# S-Net: Spoon Networks for Cloud Masking of Landsat8 Imagery

## Abstract

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。目前，很多方法基于单个像素的光谱，利用先验知识，通过设置合适的阈值，对像素进行分类。但这些方法往往没有考虑地物波段之间的关系，且无法利用图像的空间信息，对一些地物往往会造成误分，如冰雪。随着深度学习的发展，出现了一些利用神经网络做云识别的方法。但神经网络原本是处理RGB图像的，这些方法大多直接将现有的网络结构直接应用于遥感图像，没有考虑遥感图像多波段的特点。本文中，我们提出了一种新型的网络结构，叫做勺形网络，专门用于多波段的遥感图像云检测。我们用landsat8数据做实验，取得了优于Fmask的实验结果。

With the development of neural network, full convolutional neural network has been applied to remote sensing image cloud detection. However, most of the algorithms do not consider the characteristics of remote sensing image, and there is little innovation and change to the full convolution neural network. In this paper, considering the multi band characteristics of remote sensing image and the influence of complex underlying surface of cloud detection, a simple and effective network, called S-Net, is proposed.

## 1. Introduction

卫星遥感数据在当今社会的生产和生活中扮演者至关重要的角色，农业、气象、交通运输等领域的发展都离不开遥感数据的支持。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。同时，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。

云检测的任务是在遥感图像中逐像素地确定每一个像素点是否为云，若是简单地将像素分为云和非云两类，那么云检测就是一个输入一副图像，输出一个同等大小的二值化图像的过程。

传统的云检测方法往往基于阈值。算法所采用阈值的可靠性往往依赖于传感器精度和专家对所采集数据物理含义的理解。有静态阈值的，也有动态阈值的。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。随着相关技术的发展，基于云纹理和空间特征的检测方法以及模式识别等技术在云检测方向上得到了广泛的应用。这类方法在具备先验知识的条件下可以获得较好的分类效果，但人为干预极大地影响检测效率。

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。一些文章将高分辨率图像切割成一张张小图或超像素，对图片或超像素进行分类，将其分为有云、无云两类或多云、少云、无云三类。云检测其实是语义分割任务，即输入一副图像，输出也是一副图像，从Long提出FCN开始，不带全连接层的全卷积神经网络在语义分割任务上取得了巨大的成功。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，但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。需要注意的是，这些神经网络模型是针对RGB图像设计的，这些图像中的每个像素只有三个特征，对只具有三个特征的RGB图像做语义分割必然提取极其复杂的空间特征，而遥感图像往往具有多个甚至几十个波段，且光谱特征是地物最本质的特征，所以有必要针对遥感图像的这种特性，设计一种有针对性的网络。

在这篇文章中，我们提出了一个新颖的、简单的网络，叫做勺形卷积网络，主要包括三部分。第一部分，光谱特征提取部分，为了充分利用遥感图像多波段的特性，利用多层的1\*1卷积核，对原始遥感图像进行波段特征提取，输出多幅特征图。第二部分，考虑到复杂的下垫面是云检测中的难点，为了适应不同的下垫面，同时为了精简网络结构，选择方差最大的三幅特征图，输入到下一部分。这里只选择三幅是因为人眼可以根据假彩色图很好地识别云，说明三幅特征图足够得到很好的效果。第三部分，利用全卷积网络对图像进行空间特征提取。这一步我们对U-Net进行了简化，减少了深度与下采样次数。

实验结果表明，该方法可以在模型参数大大减小的情况下，明显提高云检测精度。实验所用的所有代码均在github可见。

Since 2011, Landsat8/OLI has been collecting high quality imagery of the earth at least every 16 days. Free and open to the public, with global wall-to-wall coverage of land surfaces at an ecologically meaningful spatial resolution, Landsat imagery is one of the most useful resources for ecological monitoring and wildland management. But harnessing the power of the Landsat archive to detect and describe change on the earth’s surface hinges on researchers’ ability to detect and aggregate clear-sky observations uncontaminated by clouds.

Because of this necessity, screening for clouds has always been essential for remote imagery. For a single scene, information of spectrum and space are utilized to detect cloud. A common routine is to design spectral features first, and then use spatial features for post-processing.

In the last decade, deep Convolutional Neural Networks (CNN) have revolutionized image recognition. For image segmentation, which is the process of assigning a label to every pixel in computer vision, Fully Convolutional Network(FCN) is a pioneering work, which all layers in the network are convolutional. FCN has powerful ability of spatial feature extraction. There are many methods use FCN to remote sensing image cloud detection.

However, FCN was originally designed for RGB images. Applying FCN to remote sensing image without any change will cause huge waste of computing resources. And as mentioned before, spectral features are the most essential features of ground objects, and it seems to put the cart before the horse to over emphasize spatial features.

To solve these problems, we propose a spoon convolution network, which is divided into three parts. Firstly, 1 x 1 convolution kernel is used to extract the spectral features of the image; secondly, in order to adapt to the complex underlying surface, variance is used to select the feature map generated in the first step; thirdly, a simple network similar to U-Net is used to extract and predict the spatial features of the ground objects.

## 2. Materials and Methods

### 2.1 Training and Evaluation Data

我们采用的光学遥感卫星数据集来自美国航天局的landsat卫星，该卫星分辨率可达15米，重访周期为16天。它包括两个设备，the Operational Land Imager(OLI) and the Thermal Infrared Sensor(TIRS)

The test data of this paper is from USGS official website(<https://landsat.usgs.gov/landsat-8-cloud-cover-assessment-validation-data>).This is a manually labeled data set, including 8 kinds of underlying surfaces(includding Barren, Forest, Grass/Crops, Shrubland, Snow/Ice, Urban, Water, Wetlands), each of which has 12 scenes, and each scene file has 11 bands of Landsat 8, QA band and manual annotation band labels.

Each large picture is cut into 512 \* 512 small pictures, and the ratio of training set and verification set is 2:8

### 2.2 Network Architechture

As is shown in the figure below, our model include three parts.

![img](3.png)

In the first part, we only use 1 x 1 convolution kernel, which aim is extract the spectral features of the target. This step will make our algorithm more efficient. In the second part, variance is used to select the feature map generated in the first part. As is known to all, the diversity of underlying surface will have a great impact on cloud detection. In the training process, the underlying surface of each image is the same, so we hope to select the features of the underlying surface suitable for each image through this step. Because the human eye can recognize the cloud through the color image, we choose the three characteristic images with the largest variance. In the third part, we want to utilize spatial information, a simple net similar to U-Net is used.

## 3. Experience and Evaluation

![img](7.png)

As shown in the figure above, S-Net can fit ground true well.

摘抄：

Spectral information is the most important characteristic of HSIs and plays the vital role for the classification tasks