Direct network: A new network structure of cloud detection

王树立

2020 年 2 月 11 日

摘要

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，很容易对一些地物会造成误分，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。但目前提出的神经网络大部分是为了处理灰色图或 RGB 图像的，这些方法没有考虑遥感图像多波段的特点，将现有的网络结构直接应用于遥感图像。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。本文中，我们提出了一种兼顾细节与感受野的新型网络结构，并充分考虑遥感图像多波段的优势，专门用于多波段的遥感图像云检测，但参数个数只有 unet 的十分之一。我们用 landsat8 数据做实验，取得了优于 Fmask、并且与 unet 相当的实验结果。

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。传统的实时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，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。

Long 提出 FCN [9] 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 pixel-level labeling。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 [12] 是一种结构对称的网络，丰富了 decoder 部分。虽然 FCN 用的是加操作（summation），U-Net 用的是叠操作（concatenation），但大同小异。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。以后的网络，大部分都在这个框架下。

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，又需要在边缘部位，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。Shuai Zheng 等人 [14] 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；使用并抛弃了条件随机场。也有将注意力机制应用于图像分割的方法 [8,11,13]。Zhou 等人 [15] 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。

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 [2,7]，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；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。

在这篇文章中，我们基于 encoder-decoder 结构，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 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)。Landsat-8 卫星一共有 11 个波段，波段 1-7，9-11 的空间分辨率为 30 米，波段 8 为 15 米分辨率的全色波段，卫星每 16 天可以实现一次全球覆盖。

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，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。

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

表 1: L8 Biome 数据人工标注标志位

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。

2.2 Network Architechture

我们的模型结构如下图，由三部分构成：

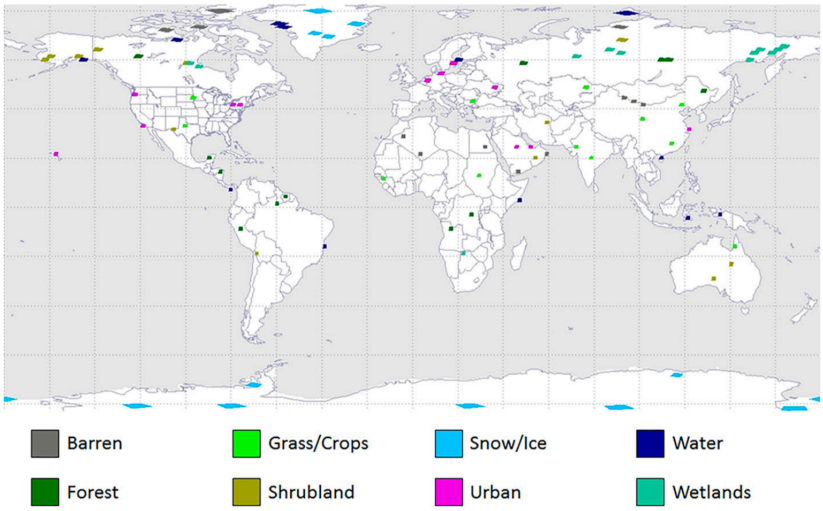

图 1: Global distribution of the 96 unique Landsat 8 Cloud Cover Assessment (CCA) scenes, sorted by International Geosphere-Biosphere Programme (IGBP) biome. Twelve scenes were selected for each of the eight biomes.

我们的模型整体上采用了 encoder-decoder 结构，与 unet 相比，主要有一下几方面改变：

1. 减少了模型层数。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，若不考虑 3*3 的卷积层，最深一层的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。

2. 增加直达通道。增加直达通道的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。由于遥感图像特有的多波段特征，使得直达通道变得可能。若是没有 CNN，即不考虑图像的空间信息，对于遥感图像，我们仅仅用 alexnet 或者 vgg 这样的网络也是可以对遥感图像进行分类的，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 landsat 官方的云掩膜方法 [16]。

3. 加入局部注意力机制。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。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。

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

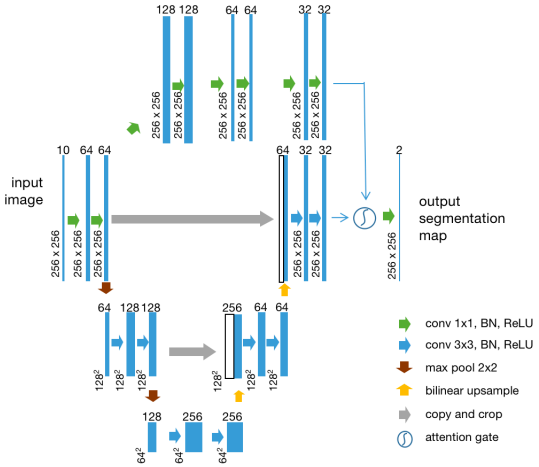

图 2: our net architecture.

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 [10]，双线性差值可以取得与反卷积相同甚至更好的效果。因此，我们选择参数少且效果好的双线性差值法。

3 Experience and Result

evaluation	Barren	Forest	Grass/Crops	Shrubland	Snow/Ice	Urban	Water	Wetlands	total
acc	96.14	95.99	94.20	94.97	87.80	95.32	94.28	95.51	94.27
recall	95.71	96.08	87.98	93.71	98.98	96.76	92.92	98.34	95.41
precision	97.95	98.13	97.17	96.87	78.82	92.85	92.25	93.95	94.06
f1	96.82	97.10	92.35	95.27	87.76	94.76	92.58	96.10	94.73

表 2: Evaluation results on the Biome dataset

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] 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.

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

[9] 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.

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

[11] 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.

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

[13] 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.

[14] 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.

[15] 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.

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