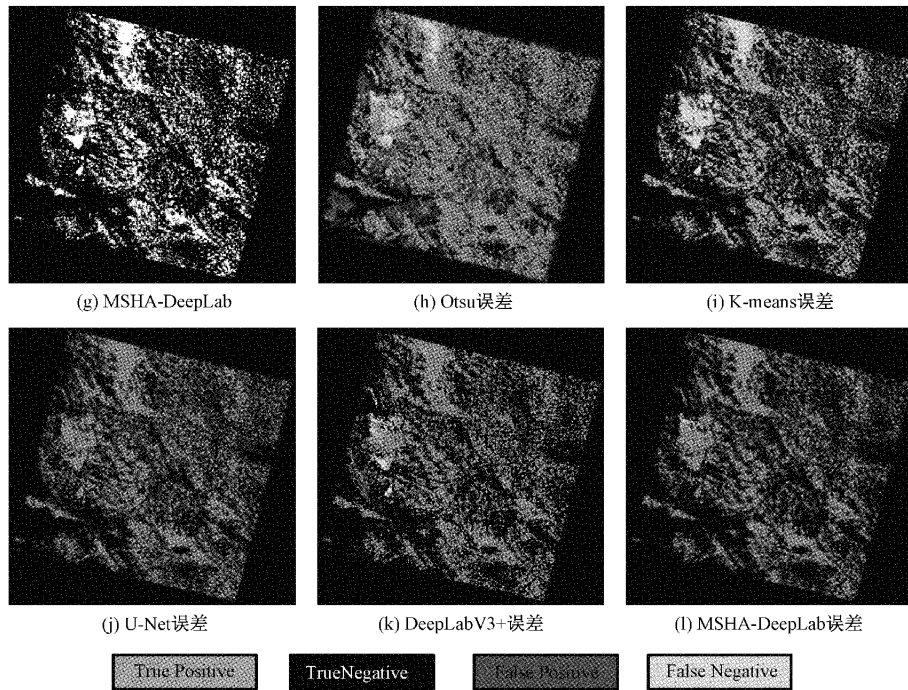


图 10 随机采样数据集的测试结果示例

3.5 不同算法云检测指标结果

5 种算法针对不同数据集所得指标结果求均值后作为

该算法的最终指标值,其结果如表 2 所示。对比表中各项指标可以看出,MSHA-DeepLab 算法在云检测中的表现更好。

表 2 不同算法的定量云检测结果

算法	查准率	召回率	特异性	交并比	精确度
Otsu	48.54	66.73	86.28	62.33	78.23
K-means	54.12	69.43	89.46	65.56	79.69
U-Net	66.83	75.81	92.99	74.22	87.97
DeepLabV3+	81.28	81.17	96.09	80.73	92.24
MSHA-DeepLab	86.38	85.90	96.92	82.85	94.61

4 结 论

本文提出了一种基于多尺度特征融合与混合注意力的 MSHA-DeepLab 算法。在 DeepLabV3+ 网络的基础上,将 ASPP 模块中空洞率为 6、12、18 的卷积操作替换为空洞率为 1、3、5、9、13 的深度可分离卷积,提升了网络对局部特征的感知能力。其次,解码层的 4 倍上采样操作替换为两次 2 倍上采样操作,能够保证像素连续性,减少语义信息的丢失。最后,MSHA-DeepLab 网络引入了通道注意力模块和空间注意力模块,前者突出重要的图像特征,后者增强图像的边界特征。结果表明,MSHA-DeepLab 算法与其他方法相比,在不同场景、不同波段组合的遥感图像中都表现出了更好的云检测效果。在随机采样数据集上的测试效果也明显好于其他算法。从检测指标可以看出,MSHA-DeepLab 算法的精确率达到了 86.376 9%,召回率达到了 85.895 9%,特异性达到了 96.915 6%,交并比

达到了 82.846 7%,精确度达到了 94.600 8%,对比其他方法都有明显提升。综上所述,MSHA-DeepLab 算法具备更好的稳定性和泛化性,能够更好地完成复杂场景下的云检测。

本文提出算法的特征融合部分参数量较大,后续研究中可对该部分进行轻量化设计,以达到实时云检测,更加符合云检测应用场景。对于测试场景和波段组合可进一步丰富,以验证本文算法的泛化性。

参考文献

[1] KING M D, PLATNICK S, MENZEL W P, et al. Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua satellites [J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 51(7): 3826-3852.
 [2] 张华,谢冰,刘煜,等. 东亚地区云对地球辐射收支和降水变化的影响研究 [J]. 中国基础科学, 2017,

- 19(5): 18-22.
- [3] XIE F, SHI M, SHI Z, et al. Multilevel cloud detection in remote sensing images based on deep learning[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(8): 3631-3640.
- [4] ZHU Z, WOODCOCK C E. Object-based cloud and cloud shadow detection in Landsat imagery [J]. Remote Sensing of Environment, 2012, 118: 83-94.
- [5] ZHU Z, WANG S, WOODCOCK C E. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel 2 images[J]. Remote Sensing of Environment, 2015, 159:269-277.
- [6] 张昊,石晓荣,倪亮. 基于分形和纹理特征的红外图像云检测[J]. 计算机仿真, 2020, 37(8):452-456.
- [7] JOSHI P P, WYNNE R H, THOMAS V A. Cloud detection algorithm using SVM with SWIR2 and tasseled cap applied to Landsat 8[J]. International Journal of Applied Earth Observation and Geoinformation, 2019, 82: 101898.
- [8] ZHU Z, WOODCOCK C E. Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change[J]. Remote Sensing of Environment, 2014, 152: 217-234.
- [9] 吴传赏. 风云卫星中分辨率数据晴空识别算法研究[D]. 成都: 电子科技大学, 2016:34-38.
- [10] 薄振华,管雪元,马训穷,等. 遥感云图纹理特征提取算法实时加速设计[J]. 电子测量技术, 2021, 44(1): 125-129.
- [11] 王伟,宋卫国,刘士兴. Kmeans 聚类与多光谱阈值相结合的 MODIS 云检测算法[J]. 光谱学与光谱分析, 2011, 31(4): 1061-1064.
- [12] 陈洋,范荣双,王竞雪. 基于深度学习的资源三号卫星遥感影像云检测方法[J]. 光学学报, 2018, 38(1): 362-367.
- [13] 张家强,李潇雁,李丽圆. 基于深度残差全卷积网络的 Landsat 8 遥感影像云检测方法[J]. 激光与光电子学进展, 2020, 57(10): 364-371.
- [14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [15] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015: 234-241.
- [16] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.
- [17] 彭龙康,刘励聪,陈学泓,等. 遥感影像云检测网络泛化性能研究:以 DeepLabv3+ 为例[J]. 遥感学报, 2021, 25(5):1169-1186.
- [18] 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.
- [19] QIN P, SHEN W, ZENG J. DSCA-Net: Indoor head detection network using dual-stream information and channel attention[J]. Chinese Journal of Electronics, 2020, 29(6): 1102-1109.
- [20] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251-1258.
- [21] MOHAJERANI S, SAEEDI P. Cloud-Net: An end-to-end cloud detection algorithm for Landsat 8 imagery[C]. IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium, IEEE, 2019: 1029-1032.

作者简介

胡威, 硕士, 主要研究方向为深度学习、图像处理等。

E-mail: 519351243@qq.com

管雪元, 硕士, 高级工程师, 主要研究方向为测试计量技术及仪器。

E-mail: 309234310@qq.com