Straight network: A new network structure of landsat imagery cloud detection

王树立

2020 年 3 月 14 日

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。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，很容易对一些高亮地物会造成误分，如冰雪。也有一些基于空间的云检测方法，但这些方法的识别结果大多过于平滑，容易丢失边缘细节信息，并且对薄云与碎云的漏警率比较高。为了兼顾并平衡光谱与空间信息，本文中，我们提出了一种新型的、轻量的网络，叫做直通网络（S-Net），一种专门用于遥感图像的深度学习模型。S-Net 主要分为三部分，第一部分，借鉴了 UNet 的 encoder-decoder 结构，但使用更少的层数，提取空间信息；第二部分，通过增加 1*1 的卷积核，使遥感图像的光谱信息可以直接到达最终的预测部分，这一部分使得图像的细节部分可以保留，使预测 mask 更加精确；第三部分，将光谱信息与空间信息以注意力的方式结合。模型在 landsat 8 iome 数据做训练并估计，取得了优于 Fmask、并且与 unet 相当的实验结果，具有更高的召回率与 f1 值，而参数个数只有 unet 的二十分之一。并且仔细观察可以发现，S-Net 在保持细节方面具有优秀的能力。

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，在农业产量估算 [19]、变化检测 [22]、灾难评估 [11] 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 [12]，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。

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，也会有很多的构造特征（如：NDVI，NDWI）。因此有很多基于单像素阈值的传统方法 [21]。Irish 等人 [9] 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 [26] 用到了 landsat 几乎所有的波段，通过设置亮度 (Brightness) 阈值、色度 (Whiteness) 阈值、温度 (Hot) 阈值、NDVI、NDSI 等，进行精确的云检测。这些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。

基于同一地区时相相近的两幅或两幅以上影像进行云检测也是一类常见的方法。该类方法原理是认为，在高频率的观测下，相对于变化较快的云，陆地表面可以被看作是静态或缓变的背景。Zhe zhu 等 [27] 提出的 TMASK 云掩膜算法首先根据 FMASK [26] 识别出历史数据中清晰的像素，再选取三个波段，运用递归最小二乘方法，对选取出的三个波段利用正余弦函数进行拟合，最后计算当前值与预测值的差值判断每个像元是否为云。 Goodwin 等 [7] 也利用时序数据构建了一种基于 Landsat TM/ETM+ 的云检测算法，该算法综合利用影像的光谱、时相和上下文信息，在云检测方面精度比 Fmask 更高，但算法较为复杂，且需要大量无云的时间序列影像作为参考。Ronggao Liu 等 [15] 提出一种从 MOD09 时间序列产品中生成云掩膜的方法。利用蓝色波段和蓝色波段与短波红外波段之间的比值，对时间序列进行排序，以拐点值作为判别云与地物的门限。但基于时序的遥感图像云检测方法往往是非实时的，且需要依赖很多的其他数据。

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，但这降低了图像的分辨率。云检测其实是图像分割任务 Pixel-level labeling tasks，图像分割可以定义为一种特定的图像处理技术，用于将图像划分为两个或两个以上有意义的区域。对于图像分割任务，输入一副图像，输出也是一副图像。

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 [2, 10]，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，而直接将为 RGB 图像设计的网络应用于多光谱遥感图像会有三个问题：1. 因为 RGB 图像只有三个波段，针对 RGB 图像提出来的算法，空间信息无疑是神经网络提取的重点，这类网络往往通过加深层数以加大感受野，而对于识别大部分地物，尤其是检测云，过于考虑全局信息会对存储与计算造成巨大的浪费；2. 感受野的增大往往意味着保持边缘细节能力的减弱，而对于云检测任务来说，我们在意的往往是这些边缘，一味地增加感受野可能会造成适得其反的效果，而且现有网络的输出往往过于平滑，相近的像素往往具有相同的预测值，这对于遥感图像碎云的检测极为不利，而碎云的检测正是云检测中的难点；3. 遥感图像有丰富的光谱信息，且光谱特征是地物最本质的特征，我们应该为光谱信息分配更多的计算资源。

随着深度学习的发展，卷积神经网可以有效提取图像空间信息，所以出现了一些利用神经网络做云识别的方法。但这些方法往往需要大量的参数与复杂的计算，且目前的神经网络算法在保持边缘细节与扩大感受野之间存在着难以协调的矛盾。

在这篇文章中，我们基于 U-Net，提出了一个新颖的、简单的、有效的网络，主要包括三部分。第一部分，考虑到复杂的下垫面是云检测中的难点，为了更好地处理复杂的下垫面，我们减少

了 encoder-decoder 层数，注重局部空间特征。第二部分，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，这样也可以保证边缘细节。第三部分，引入注意力机制，将局部空间信息以 attention 的形式与像素光谱特征结合，得到最终的云掩膜图。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。

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 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 (Operational Land Imager) 和 TIRS 热红外传感器 (Thermal Infrared Sensor)。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。

传感器类型	波段	波长范围 (μm)	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
OLI	10.TIRS1	10.60-11.19	100
	11.TIRS2	11.50-12.51	100

表 1: landsat8 波段信息

为了对模型进行训练与测试，我们利用已有的全球云和云影验证数据集“L8 Biome Cloud Validation Masks”[6]，该数据集共有 96 景图片，包含 8 个种类的下垫面 (including Barren, Forest,Grass/Crops, Shrubland, Snow/Ice, Urban, Water, Wetlands。

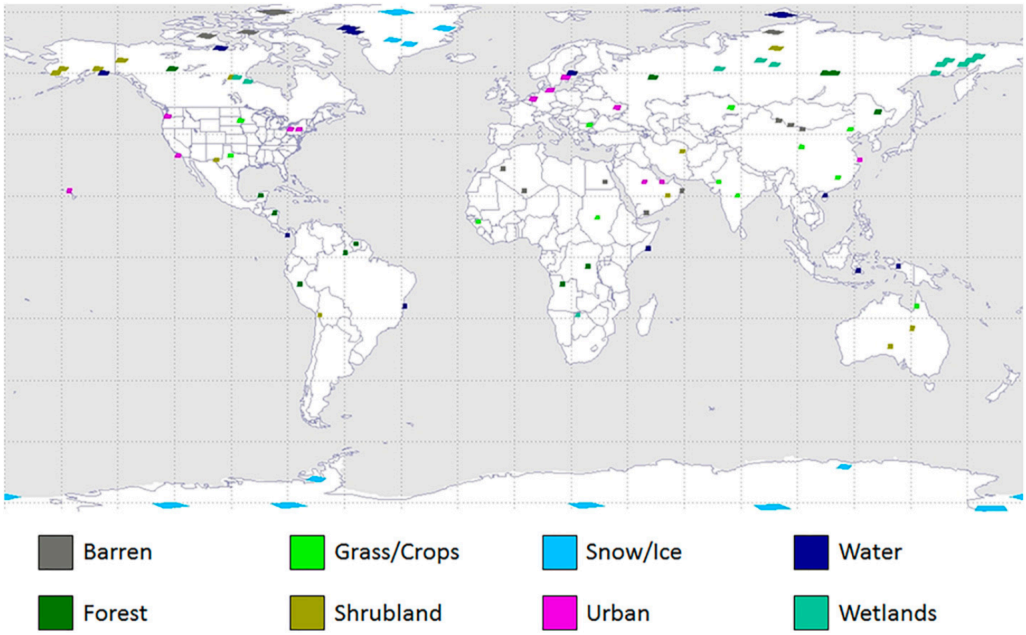


图 1: Global distribution of the 96 unique Landsat 8 Cloud Cover Assessment (CCA) scenes.

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。

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

表 2: L8 Biome 数据人工标注标志位

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*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。

2.2 Background

在提出我们的模型之前，我们先介近年来绍深度学习在语义分割方向的发展，了解模型的设计模式以及需要解决的问题。

Long 提出 FCN [16] 是 CNN 语义分割的开山之作，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 [20] 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（summation），U-Net 用的是叠操作（concatenation），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，如图 2所示。

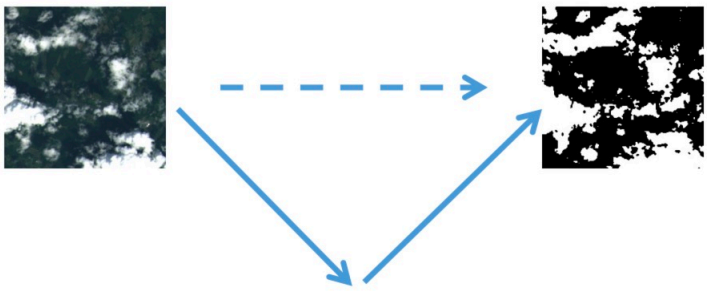


图 2: encoder-decoder 框架

需要注意的是，FCN 与 U-Net 都用到了跳层连接，这其实揭示了端到端语义分割的一个主流矛盾：既需要全局的感受野来完成分类任务，又需要用局部信息和低层的低级视觉信息来达到准确的边缘分割。为了考虑更多的空间信息，需要获得更大的感受野，因此在重构过程中会产生粗糙的输出，最大池化层的存在则进一步降低了获得精细分割输出的机会。随着卷积神经网络深度的增加，特征图的空间分辨率降低，小目标的信息逐渐丢失。

为了解决这个矛盾，有很多方法对这个结构进行了改进。Shuai Zheng 等人 [24] 提出了 FCN-CRF，利用条件随机场对分割结果进行平滑与优化。Seg-Net [1] 引入了一种新的上采样方式，叫做反池化，减少了大量参数，提升了计算性能，但准确率一般。Deeplab [3–5] 系列，用 resnet 作为基模型；利用空洞卷积在不损失分辨率和边缘信息的情况下增大感受野；引入一种类似于金字塔结构的模块–ASPP 模块，以检测不同大小的物体，并用空洞卷积替代最后几个下采样层。Zhou 等人 [25] 提出的 unet++，丰富了 decoder 结构，在每一层都加入了 decoder，使用了浅层和深层的特征。也有将注意力机制应用于图像分割的方法 [14, 18, 23]，加权地学习全局信息，并保证局部信息。Alex 等人 [13] 将图像分割看作渲染任务，只对最不确定的几个位置进行细分、预测，提出了一个有效保留图像细节的结构。

2.3 Network Architechture

上一节提到，为了兼顾感受野与细节信息，这些基于灰度图与 RGB 图像设计的网络主要通过引入金字塔型的卷积核、注意力机制、丰富跳层连接、增加模型参数等方法。我们将神经网络应用于遥感图像，首先应该思考数据源的区别–多波段是遥感图像得天独厚的优势。为了完全挖掘遥感图像的信息，更好地利用 landsat 多波段的特点，我们调整了 encoder-decoder 结构，增加了直达

通道，可以理解为一种特殊的跳层链接。我们的模型结构如图 3，可以分为三部分：

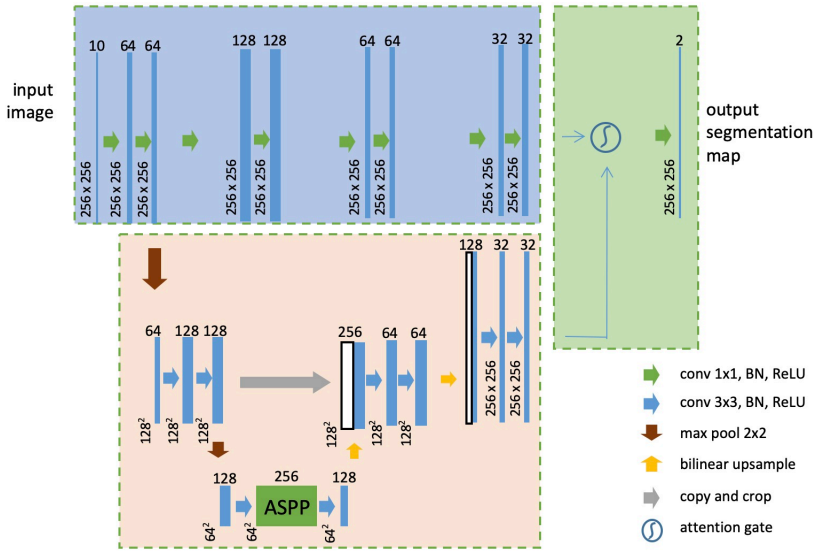


图 3: our net architecture.

我们的模型整体上采用了 encoder-decoder 结构，与 unet 相比，主要有以下几方面改变：

1. 增加直达通道。直达通道里的所有卷积核的大小均为 1*1，这其实是一个多层感知机，可以将这一部分理解为一个特殊的跳层链接。增加直达通道并不涉及空间信息的提取，专注于提取光谱信息, 目的就是得到高分辨率的云掩膜图，以解决 encoder-decoder 结构天然的缺陷–在保持边缘细节与考虑全局信息之间存在着的以协调的矛盾。

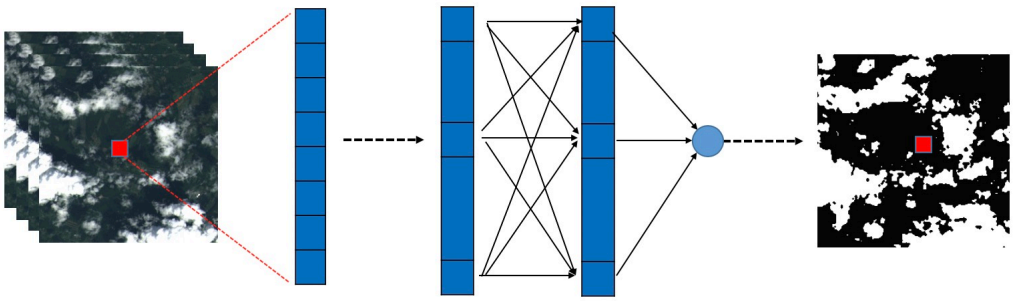


图 4: Part1. Straight access, 所有的卷积核都是 1*1 大小的，等价于多层感知机。

由于遥感图像特有的多波段特征，使得直达通道变得可能，比如构造 NDVI 这样的特征。若是不考虑图像的空间信息，对于遥感图像，我们使用决策树或者 SVM 也是可以对遥感图像的单个像素进行分类的，方法就是将每个像素单独地看作一个样本。事实上，很多方法就是这样做的，如 zhu 等人提出的 FMask [26]。因此，我们的网络也具有良好的可替换性，通过将直达通道替换

为传统的阈值法，可以将深度学习与传统经验完美地结合起来。

2. 空间信息提取。直达通道可以解析细节，但它们不包含特定于区域的信息与上下文信息，因此我们仍需要提取空间信息。图 5展示了更加清晰的 encoder-decoder 结构。unet 为了提取全局信息，获得更大的感受野，设计了 5 层网络，共进行 4 次下采样，每加深一层，模型参数都会成倍地增加。若不考虑 3*3 的卷积层，每一次下采样均由 2*2 的池化层完成，最深一层的特征图大小为输入图像的十六分之一，每个像素包含了原始图片 16*16 像素大小的空间。我们认为遥感图像云检测应该更侧重局部信息，过深的层数不仅会造成资源的浪费，还会加剧边缘模糊的程度。所以我们将层数设计为 3，意味着判断每个像素是否为云仅仅依靠其邻域 4*4 范围内的像素。

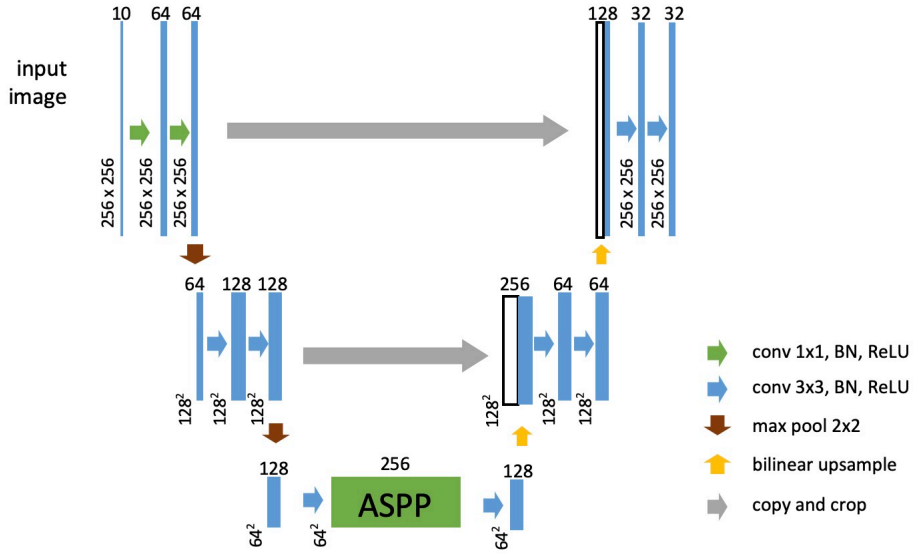


图 5: Part2. 相似于 U-Net，减小了模型深度，同时加入了 ASPP 增加感受野。

同时，为了以较小的代价捕获多尺度信息，借鉴 deeplab 的思想 [4]，引入金字塔型空洞卷积。由池化层获得的感受野信息会造成精度的损失，空洞卷积可以在不增加额外参数与下采样层的前提下提高感受野。空洞卷积实际卷积核大小：$K = k + (k-1)(r-1)$，k 为原始卷积核大小，r 为空洞卷积参数空洞率。空洞率为 1 的空洞卷积就是正常的卷积。空洞卷积可以分为串联与并联两种情况，串联的情况比较复杂，我们选择了较为简单的并联结构。

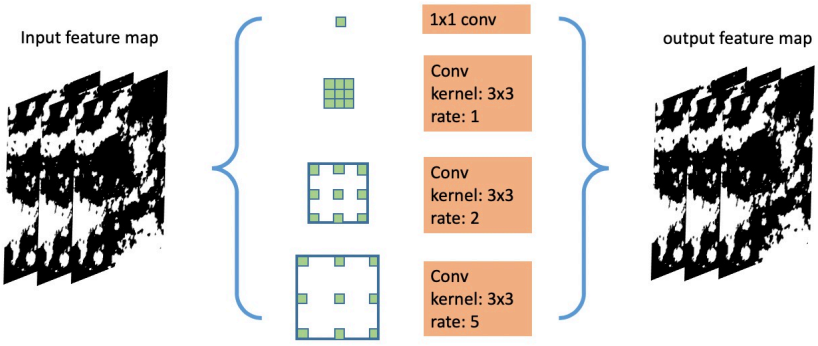


图 6: 金字塔形空洞卷积 (ASPP). 利用不同空洞率的卷积核收集不同尺度的信息。

3. 加入注意门控 (Attention Gate)。我们利用 attention gate 将光谱信息与空间信息更加有效地融合起来。近年来，有很多研究将 attention 机制加入到语义分割模型中来。在神经网络中，非线性主要来源于激活函数与池化，attention 的引入增加了非线性，一部分人以此来解释 attention 的有效性。为了获得足够大的感受野，feature map 在 CNN 中被逐渐下采样。这样，特征与像素间的相对关系可以在一个大的视野中被收集。但是，这对于一些小但具有明显特征的目标并不友好。目前大多数网络会在感受野最大的卷积位置增加 attention 结构，以此来学习全局的有效信息。我们考虑到，在云检测任务中应该更加注重下垫面与局部信息，因此，我们只在 decoder 的最后一层加入了 attention 机制。并且，这不会增加大量的额外参数。令直达网络的输出为特征 f，decoder 的输出为门控信号 g，那么我们的 AG 参考了 [18] 中的结构，可以表示为图 7。

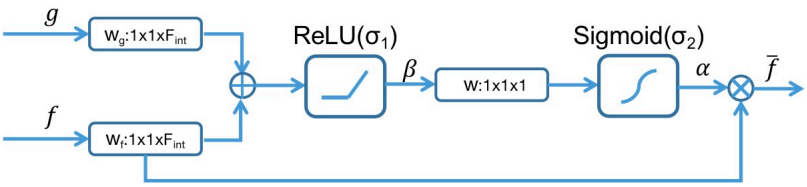


图 7: Part3. Attention gate

门控系数 α 是长宽与输入特征 f 相同，但通道数为 1 的单层权重图，$\alpha_i \in [0,1]$。$w: 1*1*1$ 意思是 $w: height, weight, out\_channelss$。$w_g$ 与 w_f 将门控信号 g 和输入特征 f 变为形状完全相同的两个张量，w 将输入张量变为 2 维的平面图，最后再由 $Sigmoid$ 函数将实域映射到 0-1 之间。注意力机制的公式表述如下，其中 $\sigma_1(x) = max(0,x)$，$\sigma_2(x) = \frac{1}{1+e^{-x}}$。

$$\beta = \sigma_1(W_g^T g + W_f^T f + b_1) \tag{1}$$

$$\alpha = \sigma_2(W^T \beta + b_2) \tag{2}$$

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 [17]，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 [8](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。

$$BN(x) = \gamma \frac{x - \mu}{\sigma} + \beta \tag{3}$$

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。

$$ReLU(x) = \begin{cases} x, & \text{if} \quad x > 0 \\ 0, & \text{if} \quad x \leq 0 \end{cases} \tag{4}$$

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。

$$L = -\sum_{i=1}^{N} y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i) \tag{5}$$

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。

3 Experience and Result

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：

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$R = \frac{TP}{TP + FN} \tag{7}$$

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$F = \frac{2 * R_{recall} * R_{precision}}{R_{recall} + R_{precision}} \tag{9}$$

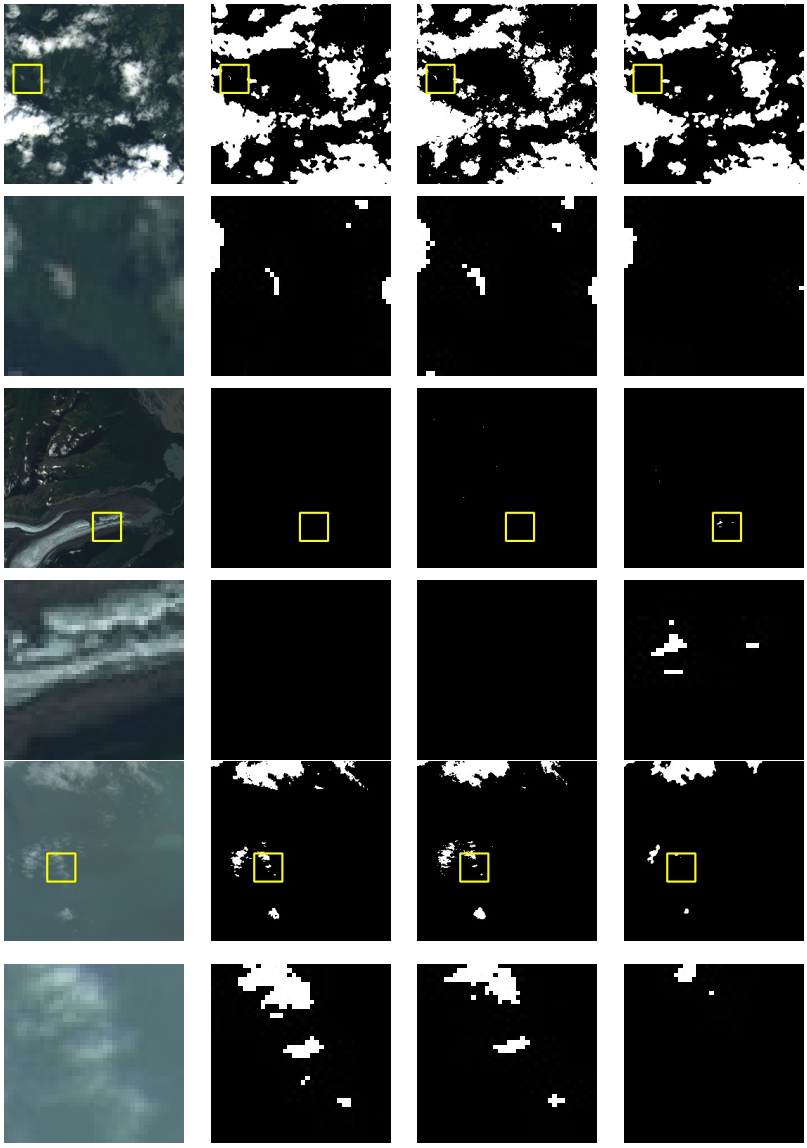
其中，TP 为真正类（True Positive），即云被判为云；TN 为真负类（True Negative），即非云被判为非云；FP 为假正类（False Positive），即非云被判为云；FN 为假负类（false Negative），云被判为非云。

本文将模型与 landsat 自带的 QA 波段、原始的 UNet 做比较。相比于 QA 波段，取得了 8 个百分点的性能提升；相比于原始的 UNet，效果十分接近，但是我们的模型参数仅不到 UNet 的十分之一，更加轻量，具有更大的应用潜力。更详细的评估结果如表 (3) 所示。具体参数，如学习率为 $1e^{-2}$，batch size 为 8，采用动量为 0.9 的随即梯度下降法训练，损失函数为交叉熵损失。所有实验均在 MacBook Pro (15-inch, 2018) 上进行，处理器为 2.2 GHz Intel Core i7，内存 16 GB 2400 MHz DDR4，Intel UHD Graphics 630 1536 MB。

model	evaluation	Barren	Forest	Grass/Crops	Shrubland	Snow/Ice	Urban	Water	Wetlands	total
our	acc	93.74	95.60	94.28	94.45	92.89	92.96	94.72	93.55	94.02
	rec.	98.64	94.26	95.17	96.59	94.76	98.63	92.62	99.29	**96.24**
	prec.	91.79	99.49	89.59	93.10	88.88	86.93	93.86	90.28	91.74
	F1	95.09	96.80	92.30	94.81	91.72	92.41	93.23	94.57	**93.87**
QA	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
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	**95.30**
	F1	96.80	95.99	88.95	94.58	90.02	94.36	91.85	95.62	93.52

表 3: Evaluation results on the Biome dataset

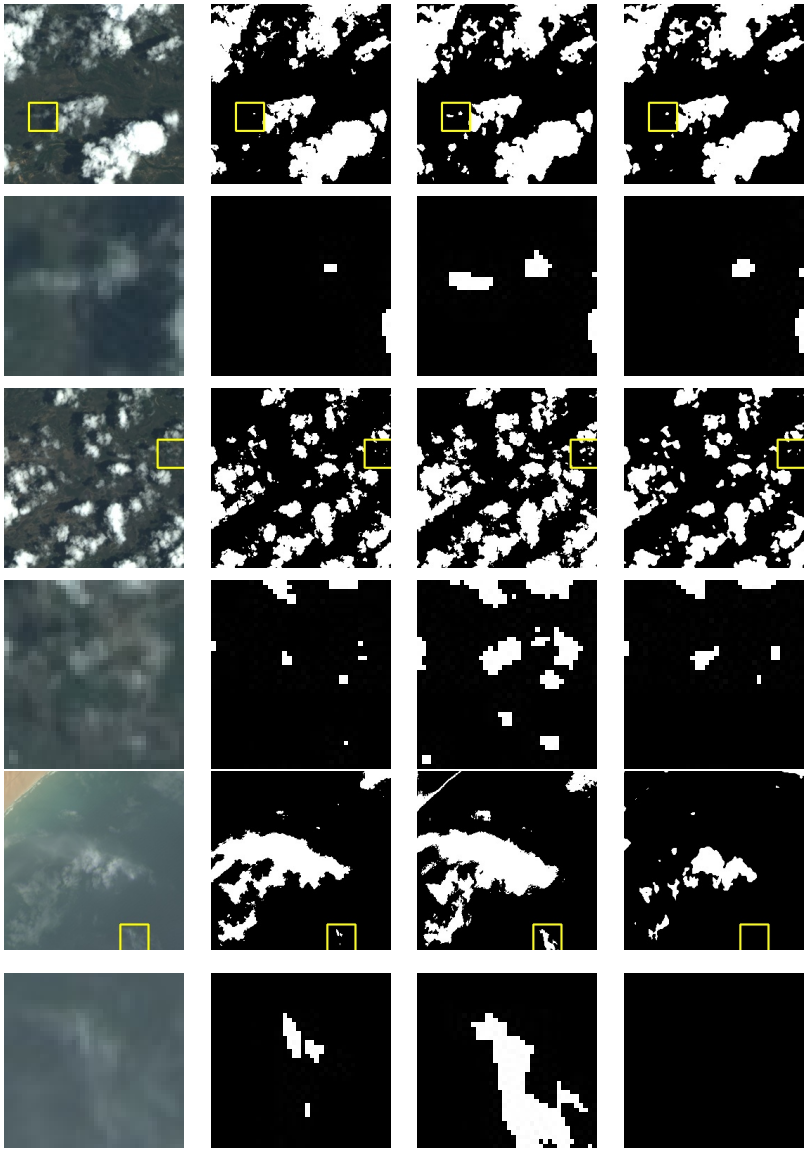
在大部分情况下，我们的模型与 U-Net 相差无几。但仔细观察可以发现，我们的模型对于碎云、细节有良好的检测与保持能力，unet 的识别更加光滑，使得一些细节被忽略，而我们的模型更加注重细节，这对于云检测是一个很重要的能力。如图 8所示，从左到右依次为真彩色图、人工标注、我们的模型预测结果、UNet 结果，第二行是第一行黄色方框的放大图，以此类推。我们的模型由于有直达通道的存在，对细节有较好的保持，对碎云有良好的识别；同时由于引入 attention，也考虑到了图像的空间信息，能以极少的参数在整体上达到与 U-Net 相近的效果。



(a) Color (b) GT (c) Ours (d) unet

图 8: Example of prediction

我们的模型由于直达通道的存在，虽然模型参数比 unet 大大降低（我们的模型有 763829 个参数，而 UNet 有 14096706），但更加注重单个像素的光谱特征，容易识别小尺寸的云，这在低分辨率的遥感图像中更加有用，因为低分辨率的碎云可能会更多。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；并且以这些标签为真值进行的训练，可能也会存在问题。



(a) Color (b) GT (c) Ours (d) unet

图 9: Example of bad GT

4 Conclusion

云检测一直是遥感领域的研究热点与难点。本文提出了一种基于 encoder-deconder 结构改进的新的遥感图像云检测模型。为了更好地保持边缘细节、增加对小的碎云的检测能力，这在云检测中非常重要，我们引入了直达通道，。直达通道全部是 1*1 的卷积核，它使得检测结果保持了“纯粹性”，没有收到其空间信息的干扰。并通过注意力机制，将空间信息与光谱信息完美地结合起来。该模型充分利用遥感图像多波段的特点，在保持边缘细节与扩大感受野之间寻找矛盾的解决方法，并在 landsat8 数据集上达到了 94% 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

直达通道与空间信息提取部分，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] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation, 2015.

[2] 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.

[3] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs, 2014.

[4] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[5] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation, 2017.

[6] 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.

[7] Nicholas R Goodwin, Lisa J Collett, Robert J Denham, Neil Flood, and Daniel Tindall. Cloud and cloud shadow screening across queensland, australia: An automated method for landsat tm/etm+ time series. *Remote Sensing of Environment*, 134:50–65, 2013.

[8] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] Alexander Kirillov, Yuxin Wu, Kaiming He, and Ross Girshick. Pointrend: Image segmentation as rendering. *arXiv preprint arXiv:1912.08193*, 2019.

[14] Hanchao Li, Pengfei Xiong, Jie An, and Lingxue Wang. Pyramid attention network for semantic segmentation. *arXiv preprint arXiv:1805.10180*, 2018.

[15] Ronggao Liu and Yang Liu. Generation of new cloud masks from modis land surface reflectance products. *Remote Sensing of Environment*, 133:21–37, 2013.

[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] Augustus Odena, Vincent Dumoulin, and Chris Olah. Deconvolution and checkerboard artifacts. *Distill*, 1(10):e3, 2016.

[18] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[19] 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.

[20] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[21] 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.

[22] 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.

[23] Hang Zhang, Kristin Dana, Jianping Shi, Zhongyue Zhang, Xiaogang Wang, Ambrish Tyagi, and Amit Agrawal. Context encoding for semantic segmentation. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 7151–7160, 2018.

[24] Shuai Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes, Vibhav Vineet, Zhizhong Su, Dalong Du, Chang Huang, and Philip HS Torr. Conditional random fields as recurrent neural networks. In *Proceedings of the IEEE international conference on computer vision*, pages 1529–1537, 2015.

[25] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11. Springer, 2018.

[26] Zhe Zhu and Curtis E Woodcock. Object-based cloud and cloud shadow detection in landsat imagery. *Remote sensing of environment*, 118:83–94, 2012.

[27] Zhe Zhu and Curtis E Woodcock. Automated cloud, cloud shadow, and snow detection in multitemporal landsat data: An algorithm designed specifically for monitoring land cover change. *Remote Sensing of Environment*, 152:217–234, 2014.