

# 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

(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** 可靠的云检测是应用遥感图像时重要的预处理步骤, 一直是遥感领域中的研究热点。本文针对目前神经网络模型存在的细节信息易损失、参数量大、计算复杂等不足, 提出了一种新型的、轻量的网络, 称为勺型网络 (Spoon-Net)。S-Net 分两个阶段分别提取光谱与空间信息, 第一阶段, 使用  $1 \times 1$  的卷积核对遥感图像进行光谱特征提取; 第二阶段, 使用  $3 \times 3$  的卷积核对遥感图像进行空间特征提取。而且, 为了更加有效地提取空间特征, 避免重复计算光谱特征, 本文引入分组卷积, 对第一阶段提取的每一层光谱通道单独进行卷积。模型在 landsat 8 biome 数据做训练并估计, 在模型参数量只有 unet 的八十分之一 (0.34M) 的情况下, 取得了全面优于 unet 的实验结果, 总准确率为 95.34%, UNet 为 94.15%; f1 值为 0.9505, UNet 只有 0.9352。

**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

(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** Reliable cloud detection is an important preprocessing step in the application of remote sensing images, and has been a research hotspot in the field of remote sensing. In this paper, a new and lightweight neural network, called spoon net, is proposed to solve the problems of large parameters, complex calculation and easy loss of details in the current neural network model. S-Net is divided into two stages to extract spectral and spatial information. In the first stage,  $1 \times 1$  convolution check is used to extract spectral features of remote sensing images; in the second stage,  $3 \times 3$  convolution check is used to extract spatial features of remote sensing images. In addition, in order to extract spatial features more effectively and avoid the repeated calculation of spectral features, this paper introduces group convolution to separate each layer of spectral channels extracted in the first stage. The model is trained and estimated with Landsat 8 biome data. When the model parameters are only one eightieth (0.35m) of UNET, the experimental results are better than UNET in all respects. The total accuracy is 95.34 %, UNET is 94.15 %; F1 is 0.9505, UNET is only 0.9352.

**Keywords** landsat, cloud detection, neural network

---

\*通讯作者: E-mail:wangshuli18@mails.ucas.ac.cn

## 0 引言

卫星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<sup>[1]</sup>、变化检测<sup>[2]</sup>、灾难评估<sup>[3]</sup>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多,且越来越容易获得。海量的多波段遥感数据也急切需要高效率和高鲁棒性的算法进行处理和数据挖掘。然而,在 landsat 数据集上,每年有高达 40% 的像素被云覆盖<sup>[4]</sup>,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,云在光谱上整体呈高亮特性,在可见光波段呈亮白色,在短波红外与亮温通道反射率会有所减小。基于光谱信息的方法<sup>[5]</sup>会人工构造一些特征(如:NDVI, NDWI),并精心设置阈值以更好地地区分地物。Irish 等人<sup>[6]</sup>提出的 ACCA 使用 Landsat7 ETM + 谱段 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<sup>[7]</sup>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<sup>[8]</sup>,再对子图或超像素进行分类,如 SVM<sup>[9]</sup>、MLP<sup>[10]</sup>,将其分为有云、无云两类或多云、少云、无云三类,但这降低了图像的分辨率。

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。提取合适的特征,选择合适的阈值是云检测任务的关

键,很多专家针对不同的云层、不同的下垫面精心设计了其特征与阈值。深度学习作为机器学习的一个分支,它可以使我们从这些繁琐的工作中解脱出来,帮助我们自动的构造特征、选择阈值。而且,精心设计的神经网络可以构造非常高维度的特征,包括光谱特征与空间特征,不像决策树只能组合一些低维度的简单特征,这将帮助我们更好的检测云。

云检测的目的是在遥感图像中逐像素地确定每一个像素点是否为云,是一个像素级的分类任务,属于图像分割,即输入是一副图像,输出是一个同等大小的二值化图像。Long 提出 FCN<sup>[11]</sup>是 CNN 图像分割的开山之作,通过将普通 CNN 分类网络后的全连接层变为卷积层,实现了像素级别的分类。从此,不带全连接层的全卷积神经网络开始在语义分割任务上大放异彩。随后提出的 U-Net<sup>[12]</sup>是一种结构对称的网络,丰富了 decoder 部分。虽然在跳层连接这一部分,FCN 用的是加操作,U-Net 用的是叠操作,但 encoder-decoder 的框架是一致的。这两个算法的提出,奠定了 encoder-decoder 结构在语义分割领域的主流位置,其中,encoder 的作用是提取空间特征,decoder 的作用是解析空间特征,并将图像还原到原来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。以后的网络,大部分都在这个框架下。

目前也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<sup>[13]</sup>将 UNet 应用于 landsat 遥感图像云检测。Chai 等人<sup>[14]</sup>将 SegNet 应用于 landsat 图像。hughes 等人<sup>[15]</sup>将 FCN 应用于云检测。这些方法都是将光谱与空间特征进行混合提取,很少有人会针对遥感图像多波段的特点对神经网络的结构进行针对性的设计。

卷积神经网络确实可以有效提取图像空间信息,所以应用神经网络做遥感图像的云检测可以表现出较好的性能。但这些方法存在两个缺点:1. 需要大量的参数与昂贵的计算,如 UNet 有 28M 个参数。2. 由于混合提取光谱特征与空间特征(我们称之为 one-stage),这些模型在保护细节与扩大感受野之间存在着难以调节的矛盾。他们从图像的输入就开始混合像元进行计算,这使得输出的掩膜很难再对单个像素提纯以至于结果非常平滑。

本文提出了一个新颖的、简单的、有效的网络,主要包括两个阶段,模型简略图 1 所示。第一部分,光谱特征提取部分。特征是分类成功的关键,而光谱

特征是遥感图像中最本质的特征。在这一部分，我们完全使用  $1 \times 1$  的卷积核，不考虑空间信息，对图像进行光谱特征提取。好的光谱特征可以减轻模型的学习压力，使得后续的空间特征更加容易被提取。并且，S-Net 将光谱特征提取的结果直达最后的分类层，使得云检测结果更加细腻。第二部分，空间特征提取部分。这一部分，采用目前主流的空间特征提取框架——encoder-decoder 框架。不同的是，在这一步，我们不会重复提取光谱特征，而是采用组卷积 (Group Conv) 的方式，对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。由于第一部分的存在，GC 使空间特征提取部分变得廉价且有效。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。

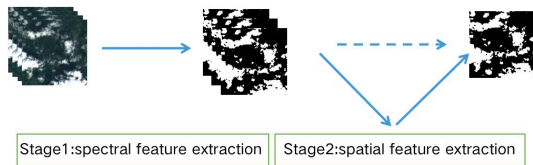


图 1: Spoon-net 框架.

Fig 1: Spoon-net framework

## 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 NASA landsat8 卫星。2013 年 2 月 11 日，美国航空航天局 (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  $0.43\mu\text{m}$  到  $2.30\mu\text{m}$ ；TIRS 提供地表温度数据，包括两个波段，波段范围从  $10.60\mu\text{m}$  到  $12.51\mu\text{m}$ ，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	波段	波长范围 ( $\mu\text{m}$ )	空间分辨率
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，本文利用已有的全球云和云影验证数据集“L8 Biome Cloud Validation Masks”<sup>[16]</sup>，该数据集共有 96 景图片，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 4 所示。

表 2: L8 Biome 数据人工标注标志位

Tab 2: flag of landsat8 biome

value	0	64	128	192	255
Interpretation	Fill	Cloud Shadow	Clear	Thin Cloud	Cloud

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。本文使用 midcloud 的所有数据，共 32 景做实验，每种地物有 4 景。数据我们将标签简单地分为云与非云两类，将每景 L8 图像均匀切割为  $256 \times 256$  大小的小图，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。波段选择了除了全色波段的所有波段，共 10 个波段，训练集与测试集的比例为 6:4，训练集有 10247 张子图，测试集有 6932 张子图。

## 1.2 网络结构

S-Net 是一个两阶段模型，将光谱特征提取与空间特征提取解藕，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。

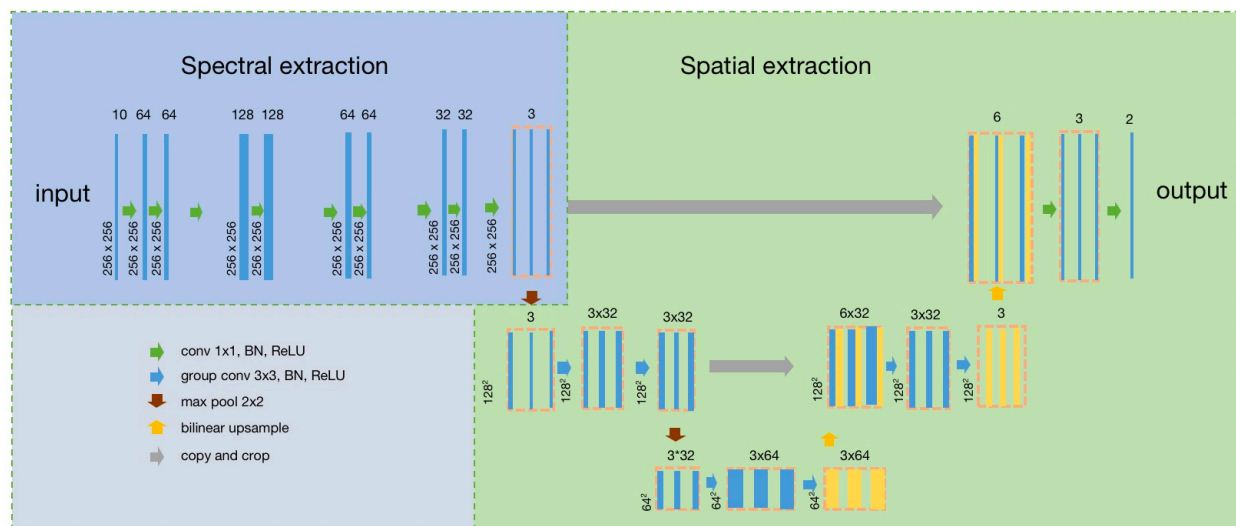


图 2: Spoon-net 详细结构.

Fig 2: Spoon-net framework

1. 光谱特征提取阶段。该阶段使用  $1 \times 1$  的卷积核对单个像素进行计算，不涉及到空间信息，可以有效保持图像的细节。多层  $1 \times 1$  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。

这一阶段生成的光谱特征，不仅会用于第二阶段的空间特征提取，还会以跳层连接的方式直接参与最终的分类，并且分类层的卷积核也是  $1 \times 1$ ，这意味着我们的网络有足够的保证分割结果的细节。实验发现，当光谱特征数为 3 时，具有最高的分类精度与 f1 值。

表 3: 光谱特征提取层数对实验结果的影响

Tab 3: Influence of the number of spectral feature extraction layers on experimental results

光谱层数	acc	f1	参数数量
2	94.38	94.11	238K
3	<b>95.34</b>	<b>95.05</b>	342K
4	94.79	94.46	450K
5	94.92	94.60	562K
6	94.84	94.61	678K

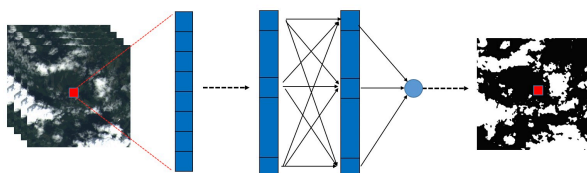


图 3: 阶段一：所有的卷积核都是  $1 \times 1$  大小的，等价于多层感知机。

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

2. 空间信息提取阶段。光谱特征虽然可以保持分辨率、解析细节，但它们不包含特定区域的信息与

上下文信息，容易造成较高的虚警率，因此我们仍需要提取空间信息去优化。

encoder-decoder 结构是图像分割领域常用的一种空间信息提取结构，它通过多层卷积与下采样层提取大感受野内的空间信息，通过多层卷积与上采样层将特征图还原到原始分辨率。丰富的空间信息需要更多的下采样层和更大的感受野，但随之而来的问题是越来越容易忽略细节，因此 encoder-decoder 结构加入了跳层连接，将高分辨率的特征图与低分辨率的空间信息结合，以缓和这一矛盾。图 4展示了每一组光谱特征的卷积过程。

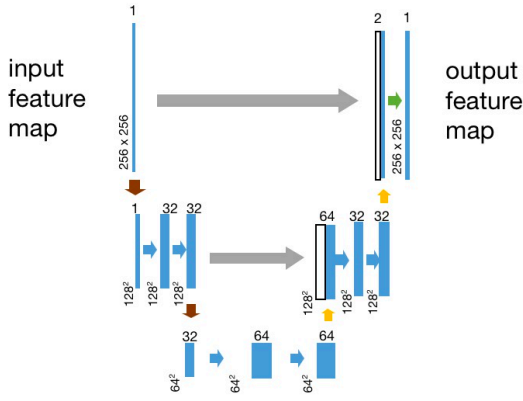


图 4: 阶段二：每组特征图的卷积过程。

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

为了更加有效提取空间信息，避免在光谱上的重复计算，我们引入组卷积的概念。图 5展示了组卷积与普通卷积的区别。通过光谱特征提取阶段我们已经得到非常有效的光谱特征，我们在这一部分不会再重复地计算光谱特征，我们对每一种光谱特征进行单独的、相同的编码解码，并将每一个光谱特征的卷积结果作为一组，进行下采样、上采样等操作。组卷积的优势在于可以专注于提取空间信息，免受光谱信息的交互与干扰，并且节省了大量参数与电脑计算力（组卷积使网络的参数与计算量减小为原来的三分之一），减少了模型过拟合的风险。

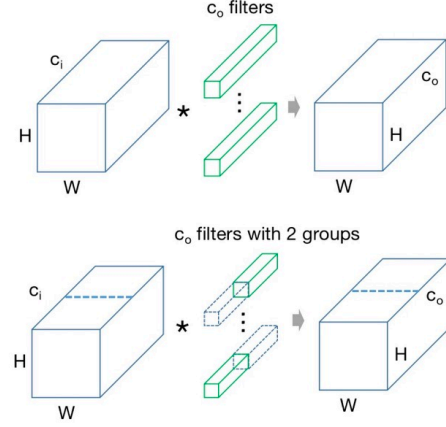


图 5: 组卷积与普通卷积的区别。

Fig 5: The difference between group convolution and ordinary convolution

同时，我们也就减少 encoder-decoder 层数。由于光谱特征提取部分的存在，在这一部分模型提取空间信息的压力已经被大大减小，所以我们并不需要很深的网络、很多的参数去拟合很复杂的空间特征，因此我们可以大胆减少 encoder-decoder 层数，而随着层数的减少，参数量几乎是几何减少的。实验发现，当 encoder-decoder 层数大于 2 时，云检测效果不会再有提高。

表 4: 光谱特征提取层数对实验结果的影响

Tab 4: Influence of the number of spectral feature extraction layers on experimental results

下采样层数	acc	f1	参数数量
1	92.33	92.12	72K
2	<b>95.34</b>	<b>95.05</b>	341K
3	95.01	94.89	1.4M
4	94.87	94.62	5.7M

上采样的方式一般有四种：插值法，反卷积，反池化，超分辨率重建领域的亚像素卷积插值。双线性插值是目前在语义分割中用的比较多的一种插值方式，比如 FCN 中就是用的这种方法。在 CNN 上下文中，反卷积是卷积的逆过程，卷积用于提取空间信息，反卷积用于解析空间信息。在实现上，反卷积是卷积的转置，所以反卷积也叫做转置卷积。反池化是池化的逆过程，在池化过程中，记录下 max-pooling 在对应 kernel 中的坐标，在反池化过程中，将一个元

素根据 kernel 进行放大, 根据之前的坐标将元素填写进去, 其他位置补 0。在下采样的时候记录 max 的位置, 上采样的时候最大值的位置还原, 其它位置填 0。反池化是速度最快的上采样操作, 计算量和参数也特别少, 但是准确率一般。虽然理论上, 由于反卷积具有更多的参数, 所以反卷积可以更好的学习特征, 但是有研究表明, 如果参数配置不当, 反卷积很容易出现输出 feature map 带有明显棋盘状的现象<sup>[17]</sup>, 双线性差值可以取得与反卷积相同甚至更好的效果。因此, 我们选择参数少且容易取得较好效果的双线性差值法。

在卷积之后, 激活函数之前, 一般会有一个批归一化操作<sup>[18]</sup>(Batch Normalization)。BN 是一种非常优雅的重参数化的方法, 它的存在类似于为网络中不同的层设置了不同的学习率。为神经网络输入的多个数据称为 batch, BN 以 batch 为单位, 先将上一层网络的输出在通道上标准化为标准正态分布, 再进行缩放与平移操作, 可以将上一层的输出调整在一个较好的范围内, 结果就是模型训练更加容易。

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

本文中几乎所有的激活函数都是 ReLU 函数,  $\sigma(x) = \max(0, x)$ 。ReLU 函数是一种分段线性函数, 把所有的负值都变为 0, 而正值不变, 这种操作被成为单侧抑制。单侧抑制使得神经网络中的神经元也具有了稀疏激活性。我们认为对于某种地物可以对某个特殊的指标有响应, 而对其他指标就反应一般。所以 ReLU 实现稀疏后的模型能够更好地挖掘相关特征, 且 ReLU 由于非负区间的梯度为常数, 因此不存在梯度消失问题, 使得模型的收敛速度维持在一个稳定状态。

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

在最后一层卷积层, 激活函数会使用 sigmoid 函数, 用于将输出映射到 0-1 之间, 代表该像素点为云的概率。为了训练模型, 使用交叉熵损失函数计算损失, 并用带动量的随即梯度下降法进行训练。

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

其中,  $N$  表示所有所有样本个数,  $y_i$  代表标签真值,  $\hat{y}_i$  代表模型预测值。

## 2 实验与结果

为客观评定算法的有效性和优越性, 采用准确率, 召回率, 精确度,  $F_1$  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。四个评价指标分别为:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

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

$$P = \frac{TP}{TP + FP} \quad (6)$$

$$F = \frac{2 * R_{recall} * R_{precision}}{R_{recall} + R_{precision}} \quad (7)$$

其中, TP 为真正类 (True Positive), 即云被判为云; TN 为真负类 (True Negative), 即非云被判为非云; FP 为假正类 (False Positive), 即非云被判为云; FN 为假负类 (false Negative), 云被判为非云。

本文将模型与 CFMask、原始的 UNet 做比较。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 UNet 的输入通道数为 10。将 ground truth 分为云与非云两类。有研究表明<sup>[14]</sup> 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 的随即梯度下降法训练。我们的损失函数包括两部分, 分别计算光谱与空间特征提取两阶段的损失, 如图 6 所示。为了使模型更加容易收敛, auxiliary loss 的权重逐渐降低, final loss 的权重逐渐增高, 分为 3 个阶段: (0.8, 0.2)、(0.2, 0.8)、(0, 1)。



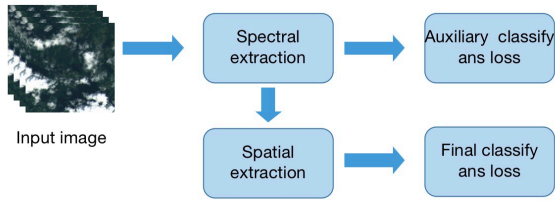


图 6: 两个损失函数。

Fig 6: Two loss function

表 5: 实验结果对比

表 5: Comparison of experimental results

model	evaluation	Barren	Forest	Grass/Crops	Shrubland	Snow/Ice	Urban	Water	Wetlands	total
SpoonNet-3	acc	96.43	96.61	94.48	95.08	96.29	94.85	94.38	94.56	<b>95.34</b>
	rec.	96.46	95.90	92.67	93.21	95.38	97.07	91.90	98.60	<b>95.15</b>
	prec.	97.89	99.38	93.10	97.41	95.61	91.57	92.82	92.34	95.01
	F1	97.17	97.61	92.89	95.26	95.50	94.24	92.36	95.37	<b>95.05</b>
U-Net	acc	96.09	94.51	92.67	94.48	91.78	95.00	93.74	94.90	94.15
	rec.	96.00	92.96	81.91	91.58	89.16	96.22	89.83	98.34	92.00
	prec.	97.60	99.22	97.31	97.77	90.90	92.56	93.97	93.05	<b>95.30</b>
	F1	96.80	95.99	88.95	94.58	90.02	94.36	91.85	95.62	93.52
CFMask	acc	87.46	95.20	85.42	90.34	60.97	83.05	91.51	90.78	86.09
	rec.	90.85	93.87	68.03	92.04	92.21	96.07	92.22	96.37	91.45
	prec.	88.99	99.31	88.85	89.83	51.74	73.25	86.97	88.38	82.89
	F1	89.91	96.51	77.05	90.92	66.29	83.13	89.52	92.21	86.96

相比于一阶段的 UNet，我们的模型具有更高的准确率与 f1 值，准确率达到 95.34%，而 UNet 只有 94.15%，比 UNet 高了一个百分点；f1 值有 0.9505，而 UNet 只有 0.9352，比 UNet 高了 1.5 个百分点。同时可以发现，我们的模型对于碎云、细节有良好的检测与保持能力。unet 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 7 所示，从左到右依次为真彩色图、人工标注、光谱特征提取结果构成的假彩色图、我们的模型预测结果、UNet 结果，白色代表云，黑色代表非云，第二行是第一行黄色方框的放大图，以此类推。我们的模型由于有光谱特征提取层的存在，并且光谱特征直达最后的分类层，所以对细节有较好的保持，对碎云有良好的识别，能以极少的参数在整体上达到优于 U-Net 的效果。

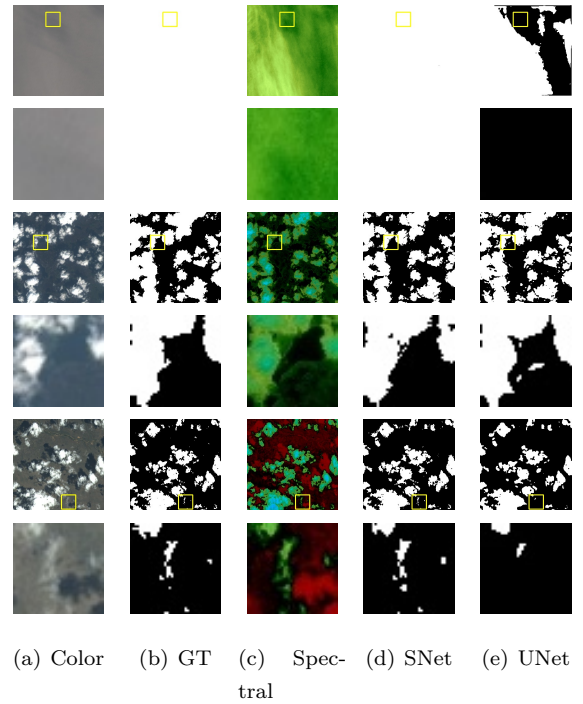


图 7: Spoon-Net 检测结果展示

Fig 7: Example of spoon-net results

我们的模型参数量有 0.34M，而主流网络如 U-Net 有 28M，FCN 有 20.5M。其他网络参数多的主要原因是具有更深的 encoder-decoder 结构，以 U-Net 为例，U-Net 最深一层就有 14M 的参数，占据了所有参数的一半。我们的模型由于更加注重单个像素的光谱特征，在提取空间信息之前先使用少量的参数 (1\*1 卷积核) 提取光谱信息，使得我们可以大胆的减少 encoder-decoder 层数，同时引入组卷积的操作，极大地简化模型，大大降低参数量。

而且，有研究表明<sup>[19]</sup> 人工标注的数据可能存在 7% 左右的错误。尤其对于一些碎云、半透明的薄云，评判标准可能会因人而异。我们发现，‘L8 Biome’数据确实存在一些问题。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。

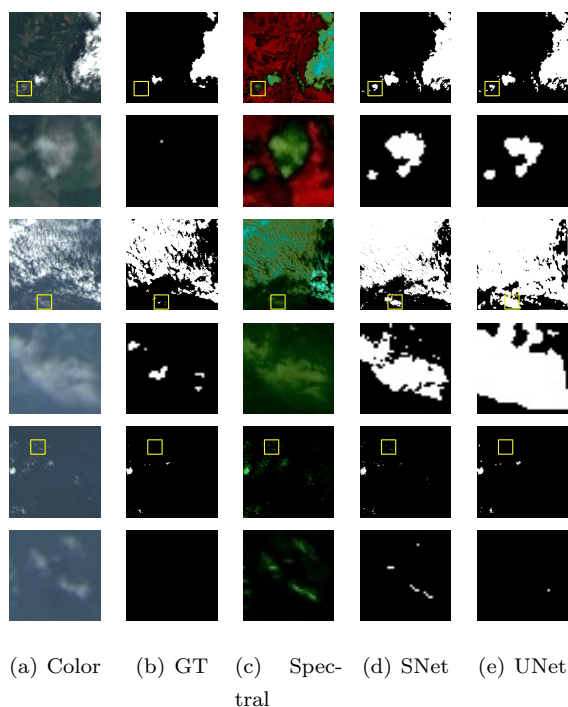


图 8: 错误标签

Fig 8: Error label

### 3 讨论

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，这在云检测中非常重要。我们首先引入光谱特征提取部分，利用 1\*1 的卷积核，使得检测结果保持了“纯粹性”，没有收到其空间信息的干扰，并使其直达最终的分类层。再通过 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 数据集上达到了 95.34% 的准确率，基本还原了输入影像的细节信息。后续将继续优化光谱特征与空间特征提取部分，并尝试深度学习与传统方法结合，以实现更加精确的遥感图像云检测。

### 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] 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.
- [12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.
- [13] Jacob Høxbroe Jeppesen, Rune Hylsberg Jacobsen, Fadil Inceoglu, and Thomas Skjodeberg Toftegaard. A cloud detection algorithm for satellite imagery based on deep learning. *Remote Sensing of Environment*, 229:247–259, 2019.
- [14] 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.
- [15] M Joseph Hughes and Robert Kennedy. High-quality cloud masking of landsat 8 imagery using convolutional neural networks. *Remote Sensing*, 11(21):2591, 2019.
- [16] 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.
- [17] Augustus Odena, Vincent Dumoulin, and Chris Olah. Deconvolution and checkerboard artifacts. *Distill*, 1(10):e3, 2016.
- [18] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- [19] 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.