

遥感影像云检测网络泛化性能研究： 以 DeepLabv3+ 为例

彭龙康，刘励聪，陈学泓，陈晋，曹鑫，邱岳安

1. 北京师范大学 地理科学学部 地表过程与资源生态国家重点实验室, 北京 100875;
2. 北京师范大学 地理科学学部 北京市陆表遥感数据产品工程技术研究中心, 北京 100875

摘要：近年来，深度学习算法得到了长足的发展，并开始应用于云检测。但是深度神经网络模型参数众多，依赖大量训练样本，因此理解其泛化性能对于深度学习在不同遥感影像的实际应用具有重要的参考价值。本文以深度语义分割算法 DeepLabv3+ 为例，以一组广泛使用的云标记数据集“L8 Biome”为验证数据，探讨该算法用于云检测时在不同的地表景观、空间分辨率和光谱波段组合的遥感影像上的泛化性能。云标记数据集“L8 Biome”包含 96 景具有全球代表性的 Landsat 8 OLI 影像及相应的人工云掩膜，被广泛用于测试云和云阴影检测算法性能。首先，利用 Landsat 8 OLI 云标记数据集“L8 Biome”，构建不同类型景观、不同空间分辨率、不同波段组合的训练影像集和测试影像集；其次基于不同训练样本集和测试集，评估了 DeepLabv3+ 算法在不同情况下的云检测精度，并与 Fmask 算法作对比分析。研究结果表明：(1) 使用全混合景观类型的训练集训练出来的云检测网络在总体检测精度 (92.81%) 与稳定度 (标准差 12.08%) 上都优于使用单一景观类型的训练集训练得到的云检测网络，也优于 Fmask 的总体精度 (88.75%) 与稳定度 (标准差 17.34%)，说明在构建深度学习算法的训练集时，应该尽可能包含多类型的地表景观；(2) 将全混合景观训练集中剔除一类景观的样本 (冰/雪景观除外) 构建的“混合-1”训练集与全混合景观训练集训练的 DeepLabv3+ 网络的云检测精度也相差不多，说明现有训练样本集已具备较强的景观泛化能力；(3) 基于 30 m 空间分辨率的全混合景观训练样本集训练得到的 DeepLabv3+ 云检测网络在不同分辨率 (30 m、60 m、120 m、240 m) 的测试集上云检测精度差异不大，都取得较好的效果，说明 DeepLabv3+ 能够泛化应用于不同空间分辨率的遥感影像，相反 Fmask 直接应用于低分辨率影像时精度明显下降；(4) DeepLabv3+ 能充分自适应不同波段的信息用于云检测，总体来说更多的光谱波段输入能够提高 DeepLabv3+ 的云检测的精度和稳定度，其中短波红外波段对于 DeepLabv3+ 区分冰/雪与云具有重要价值，而热红外波段对 DeepLabv3+ 云检测网络的性能提升很微小。以上结果说明利用现有数据集“L8 Biome”训练的 DeepLabv3+ 云检测网络能够适用于多种类型的遥感影像，并优于 Fmask 算法。

关键词：深度学习，云检测，DeepLabv3+，泛化性能，地表景观，光谱波段组合，空间分辨率

引用格式：彭龙康, 刘励聪, 陈学泓, 陈晋, 曹鑫, 邱岳安. 2021. 遥感影像云检测网络泛化性能研究: 以 DeepLabv3+ 为例. 遥感学报, 25(5): 1169-1186

Peng L K, Liu L C, Chen X H, Chen J, Cao X and Qiu Y A. 2021. Generalization ability of cloud detection network for satellite imagery based on DeepLabv3+. National Remote Sensing Bulletin, 25(5): 1169-1186 [DOI: 10.11834/jrs.20210061]

1 引言

云污染阻碍光学传感器从地表获取辐射信息，严重影响到地面目标的解译和陆地、海洋、大气相关参数的定量反演，已成为光学遥感影像应用中最重要的障碍之一 (Jeppesen 等, 2019; Zhu 和

Woodcock, 2012; Shen 等, 2015; Verbesselt 等, 2010; Prasad 等, 2006; Harshvardhan 等, 1989; Jedlovec 等, 2008)。已有研究表明全球遥感影像的年平均云量约为 66% (Zhang 等, 2004)，直接影响遥感影像质量，云检测已成为选择和有效使用遥感影像的前提和基本步骤。

收稿日期: 2020-03-09; 预印本: 2020-06-05

基金项目: 国家自然科学基金(编号:41871224)

第一作者简介: 彭龙康, 1997年生, 男, 硕士研究生, 研究方向为事遥感影像云检测。E-mail: 201921051193@mail.bnu.edu.cn

通信作者简介: 陈学泓, 1985年生, 男, 副教授, 研究方向为遥感图像融合、土地覆盖监测。E-mail: chenxuehong@bnu.edu.cn

目前,已开发的云检测方法可以分为两类:基于物理特征的经验规则算法和基于机器学习的算法。第一类算法基于云的“亮”、“白”、“冷”和“高”4个方面的物理特征,通过分析云和典型地表之间光谱差异确定阈值或规则,从而实现云检测(Zhu等,2015;Irish,2000;单娜等,2009;Chen等,2016;吕明明等,2016;Zhu等,2018a;陈曦东等,2019)。这类方法由于其原理简单,计算速度快,在各种遥感影像的业务化云检测中得到了广泛的应用,如用于AVHRR数据云检测的APOLLO算法和CLAVER算法(Saunders和Kriebel,1988;Stowe等,1999)、用于Landsat 4—8数据云检测的Fmask算法(Zhu和Woodcock,2012;Zhu等,2015)、Li等(2017)针对高分一号开发的多特征综合云检测算法等。由于地表景观的光谱复杂性和时空变异性,该类算法很难完全考虑所有影响因素,其阈值或规则的普适性问题限制了其检测性能在不同景观影像上的稳定性。此外,该类方法通常针对特定传感器设定阈值或规则,无法直接迁移应用于其他波段不同或分辨率不同的遥感影像。

基于机器学习的算法将云检测视为二类分类问题。这类算法使用由人工解译的训练数据集,训练特定分类器(监督分类)进行云识别。例如, Lee等(1990)结合纹理特征和神经网络来检测Landsat MSS影像中的云,总体云识别精度达到93%;Roy等(2010)将ACCA算法和由从全球Landsat数据集中选取的训练像元训练的分类树相结合,以降低ACCA算法对云的漏检和误检。Hayes和Hayes(2014)还研究了将空间信息作为神经网络分类器的输入,以识别Landsat影像中的云。所有这些研究都指出了基于机器学习的算法在云检测中的应用潜力。由于不需要大量经验设定的阈值和波段选择,这类方法的准确性和稳定性被认为优于基于物理规则的方法(Sun等,2020)。但是,这类方法大多数严重依赖训练样本,当训练样本的代表性不足或分布有偏时,其泛化性能可能下降,甚至在特定情况下失效(Zhu等,2018b)。近年来兴起的深度学习也属于机器学习的一种方法,由于其能够从影像中有效挖掘图像的多层次纹理特征,规避了经验性的特征选取,在各种图像分类任务中取得比传统分类器更高的精度(Schmidhuber,2015;Ball等,

2017),因此也被应用于云检测(Jeppesen等,2019;Papadomanolaki等,2016;Zi等,2018)。但是神经网络包含大量参数,因而相比于其他基于机器学习的算法对训练数据集有更大程度的依赖性。由于获取大批量训练数据通常代价较大,考查基于深度神经网络的云检测模型在不同景观影像、不同波段组合与不同空间分辨率的遥感影像上的泛化性能具有重要的意义。Li等(2019)通过多类型遥感数据训练一组新的基于深度学习的云检测网络,结果表明深度学习在不同波段组合与空间分辨率的影像上具有一定的泛化能力;Mateo-García等(2020)为解决卫星人工标记云掩膜数据不足的问题,基于物理规则将可用的Landsat 8和Proba-V的标记数据集进行转换以增强神经网络在不同类型遥感影像间的迁移能力。尽管已有部分尝试,基于深度学习的云检测模型在图像景观、空间分辨率以及波段组合3方面的泛化性还缺乏更系统,更有针对性的研究。

针对上述需求和存在问题,本文以“L8 Biome”云标记数据集为基础,构建不同景观类型、不同光谱波段组合、不同空间分辨率的云检测训练集与测试集,探究一个典型的深度学习算法DeepLabv3+在这3个方面的泛化性能。“L8 Biome”数据集是一组以Landsat 8 OLI为基础的云标记数据集,被广泛应用于云检测算法的开发与评估(Foga等,2017)。由于Fmask算法(Zhu和Woodcock,2012;Zhu等,2015)在“L8 Biome”数据集的云检测精度优于其他几种常用算法(Foga等,2017),因此将Fmask的云检测精度作为基准对比DeepLabv3+的云检测性能。

2 数据及预处理

“L8 Biome”数据集(Foga等,2017)包含具有全球代表性的96景Landsat 8 OLI影像及其相应的人工云标记掩膜数据。该数据集包括8类地表景观(冰/雪,湿地,城市,草地/农田,裸地,森林,水体,灌丛),每类景观中各有12幅不同云量的图像。所有影像都由一位解译员进行标注,将影像内像元区分为云、薄云、云阴影和清晰地表。

为更好的比较DeepLabv3+和Fmask算法的云检测性能,本研究的实验中涉及的波段和Fmask算法相同(表1)。为了聚焦云检测的效果,本研究对人工云掩膜的类型进行合并,将薄云与云合并、

将云阴影与清晰地表合并, 最终只区分云像元和无云像元两种类型。

表 1 Landsat 8 OLI/TIRS 光谱波段及其在云检测算法中的应用

Table 1 The spectral bands of Landsat 8 OLI / TIRS and their application in cloud detection algorithm

波段名称	带宽 / μm	空间分辨率 /m
Band 1 海蓝	0.43—0.45	30
Band 2 蓝	0.45—0.51	30
Band 3 绿	0.53—0.59	30
Band 4 红	0.64—0.67	30
Band 5 近红外	0.85—0.88	30
Band 6 短波红外 1	1.57—1.65	30
Band 7 短波红外 2	2.11—2.29	30
Band 8 全色	0.50—0.68	15
Band 9 卷云	1.36—1.38	30
Band 10 热红外 1	10.60—11.19	100
Band 11 热红外 2	11.50—12.51	100

注: Fmask 和 DeepLabv3+ 使用波段以粗体字突出显示。

3 方法

3.1 DeepLabv3+ 语义分割算法

DeepLabv3+ 是一个典型的基于深度神经网络的语义分割算法, 其在多个研究中展示出较好的图像分割性能 (Yang 等, 2019; 王俊强 等, 2019; 陈前 等, 2019)。图 1 表示其网络结构, 整体上是一个编码器—解码器结构 (Chen 等, 2018): 编码器部分使用 DeepLabv3 结构 (Chen 等, 2017, 2018), 通过深度卷积神经网络逐级减少特征图并获取高级语义信息, 在网络最后连接空间金字塔池化模块 ASPP (Atrous Spatial Pyramid Pooling) 获取多尺度的语义信息; 解码器部分首先对获得的多尺度语义信息进行 4 倍双线性上采样, 再结合深度卷积神经网络中同分辨率的低级特征, 通过 3×3 卷积对串联后的特征进行细化, 最后再通过一个 4 倍双线性上采样得到最终的分割结果。DeepLabv3+ 结合了 ASPP 模块与编码器—解码器结构各自的优点, 在有效捕获多尺度信息的同时, 能够获得更清晰的分割对象边界 (Chen 等, 2018)。

3.2 实验设计

图 2 是本研究的实验流程图, 主要包括: 不同地表景观、空间分辨率与波段组合的训练集和测

试集影像样本的制作, 基于训练集的 DeepLabv3+ 网络训练, 以及基于测试集的 DeepLabv3+ 云检测精度评估, 并与 Fmask 方法进行对比。

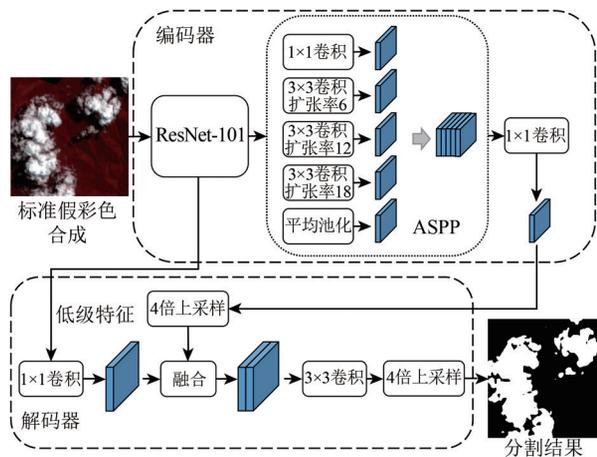


图 1 DeepLabv3+ 网络结构

Fig.1 The network structure of DeepLabv3+

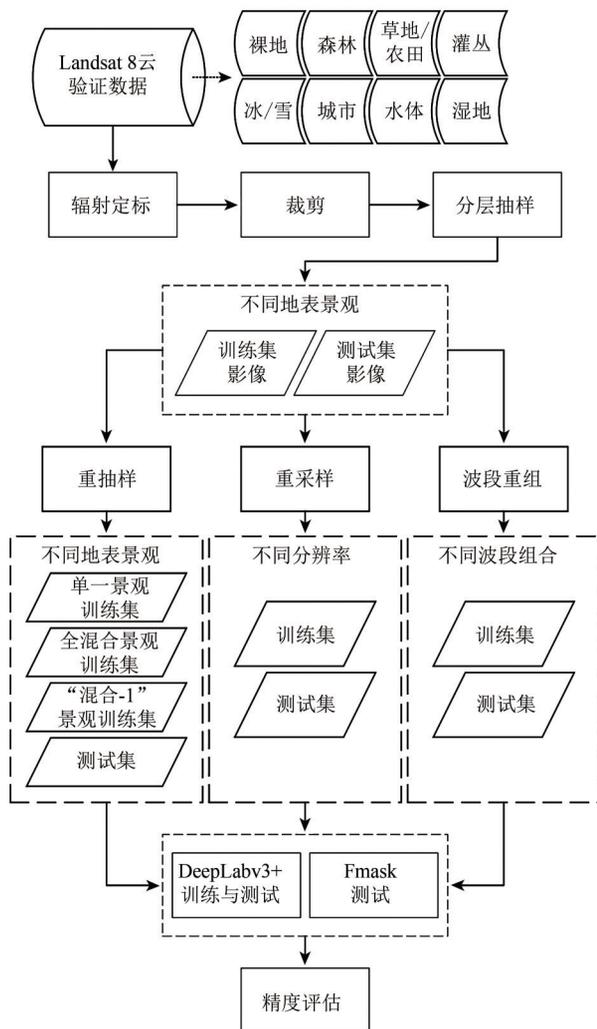


图 2 实验流程图

Fig.2 Flow chart

3.2.1 训练集和测试集影像样本制作

本文将“L8 Biome”数据集的96幅影像均分为两部分，分别组成训练集和测试集各48景。其中每类景观的训练集和测试集各包含6景影像。在样本制作时，由于计算资源的限制不可能对遥感影像进行分割，需要将其裁剪成小幅图像。为避免边界问题，对影像进行裁剪时，参照Jeppesen等(2019)的研究，基于有重叠的移动窗口（相邻窗口间设置48个像元大小的缓冲区）将每个整幅的遥感影像裁剪成约1400张256×256大小的样本图。云检测结果只取移动窗口中心208×208区域以避免边界效应。在此基础上，为测试DeepLabv3+算法在不同地表景观、波段组合和空间分辨率影像上的泛化性能，进一步将训练样本与测试样本做以下处理。

(1) 不同地表景观的影像样本制作：在本组实验中，对上述训练样本进行重新抽样，以构建不同混合程度的训练样本，从而考查DeepLabv3+算法对未知景观图像的泛化性能。首先，构建8组单一地表景观的训练样本，即每组训练样本仅由一类地表景观中的3287个256×256影像样本组成，该组样本考查单一地表景观样本训练出的网络的云检测性能。其次，构建1组全混合地表景观的训练样本，即样本中包括所有8类地表景观，每类从训练集中随机抽出的411个256×256影像样本，共计411×8=3288个样本，该组样本考查混合地表景观样本训练出的网络的云检测性能。考虑到实际应用中，混合训练样本集也不可能包括所有景观，为考查训练样本集包括的景观不完整时网络模型的云检测性能，采用留一法构建次混合景观的训练样本，即从全混合训练集中分别剔除一种景观的数据，组成8个“混合-1”景观的训练集，包括“混合-冰雪”、“混合-湿地”、“混合-城市”、“混合-草地/农田”、“混合-裸地”、“混合-森林”、“混合-水体”、“混合-灌丛”。因此，总计得到了8个单一景观、1个全混合景观、8个“混合-1”的次混合景观的训练集；将以上训练样本训练出的17个云检测网络分别用于测试集的云检测，以考查不同景观样本训练出的网络模型在不同地表景观影像上的表现。

(2) 不同空间分辨率的影像样本制作：在本组实验中，分别对48景训练影像与48景测试影像进行空间重采样，构建了30 m与240 m两种空间

分辨率的训练样本集，以及30 m、60 m、120 m以及240 m等4种空间分辨率的测试样本集，从而考查训练样本与测试图像空间分辨率不一致时，DeepLabv3+的云检测性能。本文通过均值重采样的方法改变遥感影像的空间分辨率。而对人工云掩膜数据的重采样没有采用一般分类图常用的众值重采样，而是遵循对含云像元宁缺勿漏的保守原则——粗像元内只要有一个细像元受云污染，则该粗像元即标记为云污染像元。需要注意的是，对于不同分辨率的整幅遥感影像依然以256×256的大小裁切成影像样本。因此，不同分辨率的256×256影像样本所对应的实际地面大小是不一样的，比如60 m分辨率下的256×256样本影像大小为15.36 km×15.36 km，而240 m分辨率下的256×256样本影像大小为61.44 km×61.44 km。将30 m与240 m分辨率训练集训练的两个DeepLabv3+云检测网络分别用于上述4种空间分辨率的测试影像集，以比较其不同空间分辨率影像上的泛化能力。

(3) 不同波段组合的影像样本制作：由于一种波段组合的影像无法输入另一种波段组合数据训练的网络，因此在本组实验中，对训练集和测试集同时改变其光谱波段组合，以探究DeepLabv3+在不同波段组合遥感影像下的云检测性能。本文构建了5种波段组合：1) Fmask算法使用的所有波段（All）；2) 除热红外波段外的Fmask使用的其他波段（All-NT）；3) 红光、绿光、蓝光、近红外波段（RGBN）；4) 红光、绿光、蓝光波段（RGB）；5) 绿光波段（G）。本组实验样本考查不同波段组合对DeepLabv3+算法的影响。

3.2.2 DeepLabv3+网络训练与云检测精度评估

基于3.2.1制作的不同训练集，对DeepLabv3+的网络参数进行训练，生成相应的不同云检测网络。然后，将测试集的样本影像输入云检测网络，得到256×256大小的云检测结果。为恢复裁剪前遥感影像，根据缓冲区大小，拼接每一小图像中心208×208区域的云检测结果，最后得到和原始整幅遥感影像大小一致的云检测结果。最后，使用总体精度*Acc*（Accuracy）指标对云检测结果进行定量评价：

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

式中，*TP*为被正确检测的云像元数目，*FP*为被错

误检测的云像元数目, TN 为被正确检测的无云像元数目, FN 为被错误检测的无云像元数目。此外, 本文还计算了不同影像总体精度的标准差以评价模型的稳定度。

4 结果

4.1 不同地表景观影像的云检测精度对比

图3显示了由不同地表景观样本训练出来的DeepLabv3+云检测网络及Fmask算法在所有测试影像上的云检测总体精度。其中, 全混合景观样本训练得到的DeepLabv3+网络检测的平均精度能够达到92.78%, 稳定度也较高, 总体精度标准差为12.08%。“混合-1”样本训练的网络模型与全混合景观样本训练的网络模型相差很小。相比之下, Fmask算法在所有测试影像上云检测的平均总体精度为88.75%, 总体精度标准差为17.34%, 在平均精度与稳定度两方面都劣于全混合景观与次混合景观样本训练得到的DeepLabv3+网络。但是, 单一景观样本训练得到的DeepLabv3+网络稳定度较差, 特别由冰/雪景观样本与湿地景观样本训练出的网络模型比Fmask算法更差。图4进一步显示了不同云检测网络在不同景观的测试影像的云检测精度(具体数值见表2—表4)。可以看出, 由单一景观训练集训练的云检测网络, 用于测试其他景观的影像时, 稳定度不高。而由全混合或次混合景观训练集训练得到的云检测网络稳定度高, 这表明混合景观的训练样本中缺失一种景观的样本大多不会影响云检测精度。但是缺失冰/雪景观的次混合训练集则会对冰/雪景观影像的云检测精度产生较大影响, 无法正确识别冰/雪景观的云污染像元。

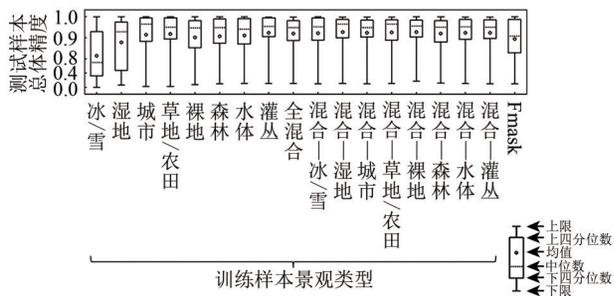


图3 不同景观样本训练得到的DeepLabv3+网络及Fmask在所有测试影像上的云检测总体精度

Fig.3 The cloud detection accuracies of all test images for different cloud detection methods, i.e. DeepLabv3+ trained by training set of different landscapes and Fmask

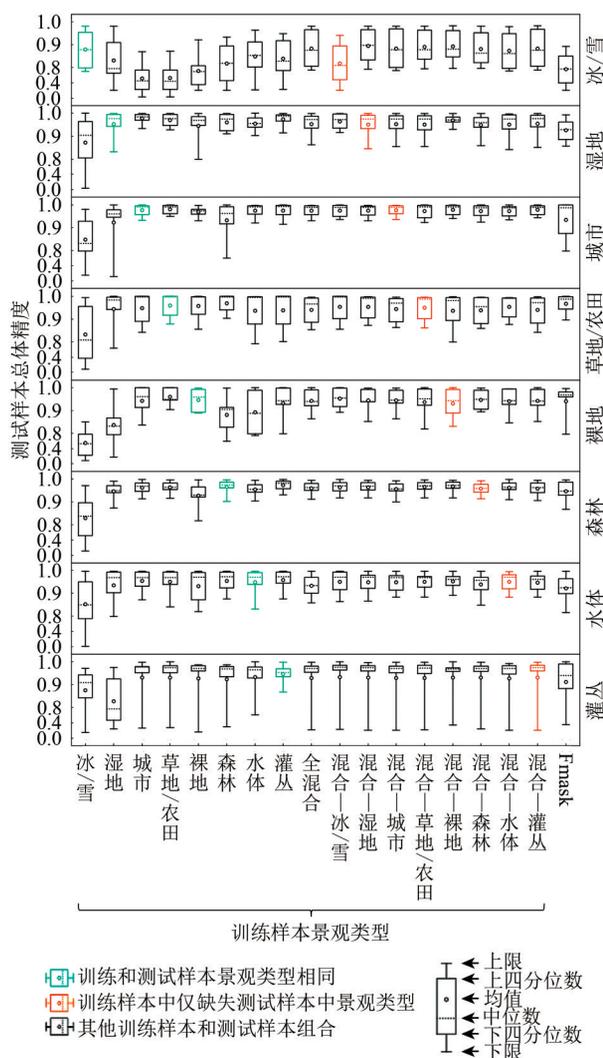


图4 不同景观样本训练得到的DeepLabv3+网络及Fmask在不同景观测试影像上的总体精度(其中每个子图表示不同方法在一种景观测试影像上的精度)

Fig.4 The cloud detection accuracies of test set with different landscapes for different cloud detection methods, i.e. DeepLabv3+ trained by training set of different landscapes and Fmask (each subplot represents the accuracy of test set with single landscape)

总体说来, 当包含冰/雪景观的训练样本时, 混合景观样本训练得到的DeepLabv3+云检测网络相比于Fmask有较高的精度与稳定性。图5展示了二者对比的两个例子, Fmask倾向于在高亮的冰/雪区域及短波红外波段反射率较高的裸地区域上误检, 在包含冰/雪景观的图像上尤为明显, 而DeepLabv3+依然能较正确地分类。这可能是因为DeepLabv3+考虑了比Fmask更多纹理特征, 能够更好的区分云和冰/雪。由于DeepLabv3+的反卷积操作, 该方法会损失部分空间细节, 使得检测出的云边缘较为光滑。但是这部分空间细节损失,

并不影响总体精度。值得注意的是,对于灌丛景观测试集中的一景影像(图6),由混合景观样本训练得到的DeepLabv3+网络与Fmask都未能正确

识别大部分薄云,分类精度都低于50%。这说明仍然存在少部分影像,现有的算法都难以正确识别其中的云像元。

表2 不同景观样本训练得到的DeepLabv3+网络及Fmask在不同景观测试影像上的总体精度的均值及标准差

Table 2 Mean and standard deviation of the overall accuracy of the DeepLabv3+ trained by training set of imagery captured over different landscapes and Fmask on test images captured over different landscapes

模型	训练样本景观		冰/雪	湿地	城市	草地/农田	裸地	森林	水体	灌丛	All
不同景观训练样本 训练得到的 DeepLabv3+ 云检测网络	冰/雪	均值	0.860	0.746	0.753	0.640	0.575	0.676	0.726	0.785	0.720
		标准差	0.105	0.338	0.280	0.350	0.332	0.308	0.338	0.290	0.313
	湿地	均值	0.708	0.953	0.827	0.923	0.771	0.946	0.934	0.678	0.843
		标准差	0.260	0.059	0.317	0.128	0.275	0.038	0.076	0.276	0.233
	城市	均值	0.463	0.978	0.976	0.948	0.790	0.962	0.954	0.856	0.866
		标准差	0.304	0.022	0.026	0.071	0.343	0.028	0.045	0.267	0.252
	草地/农田	均值	0.462	0.969	0.980	0.960	0.807	0.963	0.951	0.859	0.869
		标准差	0.305	0.027	0.021	0.054	0.340	0.026	0.054	0.260	0.250
	裸地	均值	0.599	0.944	0.968	0.957	0.953	0.928	0.932	0.837	0.890
		标准差	0.274	0.069	0.021	0.058	0.047	0.057	0.073	0.306	0.192
	森林	均值	0.665	0.961	0.905	0.969	0.866	0.965	0.954	0.854	0.892
		标准差	0.279	0.038	0.139	0.041	0.131	0.031	0.046	0.251	0.179
	水体	均值	0.766	0.955	0.974	0.929	0.887	0.955	0.949	0.907	0.915
		标准差	0.251	0.032	0.029	0.098	0.095	0.029	0.057	0.134	0.131
	灌丛	均值	0.740	0.974	0.974	0.939	0.932	0.972	0.959	0.946	0.930
		标准差	0.247	0.030	0.032	0.085	0.072	0.023	0.042	0.041	0.124
	全混合	均值	0.874	0.953	0.974	0.940	0.944	0.960	0.932	0.849	0.928
		标准差	0.095	0.047	0.026	0.059	0.045	0.027	0.048	0.283	0.120
	混合—冰/雪	均值	0.669	0.965	0.974	0.953	0.960	0.964	0.950	0.854	0.911
		标准差	0.276	0.033	0.027	0.064	0.042	0.027	0.050	0.282	0.175
	混合—湿地	均值	0.889	0.952	0.973	0.954	0.948	0.964	0.948	0.849	0.935
		标准差	0.083	0.054	0.026	0.055	0.048	0.026	0.046	0.287	0.119
	混合—城市	均值	0.868	0.953	0.976	0.943	0.947	0.956	0.949	0.846	0.930
		标准差	0.104	0.050	0.024	0.057	0.044	0.030	0.042	0.288	0.122
	混合—草地/农田	均值	0.887	0.951	0.972	0.950	0.938	0.966	0.951	0.848	0.933
		标准差	0.079	0.051	0.032	0.058	0.060	0.026	0.039	0.285	0.119
	混合—裸地	均值	0.892	0.970	0.978	0.936	0.936	0.967	0.952	0.869	0.937
		标准差	0.074	0.022	0.025	0.077	0.060	0.026	0.035	0.234	0.102
	混合—森林	均值	0.879	0.950	0.973	0.938	0.943	0.957	0.939	0.854	0.929
		标准差	0.077	0.046	0.028	0.059	0.039	0.026	0.049	0.274	0.114
混合—水体	均值	0.856	0.951	0.973	0.954	0.943	0.963	0.951	0.844	0.929	
	标准差	0.102	0.055	0.026	0.053	0.051	0.028	0.042	0.286	0.123	
混合—灌丛	均值	0.875	0.954	0.977	0.939	0.943	0.958	0.947	0.848	0.930	
	标准差	0.094	0.051	0.023	0.065	0.049	0.031	0.041	0.290	0.122	
Fmask	—	均值	0.628	0.925	0.933	0.966	0.892	0.947	0.921	0.851	0.883
		标准差	0.277	0.048	0.086	0.040	0.105	0.045	0.059	0.228	0.173

表 3 30 m 分辨率和 240 m 分辨率全混合景观样本训练得到的 DeepLabv3+ 网络及 Fmask 在不同空间分辨率、不同景观测试影像上的总体精度的均值及标准差

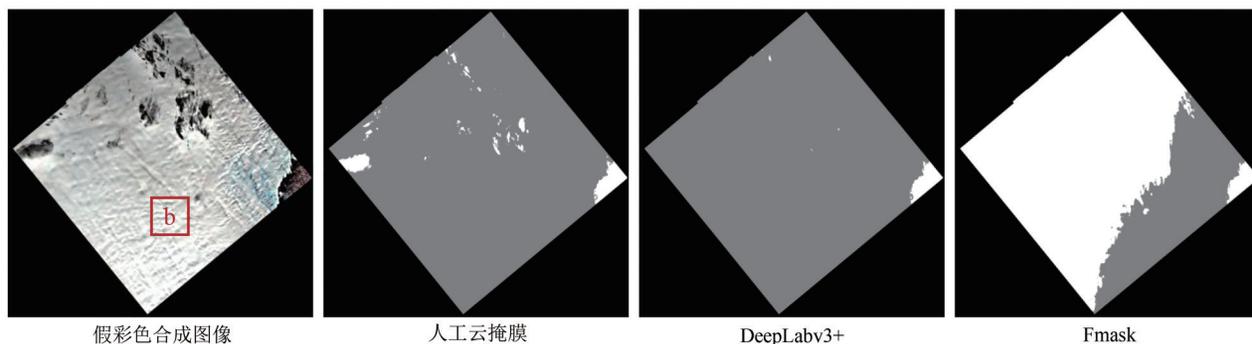
Table 3 Mean and standard deviation of the overall accuracy of the DeepLabv3+ trained by the fully mixed training set with 30 m resolution or 240 m resolution and Fmask on test images with different spatial resolutions and different landscapes

模型	测试影像空间分辨率/m		冰/雪	湿地	城市	草地/农田	裸地	森林	水体	灌丛	All
30 m 分辨率样本训练得到的 DeepLabv3+ 云检测网络	30	均值	0.874	0.953	0.974	0.940	0.944	0.960	0.932	0.849	0.928
		标准差	0.095	0.047	0.026	0.059	0.045	0.027	0.048	0.283	0.120
	60	均值	0.871	0.950	0.964	0.937	0.940	0.953	0.926	0.846	0.923
		标准差	0.095	0.046	0.039	0.060	0.047	0.034	0.048	0.279	0.119
	120	均值	0.868	0.941	0.942	0.932	0.932	0.937	0.913	0.842	0.913
		标准差	0.096	0.048	0.065	0.059	0.054	0.053	0.051	0.272	0.118
	240	均值	0.861	0.931	0.911	0.923	0.917	0.918	0.898	0.836	0.899
		标准差	0.099	0.052	0.099	0.061	0.070	0.076	0.056	0.264	0.121
240 m 分辨率样本训练得到的 DeepLabv3+ 云检测网络	30	均值	0.908	0.941	0.970	0.923	0.939	0.928	0.923	0.832	0.921
		标准差	0.055	0.064	0.026	0.079	0.069	0.036	0.057	0.305	0.126
	60	均值	0.906	0.941	0.968	0.927	0.941	0.932	0.925	0.832	0.921
		标准差	0.056	0.061	0.031	0.074	0.059	0.038	0.051	0.305	0.125
	120	均值	0.902	0.936	0.958	0.926	0.937	0.932	0.925	0.829	0.918
		标准差	0.059	0.061	0.044	0.073	0.056	0.046	0.049	0.302	0.124
	240	均值	0.892	0.930	0.939	0.920	0.925	0.923	0.924	0.824	0.910
		标准差	0.065	0.061	0.065	0.076	0.063	0.060	0.053	0.299	0.126
Fmask	30	均值	0.628	0.925	0.933	0.966	0.892	0.947	0.921	0.851	0.883
		标准差	0.277	0.048	0.086	0.040	0.105	0.045	0.059	0.228	0.173
	60	均值	0.630	0.795	0.864	0.809	0.884	0.814	0.766	0.532	0.762
		标准差	0.279	0.271	0.133	0.255	0.111	0.162	0.202	0.323	0.255
	120	均值	0.631	0.691	0.853	0.802	0.823	0.768	0.746	0.520	0.729
		标准差	0.281	0.278	0.143	0.258	0.124	0.162	0.205	0.321	0.254
	240	均值	0.627	0.669	0.694	0.791	0.802	0.776	0.759	0.476	0.699
		标准差	0.281	0.280	0.299	0.262	0.139	0.167	0.161	0.343	0.272

表 4 不同波段组合样本训练得到的 DeepLabv3+ 网络及 Fmask 在不同景观测试影像上的总体精度的均值及标准差

Table 4 Mean and standard deviation of the overall accuracy of the DeepLabv3+ trained by training set with different band combinations and Fmask on test images captured over different landscapes

模型	波段组合方式		冰/雪	湿地	城市	草地/农田	裸地	森林	水体	灌丛	All
不同波段组合训练样本训练得到的 DeepLabv3+ 云检测网络	All	均值	0.874	0.953	0.974	0.940	0.944	0.960	0.932	0.849	0.928
		标准差	0.095	0.047	0.026	0.059	0.045	0.027	0.048	0.283	0.120
	All-NT	均值	0.877	0.822	0.970	0.959	0.955	0.970	0.957	0.951	0.933
		标准差	0.058	0.320	0.034	0.055	0.046	0.023	0.042	0.067	0.132
	RGBN	均值	0.547	0.838	0.978	0.957	0.959	0.966	0.932	0.960	0.892
		标准差	0.286	0.298	0.021	0.058	0.041	0.028	0.062	0.032	0.204
	RGB	均值	0.550	0.882	0.951	0.954	0.887	0.959	0.951	0.958	0.886
		标准差	0.288	0.207	0.055	0.059	0.081	0.033	0.049	0.033	0.187
	G	均值	0.805	0.940	0.963	0.818	0.887	0.880	0.890	0.854	0.880
		标准差	0.120	0.046	0.025	0.198	0.076	0.121	0.107	0.252	0.147
Fmask	—	均值	0.628	0.925	0.933	0.966	0.892	0.947	0.921	0.851	0.883
		标准差	0.277	0.048	0.086	0.040	0.105	0.045	0.059	0.228	0.173



假彩色合成图像

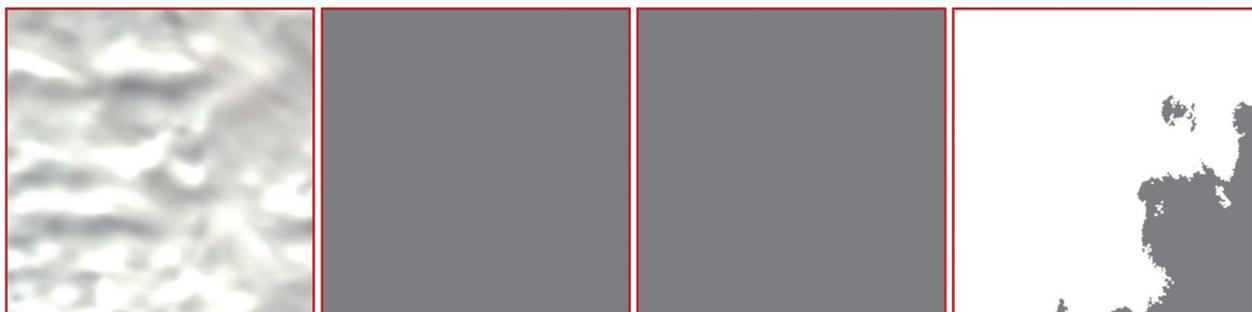
人工云掩膜

DeepLabv3+

Fmask

(a) 整幅冰/雪景观Landsat图像及其云掩膜

(a) Whole Landsat image captured over snow/ice and its cloud mask



假彩色合成图像

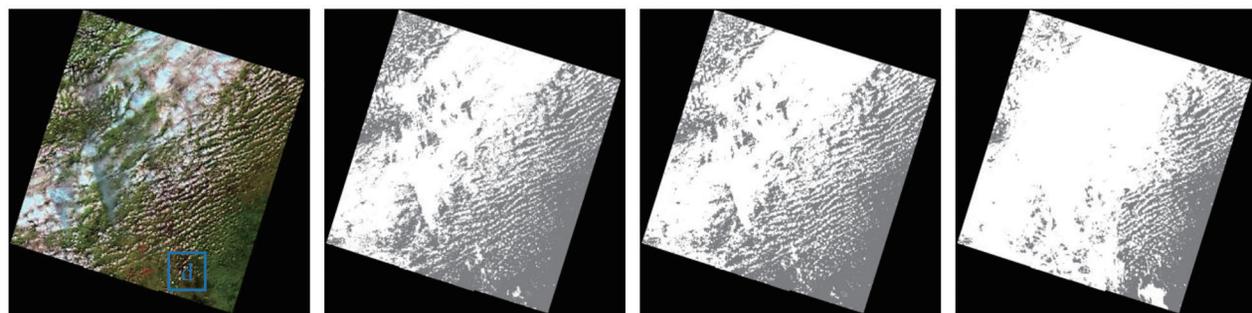
人工云掩膜

DeepLabv3+

Fmask

(b) 红色矩形区域放大图像

(b) Enlarged images of the subset area in the red rectangle



假彩色合成图像

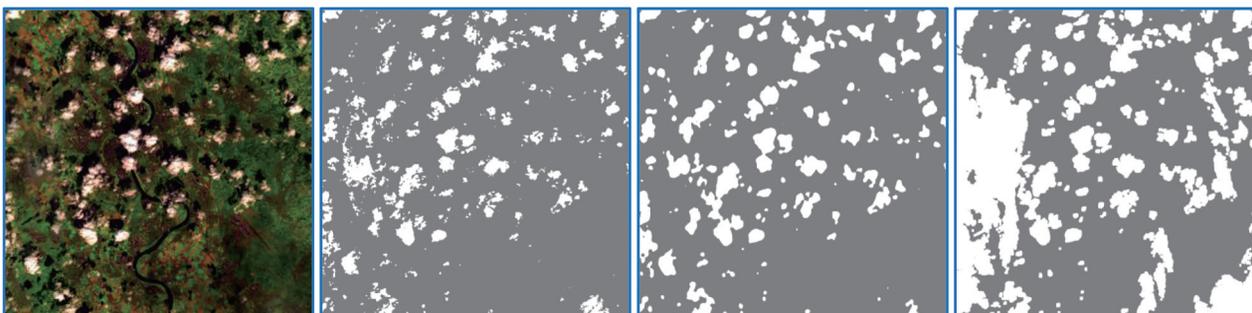
人工云掩膜

DeepLabv3+

Fmask

(c) 整幅城市景观Landsat图像及其云掩膜

(c) Whole Landsat image captured over Urban and its cloud mask



假彩色合成图像

人工云掩膜

DeepLabv3+

Fmask

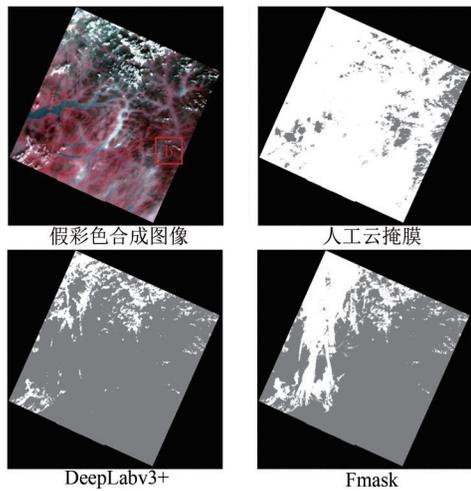
(d) 蓝色矩形区域放大图像

(d) Enlarged images of the subset area in the blue rectangle

注:(a)和(b)中图像假彩色合成使用波段5、4、3,(c)和(d)中图像假彩色合成使用波段6、5、2。云掩膜中,白色为云覆盖区域,灰色为无云区域,黑色为填充值。

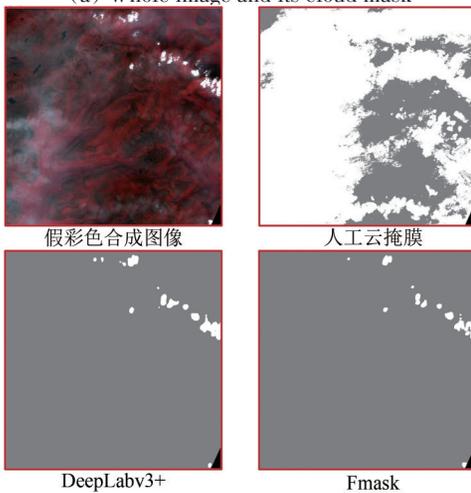
图5 两景Landsat图像的云检测结果

Fig.5 Cloud detection results of two Landsat images



(a) 整幅图像及其云掩膜

(a) Whole image and its cloud mask



(b) 红色矩形区域放大图像

(b) Enlarged images of the subset area in the red rectangle

□ 云覆盖区域 ■ 无云区域 ■ 填充值

图6 一景灌丛景观Landsat图像的云检测结果(图像假彩色合成均使用波段5、4、3)

Fig.6 Cloud detection results of one Landsat image captured over shrubland(All the images are displayed using bands 5, 4 and 3. In each cloud mask: White represents the cloud coverage area; Gray represents the cloud-free area; Black represents fill value)

4.2 不同空间分辨率影像的精度对比

图7展示了由30 m分辨率和240 m分辨率训练样本训练得到的DeepLabv3+网络及Fmask算法在不同空间分辨率测试影像上的云检测精度(具体数值见表3)。由30 m分辨率样本训练得到的DeepLabv3+云检测网络应用于更低不同空间分辨率影像时,检测精度虽略有下降,但依然保持较好的云检测精度;同样地,由240 m分辨率样本训练得到的DeepLabv3+云检测网络在低空间分辨率影像上的检测精度也仅是略低于在高空间分辨率影像上的检测精度;相比之下,Fmask算法在低分

辨率影像上精度和稳定度都很低。如图8所示,在低分辨率影像上,Fmask算法会出现很严重的误检,且影像分辨率越低,误检越严重,而DeepLabv3+算法仍能捕获到大致云形状,获得较好的云检测效果。这说明Fmask算法不能直接迁移用于其他空间分辨率的影像。这可能是由于Fmask算法所基于的一些经验规则或阈值会随着空间分辨率的改变而变化。因此试图将Fmask算法应用于新的遥感传感器获取的影像时,需要对算法进行较大的调整和重新设计;而基于机器学习的DeepLabv3+算法则具有较强的灵活性,即使训练样本和测试样本间空间分辨率不一致,也能取得较好的云检测效果。

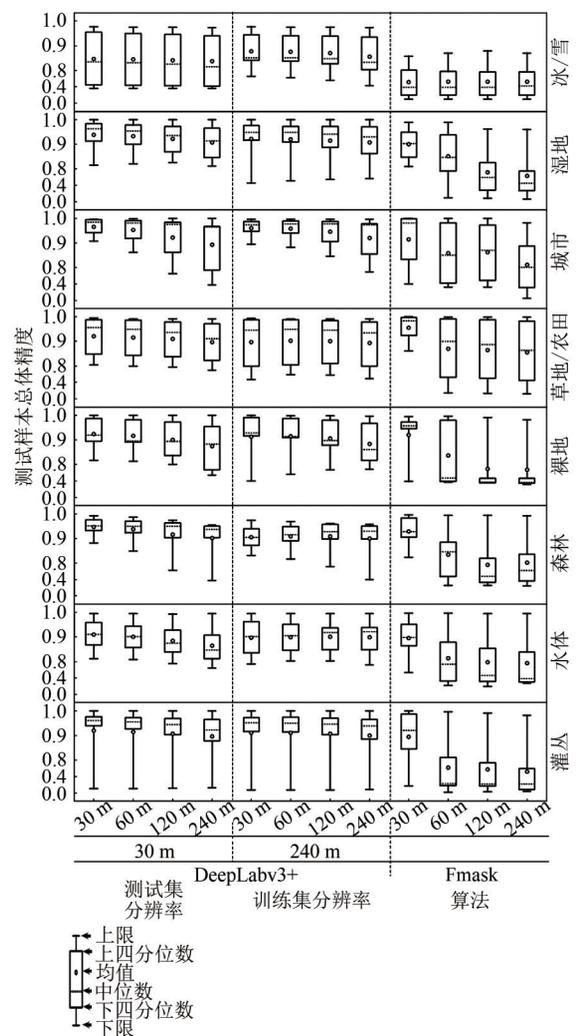


图7 30 m分辨率和240 m分辨率样本训练得到的DeepLabv3+网络及Fmask在不同空间分辨率、不同景观测试影像上的云检测总体精度(其中每个子图表示不同方法在一种景观测试影像上的精度)

Fig.7 The cloud detection accuracy of test set with different landscapes and resolutions for different cloud detection methods, i.e. DeepLabv3+ trained by training set with 30 m and 240 m resolution (each subplot represents the accuracy of test set with single landscape)

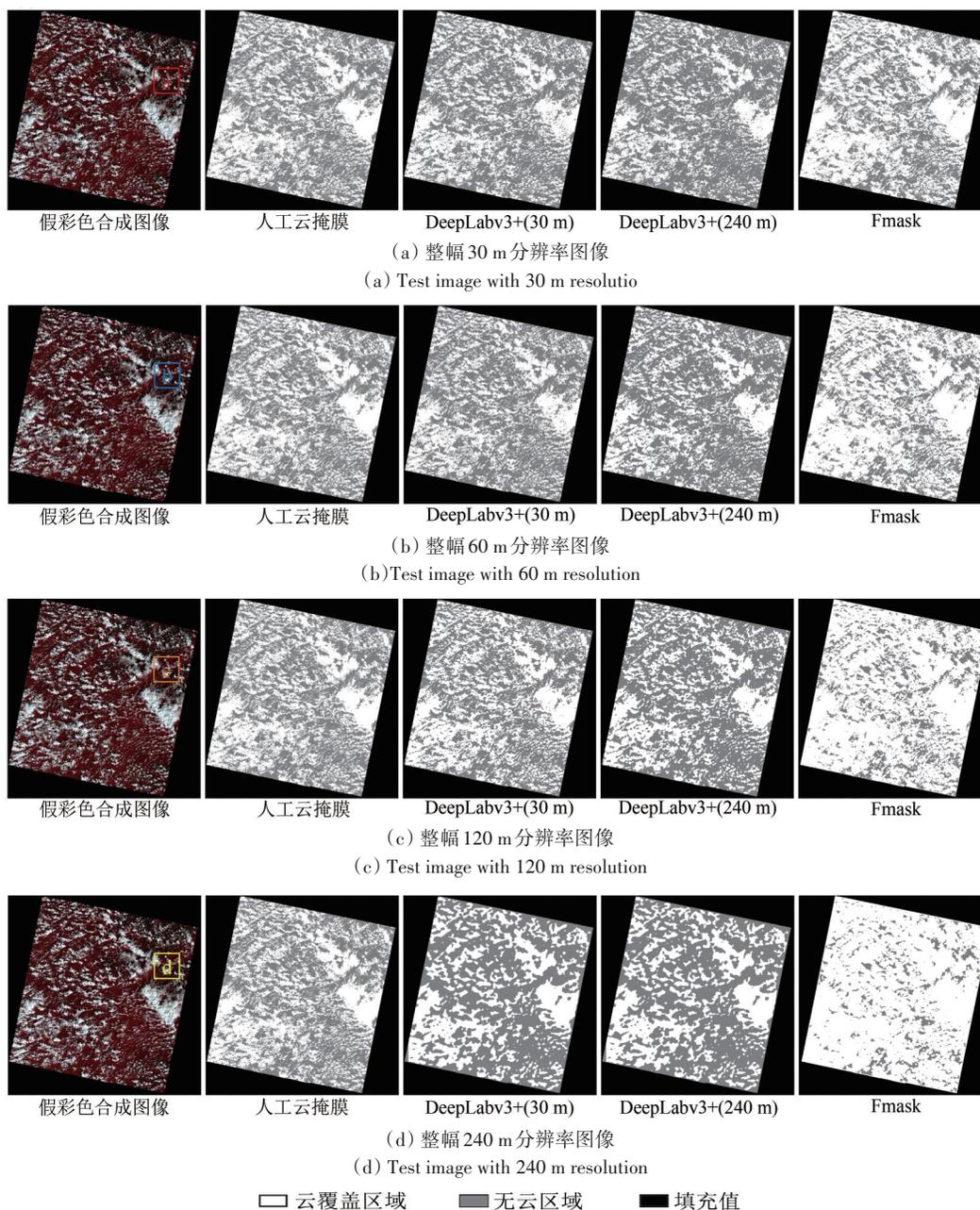


图8 30 m分辨率和240m分辨率样本训练得到的DeepLabv3+网络及Fmask算法在不同分辨率影像上的云检测结果示例 (图像假彩色合成均使用波段5、4、3)

Fig.8 The case results of DeepLabv3+ trained by the training set with 30m resolution and 240m resolution and Fmask on test images with different spatial resolutions(All the images are displayed using bands 5, 4 and 3 .In each cloud mask: White represents the cloud coverage area; Gray represents the cloud-free area; Black represents fill value)

4.3 不同波段组合影像的云检测精度对比

图10展示了由这5种波段组合方式的样本训练出来的DeepLabv3+云检测网络在不同景观测试影像上的测试精度(具体数值见表4)。整体上,随着样本中波段的增加,DeepLabv3+的云检测精度与稳定度都逐步增加。使用All和All-NT两种波段组合训练得到的网络模型精度相当,均优于Fmask的云检测精度,意味着热红外波段对

DeepLabv3+云检测网络的性能并不十分重要。使用RGBN波段组合训练得到的网络云检测精度总体相对较低,特别是在冰/雪景观测试集上精度有较大下降。但是在除冰/雪之外的其他景观测试集上,RGBN的云检测精度略优于Fmask。这说明短波红外波段对于DeepLabv3+网络区分冰/雪与云具有重要的意义。图11显示了一景冰/雪景观影像的云检测的例子,整体上看缺少热红外波段并不会影响

网络的云检测性能, 甚至在一些区域能获得更精细的检测结果 (图 11 (b)); 而缺少短波红外波段, 仅使用 RGB 或 RGBN 波段组合方式训练出的网络模型会将冰/雪误分为云。在裸地景观中 (图 12), 缺少热红外波段对网络性能影响很小; 而缺少短波红外波段影响也不大, 只会造成少量清晰像元的误检或云像元的漏检。Fmask 算法采用不同方法检测陆地和水体上的云, 该算法仅使用 NDVI 和近红外波段反射率来区分出水体像元, 这种方法在厚云区域是不能很好地分离陆地和水体像元的 (Qiu 等, 2019), 这使得 Fmask 算法在图 12 这景影

像中检测效果较差。需要注意的是, 在图 11 和图 12 这两景影像上, 使用仅由绿光波段的样本训练得到的网络模型也能获得不错的检测结果, 但是从图 10 中这一网络模型对多景冰/雪景观影像和多景裸地景观影像的测试结果看, 这一网络模型的稳定度很差。本组实验表明, 尽管从一种波段组合数据训练的 DeepLabv3+ 无法直接应用于另一种波段组合的影像, 但是可以方便的通过构建不同波段组合的训练样本, 使得 DeepLabv3+ 能够应用于不同波段组合的遥感影像的云检测。

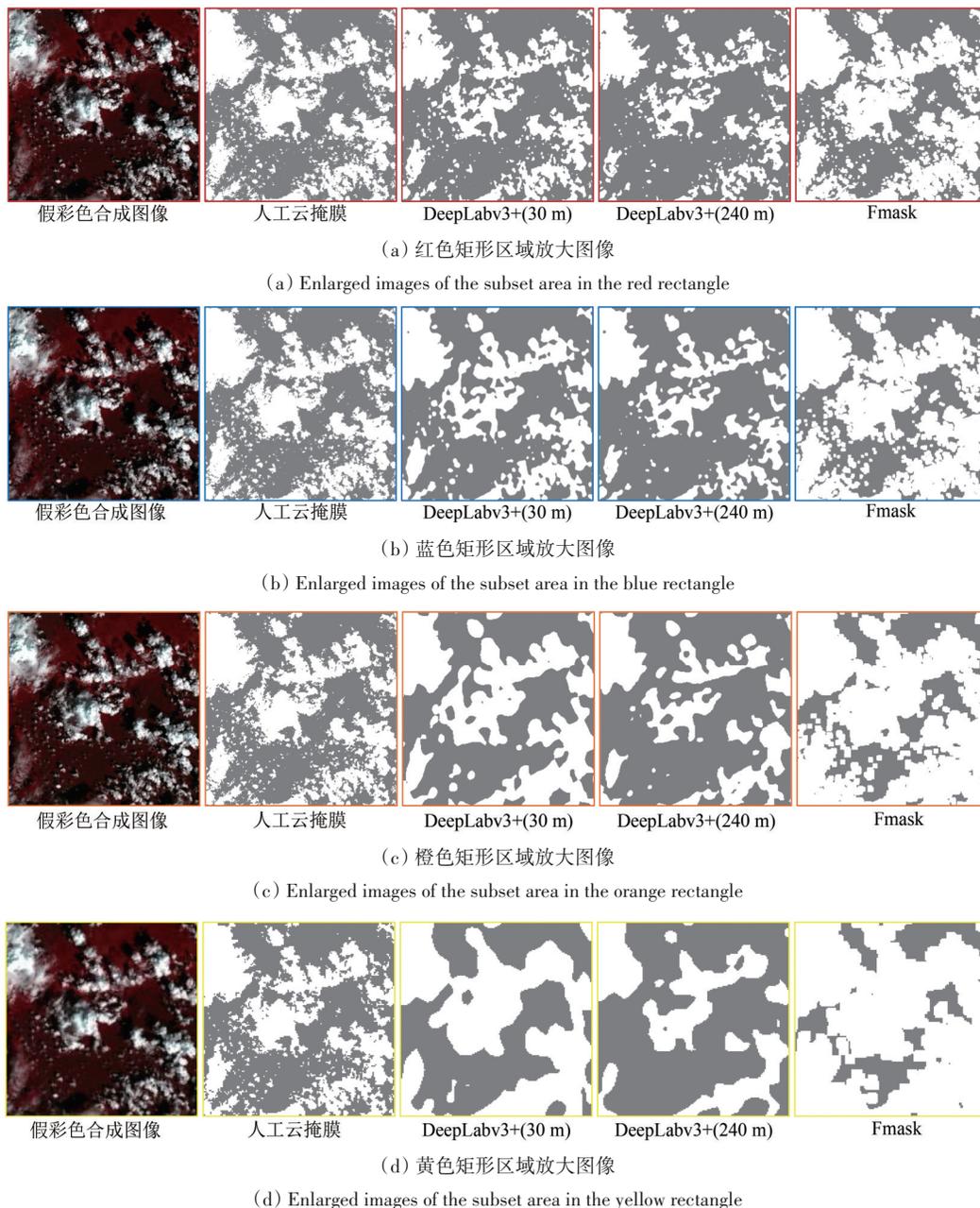


图9 图8局部区域放大图像

Fig.9 The zoomed area of Fig.8

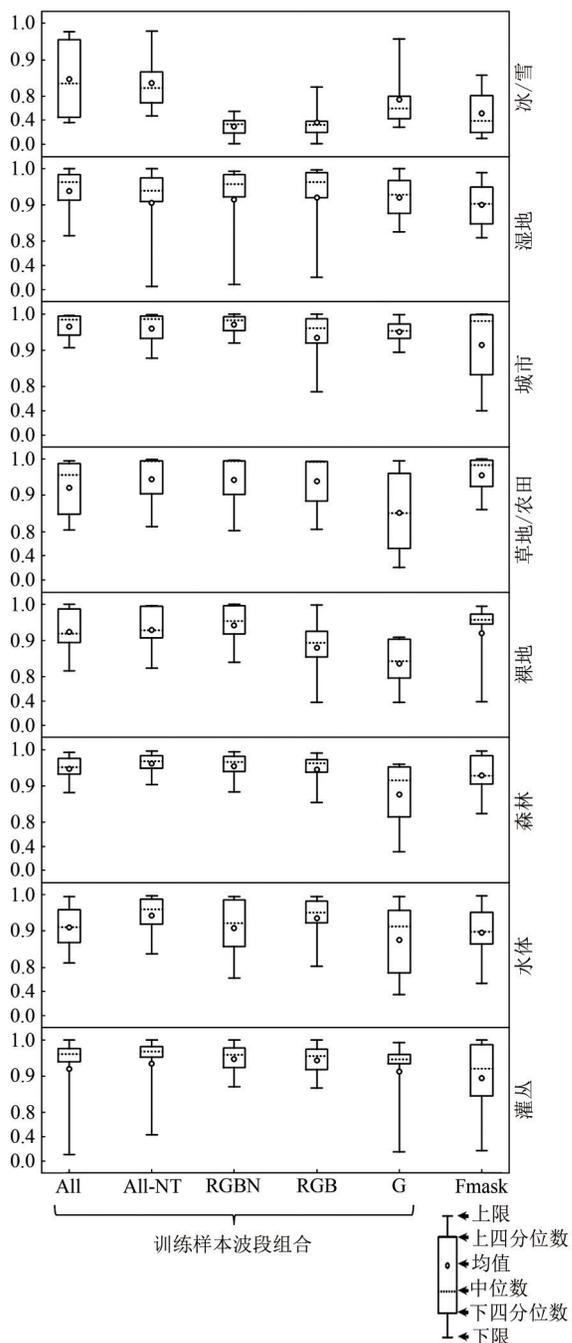


图 10 不同波段组合样本训练得到的DeepLabv3+网络及Fmask在不同景观测试影像上的云检测总体精度(其中每个子图表示不同方法在一种景观测试影像上的精度)

Fig.10 The cloud detection accuracy of test set with different landscapes for different cloud detection methods, i.e. DeepLabv3+ trained by training set with different band combinations and Fmask (each subplot represents the accuracy of test set with single landscape)

5 讨论

本文通过将Landsat影像进行空间重采样、光谱波段重组模拟不同类型的遥感影像，以考查DeepLabv3+云检测网络的泛化性能。但是，现实

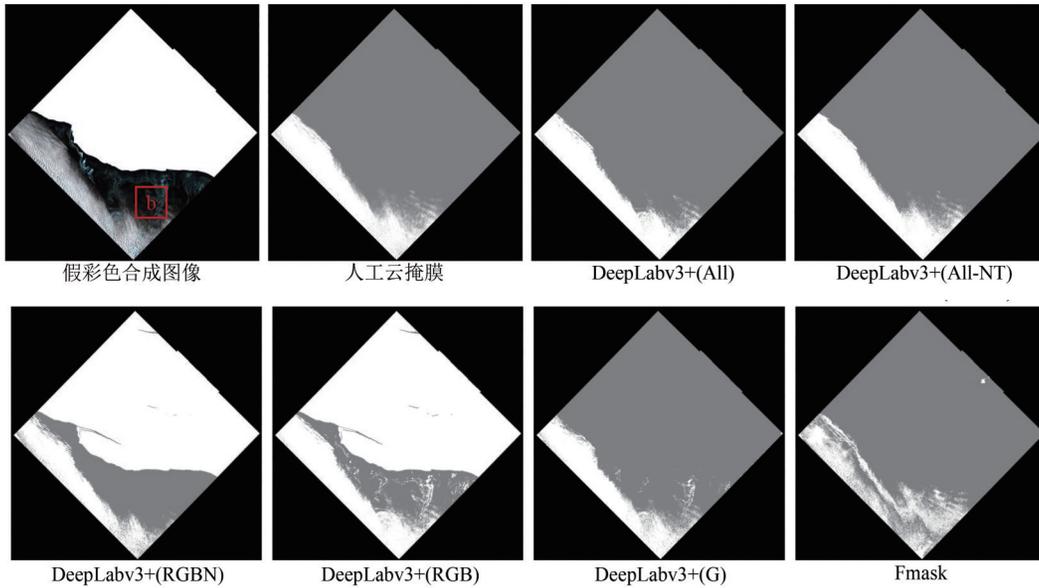
中其他遥感传感器与Landsat传感器相比，除空间分辨率、波段设置有差异外，相应的空间响应函数、光谱响应函数也可能存在差异。本文进一步尝试将已训练好的DeepLabv3+云检测网络用于1幅20 m分辨率的Sentinel-2影像与1幅1 km分辨率的MODIS影像，并将云检测结果与专门为Sentinel-2调整的Fmask算法(Zhu等, 2015)以及MODIS官方云标记数据(MOD35_L2)对比。云检测结果如图13与图14所示，该云检测网络仍可大致准确地检测出Sentinel与MODIS影像的云分布，并且在目视效果上仍略优于为这两个传感器专门设计的算法生成的云标记数据。这进一步说明，DeepLabv3+云检测在不同类型遥感影像上的泛化潜力，这与Li等(2019)的结果相似。但是，将来仍需要在更多类型数据集上进行定量测试以进一步确认云检测网络的泛化性能。

本文的实验用总体精度表征云检测的性能，但是不同用户可能对于错分和漏分性能有不同的倾向。有学者认为云检测需要适当减少漏分而容忍错分，这也是Fmask算法的设计倾向(Zhu和Woodcock, 2012)。对于DeepLabv3+网络而言，用户也可以通过调整归属概率阈值以改变错分与漏分的倾向性，灵活满足不同用户的需求。为更全面的对比DeepLabv3+与Fmask的错分与漏分性能，本文基于两种方法输出的云归属概率进一步绘制其接受者操作特性ROC(Receiver Operating Characteristic)曲线，即错分率与召回率随概率阈值的变化关系。ROC的曲线下面积AUC(Area Under Curve)越大，说明相应方法的分类性能越好。如图15所示，DeepLabv3+的ROC曲线在Fmask的ROC曲线之上，覆盖更大的AUC面积。由图15还可以看出，默认阈值设置下，DeepLabv3+在保持与Fmask相近的召回率水平下，具备更低的错分率。该结果进一步说明了DeepLabv3+比Fmask具备更好的云检测性能。

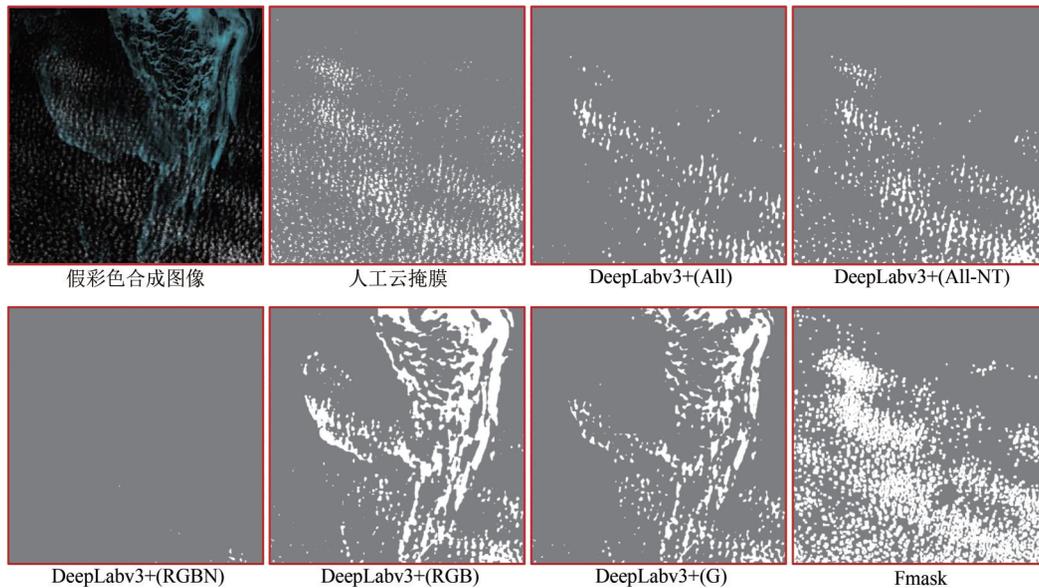
但是，本文也发现基于DeepLabv3+云检测的一些问题。首先，在一些特殊的影像中，比如在灌丛景观测试集中的一景影像(图6)上，DeepLabv3+云检测网络不能准确检测出其中的云，说明仍然存在少部分影像的云无法被正确识别；如何找出这些少数影像特例并改善其云检测精度可能是未来研究的重要方向。其次，加入热红外波段在很多情况下反而会降低DeepLabv3+网络的

云检测性能, 这与过去的经验方法不一致 (Frantz 等, 2018); 这可能是因为热红外波段和其他光学波段的性质差异较大, 导致本文使用的神经网络未能正确利用热红外波段的信息。最后, DeepLabv3+ 尽管结合低级特征上采样恢复空间信

息, 但仍存在空间细节信息的丢失, 这导致 DeepLabv3+ 的云检测结果的边缘过于平滑; 相比之下, 基于像元的传统识别方法对空间细节的表达更好, 结合像元级的识别方法与图像分割算法可能有助于该问题的解决。



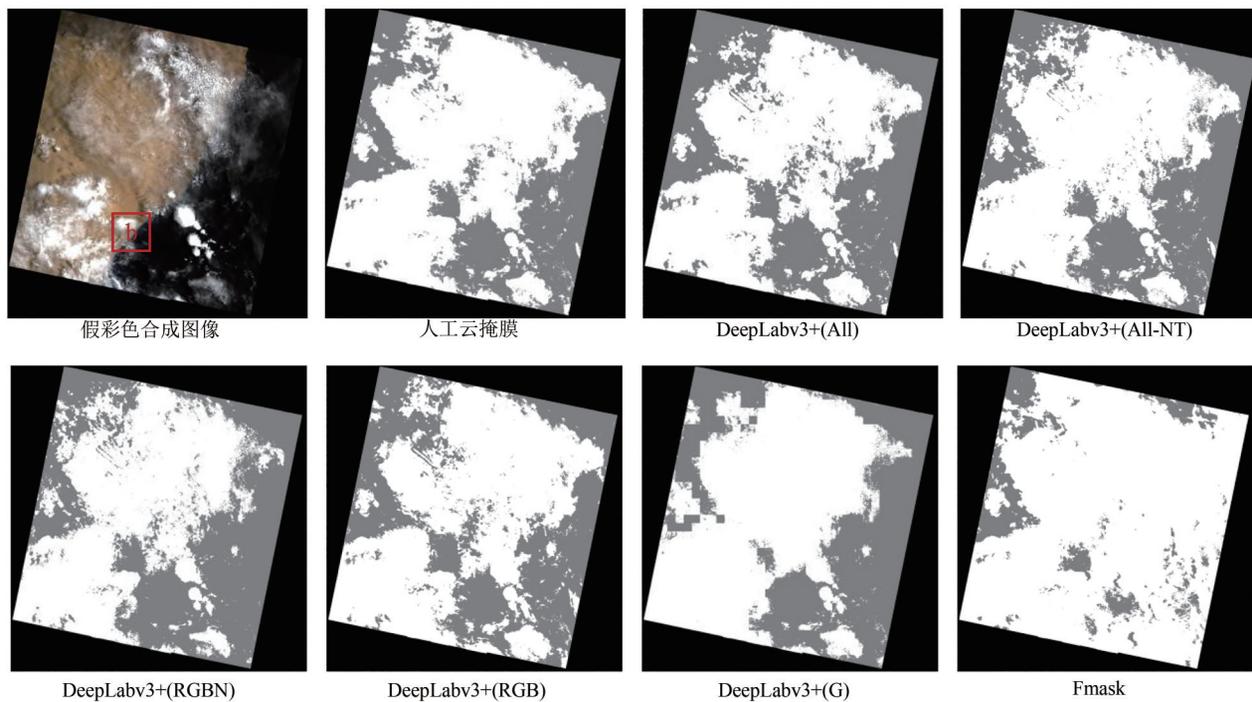
(a) 整幅图像
(a) Whole image



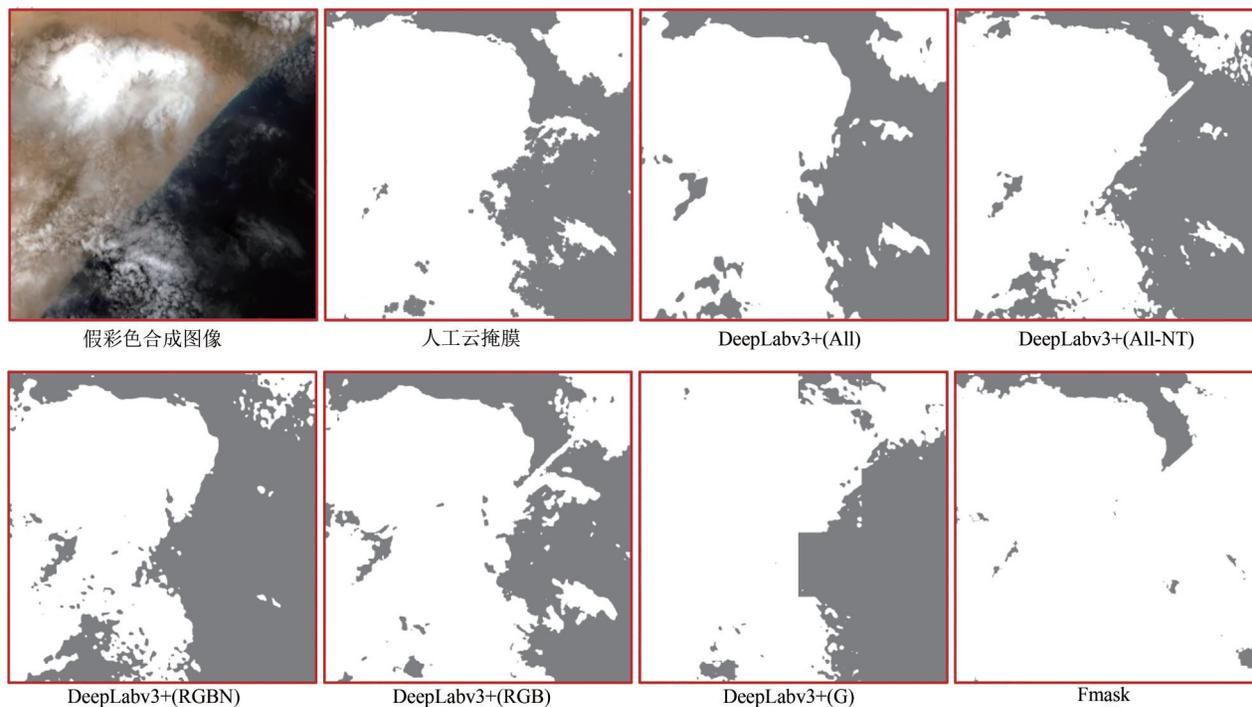
(b) 红色矩形区域放大图像
(b) Enlarged images of the subset area in the red rectangle
□ 云覆盖区域 ■ 无云区域 ■ 填充值

图 11 由不同波段组合样本训练得到 DeepLabv3+ 网络在一景冰/雪景观 Landsat 图像上的云检测结果 (图像假彩色合成均使用波段 5、4、3)

Fig. 11 Cloud detection results of the DeepLabv3+ trained by training set with different band combinations and Fmask on test images captured over snow/ice (All the images are displayed using bands 5, 4 and 3 In each cloud mask: White represents the cloud coverage area; Gray represents the cloud-free area; Black represents fill value)



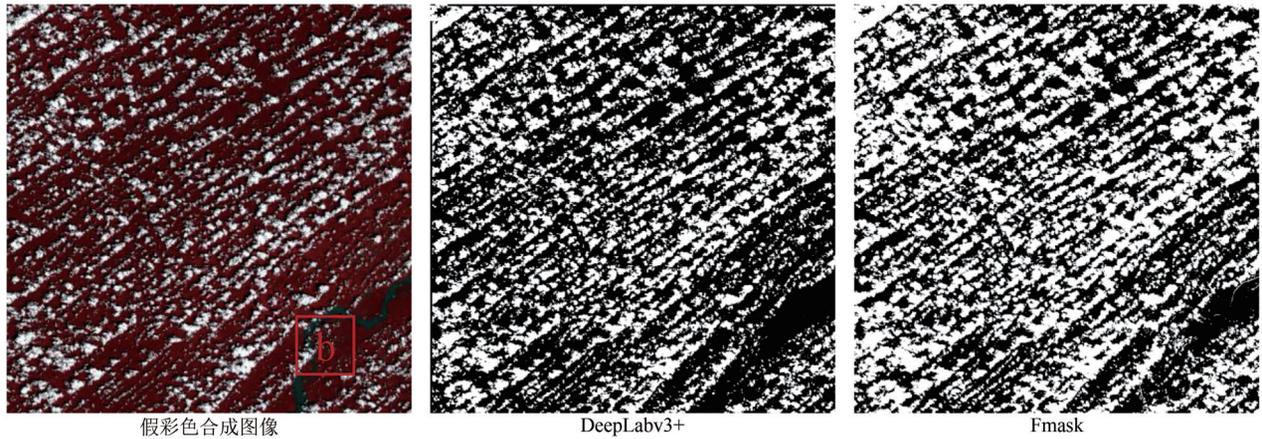
(a) 整幅图像
(a) Whole image



(b) 红色矩形区域放大图像
(b) Enlarged images of the subset area in the red rectangle
□ 云覆盖区域 ■ 无云区域 ■ 填充值

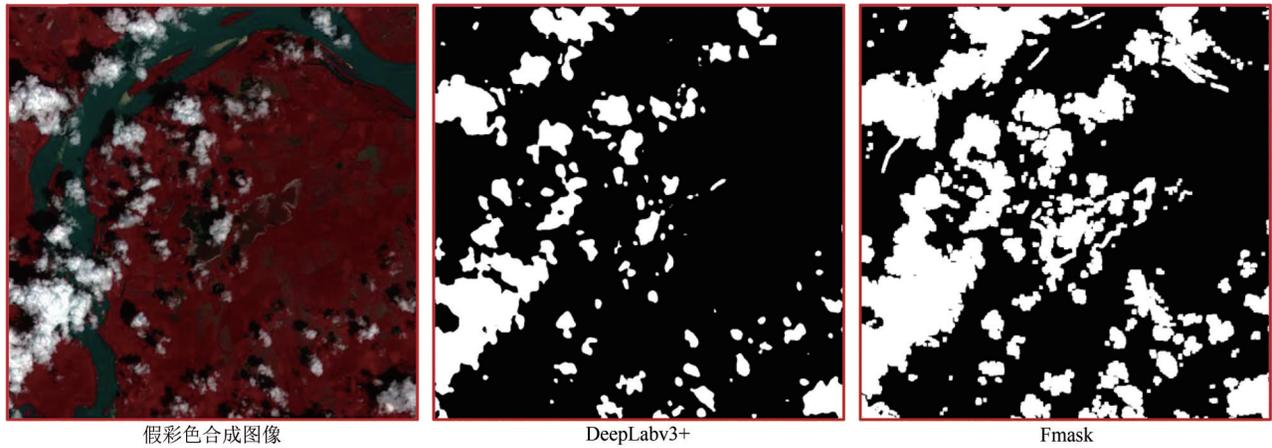
图 12 由不同波段组合样本训练得到 DeepLabv3+ 网络在一景裸地景观 Landsat 图像上的云检测结果(图像假彩色合成均使用波段 5、4、3)

Fig. 12 Cloud detection results of the DeepLabv3+ trained by training set with different band combinations and Fmask on test images captured over barren(All the images are displayed using bands 5, 4 and 3)



(a) 由 All-NT 波段组合方式样本训练得到的 DeepLabv3+ 网络在 20 m 分辨率 Sentinel-2A 影像上的云检测结果

(a) The cloud detection result of the DeepLabv3+ trained by training set with All-NT band combination on Sentinel-2A image with 20 m resolution



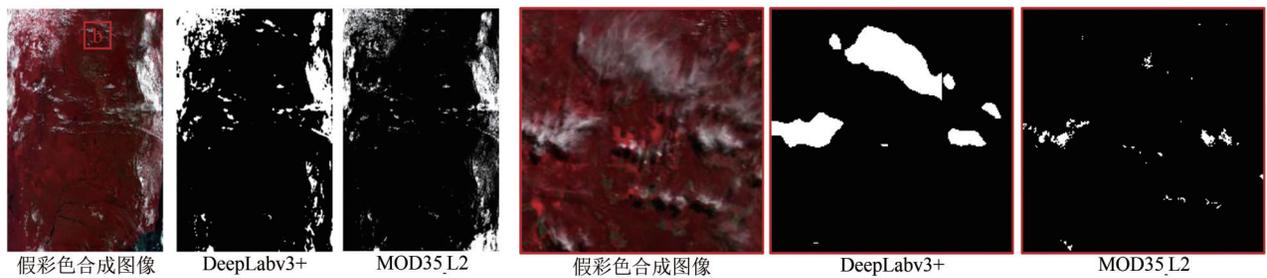
(b) 红色矩形区域放大图像

(b) Enlarged images of the subset area in the red rectangle

□ 云覆盖区域 ■ 无云区域 ■ 填充值

图 13 DeepLabv3+ 网络在 Sentinel-2A 影像 (ID: L1C_T21MWR_A023130_20191126T141046) 上的应用

Fig.13 Cloud detection examples of DeepLabv3+ on Sentinel-2A image (ID: L1C_T21MWR_A023130_20191126T141046)



(a) 30m 分辨率全混合景观样本训练得到的 DeepLabv3+ 网络在 1 km 分辨率 MODIS 影像上的云检测结果
(a) The cloud detection result of the DeepLabv3+ trained by the fully mixed training set with 30m resolution on MODIS image with 1 km resolution

(b) 红色矩形区域放大图像, 其中 MOD35_L2 为 MODIS 的云掩膜产品
(b) Enlarged images of the subset area in the red rectangle. MOD35_L2 is a cloud mask product of MODIS

□ 云覆盖区域 ■ 无云区域 ■ 填充值

图 14 DeepLabv3+ 网络在 MODIS 影像 (ID: MOD021KM.A2014141.1330.061.2017311185050) 上的应用

Fig.14 Cloud detection examples of DeepLabv3+ on MODIS image (ID: MOD021KM.A2014141.1330.061.2017311185050)

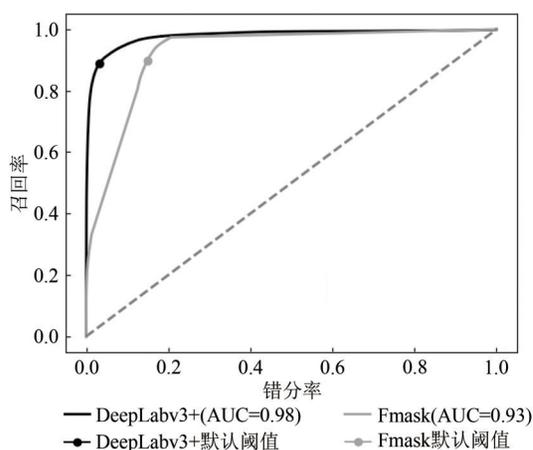


图 15 30 m 分辨率全混合景观样本训练得到的 DeepLabv3+ 网络及 Fmask 算法在所有测试影像上的 ROC 曲线

Fig.15 The receiver operating characteristic curves (ROC) of the DeepLabv3+ trained by the fully mixed training set with 30m resolution and Fmask on all the test images

6 结 论

本研究使用“L8 Biome”数据集构建不同的训练集和测试集，分别训练 DeepLabv3+ 网络并测试其云检测的精度，探究了该算法在不同地表景观、不同空间分辨率以及不同光谱波段组合影像上的泛化性能。主要结果如下：(1) 使用全混合景观类型的训练集训练出来的云检测网络在总体检测精度 (92.81%) 与稳定度 (标准差 12.08%) 上都优于使用单一景观类型的训练集训练得到的云检测网络，也优于 Fmask 的总体精度 (88.75%) 与稳定度 (标准差 17.34%)；(2) 将全混合景观训练集中剔除一类景观的样本 (冰/雪景观除外) 构建的“混合-1”训练集与全混合景观训练集训练的 DeepLabv3+ 网络的云检测精度也相差不多；(3) 基于 30 m 空间分辨率的全混合景观训练样本集训练得到的 DeepLabv3+ 云检测网络在不同分辨率 (30 m、60 m、120 m、240 m) 的测试集上云检测精度差异不大，都取得较好的效果，相反 Fmask 直接应用于低分辨率影像时精度明显下降；(4) DeepLabv3+ 能充分自适应不同波段的信息用于云检测，并且更多的光谱波段输入总的来说能够提高 DeepLabv3+ 的云检测的精度和稳定度，其中短波红外波段对于 DeepLabv3+ 区分冰/雪与云具有重要价值，而热红外波段对 DeepLabv3+ 云检测网络的性能提升很微小。以上结果表明训练 DeepLabv3+ 云检测网络，最好包括多种地表景观

的影像样本，一定要包括冰/雪景观的影像样本。但是缺少一部分除冰/雪外的其他景观影像的训练样本并不会影响云检测精度，因此本文认为对于现有样本库“L8 Biome”数据集，已能覆盖绝大多数情况，进一步增加样本对云检测的精度提升有限。此外，基于“L8 Biome”数据训练的 DeepLabv3+ 云检测网络能够直接应用于不同空间分辨率的遥感影像并取得较高的云检测精度；也能通过对训练集进行波段重组的方式迁移应用到不同波段组合的遥感影像以充分发挥不同波段的信息。综上，DeepLabv3+ 具备比 Fmask 算法更高的云检测精度与灵活度，能够适用于不同景观、不同空间分辨率与不同波段组合的遥感影像。

参考文献 (References)

- Ball J E, Anderson D T and Chan C S. 2017. Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community. *Journal of Applied Remote Sensing*, 11(4): 042609 [DOI: 10.1117/1.JRS.11.042609]
- Chen L C, Papandreou G, Schroff F and Adam H. 2017. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-12-05). <http://arxiv.org/pdf/1706.05587v3>
- Chen L C, Zhu Y K, Papandreou G, Schroff F and Adam H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 833-851 [DOI: 10.1007/978-3-030-01234-2_49]
- Chen Q, Zheng L J, Li X J, Xu C B, Wu Y, Xie D H and Liu L. 2019. Water body extraction from high-resolution satellite remote sensing images based on deep learning. *Geography and Geo-information Science*, 35(4): 43-49 (陈前, 郑利娟, 李小娟, 徐崇斌, 吴侯, 谢东海, 刘亮. 2019. 基于深度学习的高分遥感影像水体提取模型研究. *地理与地理信息科学*, 35(4): 43-49)[DOI: 10.3969/j.issn.1672-0504.2019.04.007]
- Chen S L, Chen X H, Chen J, Jia P F, Cao X and Liu C Y. 2016. An iterative haze optimized transformation for automatic cloud/haze detection of Landsat imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 54(5): 2682-2694 [DOI: 10.1109/TGRS.2015.2504369]
- Chen X D, Zhang X, Liu L Y and Wang X F. 2019. Enhanced multi-temporal cloud detection algorithm for optical remote sensing images. *Journal of Remote Sensing*, 23(2): 280-290 (陈曦东, 张肖, 刘良云, 汪晓帆. 2019. 增强型多时相云检测. *遥感学报*, 23(2): 280-290) [DOI: 10.11834/jrs.20198017]
- Foga S, Scaramuzza P L, Guo S, Zhu Z, Dilley Jr R D, Beckmann T, Schmidt G L, Dwyer J L, Joseph Hughes M and Laue B. 2017. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote Sensing of Environment*,

- 194: 379-390 [DOI: 10.1016/j.rse.2017.03.026]
- Frantz D, Haß E, Uhl A, Stoffels J and Hill J. 2018. Improvement of the Fmask algorithm for Sentinel-2 images: separating clouds from bright surfaces based on parallax effects. *Remote Sensing of Environment*, 215: 471-481 [DOI: 10.1016/j.rse.2018.04.046]
- Harshvardhan, Randall D A and Corsetti T G. 1989. Earth radiation budget and cloudiness simulations with a general circulation model. *Journal of the Atmospheric Sciences*, 46(13): 1922-1942 [DOI: 10.1175/1520-0469(1989)046<1922:ERBACS>2.0.CO;2]
- Hughes M J and Hayes D J. 2014. Automated detection of cloud and cloud shadow in single-date Landsat imagery using neural networks and spatial post-processing. *Remote Sensing*, 6(6): 4907-4926 [DOI: 10.3390/rs6064907]
- Irish R R. 2000. Landsat 7 automatic cloud cover assessment//Proceedings of SPIE 4049, Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI. Orlando, USA: SPIE: 348-355 [DOI: 10.1117/12.410358]
- Jedlovec G J, Haines S L and LaFontaine F J. 2008. Spatial and temporal varying thresholds for cloud detection in GOES imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 46(6): 1705-1717 [DOI: 10.1109/TGRS.2008.916208]
- Jeppesen J H, Jacobsen R H, Inceoglu F and Toftgaard T S. 2019. A cloud detection algorithm for satellite imagery based on deep learning. *Remote Sensing of Environment*, 229: 247-259 [DOI: 10.1016/j.rse.2019.03.039]
- Lee J, Weger R C, Sengupta S K and Welch R M. 1990. A neural network approach to cloud classification. *IEEE Transactions on Geoscience and Remote Sensing*, 28(5): 846-855 [DOI: 10.1109/36.58972]
- Li Z W, Shen H F, Cheng Q, Liu Y H, You S C and He Z Y. 2019. Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors. *ISPRS Journal of Photogrammetry and Remote Sensing*, 150: 197-212 [DOI: 10.1016/j.isprsjprs.2019.02.017]
- Li Z W, Shen H F, Li H F, Xia G S, Gamba P and Zhang L P. 2017. Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery. *Remote Sensing of Environment*, 191: 342-358 [DOI: 10.1016/j.rse.2017.01.026]
- Lyu M M, Han L J, Tian S F, Zhou W Q, Li W F and Qian Y G. 2016. Cloud detection under varied surfaces and atmospheric conditions with MODIS imagery. *Journal of Remote Sensing*, 20(6): 1371-1380 (吕明明, 韩立建, 田淑芳, 周伟奇, 李伟峰, 钱雨果). 2016. 多样地表和大气状况下的MODIS数据云检测. *遥感学报*, 20(6): 1371-1380 [DOI: 10.11834/jrs.20165281]
- Mateo-García G, Laparra V, López-Puigdollers D and Gómez-Chova L. 2020. Transferring deep learning models for cloud detection between Landsat-8 and Proba-V. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160: 1-17 [DOI: 10.1016/j.isprsjprs.2019.11.024]
- Papadomanolaki M, Vakalopoulou M, Zagoruyko S and Karantzas K. 2016. Benchmarking deep learning frameworks for the classification of very high resolution satellite multispectral data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3(7): 83-88 [DOI: 10.5194/isprs-annals-III-7-83-2016]
- Prasad A K, Chai L, Singh R P and Kafatos M. 2006. Crop yield estimation model for Iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation*, 8(1): 26-33 [DOI: 10.1016/j.jag.2005.06.002]
- Qiu S, Zhu Z and He B B. 2019. Fmask 4.0: improved cloud and cloud shadow detection in Landsats 4-8 and Sentinel-2 imagery. *Remote Sensing of Environment*, 231: 111205 [DOI: 10.1016/j.rse.2019.05.024]
- Roy D P, Ju J C, Kline K, Scaramuzza P L, Kovalsky V, Hansen M, Loveland T R, Vermote E and Zhang C S. 2010. Web-enabled Landsat Data (WELD): landsat ETM+ composited mosaics of the conterminous United States. *Remote Sensing of Environment*, 114(1): 35-49 [DOI: 10.1016/j.rse.2009.08.011]
- Saunders R W and Kriebel K T. 1988. An improved method for detecting clear sky and cloudy radiances from AVHRR data. *International Journal of Remote Sensing*, 9(1): 123-150 [DOI: 10.1080/01431168808954841]
- Schmidhuber J. 2015. Deep learning in neural networks: an overview. *Neural Networks*, 61: 85-117 [DOI: 10.1016/j.neunet.2014.09.003]
- Shan N, Zheng T Y and Wang Z S. 2009. High-speed and high-accuracy algorithm for cloud detection and its application. *Journal of Remote Sensing*, 13(6): 1138-1155 (单娜, 郑天垚, 王贞松). 2009. 快速高准确度云检测算法及其应用. *遥感学报*, 13(6): 1138-1155 [DOI: 10.11834/jrs.20090612]
- Shen H F, Li X H, Cheng Q, Zeng C, Yang G, Li H F and Zhang L P. 2015. Missing information reconstruction of remote sensing data: a technical review. *IEEE Geoscience and Remote Sensing Magazine*, 3(3): 61-85 [DOI: 10.1109/MGRS.2015.2441912]
- Stowe L L, Davis P A and McClain E P. 1999. Scientific basis and initial evaluation of the CLAVR-1 global clear/cloud classification algorithm for the advanced very high resolution radiometer. *Journal of Atmospheric and Oceanic Technology*, 16(6): 656-681 [DOI: 10.1175/1520-0426(1999)016<0656:SBAIEO>2.0.CO;2]
- Sun L, Yang X, Jia S F, Jia C, Wang Q, Liu X Y, Wei J and Zhou X Y. 2020. Satellite data cloud detection using deep learning supported by hyperspectral data. *International Journal of Remote Sensing*, 41(4): 1349-1371 [DOI: 10.1080/01431161.2019.1667548]
- Verbesselt J, Hyndman R, Newnham G and Culvenor D. 2010. Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment* 114(1): 106-115 [DOI: 10.1016/j.rse.2009.08.014]
- Wang J Q, Li J S, Zhou H C and Zhang X. 2019. Typical element extraction method of remote sensing image based on Deeplabv3+ and CRF. *Computer Engineering*, 45(10): 260-265, 271 (王俊强, 李建胜, 周华春, 张旭). 2019. 基于 Deeplabv3+ 与 CRF 的遥感影像典型要素提取方法. *计算机工程*, 45(10): 260-265, 271 [DOI: 10.19678/j.issn.1000-3428.0053359]
- Yang W, Zhang J L, Xu Z Y and Hu K. 2019. Real-time DeepLabv3+ for pedestrian segmentation. *Journal of Optical Technology*, 86

- (9): 570-578 [DOI: 10.1364/JOT.86.000570]
- Zhang Y C, Rossow W B, Laciš A A, Oinas V and Mishchenko M I. 2004. Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: refinements of the radiative transfer model and the input data. *Journal of Geophysical Research*, 109(D19): D19105 [DOI: 10.1029/2003JD004457]
- Zhu X L and Helmer E H. 2018a. An automatic method for screening clouds and cloud shadows in optical satellite image time series in cloudy regions. *Remote Sensing of Environment*, 214: 135-153 [DOI: 10.1016/j.rse.2018.05.024]
- Zhu Z, Qiu S, He B B and Deng C B. 2018b. Cloud and cloud shadow detection for Landsat images: the fundamental basis for analyzing Landsat time series//Weng Q H, ed. *Remote Sensing Time Series Image Processing*. Boca Raton: CRC Press: 25-46
- Zhu Z, Wang S X and Woodcock C E. 2015. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel 2 images. *Remote Sensing of Environment*, 159: 269-277 [DOI: 10.1016/j.rse.2014.12.014]
- Zhu Z and Woodcock C E. 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, 118: 83-94 [DOI: 10.1016/j.rse.2011.10.028]
- Zi Y, Xie F Y and Jiang Z G. 2018a. A cloud detection method for Landsat 8 images based on PCANet. *Remote Sensing*, 10(6): 877 [DOI: 10.3390/rs10060877]

Generalization ability of cloud detection network for satellite imagery based on DeepLabv3+

PENG Longkang, LIU Licong, CHEN Xuehong, CHEN Jin, CAO Xin, QIU Yuean

1. State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China;

2. Beijing Engineering Research Center for Globe Land Remote Sensing Products, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

Abstract: Deep learning algorithms have been developed and applied in detecting clouds for satellite imagery in recent years. However, deep neural network models consist of thousands or millions of parameters, thus usually requiring large amounts of training data. Therefore, understanding the generalization ability of deep learning techniques is vital in their application to cloud detection of different types of satellite imagery. Taking DeepLabv3+, a typical deep semantic segmentation algorithm, as an example, this study explored the generalization ability of the algorithm on the cloud detection of satellite imagery with different landscapes, spatial resolutions, and spectral band combinations based on the cloud labeled dataset "L8 Biome." The "L8 Biome" dataset consists of 96 typical Landsat 8 OLI images and the corresponding manual cloud mask, which has been widely used for evaluating the performance of cloud and cloud shadow detection methods. First, the cloud labeled dataset "L8 Biome" was used to generate different training and test samples with different landscapes, spatial resolutions, and band combinations. Then, the performance of DeepLabv3+ was evaluated based on different training and test sets and compared with that of the typical Function of Mask (Fmask) algorithm. Results show the following: (1) the DeepLabv3+ trained by a fully mixed training set (consisting of imagery captured over all landscapes) has higher overall cloud accuracy (92.81%) and stability (standard deviation 12.08%) than that trained by the training set of imagery captured over a single landscape and performed better than the Fmask algorithm with an overall cloud accuracy of 88.75% and stability of 17.34%, indicating that the training set of the deep learning algorithm should include images captured over various landscape types; (2) the DeepLabv3+ trained by "mixed-1" training sets, which were built by removing the images captured over a single landscape type (except of snow/ice) from the fully mixed training set, achieved comparable accuracies with that trained by a fully mixed training set, indicating that the available training set is diverse enough and has a good generalization ability on imagery captured over different landscapes; (3) the DeepLabv3+ trained by a fully mixed training set with a 30-m resolution achieved similar cloud detection accuracies on the satellite images with different spatial resolutions (i.e., 30, 60, 120, and 240 m), indicating that the trained DeepLabv3+ could be directly applied on satellite imagery with different spatial resolution, whereas the Fmask algorithm performed poorly on the images with coarse resolutions; (4) DeepLabv3+ could explore the effective information from different spectral bands for cloud detection, and more spectral band input can generally improve the overall cloud accuracy and stability of DeepLabv3+. Among all input bands, shortwave infrared bands are greatly helpful for distinguishing snow/ice from clouds, whereas the thermal infrared band marginally improves the cloud detection accuracy for DeepLabv3+. Results indicate that the DeepLabv3+ cloud detection network trained by the "L8 Biome" dataset can be applied to various types of satellite imagery and outperforms the Fmask algorithm.

Key words: deep learning, cloud detection, DeepLabv3+, generalization ability, landscape, spectral band combination, spatial resolution

Supported by National Natural Science Foundation of China (No. 41871224)