引文格式：甘文祥, 张远谊, 李欣园. 一种航空影像建筑物检测的轻量化CNN建模方法[J]. 地理空间信息, 2023, 21(6): 24–27. doi:10. 3969/j. issn. 1672–4623.2023.06.006

一种航空影像建筑物检测的轻量化CNN建模方法

甘文祥[1]，张远谊[1]，李欣园[1]
（1. 中国地质大学（武汉）地理与信息工程学院，湖北 武汉 430074）

摘　要：以卷积神经网络为代表的深度学习方法大幅提高了遥感影像建筑物自动检测精度，但由于建筑物复杂多样，为了提取区分能力更强的图像特征，现有卷积神经网络方法往往倾向于构建层次复杂、参数庞大的深度模型。这使得模型的存储和内存开销都较高、检测速率也容易受到影响，一定程度上造成在移动设备平台或灾害应急等场合的应用受限。针对此问题，提出一种用于航空影像建筑物检测的轻量化卷积神经网络建模方法，采用深度可分离卷积方法对复杂网络进行简化，大幅减少了计算量，并较好地维持了原有精度。实验表明新方法相比改进前，在计算量和参数量分别减少86%和87%、训练时间缩短10%的情况下，建筑物检测的精度仅降低3%。

关键词：航空影像；建筑物检测；CNN；深度可分离卷积；轻量化网络
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672–4623（2023）06–0024–04

A Lightweight CNN Modeling Method for Building Detection from Aerial Images

GAN Wenxiang[1], ZHANG Yuanyi[1], LI Xinyuan[1]
(1. School of Geography and Information Engineering, China University of Geosciences (Wuhan), Wuhan 430074, China)

Abstract: In recent years, represented by convolutional neural network(CNN) method, the deep learning method has greatly improved the automatic detection accuracy of buildings in remote sensing images. However, due to the complexity and diversity of buildings, in order to construct and extract image features with stronger distinguishing ability, existing CNN methods tend to construct deep models with complex layers and large parameters. This makes high storage and memory overhead of the model, and the detection rate easy to be affected, which to some extent limits the application in mobile device platform or disaster emergency. To solve this problem, we proposed a lightweight CNN modeling method for building detection from aerial images. We used the depthwise separable convolutional method to simplify the complex network, which could greatly reduce the computation and maintain the original accuracy. The test results show that compared with the method before the improvement, the accuracy of building detection is only reduced by 3% when the calculation and reference quantity are reduced by 86% and 87% respectively and the training time is shortened by 10%.

Key words: aerial image, building detection, CNN, depthwise separable convolution, lightweight network

现阶段，建立轻量化网络模型的方法通常可分为模型压缩和模型简化两类。模型压缩类方法主要通过压缩、分解等方式以降低预训练模型文件大小，具体算法包括量化压缩、哈希、和霍夫曼编码等[1–3]。模型简化类方法则通过重新设计模型架构实现网络参数的轻量化，其中具有代表性的MobileNet模型通过应用深度可分离卷积方法，在减少模型计算量、降低网络计算复杂度等方面取得了良好效果[4–5]。在遥感领域，轻量化CNN模型设计的研究尚未引起较多关注，其中张晓玲[6]等提出了一种基于DSC的CNN模型，并结合注意力机制实现了高速和高精度的SAR舰船检测，但目前面向建筑物检测的轻量化CNN建模方法尚不多见。

本文以一个覆盖面积约34 km^2，涵盖10 000余个建筑物目标的航空影像建筑物数据集为研究对象，通过应用DSC技术实现了对深度复杂网络模型的轻量化处理，并在此基础上提出了一种用于航空影像建筑物检测的轻量化CNN建模方法。实验表明，本文方法可在建筑物检测精度仅降低3%的前提下，将基准模型的计算量和参数量分别减少86%和87%，训练时间缩短10%。

1　面向航空影像建筑物检测的轻量化CNN建模

下文首先对CNN中的经典卷积操作进行简介，在此基础上描述深度可分离卷积操作的过程，并对二者的计算量进行比较。随后对本文构建的轻量化CNN模

收稿日期：2021–09–12；修回日期：2022–09–01。
项目来源：国家自然科学基金资助项目（41601506）。
第一作者简介：甘文祥（1999—），主要研究方向为基于深度学习的遥感影像建筑物提取，E–mail：gwxhagan@163.com。

型的具体实现进行介绍，并给出建筑物检测的技术流程。

1.1 经典卷积神经网络

CNN模型的基本原理是通过一次或多次的卷积、池化操作实现输入数据的特征提取，随后基于全连接操作完成特征的自动分类。

CNN对建筑物进行检测包含训练与预测2个部分。训练是指遥感影像输入到模型后，CNN模型滑动窗口式地对每个像元进行一次或多次的卷积、池化，对输入的数据的特征进行提取，并通过全连接的操作，完成特征的自动分类，实现对建筑物的检测。

卷积操作是指在输入的影像或特征图中利用滑动卷积核对每个像元进行统一的非线性变换，并实施叠置变换后，得到卷积层输出的特征图。如图1所示，经典卷积中，假设输入的特征映射F的尺寸为 $D_{\text{in}}\times D_{in}\times M$，输出的特征图G的尺寸为 $D_{\text{out}}\times D_{\text{out}}\times M$，采用的经典卷积核 K 的尺寸为 $D_k\times D_k\times M\times N$，其中，$D_{\text{in}}$ 为一个输入影像的尺寸，M 为输入影像的通道数，D_{out} 为一个输出影像的尺寸，N 为输出影像的通道数，D_k 为卷积核的尺寸。

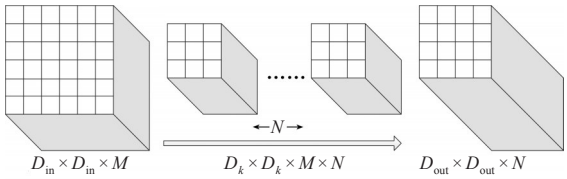


图1　经典卷积示意图

假设卷积步长为1且有填充，则经典卷积的计算量 I_s 为卷积核的大小 $D_k\times D_k$ 与输入通道的个数 M 与输出通道的个数 N 再与输入特征图的大小 $D_{\text{in}}\times D_{\text{in}}$ 直接相乘：

$$I_S=D_k\times D_k\times M\times N\times D_{\text{in}}\times D_{\text{in}} \tag{1}$$

1.2 深度可分离卷积

DSC通过将一次经典卷积分解为2个步骤。第一步，深度卷积，如图2所示，对每个输入通道进行一次卷积，提取特征。

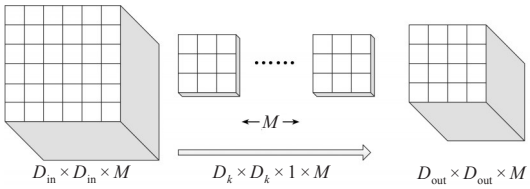


图2　深度卷积示意图

深度卷积的计算量 I_D 为：

$$I_D=D_k\times D_k\times M\times 1\times D_{\text{in}}\times D_{\text{in}} \tag{2}$$

第二步，逐点卷积，如图3所示。通过1×1卷积组合深度卷积得到的特征，并将这些特征转换通道，创建一个深层的线性的输出层。同时，深度卷积与逐点卷积2个层每层之后都使用批量标准化[7]（batch-norm）和ReLU。

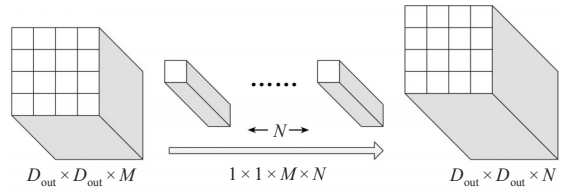


图3　逐点卷积示意图

逐点卷积的计算量 I_P 为：

$$I_P=1\times 1\times M\times N\times D_{\text{in}}\times D_{\text{in}} \tag{3}$$

深度卷积仅仅对输入通道进行了卷积，但还没有将这些通道组合起来构建新的特性。因此，下一步逐点卷积则将深度卷积输出的线性组合来生成新特性。这种分解方法的优点是大大减少了计算量和模型大小，打破了输出通道数量与内核大小之间的交互。

DSC由深度卷积与逐点卷积的组合而成，它的计算量 I_{DSC} 为2种卷积方式的计算量之和，公式如下：

$$I_{\text{DSC}}=I_D+I_P=D_k\times D_k\times M\times D_{\text{in}}\times D_{\text{in}}+M\times N\times D_{\text{in}}\times D_{\text{in}} \tag{4}$$

对比经典卷积，使用DSC后计算量的减少公式如下：

$$\frac{I_{\text{DSC}}}{I_S}=\frac{D_k\times D_k\times M\times D_{\text{in}}\times D_{\text{in}}+M\times N\times D_{\text{in}}\times D_{\text{in}}}{D_k\times D_k\times M\times N\times D_{\text{in}}\times D_{\text{in}}}=\frac{1}{N}+\frac{1}{D_k^2} \tag{5}$$

由上式可知，在输出通道数 N，或者卷积核窗口大小 D_K 足够大，即使用更大的卷积核获取更多信息和构造更深的特征图的情况下，DSC可以使网络的计算量大幅减少。

1.3 模型实现

本文采用的基准模型为深度残差网络[8]，同时应用了特征金字塔网络[9]模块。该框架具有很强的图像特征提取能力和较高的语义分割精度，适合作为本文实验的基准模型。ResNet通过增加在卷积层之间加入跳转连接，跳过一个或多个层，通过构建浅层的恒等映射来增加更多的层，使网络拥有比较深的深度，并让这些层来拟合残差映射，解决网络退化的问题并优化网络模型。

FPN主要由自底向上的卷积部分、自顶向下的上采样反卷积部分、横向连接部分组成。自底向上的卷积提取遥感影像的低分辨率高维特征图，自顶向下的上采样则产生空间上更粗但语义上更强的特征图，最终生成与输入影像分辨率相同的特征图。自底向上的

卷积与自顶向下的反卷积通过横向连接，构成了特征金字塔。详细的网络结构将在表1中展示，使用DSC的基准模型，将基准模型的3×3经典卷积、7×7经典卷积替换为3×3深度可分离卷积、7×7深度可分离卷积，用于转换通道的1×1卷积则保持经典卷积。

1.4 建筑物检测

本文的遥感影像建筑物检测，通过将实验数据分为训练区域和测试区域，把流程分为训练和测试2个阶段：①训练阶段：将训练区的遥感影像按照固定的间隔与重叠度裁剪为大小相同的图像，并将图像中建筑物像素占比过低的筛除，选为训练数据，其余数据则为验证和测试数据。将数据取小批量输入模型中，采用反向传播和随机梯度下降方法对模型进行训练，得到模型参数。根据模型在验证数据集中的检测结果，对模型参数进行调试，对模型结构进一步优化，重复这些调试，最终得到稳定的模型参数。②测试阶段：将测试数据取一定的大小，输入训练得到的稳定的模型中，最终得到建筑物的预测结果，对模型输出的sigmoid结果使用0.5的阈值进行二值化，输出建筑物和非建筑物的二值分割图像。同时对分割结果采用形态学运算，利用先膨胀后腐蚀的闭运算将建筑物内部的微小空洞填充，并删除过小的非建筑目标，通过地表真值数据采用一定的指标进行定量评价，得到最终的结果与精度评价。

2 实验结果与分析

2.1 数据集介绍

本文实验数据使用的是新西兰基督城地区的开源航空影像数据集[10]，如图4所示。该研究区域覆盖面积约34 km^2，包含了基督城城区10 000余个建筑物目标，同时还含有农田、河道、丘陵等不同地形，其中红色区域为经过校对后的地面真实的建筑物矢量数据，并且以图中黑线为基准，左边的区域被选为训练集数据，分辨率为27 750×27 750，用于模型训练这些影像内建筑物的对应屋顶轮廓矢量；右边的区域被选为测试集数据，分辨率为27 750×27 750，用于评估模型的性能和参数调试。该数据已经过正射纠正且均为正射影像，投影差的存在不会给算法设计带来歧义，使本文的研究更专注于检测建筑物。

本文实验以Ubuntu 16.04（64位）操作系统为平台、装备NVIDIA GeForce GTX 1080 Ti GPU，使用Python语言，以Pytorch为基础库，对实验进行设计与测试。训练阶段，考虑到GPU的内存限制，训练区域影像将被裁切为401像素×401像素、相邻切片间有10%重叠度的46 657张小切片并输入模型，训练批量大小设置为4，使用Adam算法[11]对网络参数迭代更新，设置学习率为0.001，训练20个epoch。在测试阶段，将测试区的影像切割成5 000像素×5 000像素、相邻切片间有10%重叠度的18张切片并输入模型。在计算总体精度的过程中，对于切片之间的重叠区域，取平均精度为最终总体精度。

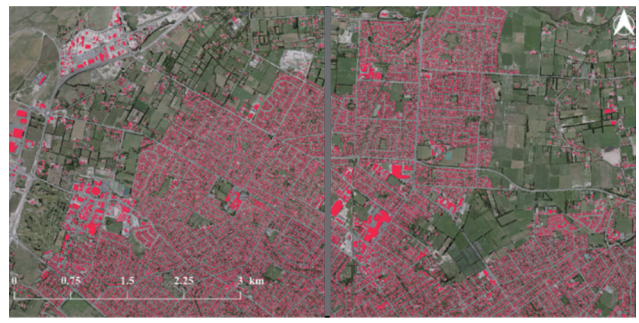


图4 实验区

2.2 实验结果

如表1所示，轻量化模型的浮点运算数FLOPs，即该模型的计算量，减少了85%，参数量减少了87%，在数量级上减少了一位。

表1 各网络模型的计算量与参数量对比

CNN模型	FLOPs/(Million)	参数量/(Million)
基准模型	753.85	11.35
轻量化模型	116.49	1.52

如表2所示，改良后的模型，将86%的计算量、98%的参数量花费在1×1卷积上，而使用经典卷积的基准模型，将94%的计算量、98%的参数量花费在3×3卷积上。

表2 各模型各类型卷积的计算量与参数量的占比

模型	卷积类型	FLOPs/%	参数量/%
基准模型	Conv 7×7	1.20	0.008
	Conv 3×3	93.78	97.82
	Conv 1×1	4.92	2.17
轻量化模型	Conv 7×7	1.86	0.007
	Conv 3×3	12.33	2.4
	Conv 1×1	85.81	97.59

本文通过计算训练区域所有预测值与地面实际真值的平均IoU[12]与F_1-score[13]，比较使用DSC前后的IoU与F_1-score来了解改进后的模型精度变化情况，并计算模型改进前后训练10个epoch所需要的时间与测试效率，来得出效率上的变化。其中，测试效率的计算方式为通过输入一张以随机数生成的尺寸与输入影像一致的图片进入模型，计算从输入到输出该过程所需要的时间，该时间的倒数为模型的测试效率（帧/s）。

实验的结果在表3所示。

表3　各模型检测精度、训练效率、测试效率比较

模型	IoU/%	F_1-score/%	Precision/%	Recall/%	训练时间/h	测试效率/（帧/s）
基准模型	88.9	94.1	95.4	92.9	9.2	9.43
轻量化模型	86.4	91.9	93.8	90.0	8.3	10.57

由表3可知，DSC通过减少模型计算量与参数量，使基准模型训练时间减少了10%，测试效率提升了12%。并且IoU与F_1-score基本能够维持在原有水平，IoU下降了2.8%，F_1-score下降了2.3%。比起效率上的提升，下降的幅度在可以接受的范围之内，说明深度可分离卷积的使用效果良好。改良后的模型，可以在维持精度的情况下，提高检测效率。

3　结　论

本文通过使用深度可分离卷积，提出了一种用于航空影像建筑物检测的轻量化卷积神经网络建模方法。在新西兰基督城的数据集的实验结果表明，本文改良后的方法，可大幅减少基准模型的计算量、参数量，并小幅减少训练时间，同时建筑检测的精度能够维持原有水平，使得模型在相对低配置的环境下也能训练出相对良好的结果。后续的研究，就采用更大规模的数据集，使用更深、更复杂的基准模型，以提升该方法的泛用性，更进一步关注在保持减少训练时间的同时，维持精度不变甚至提高精度的问题。

参考文献

[1] Wu J, Leng C, Wang Y, et al. Quantized Convolutional Neural Networks for Mobile Devices[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016

[2] Chen W, Wilson J, Tyree S, et al. Compressing Neural Networks with the Hashing Trick[C]// International Conference on Machine learning, 2015

[3] Han S, Mao H, Dally W J. Deep Compression: Compressing Deep Neural Networks with Pruning Trained Quantization and Huffman Coding[J]. arXiv Preprint arXiv:1510.00149, 2015

[4] Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. arXiv Preprint arXiv:1 704.04861, 2017

[5] Sifre L, Mall A T S. Rigid-motion Scattering for Texture Classification[J]. arXiv Preprint arXiv:1 403.1 687, 2014

[6] 张晓玲,张天文,师君,等. 基于深度分离卷积神经网络的高速高精度SAR舰船检测[J]. 雷达学报,2019,8(6):841-851

[7] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J]. arXiv Preprint arXiv:1 502.03167, 2015

[8] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016

[9] Lin T Y, Dollár P, Girshick R, et al. Feature Pyramid Networks for Object Detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017

[10] Qi C, Wang L, Wu Y F, et al. Waslander Teray Rremoval: Aerial Imagery for Roof Segmentation: A Large-scale Dataset Towards Automatic Mapping of Buildings, ISPRS Journal of Photogrammetry and Remote Sensing, Volume 147, 2019, Pages 42-55, ISSN 0924-2716

[11] Kingma D P, BA J, Kingma D P. Adam: A Method for Stochastic Optimization[J]. arXiv Preprint arXiv:1412 6980, 2014

[12] Jaccare P. The Distribution of the Flora in the Alpine Zone [J]. New Phytologist, 1912, 11(2): 37-50

[13] Powers D M W. Evaluation: From Predcision Recall and F-Factor to ROC Informedness Markedness & Correlation[J]. J Mach Learn Technol, 2011, 2(1): 37-63

（上接第23页）

效率和精度。

参考文献

[1] 陈军,陈晋,廖安平,等. 全球30 m地表覆盖遥感制图的总体技术[J]. 测绘学报,2014,43(6):551-557

[2] 张兵. 遥感大数据时代与智能信息提取[J]. 武汉大学学报(信息科学版),2018,43(12):1 861-1 871

[3] 郑远攀,李广阳,李晔. 深度学习在图像识别中的应用研究综述[J]. 计算机工程与应用,2019,55(12):20-36

[4] 张军阳,王慧丽,郭阳,等. 深度学习相关研究综述[J]. 计算机应用研究,2018,35(7):1 921-1 928

[5] Fu G,Liu C J,Zhou R,et al. Classification for High Resolution Remote Sensing Imagery Using a Fully Convolutional Network[J]. Remote Sensing,2017,9(5):498

[6] 吕玉娜. 面向深度学习的甘肃省遥感影像样例库建设[D]. 兰州:甘肃农业大学,2019

[7] 林晓萍. 基于深度学习的耕地新增建房监测研究[J]. 地理空间信息,2021,19(11):88-91