引 用 格 式：Wang Yu, Yang Yi, Wang Baoshan, *et al*.Building Segmentation in High Resolution Remote Sensing Imageby Deep ResNet[J].Remote Sensing Technology and Application,2019,34(4):736-747.[王宇,杨艺,王宝山,等.深度残差神经网络高分辨率遥感图像建筑物分割[J].遥感技术与应用,2019,34(4):736-747.]
doi:10.11873/j.issn.1004-0323.2019.4.0736

深度残差神经网络高分辨率遥感图像建筑物分割

王 宇[1,2],杨 艺[3],王宝山[1],王 田[4],卜旭辉[3],王传云[5]
(1.河南理工大学 测绘与国土信息工程学院,河南 焦作 454000;
2.河南理工大学 国土资源部野外科学观测研究基地,河南 焦作 454000;
3.河南理工大学 电气工程与自动化学院,河南 焦作 454000;
4.北京航空航天大学 自动化科学与电气工程学院,北京 100191;
5.沈阳航空航天大学 计算机学院,辽宁 沈阳 110136)

摘要：针对高分辨率遥感图像建筑物分割问题,提出一种Encoder-Decoder的深度学习框架,建立输入图像到分割结果之间的端对端的分割模型。其中Encoder以残差网络为基础,自动提取建筑物的特征;Decoder采用反卷积实现对特征图的上采样,从而完成对建筑物的分割;同时引入批量规范化处理,降低了神经网络权重训练过程中的梯度竞争,从而减小了神经网络的训练难度。实验表明:提出的建筑物分割算法能有效提取建筑物的块状特征和边缘信息,降低复杂道路等干扰的影响,提升建筑物的分割精准度,算法对邻近复杂道路的建筑物、规律性建筑物、单体复杂建筑物等3种典型建筑物的分割精度分别为:0.837、0.892和0.630;F值分别为:0.851、0.879和0.730。同时,多分辨率条件下的分割实验结果表明,该算法对于一定范围内的多分辨率遥感图像具有较好的泛化能力。

关 键 词：高分辨率遥感图像;建筑物分割;深度学习;残差神经网络;批量规范化
中图分类号：P237 **文献标志码**：A **文章编号**：1004-0323(2019)04-0736-12

1 引 言

遥感图像是获取建筑物信息的重要渠道之一。实现遥感图像中建筑物的识别、分割及面积统计,对于国土资源管理、土地利用规划、灾害监控、军事侦察及导航等领域具有及其重要的意义。随着分辨率的逐步提高,遥感图像提供的地物信息更加丰富,这为建筑物的精确分割奠定了重要的信息基础,使得建筑物的自动识别、高精度分割及面积统计成为可能。

遥感图像建筑物分割的实质是:在遥感图像域内,通过有效的图像特征,构建遥感图像与建筑物之间的映射模型。因此特征空间的确立和映射模型的构建是实现建筑物高精度分割的关键。

在传统的遥感图像分割方法中(如分水岭算法[1]、主动轮廓模型[2]、统计区域融合[3]等),往往通过实验、归纳和总结,首先建立一个低秩的特征空间,并在此基础上应用数学工具建立分割模型,该模型一般具有显性的代数表达形式。如在分水岭算法中,集水盆函数是建立在区域内像素灰度基础之上的[1];在主动轮廓模型进化过程中,能量泛函是基于区域灰度或信息熵[2];统计区域融合模型的基础则是图像的区域同质性信息[3]。由于遥感图像到建筑物之间的映射关系往往呈现高维非线性,甚至难以用简单的代数关系描述。为此,神经网络[4]、支

收稿日期:2018-04-25;**修订日期**:2019-06-15
基金项目:国家自然科学基金项目(61503017、61703287、61573129),航空科学基金项目(2016ZC51022)。
作者简介:王 宇(1978—),女,河南焦作人,讲师,主要从事遥感图像处理和土地利用规划方面的研究。E-mail:wangyu@hpu.edu.cn。
通讯作者:杨 艺(1980—),男,河南焦作人,讲师,主要从事数字图像处理和智能控制方面的研究。E-mail:yangyi@hpu.edu.cn。

持向量机[5]等机器学习方法被用来代替人工建模过程,但由于网络层数少、计算量大等问题导致分割精度不高。

近几年,由于深度学习[6]能够自动提取所需的有效特征,并通过深层神经网络结构建立高维强非线性模型,使得图像处理的研究及应用发生了巨大飞跃[7-9]。在图像分割领域,2015年Evan Shelhamer等[10-11]首次提出全卷积神经网络(Fully Convolutional Networks-FCN),将VGG(Visual Geometry Group)深度神经网络[12]的全连接层改造为1×1×N的全卷积层,并成功应用于图像的语义分割,实现了目标特征的自动提取。为了减少网络结构中池化(Pool)操作降低特征图尺寸的影响,Badrinarayana[13]改变了FCN中全卷积层中1×1×N的卷积核结构,采用编码—解码(Encoder-decoder)架构完成分割目标特征的自动获取和分割图像的复原。为了提高FCN的分割精度,Fisher等[14]引入空洞卷积,通过特征图的感受野(Receptive Field)来提升输出特征的稠密度,从而使目标分割更加精准。基于相似原理并结合条件随机场原理,Chen等[15]提出DeepLab V1[16]图像分割算法,目前已经发展到了DeepLabV3,使得目标分割的精度得到了进一步提高。

基于深度学习的巨大优势,遥感图像处理领域正逐步引进深度学习方法。Vakalopoulou等[17]采用深度神经网络提取4 096维特征向量,在此基础上采用支持向量机实现像素的分类,并结合条件随机场完成建筑物的分割。Huang等[18]将遥感图像的RGB和NRG信息分开,构建两个10层的卷积神经网络提取各自的特征,通过最小化特征融合RGB和NRG的特征输出,并采用反卷积实现建筑物分割和输出复原。Saito等[19]建立一种由3个卷积层和2个全连接层组成的深度神经网络,用于遥感图像中建筑物和道路的分割。Yuan等[20]构建了一种多个特征图融合的深度神经网络,并在样本标签中引入建筑物的轮廓信息、区域信息和建筑物区域内至边缘的距离信息,提高了建筑物的分割精度。Bittner等[21]采用FCN神经网络和条件随机场相结合,实现遥感图像中建筑物的分割,其中FCN的输出作为条件随机场的一元势函数的输入值,而成对势函数的建立则是基于位置和色彩信息。Bischke等[22]采用VGG16网络构建了Encoder-Decoder的深度神经网络结构,在训练过程中使用了建筑物边缘的距离信息,并据此构建了多任务能量损失函数,加快了神经网络的训练速度,实现了建筑物的有效分割。Wang等[23]将FCN应用于高分3号极化雷达图像和H-A-α极化分解,对水域、植被和建筑物进行分类识别。Alshehhi等[24]提出一种Single Patch-based卷积神经网络结构,用于提取遥感图像中道路和建筑物的特征。Lin等[25]将FCN应用于遥感图像中近海岸线的舰船检测,其卷积层被分为浅卷积层和深卷积层两种,浅卷积层初步实现目标检测,深卷积层提取特征,再结合FCN实现近海岸线舰船检测与分割。Jiao等[26]针对高光谱遥感图像的分割问题,提出基于FCN的一种深度多尺度空间-频谱特征提取方法,实现了高光谱遥感图像中目标特征的提取和分割。

综上基于深度神经网络的遥感图像分割方法可知,特征空间的构建过程就是深度神经网络提取遥感图像特征的过程,其准确度是由深度神经网络的结构决定的,也必将影响建筑物分割的精度。同时,为进一步逼近分割模型的强非线性,神经网络的深度一般会设置得较深。这常常导致网络训练过程中的误差反向传播出现梯度弥散或梯度爆炸[27],使得深度神经网络难以训练。因此,实现建筑物端对端的高精度分割的关键在于构建合理的深度学习的框架和易于训练的深度神经网络的结构。

为此,本文以深度残差网络(ResNet)[28]为基础,构建Encoder-Decoder深度学习框架,完成遥感图像特征的自动提取和分割模型的学习建立,从而实现从输入图像到分割结果间端对端映射模型的精确逼近。同时,在每个卷积操作之前采用批量规范化(Batch Normalization)[27]技术实现数据的规范化处理,从而摈弃了Dropout、Weight Decay等增强神经网络收敛的方法,提高了神经网络的训练精度。最后针对IAILD(Inria Aerial Image Labeling Dataset)遥感图像数据库[29-30]中的建筑物开展验证实验研究。

2　高分辨率遥感图像建筑物分割的挑战性问题

设遥感图像为I,像素点为$g_h \times g_w = g$,g_h表示I的高度、g_w表示I的宽度,信息通道个数为c;若将图像拉直,则有$I=\{x_1,x_2,\cdots,x_g\} \subset R^{g\times c}$;遥感图像建筑物分割结果为$I_s=\{y_1,y_2,\cdots,y_g\} \subset R^{g\times 2}$。其分割流程如图1所示。

建筑物分割的实质是建立一个从遥感图像I到分割结果I_s之间的端对端的(End-to-End)映射关系$f(\cdot)$。由于输入图像I的维度高、背景复杂且建筑物外形结构复杂多变,一般情况下$f(\cdot)$是一个高维非线性模型,难以一次性建立$I_s=f(I)$的映射关系,$f:g\times c\rightarrow g\times 2$。因此,将该过程分为两个步骤:首先提取特征,建立特征空间;然后在特征空间的基础上构建合理的分割映射模型$f_H(\cdot)$,完成建筑物的分割。

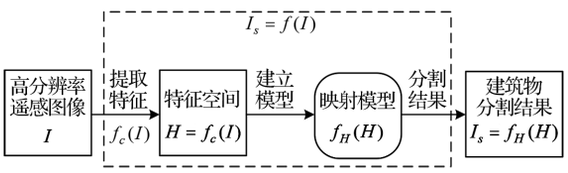

图1 建筑物分割流程图

Fig.1 Flow diagram of building segmentation

设建立分割模型的特征空间为$H=h_1,h_2,\cdots,h_n\subset R^{n\times m}$,则从I到H之间同样存在一个映射关系$f_c(\cdot)$,$f_c:g\times c\rightarrow n\times m$,完成图像的特征提取,且有$H=f_c(I)$。显然从特征空间到分割结果之间的映射模型为$f_H(\cdot)$,$f_H:n\times m\rightarrow g\times 2$,实现建筑物分割$I_s=f_H(H)$。

随着遥感图像分辨率的不断提高,建立图像特征空间与建筑物之间的映射模型存在两个挑战性问题:

(1)特征空间的维度大幅提升。对于同一建筑物而言,遥感图像分辨率的提高意味着表示该目标的像素必然增加。若$I\subset R^{g\times c}$仅表示建筑物,则g一定随分辨率提高而增大。这必然使得遥感图像到分割结果之间的映射关系$I_s=f(I)$呈现高维特性。

同时,分辨率的提升,使得遥感图像的结构、纹理、光谱等信息更加丰富,可被用于有效表征建筑物的特征数量必然增多,即遥感图像特征空间H的维度大幅增加。这必然使得人工构建$f_c(\cdot)$和$f_H(\cdot)$更加困难。这也是传统遥感图像分割方法分割精度不高、泛化能力低,甚至失效的主要原因之一。

(2)映射模型的非线性关系更加复杂。在低分辨率遥感图像中,一个建筑物可能只有几个或者几十个像素点表征,可通过特定的光谱信息(或灰度信息)直接建立简单的分割模型$f_H(\cdot)$。而在高分辨率遥感图像中,特征空间H维度增大,特征变量(如灰度、信息熵、边缘信息等)之间往往相互耦合、特征变量与分割目标之间呈现强非线性关系。这必然导致手工建立的分割模型$f_H(\cdot)$无法准确描述图像与目标之间的关系。

3 本文算法架构及实施技术

3.1 算法总体框架

深度学习是通过大量样本的训练,使被训深度神经网络逼近真实模型$I_s=f(I)$,无需中间过程,从而可实现从输入图像到分割结果间端对端的任务模式。本文提出的高分辨率遥感图像建筑物分割深度学习框架如图2所示。

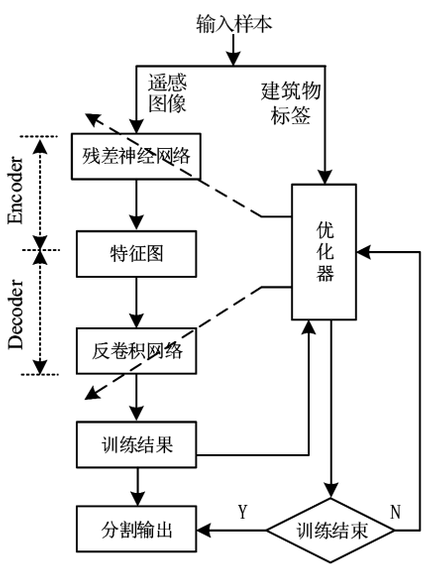

图2 遥感图像建筑物分割深度学习框架

Fig.2 Deep learning framework of building segmentation in remote sensing image

每个批次输入样本包含遥感图像和建筑物标签两个部分。遥感图像经过深度残差神经网络得到特征图(由此扩张成特征空间H),实现建筑物特征的编码功能(Encoder)。然后采用反卷积构成解码器,通过对特征图的上采样完成建筑物分割,并将输出尺寸还原至遥感图像的原始尺寸。该过程实现解码的功能(Decoder)。解码结果与建筑物标签一起送到优化器,采用随机梯度下降法对残差神经网络和反卷积网络中的权值进行优化训练。当训练结束后,固定神经网络中的权值不变,此时的输出结果即为建筑物的分割结果。

残差神经网络[28]是为降低深度神经网络训练难度而提出的一种网络结构。其训练对象不再是真实模型$R(x)$,而是真实模型与输入样本之间的差$R(x)-x$。残差神经网络的框架结构如图3(b)所示。其中:被训模型为残差模型$F(x)=R(x)-x$,在模型输出端引入样本的前馈通道构成闭环,使得$y=$

$F(x)+x=R(x)$,则最终输出仍然为真实模型。在该框架下,神经网络的权值收敛更加有效[28]。

3.2　基于残差深度神经网络构建的Encoder结构

设样本输入为x,训练输出为$y=R(x)$,常规的卷积神经网络如图3(a)所示,训练结果即为真实模型的直接逼近。为了使输出模型更加逼近真实模型,深度学习网络通常设计得很深,以此来获得高维和强非线性映射。但这常常使得网络训练困难,从而导致预测精度降低甚至训练失败。为此,He等[28]提出一种残差神经网络,能较好地逼近系统的真实模型,其原理如图3所示。

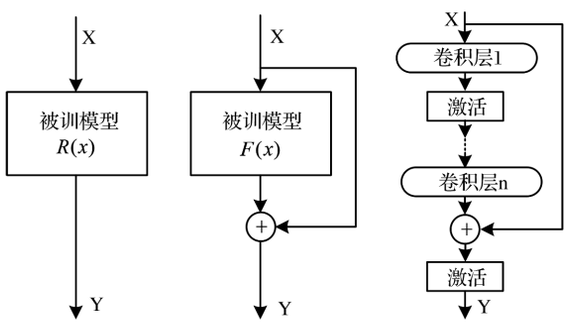

图3　深度学习框架

Fig.3　Deep learning framework

单层残差神经网络的实施框架如图3(c)所示。前端卷积层之后为激活层,最后一个卷积层输出与前馈信号叠加后激活输出。在前馈通道闭环中,可以包含多个卷积层和激活层,本文采用典型的2层卷积形式。

在文献[28]核心思想的指引下,本文借用其深度残差神经网络为基础,构建Encoder结构,以自动提取建筑物分割所需的特征。然而,典型残差神经网络的层次一般设计的较深,通常有50层、101层、200层甚至更多[28],网络结构越深,势必增加计算量。经过大量的实验,在保证分割精度的前提下,设计了一种31层结构的残差神经网络,以降低计算量。设计的深度残差网络共分为5种卷积类型,每种卷积层的数量分别设置为1、6、8、8、8。该卷积层数量既能较为准确地提取建筑物分割所需的特征,又能明显降低计算量。除了第1类卷积层以外,其余卷积类的区别主要在于卷积核的数量,如图4所示。

其中,K表示卷积核大小,S表示卷积步幅,C表示卷积输出通道数。每个卷积层均包含卷积、Relu激活[31]和批量规范化[27]处理,卷积过程中的补零操作(Padding)选用“SAME”模式。

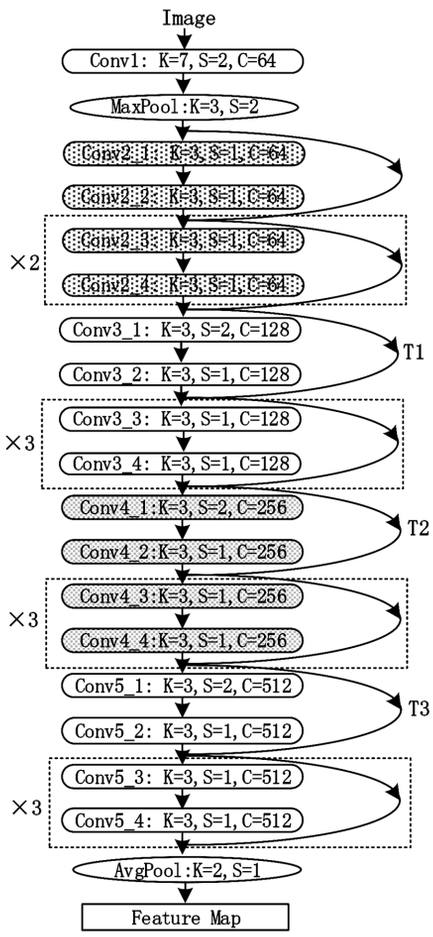

图4　残差神经网络结构图

Fig.4　Architecture of ResNet neural network

为了进一步降低计算量,采用文献[28]中前馈闭环的改进方式,将前馈闭环中卷积核大小为3的两层卷积替换成三层卷积,卷积核大小分别是1、3和1,同时通过增加输入和输出通道数量来确保逼近精度。以Conv2_2和Conv2_3为例,替换方式如图5所示。

当残差神经网络中卷积层的输出通道发生变化时,如Conv2_4到Conv3_1的输出通道数量从64变成了128,前馈通道的输入和输出不一致,如图4所示T1的输入通道为64,输出通道为128[28]。因此,前馈通道与尾部卷积输出不能直接相加并激活,需要在前馈的卷积操作时对输出通道进行扩容。以T1层为例,其前馈闭环及通道扩容结构替换如图6所示。类似的结构替换还包括T2和T3。

对比图5和图6两类替换后的结构,图5中前馈卷积操作的输入和输出通道数量均为256。图6中前馈卷积操作的输入通道为256,但是输出通道为

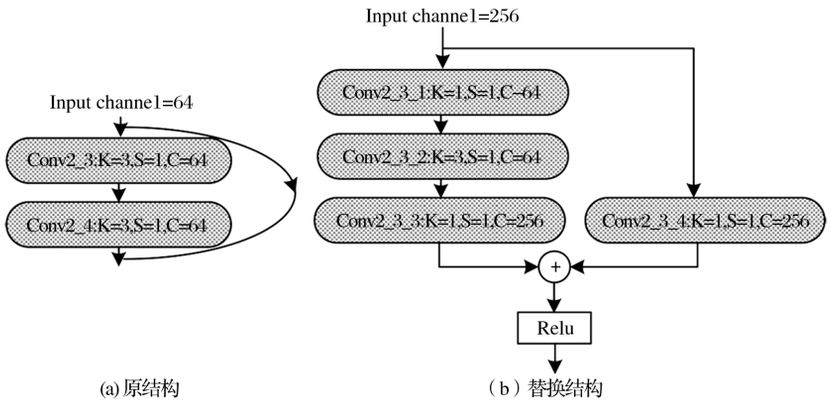

图 5 前馈闭环结构替换方法

Fig.5 Substitute structure of feed-forward loop

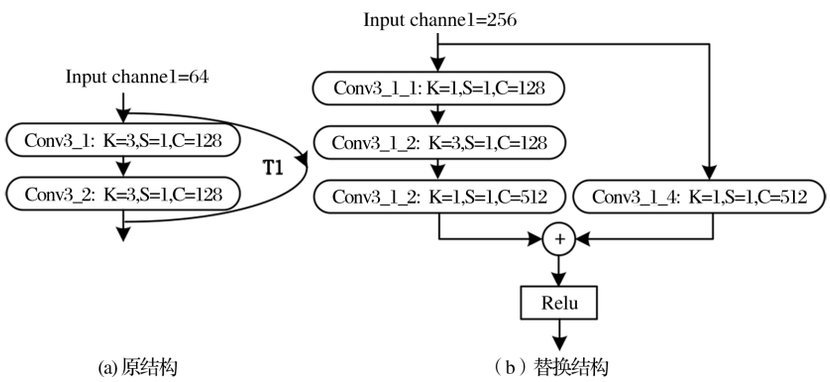

图 6 前馈闭环及通道扩容替换方法

Fig. 6 Substitute structure of feed-forward loop with expanded channel

512,通道扩容后与Conv3_1_3的输出通道相同。

3.3 反卷积构建的Decoder结构

遥感图像通过深度残差神经网络提取得到特征图,实现了编码的功能。在3×3的图像上采用1个2×2的卷积核,步幅S=1的卷积过程如图7所示。

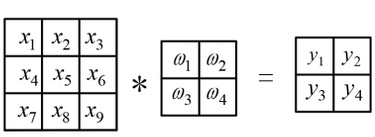

图 7 图像卷积过程

Fig.7 Image convolution processing

输入图像为向量$X=x_1,x_2,\cdots,x_9^{\mathrm{T}}$,输出特征图$Y=y_1,y_2,y_3,y_4^{\mathrm{T}}$,卷积过程可表示为:

$$CX=Y \tag{1}$$

$$C=\begin{pmatrix}\omega_1&\omega_2&0&\omega_3&\omega_4&0&0&0&0\\0&\omega_1&\omega_2&0&\omega_3&\omega_4&0&0&0\\0&0&0&\omega_1&\omega_2&0&\omega_3&\omega_4&0\\0&0&0&0&\omega_1&\omega_2&0&\omega_3&\omega_4\end{pmatrix}$$

深度神经网络中,反卷积的过程是卷积的逆过程[32]。因此,式(1)的反卷积可看作是从Y到X的传播过程。设输出损失函数为Q,根据BP(Back Propagation)算法求卷积的反向传播,有:

$$\frac{\partial Q}{\partial X}=\left(\frac{\partial Q}{\partial x_1}\ \cdots\ \frac{\partial Q}{\partial x_9}\right)^{\mathrm{T}} \tag{2}$$

根据式(1)有:

$$\begin{aligned}\frac{\partial Q}{\partial x_i}&=\sum_{j=1}^{4}\frac{\partial Q}{\partial y_j}\ \frac{\partial y_j}{\partial x_i}\\&=C_{1i}\frac{\partial Q}{\partial y_1}+C_{2i}\frac{\partial Q}{\partial y_2}+C_{3i}\frac{\partial Q}{\partial y_3}+C_{4i}\frac{\partial Q}{\partial y_4}\\&=C_{all,i}{}^{\mathrm{T}}\frac{\partial Q}{\partial y}\end{aligned} \tag{3}$$

其中:C_{ji}表示矩阵C中第j行、第i列的元素,$C_{all,i}{}^{\mathrm{T}}=C_{1,i}\ \cdots\ C_{4,i}$,,所以有:

$$\frac{\partial Q}{\partial Y}=\left(\frac{\partial Q}{\partial y_1}\ \cdots\ \frac{\partial Q}{\partial y_4}\right)^{\mathrm{T}}\ \frac{\partial Q}{\partial X}=C^{\mathrm{T}}\frac{\partial Q}{\partial Y} \tag{4}$$

根据式(4)可知,反卷积实质是对输入左乘C^{T},因此反卷积也被称为转置卷积[32]。

采用反卷积实现Decoder功能。Decoder不仅需要通过提取的特征实现建筑物的分割,同时也需要将分割结果复原为输入图像的原始大小。因此,反卷积的信息来源不能仅限于Encoder的特征输

出。在确保分割精度的基础上,为了降低计算量,经过大量实验,将Decoder的信息来源确定为Encoder的输出特征图和Conv4_4的输出结果,并将两个反卷积融合,实现建筑物的分割。反卷积实现Decoder功能的实施方案如图8所示。

首先对Encoder的输出特征图进行反卷积,设输出通道为2(对应于2个分割种类);输出的尺寸对应于Conv4_4卷积层的特征输出尺寸。在融合Conv4_4特征输出之前,引入一个卷积核K=1,S=1,C=2的卷积操作,将特征图的通道都变为2。在融合这两个特征信息之后再次进行反卷积,设置输出通道为2,输出尺寸为遥感图像原始尺寸。此时输出的规格为[g_h g_w 2],即两个与输入图像尺寸相同的矩阵,若在此基础上引入softmax操作,两个矩阵分别表示每个像素点属于建筑物和背景的概率。当训练结束后,对输出结果的2个通道进行Argmax运算,得到建筑物的分割结果。

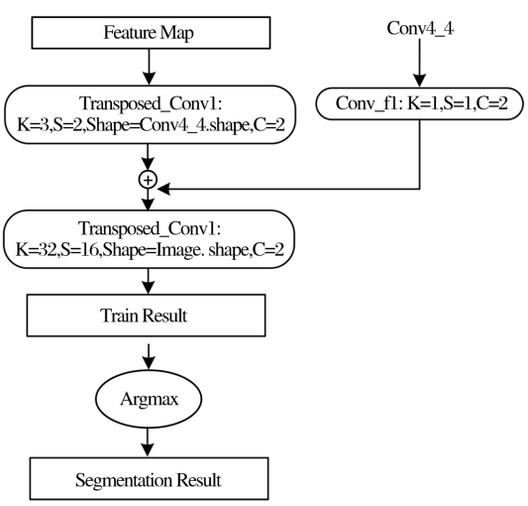

图8　Decoder的反卷积实现

Fig. 8　Deconvolution for Decoder

3.4　批量规范化

深度神经网络训练的过程是通过误差反向传播,采用随机梯度下降法等优化方法使权值收敛到最优值。随着神经网络层数的不断增加,可能引发梯度弥散或爆炸、过拟合以及权值震荡等问题,从而导致神经网络的训练更加困难。1.3节中所采用的残差神经网络不仅能更加精确地逼近真实模型,而且在一定程度上能抑制梯度弥散或爆炸。针对训练模型的过拟合问题,通常是采用Dropout技术,在每次训练过程中随机挑选一定比例的权值不参与此次训练,从而降低过拟合。针对权值震荡问题,主要是通过weight decay技术,使梯度下降过程中的步长逐渐衰减,以精确逼近权重的最优值。

在训练神经网络的过程中,数据流逐层传递,低层网络权值的变化必然引起输出数据的分布发生变化,这也是引发深度神经网络训练困难的原因之一。在每个卷积操作输入前,引入批量规范化(Batch Normalization)[27]技术对数据进行规范化处理,替代了Dropout、weight decay等技术,降低了神经网络的训练难度。

设所有样本为$X=\{x_1,x_2,\cdots,x_N\}$,若每次训练前对所有样本进行规范化处理。

$$\hat{x}_i=\frac{x_i-E(X)}{\sqrt{Var(x_i)}} \tag{5}$$

其中:$E(\cdot)$表示均值,$Var(\cdot)$表示方差。则每一个卷积操作过程中,所有的训练样本都服从同一分布。然而,对于深度神经网络庞大的训练样本而言,求取所有样本的均值和方差计算量十分巨大。因此,采用批量规范化操作代替全局均值和方差。即计算每个训练批次样本的均值和方差,再对所有均值和方差取平均值,并以此替代全局均值和方差。设每次训练样本的数量为m,批量规范化的实现过程如算法1所示(表1)。

表1　批量规范化

Table 1　Batch normalization

步骤	操作
输入	获取当前批次训练样本: $X=\{x_1,x_2,\cdots,x_m\}$,容量m
1	计算均值:$\mu_k=\frac{1}{m}\sum_{i=1}^{m}x_i$
2	计算方差:$\sigma_k^2=\frac{1}{m}\sum_{i=1}^{m}x_i-\mu_k^2$
3	样本规范化:$\hat{x}_i=\frac{x_i-\mu}{\sqrt{\sigma_k^2+\varepsilon}}$,$\varepsilon$为小数值正常数
4	尺度变换:$y_i=\gamma\hat{x}_i+\beta$,$\gamma$、$\beta$为可学习参数
5	规范化输出:$BN(x_i)=y_i$

算法1的第4步是对规范化的信号进行重构,其中γ、β是需要学习的参数。设输出的损失值为Q,参数γ、β的初始值分别为1和0,根据BP算法和链式法则,γ、β变化率如下所示。

$$\frac{\partial Q}{\partial\hat{x}_i}=\frac{\partial Q}{\partial y_i}\cdot\gamma$$

$$\frac{\partial Q}{\partial\sigma_i^2}=-\frac{1}{2}\sum_{i=1}^{m}\frac{\partial Q}{\partial\hat{x}_i}\cdot xi-\mu k\cdot\sigma_k^2+\varepsilon^{-2/3}$$

$$\frac{\partial Q}{\partial \mu_k}=\sum_{i=1}^{m}\frac{\partial Q}{\partial \hat{x}_i}\cdot\frac{-1}{\sqrt{\sigma_k^2+\varepsilon}}$$

$$\frac{\partial Q}{\partial x_i}=\frac{\partial Q}{\partial \hat{x}_i}\cdot\frac{-1}{\sqrt{\sigma_k^2+\varepsilon}}+2\frac{\partial Q}{\partial \sigma_k^2}\cdot\frac{x_i-\mu_k}{m}+\frac{\partial Q}{\partial \mu_k}\cdot\frac{1}{m}$$

$$\frac{\partial Q}{\partial \gamma}=\sum_{i=1}^{m}\frac{\partial Q}{\partial y_i}\cdot\hat{x}_i$$

$$\frac{\partial Q}{\partial \beta}=\sum_{i=1}^{m}\frac{\partial Q}{\partial y_i} \tag{6}$$

4 实验及结果分析

4.1 数据资源及实验平台

以IAILD(Inria Aerial Image Labeling Dataset)遥感图像数据库[29]为对象,展开验证实验研究。IAILD遥感图像数据库提供地表面积为810 km^2,分辨率为0.3 m的遥感图像,其中405 km^{2}已经对建筑物做过精确标记,可作为训练样本。每个样本像素为5 000×5 000,共计180个。由于受GPU存储单元的限制,将每个样本裁剪成1 000×1 000像素大小,则共有样本4 500个,其中4 480个作为训练样本,20个作为测试样本。样本中的遥感图像为RGB 3通道图像,张量表示形式为[N, 1 000, 1 000, 3];标记图像为灰度图像,张量表示形式为[N, 1 000, 1 000],N为批处理时样本的个数。

实验平台搭载Intel-i7-7700K四核CPU处理器、32G内存、ASUS STRIX-GTX1080TI-11G显卡(GPU处理单元),深度学习框架采用Google公司的TensorFlow。

4.2 数据预处理及结果评价指标

在训练神经网络之前,对样本数据进行预处理:

(1)求取所有遥感图像的RGB(Red-Green-Blue)3个通道均值;

(2)每一幅遥感图像的RGB通道减去上述均值;

(3)将标记图像转换成2通道张量,[N,1 000,1 000,2],其中第1通道中建筑物位置为1,其余位置为0,第2通道数值相反。

为了量化评价分割结果,采用召回率(Recall Rate)、精确率(Precision Rate)和F值(F-measure)来评价分割结果,其计算方式如下所示。

$$\text{Recall}=\frac{B_{seg}}{B_{seg}+I_{unseg}}$$

$$\text{Precision}=\frac{B_{seg}}{B_{seg}+I_{wseg}}$$

$$\text{F-measure}=\frac{2\text{Recall}\times\text{Precision}}{\text{Recall}+\text{Precision}} \tag{7}$$

其中:B_{seg}表示分割结果中建筑物分割正确的像素点数,I_{unseg}表示图像中是建筑物但未被分割为建筑物的像素点数,I_{wseg}表示图像中误将背景分割为建筑物的像素点数。

召回率表示被分割为建筑物的像素点与真实建筑物像素点的比率。表征在不考虑遥感图像背景的情况下,建筑物分割的准确度。精确率表示被正确分割为建筑物的像素点与所有分割为建筑物的像素点的比率。精确率高表示能够将建筑物提取出来。F值是综合召回率和精确率这两项指标的评估指标,是用于综合反映整体的指标。

综合Vakalopoulou等[17-22]采用的深度学习框架,具有典型代表意义的是VGG全卷积神经网络、VGG与全连接条件随机场相结合的网络结构两种。为了验证本文提出算法的对遥感图像建筑物分割的有效性,在IAILD数据库上开展了与VGG全卷积神经网络(VGG)、VGG条件随机场网络(VGGCRF)的对比实验。为了表述简单,本文提出的算法采用ResNet表示。实验中,VGG采用文献[33]所示的结构,并且前13层神经网络的卷积核参数调用已经训练好的数值,全连接层F6、F7、F8的卷积核分别设置为[16, 16, 512, 1 024]、[1,1,1 024, 2 048]、[1,1,2 048,2],最后融合多层反卷积结果实现建筑物分割结果输出。VGGCRF是在VGG的最后引入全连接条件随机场,具体结构见文献[22]。

4.3 对比实验及结果分析

深度神经网络的训练过程是通过对样本数据的学习,使得深度神经网络权值收敛的过程。3种网络结构均采用交叉熵作为训练的损失函数,训练过程如图9所示。由于残差神经网络的特殊结构和批量规范化技术使得神经网络权值更加容易训练,权重的收敛性能也更好。

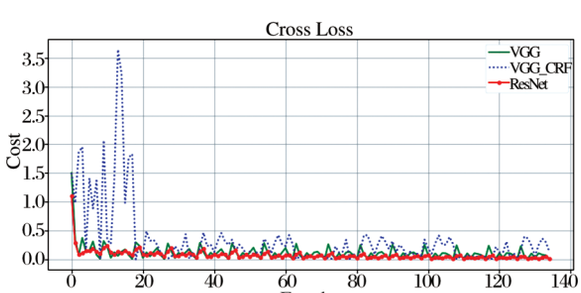

图9 深度神经网络训练误差

Fig.9 Train error of deep neural network

需要注意的是,即使在使用了批量规范化技术的情况下,VGGCRF 的训练仍然较为困难。经过反复试验,笔者认为 VGGCRF 可采用两个步骤进行训练,可使网络的权值收敛。首先直接采用 VGG 网络训练,待建筑物出现分割迹象后,再引入全连接条件随机场一起参与训练,直到满足训练结束条件。三种网络的训练耗时如表 2 所示。

表 2　深度神经网络训练耗时对比

Table 2　Time for training the deep neural networks

时间	VGG	VGGCRF	ResNet
网络调用时间/s	25.26	31.23	137.33
单个样本时间/s	0.69	5.11	0.23
样本集一次训练时间/s	3 328.94	23 222.35	1 418.86

由于 ResNet 网络结构比 VGG、VGGCRF 复杂,因此网络调用时间最长,但是单个样本的训练时间最短,则对应于样本集一次训练时间最短。VGGCRF 中引入了全连接条件随机场,需要迭代计算所有像素对之间的成对势函数的值,因此训练耗时最长。

4.4　对比实验及结果分析

4.4.1　建筑物分割对比实验

实验选择包含复杂道路影响的建筑物、单体复杂建筑物和规律性分布建筑物等三种典型的遥感图像建筑物,采用上述深度神经网络框架进行分割实验,并对实验结果对比分析。其中,图 10 所示的遥感图像中包含较为复杂的道路。

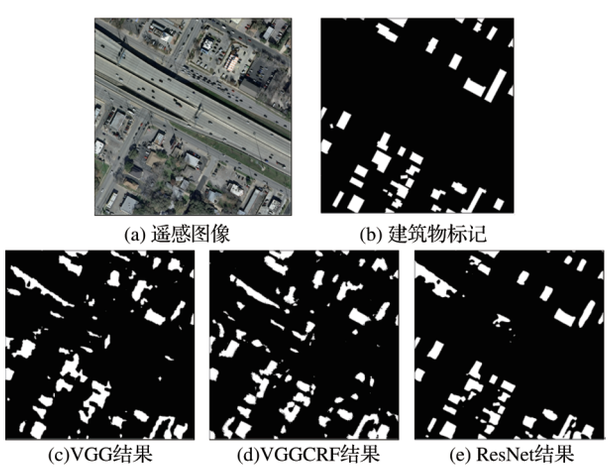

图 10　包含复杂道路的建筑物分割结果

Fig.10　Building segmentation result with intricate road

分析三种深度神经网络分割结果可知,三种网络均能分割建筑物,其中 VGG 能大致将建筑物确定在一定的范围内,但建筑物边缘信息准确性不高。VGGCRF 对建筑物边缘的提取较 VGG 有所提升。ResNet 能较好地提取建筑物的块状信息和边缘特征。同时三种网络对立交桥和阴影都出现一定程度的误分。其中 VGG 以图像轮廓信息为主导,误分结果表现为条状信息;VGGCRF 在 VGG 基础上缓解了误分的发生;而 ResNet 提取的立交桥的片状信息。

分割性能评价指标如图 11 所示,可见 ResNet 能较大幅度提高这类遥感图像中建筑的召回率、精确率和 F 值。

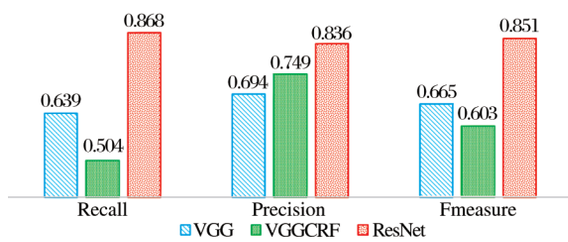

图 11　包含复杂道路的建筑物分割性能指标

Fig.11　Performance index of building segmentation with intricate road

图 12 所示遥感图像中建筑物分布较为规律,建筑物分割的干扰主要来自于植物的影响。ResNet 的网络结构对建筑物轮廓的检测更加准确,对较小的建筑物仍然能够实现有效分割。根据图 13 所示的分割结果评价指标可知,ResNet 的召回率、精确率和 F 值均有较大幅度提高。

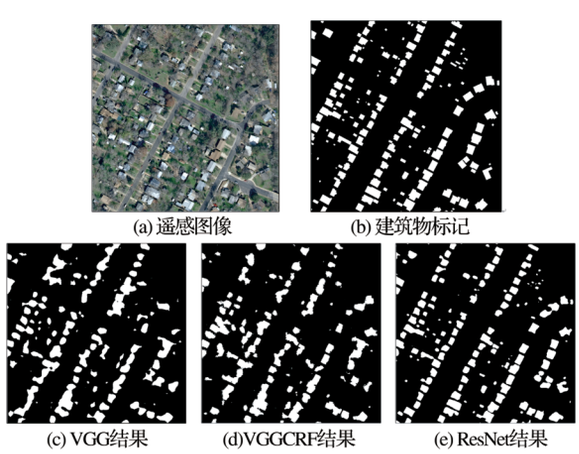

图 12　规律性建筑物分割结果

Fig.12　Ordered Building segmentation result

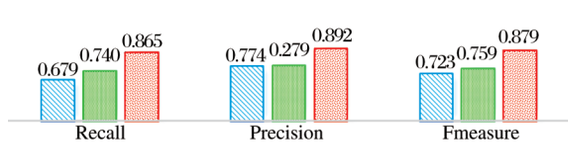

图 13　规律性建筑物分割性能指标

Fig.13　Performance index of ordered building segmentation

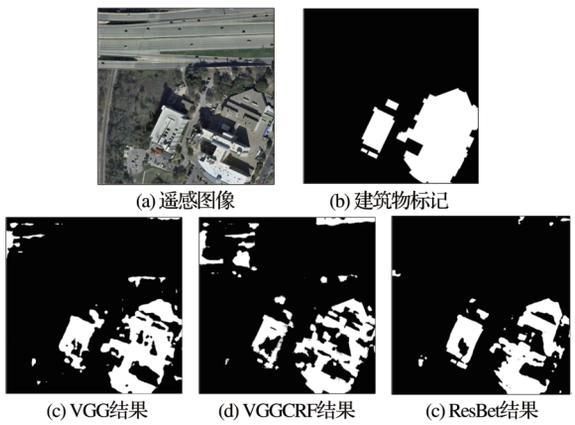

图 14 单体复杂建筑物分割结果

Fig.14 Complicated single building segmentation result

图 14 遥感图像中为单体建筑物，建筑物顶层有错层结构和大面积的阴影。此外，部分建筑物的色彩信息与其上方的道路近似，因此道路的影响更加明显。由分割结果可知，3 种网络结构分割结果都出现了较大面积的误分。然而 ResNet 成功避开了左上角立交桥的影响，且对建筑物的边缘检测更为准确。但是相对于 VGG 而言 ResNet 对阴影干扰的鲁棒性不强，将大面的错层阴影误分为了背景，从而导致分割精确率较低，且 F 值不高。

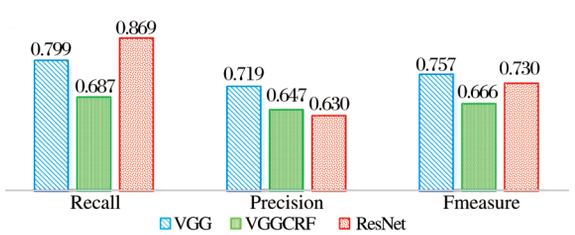

图 15 单体复杂建筑物分割性能指标

Fig. 15 Performance index of complicated single building segmentation

上述 3 种网络对建筑物的分割细节如图 16 所示。对比 3 种不同网络的分割结果可知，VGG 能大致指示建筑物的范围，同时也能较好的提取受干扰较小的建筑物边缘；VGGCRF 对直线型且无干扰的建筑物边缘提取十分准确，但是对于其他类型边缘存在明显的散点式误分。ResNet 对建筑物的边缘提取较为准确，且分割结果成块状形态，相对于 VGG 和 VGGCRF 而言，对建筑物的分割更为精确。

4.4.2 多分辨率遥感图像建筑物分割实验

采用的 IAILD 遥感图像数据库分辨率为 0.3m。为了检测遥感图像分辨率对本文算法的影响，采用双三次插值法将原始遥感图像进行压缩，压缩比例

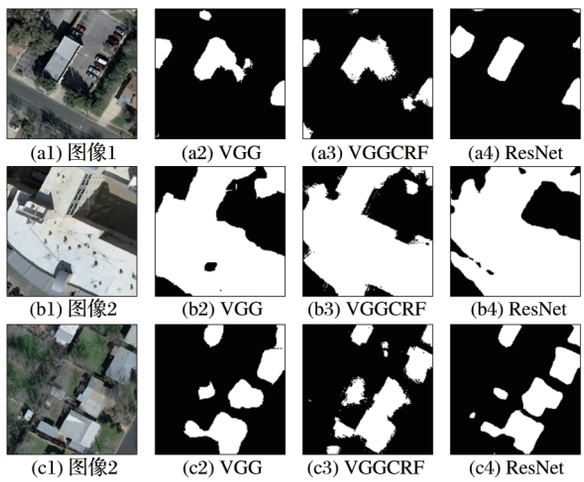

图 16 建筑物分割细节

Fig.16 Details of building segmentation

分别为 R=0.2、0.5 和 0.8，则对应的分辨率分别近似为 0.375、0.6 和 1.5 m。实验验证时，建筑物标签也做相应的压缩。实验结果如图 17 所示，其中(b1)遥感图像为(a1)中白色方框所示的部分；(c1)遥感图像为(b1)中白色方框内部分。分割性能指标如图 18 所示。

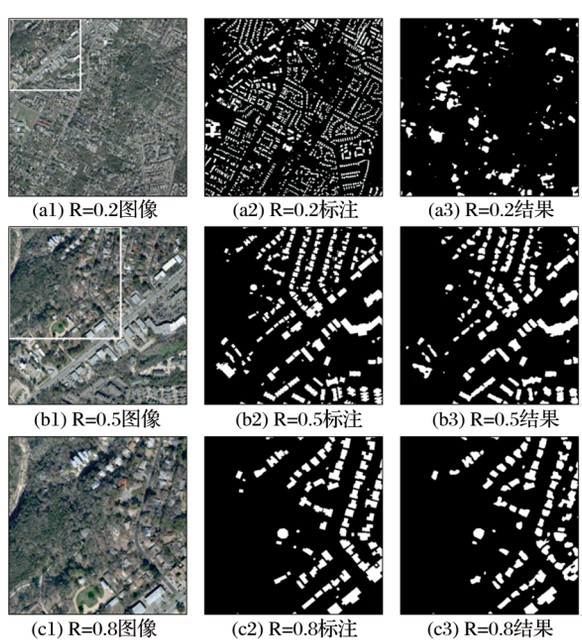

图 17 多分辨率感图像建筑物分割细节

Fig.17 Building segmentation in remote sensing image with multi-resolution

对遥感图像进行压缩使得表示建筑物的像素点减少，必然导致获取的建筑物的信息减小。当压缩比例低至 0.5 时，本文算法仍然能对建筑物实现准确分割，召回率、精确度和 F 值均较高。但当压缩

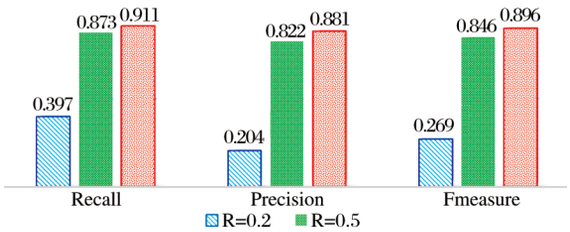

图 18　多尺度遥感图像建筑物分割性能指标

Fig. 18　Performance index of building segmentation in remote sensing image with multi scale

比例达到 0.2 时,本文提出分割算法对建筑物的分割精准度严重降低。

5　结　语

本文针对高分辨率遥感图像中建筑物的自动精确分割问题提出以残差深度神经网络为基础,构造 Encoder-Decoder 特征提取与分割复原输出的框架,并运用批量规范化技术手段,在 IAILD 遥感数据库上开展实验验证。实验结果表明,本文提出的算法计算量较小,对单一样本训练时间为 0.23 s,样本集一次训练时间为 1 418.862 s;在对邻近复杂道路的建筑物、规律性建筑物、单体复杂建筑物等 3 种典型建筑的分割实验结果中,分割精度分别达到了 0.837、0.892 和 0.630;F 值分别为:0.851、0.879 和 0.730。与 VGG 全卷积网络和 VGG 全连接条件随机场网络相比,本文提出的算法能有效避免复杂道路的影响,对建筑物的块状特征和边缘信息的提取更为准确,能获取更加精准的建筑物分割结果。此外,在多分辨率遥感图像分割中,对分辨率压缩比为 0.2、0.5 和 0.8 的 3 种不同分辨率的遥感图像分割精度能达到 0.873、0.822 和 0.846;F 值能分别达到 0.911、0.881 和 0.896,表明本文算法对一定范围内的多分辨率遥感图像具有较好的泛化能力。

然而,对于楼顶结构存在错层和大面积阴影等干扰的复杂建筑物,本文提出的算法仍然存在边缘检测错误、分割精度不高的问题。在后续工作中,将重点研究如何消除复杂建筑物中错层及阴影的干扰,进一步提高建筑物的分割精准度。

参考文献(References):

[1] Chen Jie , Deng Min, Xiao Pengfeng, *et al*. Object-oriented Classification of High Resolution Imagery based on Watershed Transform and Sptial Clustering[J]. Remote Sensing Technology and Application, 2010, 25(5):597-603.[陈杰,邓敏,肖鹏峰,等. 基于分水岭变换与空间聚类的高分辨率遥感影像面向对象分类[J]. 遥感技术与应用, 2010, 25(5):597-603.]

[2] Wang Yu, Wang Baoshan, Wang Tian, *et al*. Image Entropy Active Contour Models towards Water Area Segmentation in Remote Sensing Image[J]. Optics and Precision Engineering, 2018[王宇,王宝山,王田,等. 面向遥感图像水域分割的图像熵主动轮廓模型[J]. 光学精密工程,2018, 26(3):698-707.]

[3] Wu Jun, Wang Yuanyuan, Chen Yue, *et al*. Speckle Reduction of Ultrasound Images with Anisotropic Diffusion based on Homogeneous Region Automatic Selection [J]. Optics and Precision Engineering, 2014, 22(5):1312-1321.[吴俊,汪源源,陈悦,等. 基于同质区域自动选取的各向异性扩散超声图像去噪[J]. 光学精密工程,2014, 22(5):1312-1321.]

[4] Chen Kuntang, Dong Xiaolong, Xu Xing' ou, *et al*. The Study on Oceanic Vector Wind Field Retrieve Technique based on Neural Networks of Microwave Scatterometer [J]. Remote Sensing Technology and Application, 2017, 32 (4): 683-690.[陈坤堂,董晓龙,徐星欧,等. 微波散射计反演海面风场的神经网络方法研究[J]. 遥感技术与应用, 2017, 32 (4): 683-690.]

[5] Wang Yu, Li Yu, Zhao Quanhua. Region-based Multiscale Segmentation of Panchromatic Remote Sensing Image [J]. Control and Decision, 2018, 33(3): 535-541.[王玉,李玉,赵泉华. 基于区域的多尺度全色遥感图像分割[J]. 控制与决策, 2018, 33(3): 535-541.]

[6] Hinton G E, Salakhutdinov R R. Reducing the Dimensionality of Data with Neural Networks [J]. Science, 2006, 313 (5786): 504-507.

[7] Zeng N Y, Zhang H, Song B Y, *et al*. Facial Expression Recognition via Learning Deep Sparse Autoencoders[J]. Neurocomputing, 2018, 273(17): 643-649.

[8] Xu X Y, Pan J S, Zhang Y J, *et al*. Motion Blur Kernel Estimation via Deep Learning[J]. IEEE Transactions on Image Processing, 2018, 27(1): 194-205.

[9] Shao H D, Jiang H K, Zhang H Z, *et al*. Rolling Bearing Fault Feature Learning Using Improved Convolutional Deep Belief Network with Compressed Sensing[J]. Mechanical Systems and Signal Processing, 2018, 100: 743-765.

[10] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation[C]// IEEE Transactions on Pattern Analysis & Machine Intelligence,2015:3431-3440.

[11] Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (4): 640-651.

[12] Simonyan K, Zisserman A. Visual Geometry Group [EB/OL]. http: //www. robots. ox. ac. uk/~vgg/research/very_deep/,2014.

[13] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-decoder Architecture for Scene Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.

[14] Yu F, Koltun V, Multi-scale Context Aggregation by Dilated Convolutions[C]// International Conference on Learning Rep-

resentations,2016.

[15] Chen L C, Papandreou G, Kokkinos I, *et al*. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs[C]//International Conference on Learning Representations, 2015.

[16] Deeplab-public [EB/OL], https://bitbucket.org/deeplab/deeplab-public/,2017,2018.

[17] Vakalopoulou M, Karantzalos K, Komodakis N, *et al*. Building Detection in very High Resolution Multispectral Data with Deep Learning Features[C]//IEEE. Geoscience & Remote Sensing Symposium, 2015:1873-1876.

[18] Huang Z M, Cheng G L, Wang H Z, *et al*. Building Extraction from Multi-source Remote Sensing Images via Deep Deconvolution Neural Networks[C]//IEEE. Geoscience and Remote Sensing Symposium, 2016:1835-1838.

[19] Saito S, Aoki Y. Building and Road Detection from Large Aerial Imagery[C]//Image Processing: Machine Vision Applications VIII, 2015:1814-1821.

[20] Yuan J. Learning Building Extraction in Aerial Scenes with Convolutional Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 40(11): 2793-2798.

[21] Bittner K, Cui S Y, Reinartz P. Building Extraction from Remote Sensing Data Using Fully Convolutional Networks[C]//ISPRS Hannover Workshop: Hrigi, 2017:481-486.

[22] Bischke B, Helber P, Folz J, *et al*. Multi-Task Learning for Segmentation of Building Footprints with Deep Neural Networks[EB/OL].https://arxiv.org/abs/1709.05932,2017,2017.

[23] Wang Y, Wang C, Zhang H. Integrating H-A-α with Fully Convolutional Networks for Fully PolSAR Classification[C]//IEEE. International Workshop on Remote Sensing with Intelligent Processing, 2017:1-4.

[24] Alshehhi R, Marpu P R, Woon W L, *et al*. Simultaneous Extraction of Roads and Buildings in Remote Sensing Imagery with Convolutional Neural Networks[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2017, 130: 139-149.

[25] Lin H, Shi Z, Zou Z. Fully Convolutional Network With Task Partitioning for Inshore Ship Detection in Optical Remote Sensing Images[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(10): 1665-1669.

[26] Jiao L, Liang M, Chen H, *et al*. Deep Fully Convolutional Network-based Spatial Distribution Prediction for Hyperspectral Image Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(10): 5585-5599.

[27] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[EB/OL], https://arxiv.org/abs/1502.03167,2015,2018.

[28] He K M, Zhang X Y, Ren S Q, *et al*. Deep Residual Learning for Image Recognition [J]. https://arxiv.org/abs/1512.03385,2015.

[29] Nirvana Inria Aerial Image Labeling Dataset [EB/OL]. https://project.inria.fr/aerialimagelabeling/,2016,2017..

[30] Maggiori E, Tarabalka Y, Charpiat G, *et al*. Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark[C]//IEEE International Symposium on Geoscience and Remote Sensing, 2017:3226-3229.

[31] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[C]//Neural Information Processing Systems Conference, 2012:1097-1105

[32] Dumoulin V, Visin F. A Guide to Convolution Arithmetic for Deep Learning [EB/OL], https://arxiv.org/abs/1603.07285, 2016,2018.

[33] Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Trans Pattern Anal Mach Intell, 2017, 39(4): 640-651.

Building Segmentation in High Resolution Remote Sensing Image by Deep ResNet

Wang Yu[1,2], Yang Yi[3], Wang Baoshan[1], Wang Tian[4], Bu Xuhui[3], Wang Chuanyun[5]

(1.*School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo* 454000, *China*;
2.*Field Scientific Observation and Research base of Ministry of Land and Resources, Henan Polytechnic University, Jiaozuo* 454000, *China*;
3.*School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo* 454000, *China*;
4.*School of Automation Science and Electrical Engineering, Beihang University, Beijing* 100191, *China*;
5.*School of computer Science, Shenyang Aerospace University, Shenyang* 110136, *China*)

Abstract: This paper addresses the buildings segmentation in high resolution remote sensing image and proposes an Encoder-Decoder architecture of deep learning with End-to-End model, in which Encoder is based on ResNet, and the features needed by segmentation are exacted automatically, and the Decoder produces the segmentation result by deconvolution. Furthermore, in the training process, batch normalization is employed to decrease the gradient competition, so as to reduce the difficulty of training the deep neural network. The experiment results show that the algorithm effectively exacts the bulk feature and edge information of building in the high resolution remote sensing image, therefore the complex road disturbance is suppressed convincingly, and the building segmentation precision is improved effectively, the segmentation precision for three typical buildings, the building besides complex road, the ordered buildings and the complex single building, are 0.836 5, 0.892 4, and 0.629 7 respectively; and the F-measure are 0.851 4, 0.878 6 and 0.729 8, respectively. Meanwhile, the experiment results for multi-resolution remote sensing images show that the method can be generalized to the multi-resolution image within limits.

Key words: High resolution remote sensing image; Building segmentation; Deep learning; ResNet; Batch normalization