

勺型网络：用于 Landsat 遥感图像云检测的新型网络

王树立*, 唐海蓉, 计璐艳

(中国科学院空天信息创新研究院, 北京 100094; 中国科学院大学, 北京 100049;
中国科学院空间信息处理与应用系统技术重点实验室, 北京 100190)

摘要 针对目前用于遥感图像云检测的神经网络模型存在光谱信息未能充分利用而导致的边缘平滑、碎云漏检率大、计算复杂等不足, 本文提出了一种新型且轻量的网络, 称为勺型网络 (Spoon-Net, 简称 S-Net), 应用于 Landsat 遥感图像的云检测。S-Net 分为两个阶段, 第一阶段, 使用 1×1 的卷积核提取图像光谱特征, 避免图像细节被模糊; 第二阶段, 使用 encoder-decoder 框架提取图像空间特征, 并引入分组卷积, 对第一阶段提取的每一层光谱通道单独进行卷积, 保持光谱特征并减少模型参数。模型在 Landsat8 Biome 数据训练测试并评估, 结果表明模型在内存与时间上具有较大优势, 并达到了 95% 的准确率。

关键词 Landsat, 云检测, 神经网络

中图分类号 TP751.2

Spoon network: A new network structure of Landsat imagery cloud detection

Wang Shuli, Tang Hairong, Ji Luyan

(Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;
University of Chinese Academy of Sciences, Beijing 100049, China; Key Laboratory of Technology in Geo-spatial
Information Processing and Application System, Beijing 100190, China)

Abstract In view of the shortcomings of the current neural network model for remote sensing image cloud detection, such as edge smoothness, large rate of missing detection of broken cloud, and complex calculation, which is caused by the insufficient use of spectral information, this paper proposes a new and lightweight network, called spoon net (S-Net for short), which is applied to the cloud detection of Landsat remote sensing image. S-Net is divided into two stages. In the first stage, the convolution kernel of 1×1 is used to extract image spectral features to avoid image details being blurred; in the second stage, the encoder decoder framework is used to extract image spatial features, and group convolution is introduced to convolute each layer of spectral channels extracted in the first stage separately to maintain spectral features and reduce model parameters. The model is tested and evaluated in landsat8 biome data training, and the results show that the model has a great advantage in memory and time, and achieves an accuracy of 95 %.

Keywords Landsat, cloud detection, neural network

引言

以 Landsat 卫星为代表的遥感数据在当今社会的生产和生活中扮演着至关重要的角色, 在农业产

量估算^[1]、变化检测^[2]、灾难评估^[3]等方面发挥着重要的作用。随着科技的发展, 遥感数据变得越来越, 且越来越容易获得。海量的多波段遥感数据也迫切需要高效率和高鲁棒性的算法进行处理和数据挖

*通讯作者: E-mail:wangshuli18@mailsucas.ac.cn

掘。然而,在 Landsat 数据集上,每年有高达 40% 的像素被云覆盖^[4],云层作为光学遥感图像的主要污染源,对遥感图像的应用造成了极大的限制。所以对云检测算法的研究一直是遥感领域中的热点。

实时的云检测算法目前可以分为两类:基于光谱信息的和基于空间信息的。光谱是地物最本质的特征之一,不同的地物有不同的辐射与反射特性,云在反射波段表现为亮目标,发射波段表现为暗目标。有些基于光谱信息的方法^[5]直接利用波段特征,也有些通过非线性映射构造新的特征(如波段比值、波段指数等),并精心设置阈值以更好地地区分地物。Irish 等人^[6]提出的 ACCA 使用 Landsat7 ETM+ 波段 2-6 的信息,获得暖云掩码,冷云掩码,非云掩码和雪掩码。Zhu 等人提出的 FMask 算法^[7]用到了 Landsat 几乎所有的波段,通过设置亮度阈值、色度阈值、温度阈值、NDVI、NDSI 等,通过决策树选择出两个潜在云掩膜,并组合成最终的结果。此类方法实现简单,便于理解,可解释性强,在一般情况下可以取得较好的效果,但当地面覆盖了冰、雪、沙漠,或云为薄卷云、小积云时,云和地面难以区分。

在空间上,云的表现则更加多样,有小面积的碎云,一大片的层云,有较厚的积云,也有较薄的卷云,但云是由水汽聚集而成,处于中心位置的云更加容易识别,边缘部分或者较为模糊的云可以利用空间分布进行识别。一些方法通过提取图像纹理特征,如 LBP 特征、HOG 特征、haar 特征等,利用云与地物在空间上结构的不同进行区分。还有一些文章将高分辨率图像切割成一张张子图或超像素,如 SLIC^[8],再利用机器学习的方法对子图或超像素进行分类,如 SVM^[9]、MLP^[10]。这些方法一方面降低了图像的分辨率,另一方面受到云多样性的影响,在处理薄云、小面积云时效果并不佳。

近年来,深度学习在自然语言处理、降维、图像分类、目标检测、语义分割等方面取得了诸多成果。从 AlexNet 开始,深度学习开始席卷图像处理领域。相较许多传统方法需要人工构造特征和精心选择阈值,深度学习通过构造多层神经网络,自动提取特征和阈值。而且,精心选择的神经网络可以构造出高维特征,更加有效地区分云与地物。

目前有很多方法将全卷积网络应用于遥感图像的云检测。Jeppesena 等人^[11]将 U-Net 应用于 Landsat 遥感图像云检测。Chai 等人^[12]将 SegNet 应用于 Landsat 图像。Hughes 等人^[13]将 FCN 应

用于云检测。这些方法是对已有模型在遥感图像云检测任务上的应用,都取得了不错的效果。Zhang 等人^[14]在原有通道的基础上增加了并行的 1×1 卷积核,增强了对光谱的重视,取得了不错的效果。但 these 方法存在一些不足。如:没有充分利用地物的光谱特征;过于重视遥感图像的空间特征。这不仅会导致训练过程中容易出现过拟合、训练不稳定等情况,也会使小的碎云容易被忽视,边缘细节容易被丢失,分类结果过于平滑等。

为了充分利用云的光谱属性,保持云的边缘细节,本文提出了一个新颖的、简单的、有效的网络,称为勺型网络(Spoon-Net,简称 S-Net)用于 Landsat 图像云检测。网络主要包括两个阶段。第一阶段,光谱特征提取阶段。在这一阶段,我们完全使用 1×1 的卷积核,不受空间信息干扰,专门对图像进行光谱特征提取。而且 1×1 卷积核可以在不降低图像分辨率的前提下,使得后续的空间特征更加容易被提取。第二阶段,空间特征提取阶段。这一阶段,采用轻量化的 encoder-decoder 框架。进一步,我们不会破坏已提取的光谱特征,而是采用组卷积(Group Conv)的方式,将第一部分得到的每一种光谱特征作为一组进行单独的卷积。由于光谱特征提取阶段的存在,组卷积在大大减少参数数量的同时,有效提取空间特征。最终的分类结果将像素分为云与非云两类。实验结果表明,S-Net 可以在模型参数大大减小的情况下,明显提高云检测精度。

1 数据与方法

1.1 数据

本文采用的光学遥感卫星数据集来自美国航空航天局(NASA) Landsat8 卫星。2013 年 2 月 11 日,NASA 成功发射 Landsat-8 卫星。Landsat-8 卫星上携带两个传感器,分别是 OLI 陆地成像仪和 TIRS 热红外传感器。OLI 提供 9 个波段,波段范围从 0.43um 到 2.30um; TIRS 提供地表温度数据,包括两个波段,波段范围从 10.60um 到 12.51um,具体信息见表 1。Landsat 系列卫星每 16 天可以实现一次全球覆盖。

表 1: Landsat8 波段信息

Tab 1: Landsat8 band information

传感器类型	编号	波段	波长范围 (μm)	空间分辨率 (m)
OLI	1	Coastal	0.433-0.453	30
	2	Blue	0.450-0.515	30
	3	Green	0.525-0.600	30
	4	Red	0.630-0.680	30
	5	NIR	0.845-0.885	30
	6	SWIR1	1.56-1.66	30
	7	SWIR2	2.1-2.3	30
	8	Pan	0.5-0.68	15
	9	Cirrus	1.36-1.39	30
TIRS	10	TIRS1	10.60-11.19	100
	11	TIRS2	11.50-12.51	100

为了对模型进行训练与测试,本文利用带人工标注标签的全球云和云影验证数据集“L8 Biome Cloud Validation Masks”^[15],该数据集共有 96 景图片,每景约 7000 * 7000 大小,包含 8 个种类的下垫面。

每景图片的标签均是人工标注,可信度较高。每个文件包含.TIF 格式的 Landsat 8 Level-1 数据文件、质量文件和.img (ENVI) 格式的真值标签,人工标志位如表 2所示。

表 2: L8 Biome 数据人工标注标志位

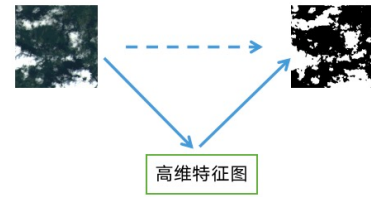
Tab 2: flag of Landsat8 biome

数值	0	64	128	192	255
解释	填充值	云阴影	非云像素	薄云	云

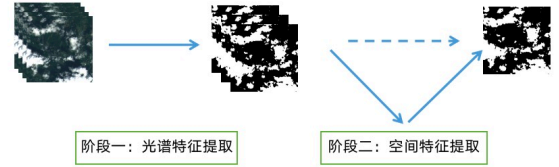
根据云量百分比的多少,‘L8 Biome’中 96 景分为 clear, midcloud, cloud 三种,每种各占三分之一,云量低于 35% 的为 clear,云量高于 65% 的为 cloud,云量介于 35% 与 65% 之间的为 midcloud。我们将数据标签简单地分为云与非云两类,将每景图像均匀切割为 256 * 256 大小的子图(共约 5 万张),切割时过滤掉带填充值的图片,因此图像边缘的填充像素并不会出现在训练与测试的步骤中。选择除了全色波段的所有波段(TIRS 数据的空间分辨率已重采样到 30m),共 10 个波段,训练集与测试集的比例为 6:4。

1.2 Encoder-decoder 框架与改进

Encoder-decoder 框架是图像分割领域的主流框架。Encoder 的作用是提取空间特征,decoder 的作用是解析空间特征,并将图像还原到原图的大小,以获得像素级别的分类,跳层连接统筹兼顾感受野与空间分辨率。FCN^[16]与 U-Net^[17]均采用这种框架。但该框架一般应用于灰度图像与彩色图像,若应用于遥感图像云检测,虽然可以涉及光谱维度的计算,但主要是在提取空间特征,对光谱特征的提取能力不足。本文考虑遥感图像多波段的特点,对 encoder-decoder 框架进行改进,将其扩充为两阶段,分步提取光谱特征与空间特征。框架对比如图 1所示。



(a) encoder-decoder 框架.



(b) Spoon-Net 框架.

图 1: encoder-decoder 框架与 Spoon-net 对比

Fig 1: Comparison of encoder-decoder framework and Spoon-net

1.3 网络结构

本文提出的 S-Net 是一个两阶段模型,第一阶段是光谱特征提取阶段,第二阶段是空间特征提取阶段。详细模型结构如图 2所示,每个矩形框上面的数字代表特征图的个数。

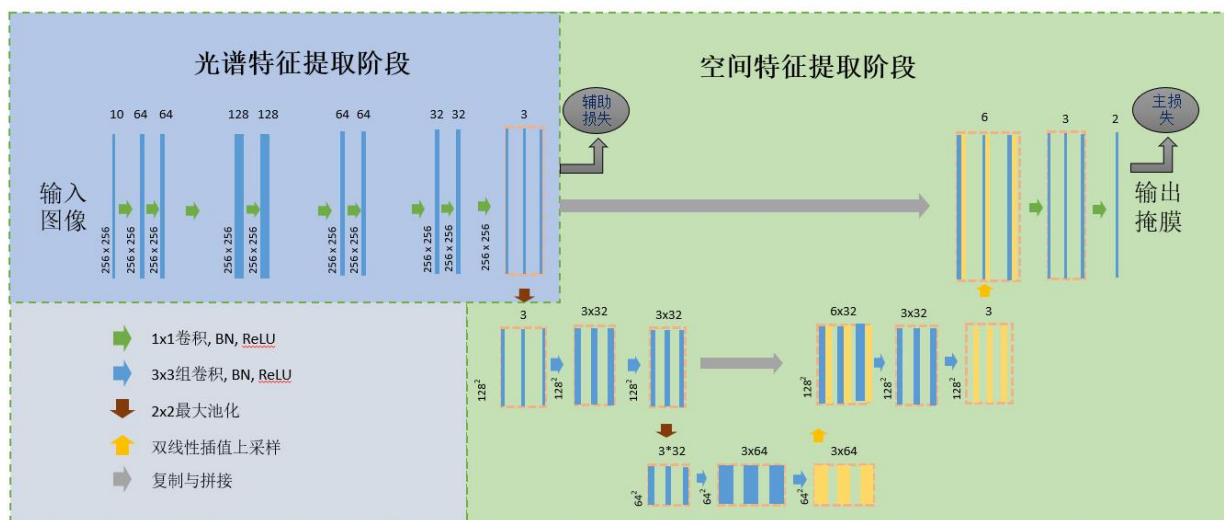


图 2: Spoon-net 详细结构.

Fig 2: Spoon-net framework

1. 光谱特征提取阶段。该阶段使用 1×1 的卷积核单个像素进行计算，提取出三个最优的光谱特征。因为不涉及到空间信息，所以可以有效保持图像的细节。多层 1×1 的卷积操作等价于一个多层感知机（MLP），MLP 可以有效提取非线性光谱特征，为后续分类提供良好的基础，并且不会降低图像分辨率。这一阶段输出的每一层特征图都是一种有效的光谱特征，类似于 NDVI、NDWI 等，但远比它们复杂得多，更加具有非线性，也更有效。详细结构如图 3 所示。

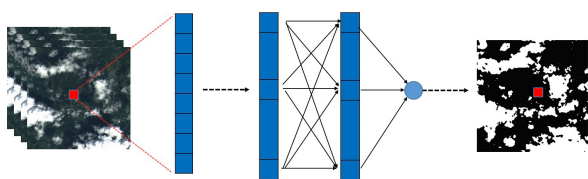


图 3: 阶段一：所有的卷积核都是 1×1 大小的，等价于多层感知机。

Fig 3: stage 1: all convolutional kernels are 1×1 , which is equivalent to multi-layer perceptron.

这一阶段生成的光谱特征，不仅会用于第二阶段的空

间信息提取阶段。在第一阶段提取的光谱特征基础上，通过进一步的区域空间信息与上下文信息的提取，充分利用空间信息，有效减少了云检测的虚警率。同时为了使模型更加轻量化，我们采用浅层神经网络与组卷积。详细的空间特征提取过程如图 4 所示。

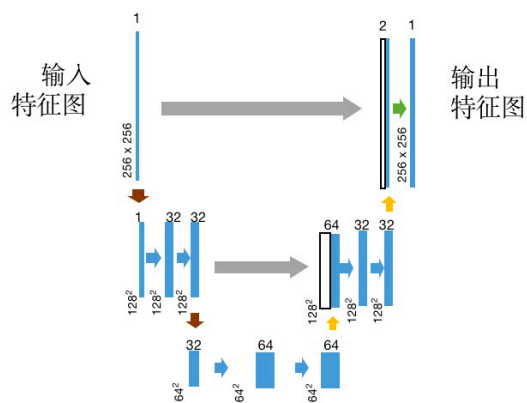


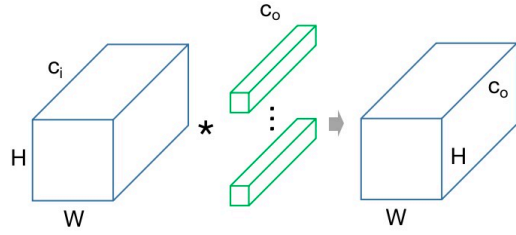
图 4: 阶段二：每组特征图的卷积过程。

Fig 4: stage 2: Convolution process of each group.

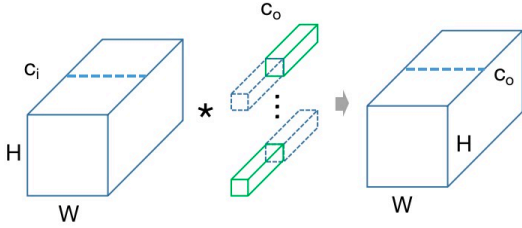
空间特征提取采用浅层（2 层）的 encoder-decoder 结构，除了最后的分类层是 1×1 的卷积核，其余卷积核为 3×3 ；利用最大池化层降低图像分辨率，扩大感受野；将高分辨率的特征图通过跳层连接

与低分辨率的特征图进行拼接，以检测不同大小的目标。

同时，本文引入组卷积的概念，避免在光谱上的重复计算，更加有效提取空间信息。图 5 展示了组卷积与普通卷积的区别。普通卷积会使所有的特征图参与计算，而组卷积将特征图与卷积核分组，每组卷积核只会与该组内的特征图进行卷积。在本文中，组卷积具有明确的意义，将第一阶段提取的每种特征作为单独的一组，针对性地提取空间信息，也意味着这一阶段不会重复提取或破坏已有的光谱特征，专注于空间特征的提取。



(a) 普通卷积， c_o 个卷积核。



(b) 组卷积， c_o 个卷积核分为 2 组。

图 5: 组卷积与普通卷积的区别。

Fig 5: The difference between group convolution and ordinary convolution

组卷积也是轻量化模型的重要手段。假设输入有 c_i 层特征图，每个特征图都是 $H \times W$ ， 3×3 的卷积核 c_o 个，普通卷积将进行 $H \times W \times 3 \times 3 \times c_i \times c_o$ 次乘法计算；若将其分为 g 组，整个卷积过程只会进行 $H \times W \times 3 \times 3 \times (c_i/g) \times c_o$ 次乘法，计算量变为原来的 $\frac{1}{g}$ 。同样，参数个数也变为原来的 $\frac{1}{g}$ ，减少了模型过拟合的风险。

为了保证稳定性和精度，S-Net 的网络设计还包括：

a) 上采样方式选择双线性插值法。

b) 在卷积之后，激活函数之前，加入批归一化层^[18] 加快模型收敛 (Batch Normalization, 简称 BN)。

$$BN(x) = \gamma \frac{x - \mu}{\sigma} + \beta \quad (1)$$

c) 激活函数以 *ReLU* 函数为主。本文中几乎所有的激活函数都是 *ReLU* 函数，即：

$$ReLU(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \leq 0 \end{cases} \quad (2)$$

ReLU 由于非负区间的梯度为常数，可以缓解梯度消失问题，使得模型的收敛速度维持在一个稳定状态。

在最后一层卷积层，激活函数会使用 *Sigmoid* 函数，即： $Sigmoid(x) = \frac{1}{1+e^{-x}}$ 用于将输出映射到 0-1 之间，代表该像素点为云的概率。

d) 损失函数采用交叉熵损失函数，用 L 表示，即：

$$L = - \sum_{i=1}^N y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \quad (3)$$

其中， N 表示所有样本个数， y_i 代表标签真值， \hat{y}_i 代表模型预测值。并且我们的损失函数包括两部分，分别计算光谱特征提取与空间特征提取阶段的损失。

若用 L_a 表示辅助损失， L_m 表示主损失，则整体损失 L_{total} 为：

$$L_{total} = w * L_a + (1 - w) * L_m, w \in [0, 1] \quad (4)$$

2 实验与分析

本文将 S-Net 的分割结果与 CFMask (FMask^[7] 的 C 语言实现)、U-Net^[17]、SegNet^[19] 的分割结果作比较。Foga 等人^[15] 使用多种传统方法在 Landsat8 biome 数据集上进行实验，结果表明 FMask 是最优秀的方法。同时 FMask 也是 Landsat 官方生成质量评估 (QA) 波段所用的方法。U-Net 是图像分割领域中经典的深度学习方法，Jeppesena 等人^[11] 使用 U-Net 在 Landsat8 biome 数据集上进行的实验取得了很好的效果。而且 S-Net 的空间特征提取部分也借鉴了 U-Net 的思想。Chai 等人^[12] 的实验表明，SegNet 的云检测结果优于多种神经网络模型。因此，将 S-Net 的分割结果与 CFMask、U-Net、SegNet 的分割结果做比较具有较高的说服力。

输入是除了全色波段的其余 10 个波段，将所有图像按 6:4 随机划分为训练集和测试集，并调整 U-Net 的输入通道数为 10。将地物真实标签（Ground Truth，简称 GT）分为云与非云两类。Chai 等人^[12]的研究表明输入 DN 值 (Digital number) 或大气顶部反射率 (ToA) 数据，会取得相似的结果，我们选择使用 DN 值作为模型输入，并为了使训练更加稳定，对输入进行归一化。具体参数，如学习率为 $1e^{-2}$ ，批训练大小为 8，采用动量为 0.9 的随机梯度下降法训练，动态调整辅助损失与主损失的比例，辅助损失的权重逐渐降低，主损失的权重逐渐增高，两者权重的变化分为 3 个阶段：(0.8, 0.2)、(0.2, 0.8)、(0, 1)。本文所有实验均在深度学习框架 Pytorch 上进行，操作系统为 Ubuntu 16.04，处理器为 NVIDIA Titan XP，内存 16 GB。

为客观评定算法的有效性和优越性，采用准确率 (Acc)、召回率 (Rec)、精确度 (Prec)、 F_1 值对结果进行评估。其中，准确率衡量像素分类正确的概率；召回率衡量属于云的像素中被分类正确的概率，是漏警率的相反数；精确度衡量被识别为云的像素中真正是云的概率，是虚警率的相反数； F_1 值是召回率与精确度的调和平均数，常被用于二分类问题，可以有效

衡量样本不均衡时检测结果的好坏。

2.1 整体评估

在 Landsat8 Biome 数据集上的实验结果如表 (3) 所示，一共包括 8 种下垫面（即：裸土、森林、草地/农田、灌木、冰雪、城市、水、湿地）。结果表明，S-Net 在几乎所有下垫面上的检测结果均优于其他算法（除了湿地稍落后于 U-Net）。S-Net 检测结果的平均 F_1 值达到了 0.9515，高于 U-Net 的 0.9451、SegNet 的 0.938 和 CFMask 的 0.87；平均准确率 (Acc) 达到了 95.04%，高于 U-Net 的 94.51%、SegNet 的 93.71% 和 CFMask 的 86.16%。虽然平均精确度 (Prec) 略低于 U-Net，但平均召回率 (Rec) 达到了 95.88%。需要强调的是，我们的模型非常轻量，参数量只有 0.34M 个，而 U-Net 有 28M 个参数，SegNet 有 30M 个参数。

S-Net 在计算时间上也具有较大优势。我们将 S-Net、U-Net、SegNet 三个模型在 $256 * 256$ 大小的图像上比较检测时间，取遍历 1000 次的平均结果，S-Net 的运行时间最少，SegNet 最耗时，三个模型的消耗时间分别为：42ms、57ms、63ms。

表 3: 在 Landsat8 Biome 数据集上，S-Net、U-Net、CFMask、SegNet 实验结果对比

表 3: Comparison of S-Net, U-Net, CFMask, SegNet experimental results on Landsat8 Biome dataset

模型	指标	裸土	森林	草地/农田	灌木	冰雪	城市	水	湿地	所有下垫面
S-Net	Acc	96.36	96.46	95.00	95.10	93.19	95.16	94.85	94.53	95.04
	Rec.	96.56	95.95	90.24	93.51	96.19	97.01	93.69	99.07	95.88
	Prec.	97.50	99.01	95.63	97.05	92.28	92.25	93.23	91.90	94.43
	F_1	97.03	97.46	92.85	95.25	94.21	94.57	93.46	95.35	95.15
U-Net	Acc	95.68	94.41	94.66	95.07	90.87	95.04	94.27	94.93	94.51
	Rec.	96.23	92.42	88.48	92.54	91.32	96.21	91.20	98.56	93.14
	Prec.	96.74	99.66	96.40	97.96	87.56	92.64	94.03	92.90	95.92
	F_1	96.48	95.90	92.27	95.17	89.40	94.40	92.59	95.65	94.51
CFMask	Acc	87.62	95.16	85.45	90.40	61.00	83.14	91.58	90.78	86.16
	Rec.	91.04	93.85	68.44	91.98	92.50	96.09	92.28	96.32	91.55
	Prec.	89.10	99.34	88.81	89.96	51.75	73.48	87.23	88.38	83.02
	F_1	90.06	96.51	77.31	90.96	66.37	83.28	89.68	92.18	87.08
SegNet	acc	94.89	94.84	91.12	93.16	93.27	94.18	93.78	93.73	93.71
	rec.	96.39	93.87	80.55	90.05	92.94	97.67	91.32	98.54	93.69
	prec.	95.35	98.77	93.92	96.72	91.08	89.82	92.74	91.11	93.90
	F_1	95.86	96.26	86.73	93.27	92.00	93.58	92.02	94.68	93.80

S-Net 对于碎云、细节也有良好的检测与保持能力。U-Net与SegNet的识别更加光滑，使得一些细节被忽略，而我们的模型更加注重细节，这对于云检测是一个很重要的能力。如图 6所示，从左到右依次为真彩色图、人工标注、S-Net 提取的光谱特征构成的假彩色图、我们的模型预测结果、U-Net 结果、SegNet 结果、CFMask 结果，白色代表云，黑色代表非

云，偶数行的图像是奇数行图像中黄色方框部分的放大结果，黄色方框的大小为 20×20 。图 6 共展示了 4 张子图，下垫面各不相同，分别为城市、农田、冰雪、裸土。在多种下垫面上，S-Net 对于碎云均有良好的检测能力，同时对于冰雪与裸土这样的高亮地物，也有较低的虚警率。

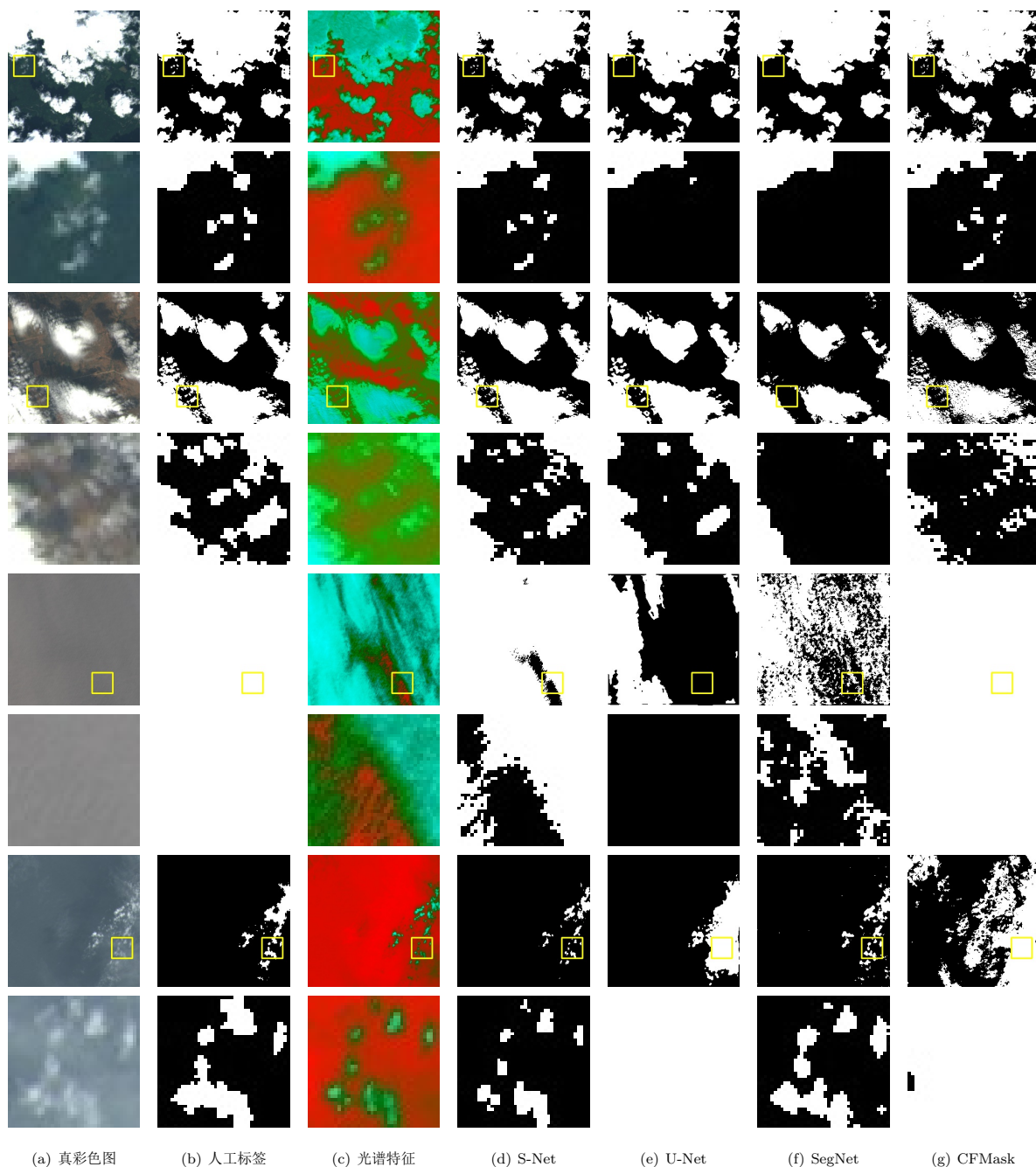


图 6: Spoon-Net 检测结果展示，偶数行是奇数行黄色方框部分的放大图。

Fig 6: Example of spoon-net results, even lines are enlarged views of the yellow boxes of odd lines

2.2 碎云检测评估

为充分体现 S-Net 的优势, 本文进一步比较模型对碎云的检测能力, 并定义碎云为: 面积 (四连通) 小于 30 个像素的云。由于此时计算准确率 (Acc) 已无

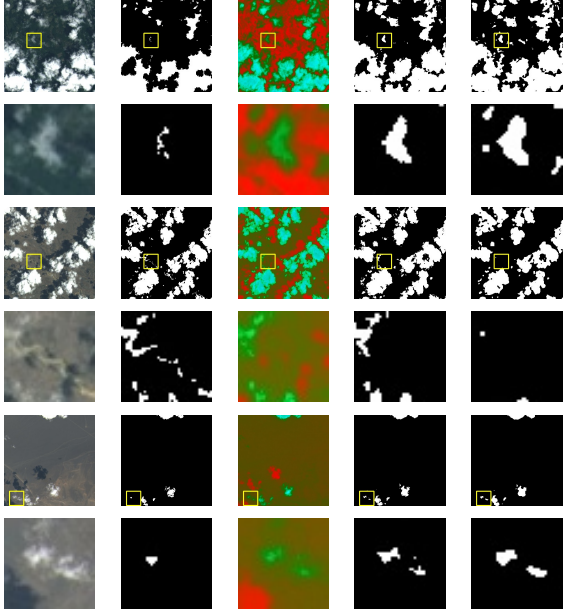
意义, 所以我们只选择召回率 (Rec)、精确度 (Prec) 和 F_1 值作为衡量准则。实验表明, 在识别碎云时, S-Net 拥有 0.1802 的平均 F_1 值, 显著高于 U-Net 的 0.1547、SegNet 的 0.0717 和 CFMask 的 0.091, 并且在除了城市和农田外的所有下垫面上均优于 U-Net。

表 4: 在 Landsat8 Biome 数据集上, S-Net、U-Net、CFMask、SegNet 对碎云的检测结果对比

表 4: On the Landsat8 Biome data set, S-Net, U-Net, CFMask, SegNet compare the detection results of broken clouds

模型	指标	裸土	森林	草地/农田	灌木	冰雪	城市	水	湿地	所有下垫面
S-Net	Rec.	10.15	18.30	19.88	18.63	13.54	24.28	18.76	21.23	19.19
	Prec.	11.44	18.56	16.67	18.78	7.23	13.74	17.12	28.03	16.99
	F_1	10.76	18.43	18.13	18.70	9.42	17.54	17.90	24.16	18.02
U-Net	Rec.	6.75	10.11	15.26	15.79	2.88	19.50	11.93	16.23	13.97
	Prec.	16.56	14.53	27.26	20.77	4.79	17.32	12.34	27.26	17.34
	F_1	9.59	11.92	19.57	17.94	3.60	18.34	12.15	20.34	15.47
CFMask	Rec.	5.14	48.82	15.32	18.20	3.45	16.98	20.18	13.54	17.59
	Prec.	1.66	32.98	7.50	5.96	0.25	5.95	3.56	5.69	6.14
	F_1	2.51	39.37	10.07	8.98	0.47	8.81	6.05	8.01	9.10
SegNet	rec.	4.60	3.93	9.10	8.36	4.27	10.29	9.41	10.97	8.28
	prec.	1.67	5.11	12.06	8.52	0.58	9.19	7.20	11.38	6.31
	F_1	2.45	4.44	10.38	8.44	1.02	9.71	8.16	11.17	7.17

虽然在检测碎云时, S-Net 的表现显著优于其他算法, 但 F_1 也仍不到 0.2。原因可能有两点: 1. 检测碎云具有较高的难度; 2. 人工在标注碎云时存在困难, 数据存在错误。Scaramuzza 等人^[20]的研究表明人工标注的数据可能存在 7% 左右的错误。尤其对于碎云, 人工标注更加不准确, 且碎云数量较小, 对错误标签更加敏感。我们发现, 'L8 Biome' 数据确实存在一些问题, 图 7 展示了一些存在问题的图像样本。同时, 由于人工标注的不稳定性, 简单地依靠评价指标可能并不能真实反应模型的优劣, 因为大部分模型都很容易对于大面积的厚云 (低下垫面信息的) 有较好的识别能力; 并且以这些标签为真值进行的训练, 可能也会存在问题。



(a) 真彩色 (b) 人工标 (c) 光谱特 (d) S-Net (e) U-Net
图 签 征

图 7: 问题标签

Fig 7: Error label

3 结论

云检测一直是遥感领域的研究热点与难点。本文提出了一种两阶段的遥感图像云检测模型。相比于现有的深度学习模型,我们的模型更加轻量,并且具有更好的保持边缘细节的能力和对小的碎云的检测能力。我们注重遥感图像光谱特征的提取,在第一阶段利用 1×1 的卷积核专门提取光谱特征,使得检测结果保持了“纯粹性”,没有受到其空间信息的干扰,并使其直达最终的分类层。再利用浅层 encoder-decoder 结构,引入组卷积,对每个光谱特征分别计算空间信息。该模型充分利用遥感图像多波段的特点,有效解决了现有方法在保持边缘细节与扩大感受野之间的矛盾,并在 Landsat8 Biome 数据集上达到了 95% 的准确率,基本还原了输入影像的细节信息。

值得强调的是, Landsat8 遥感卫星可以提供丰富的光谱信息,这是 S-Net 能表现优异的前提。当输入的遥感图像具有较少的波段时,光谱信息在云检测任务中的权重就会降低, S-Net 的性能会受到影响。

虽然 S-Net 表现优异,但仍有一些方向值得研究。1.S-Net 采用了较为简单的空间特征提取结构,后续将探索更加适合云检测的空间特征提取算法。2. 传统方法对云检测有大量的先验知识,如何将这些先验知识与深度学习方法结合也是后续的一个研究方向。

参考文献

- [1] Anup K Prasad, Lim Chai, Ramesh P Singh, and Menas Kafatos. Crop yield estimation model for iowa using remote sensing and surface parameters. *International Journal of Applied Earth Observation and Geoinformation*, 8(1):26–33, 2006.
- [2] Jan Verbesselt, Rob Hyndman, Glenn Newnham, and Darius Culvenor. Detecting trend and seasonal changes in satellite image time series. *Remote sensing of Environment*, 114(1):106–115, 2010.
- [3] Karen E Joyce, Stella E Belliss, Sergey V Samsonov, Stephen J McNeill, and Phil J Glassey. A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters. *Progress in Physical Geography*, 33(2):183–207, 2009.
- [4] Junchang Ju and David P Roy. The availability of cloud-free landsat etm+ data over the conterminous united states and globally. *Remote Sensing of Environment*, 112(3):1196–1211, 2008.
- [5] Lin Sun, Xinyan Liu, Yikun Yang, TingTing Chen, Quan Wang, and Xueying Zhou. A cloud shadow detection method combined with cloud height iteration and spectral analysis for landsat 8 oli data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 138:193–207, 2018.
- [6] Richard R Irish, John L Barker, Samuel N Goward, and Terry Arvidson. Characterization of the landsat-7 etm+ automated cloud-cover assessment (acca) algorithm. *Photogrammetric engineering & remote sensing*, 72(10):1179–1188, 2006.
- [7] Zhe Zhu and Curtis E Woodcock. Object-based cloud and cloud shadow detection in landsat imagery. *Remote sensing of environment*, 118:83–94, 2012.
- [8] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.
- [9] Yoonkyung Lee, Grace Wahba, and Steven A Ackerman. Cloud classification of satellite radiance data by multicategory support vector machines. *Journal of Atmospheric and Oceanic Technology*, 21(2):159–169, 2004.
- [10] Bin Tian, Mukhtiar A Shaikh, Mahmood R Azimi-Sadjadi, Thomas H Vonder Haar, and Donald L Reinke. A study of cloud classification with neural networks using spectral and textural features. *IEEE transactions on neural networks*, 10(1):138–151, 1999.

- [11] Jacob Høxbroe Jeppesen, Rune Hylsberg Jacobsen, Fadil Inceoglu, and Thomas Skjødeberg Toftegaard. A cloud detection algorithm for satellite imagery based on deep learning. *Remote Sensing of Environment*, 229:247–259, 2019.
- [12] Dengfeng Chai, Shawn Newsam, Hankui K Zhang, Yifan Qiu, and Jingfeng Huang. Cloud and cloud shadow detection in landsat imagery based on deep convolutional neural networks. *Remote sensing of environment*, 225:307–316, 2019.
- [13] M Joseph Hughes and Robert Kennedy. High-quality cloud masking of landsat 8 imagery using convolutional neural networks. *Remote Sensing*, 11(21):2591, 2019.
- [14] Xiaodong Zhang, Tong Wang, Guanzhou Chen, Xiaoliang Tan, and Kun Zhu. Convective clouds extraction from himawari-8 satellite images based on double-stream fully convolutional networks. *IEEE Geoscience and Remote Sensing Letters*, 11(21):2591, 2019.
- [15] Steve Foga, Pat L Scaramuzza, Song Guo, Zhe Zhu, Ronald D Dilley Jr, Tim Beckmann, Gail L Schmidt, John L Dwyer, M Joseph Hughes, and Brady Laue. Cloud detection algorithm comparison and validation for operational landsat data products. *Remote sensing of environment*, 194:379–390, 2017.
- [16] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.
- [18] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- [19] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation, 2015.
- [20] Pasquale L Scaramuzza, Michelle A Bouchard, and John L Dwyer. Development of the landsat data continuity mission cloud-cover assessment algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 50(4):1140–1154, 2011.