

融合特征和决策的卷积-反卷积图像分割模型

冯晨霄¹, 汪西莉^{1,2*}

¹陕西师范大学计算机科学学院, 陕西 西安 710119;

²陕西师范大学现代教育技术教育部重点实验室, 陕西 西安 710119

摘要 基于全卷积网络提出了一种图像分割模型以获取目标分割结果,模型包含两个结构相同的深层神经网络分支,每个分支采用卷积-反卷积的结构实现特征提取和从特征恢复目标区域;两个分支接收不同类型图像输入,将来源于两个分支的结果通过加权融合得到最终的分割结果。模型融合了不同图像源的多级尺度特征,在训练样本数有限的情况下,通过数据增强使训练得到的模型稳健性更强。在光学图像数据集 Weizmann horse 和遥感影像数据集 Vaihigen 上进行实验,并与相关文献进行比较,结果表明,所提模型具有更高的目标分割完整度和最优的分割性能,在训练数据有限、形态各异、尺度变化较大等的遥感影像建筑物提取中取得了理想的结果,表明该模型可应用于复杂的遥感影像目标分割。

关键词 图像处理; 图像分割; 深层卷积神经网络; 多尺度特征融合; 多源输入; 决策级融合

中图分类号 O436

文献标识码 A

doi: 10.3788/LOP56.011008

Convolution-Deconvolution Image Segmentation Model for Fusion Features and Decision

Feng Chenxiao¹, Wang Xili^{1,2*}

¹ School of Computer Science, Shaanxi Normal University, Xi'an, Shaanxi 710119, China;

² Key Laboratory of Modern Teaching Technology, Ministry of Education, Shaanxi Normal University, Xi'an, Shaanxi 710119, China

Abstract Based on the full convolutional network, an image segmentation model is proposed to obtain the target segmentation results. This model consists of two deep neural network branches with the same structures. As for each branch, a convolution-deconvolution structure is adopted to implement the feature extraction and to recover the target area from the features. These two branches receive different image inputs, and then the final segmentation results are obtained via the weighted fusion of the results from these two branches. This model combines the multi-level scale features of different image sources, and the training model is more robust through data enhancement when the number of training samples is limited. The experiments are carried out on the optical image dataset of Weizmann horse and the remote sensing image dataset of Vaihigen. The comparison with the related literatures is also made. The results show that the proposed model has a higher target segmentation integrity and an optimal segmentation performance. The ideal extraction results of remote sensing image buildings under the conditions of limited training data, various shapes, large scale changes and so on, indicate that the proposed model can be applied to the complex remote sensing image object segmentation.

Key words image processing; image segmentation; deep convolutional neural network; multi-scale feature fusion; multi-source input; decision-level fusion

OCIS codes 100.3008; 100.4996; 150.1830

收稿日期: 2018-08-08; 修回日期: 2018-09-12; 录用日期: 2018-09-18

基金项目: 国家自然科学基金(41471280,61401265,61701290,61701289)

* E-mail: wangxili@snnu.edu.cn

1 引言

2015年,Long等^[1]提出了全卷积网络(FCN)用于图像的语义分割,网络的输出结果为像素对应的类别的概率图,将传统的分类神经网络模型改为分割模型。而后,图像分割领域展示出新的生机与活力^[1-2]。目前采用深度卷积神经网络(CNN)进行图像分割的挑战主要有:1)卷积阶段的池化操作降低了特征分辨率,由其通过上采样进行预测结果比较粗略,难以精确保留目标区域边缘、位置等细节特征;2)图像和目标自身的复杂性,如遥感影像目标混杂、存在遮挡、阴影等,进一步增加了分割的困难性;3)研究表明,加深网络层次有利于提取特征,但将使参数增多,需要更多的标记样本来训练网络,而对于多数实际应用难以提供充足的训练样本。

现有研究针对上述问题,提出了一些解决手段。第一类研究方法:改进全卷积网络内部结构,如卷积结构和池化方式等改变感受野,提升特征分辨率以改善分割结果。Liu等^[2]将全局均值池化得到的全局特征与全卷积网络的局部特征相结合,来纠正局部分类错误。Wang等^[3]对上采样和膨胀卷积进行修改,提出密集向上采样卷积(DUC)来捕获和解码更详细的信息。第二类研究方法:采用识别性能更强且更深层次的卷积网络来提高分割性能,如Noh等^[4]提出DeconvNet,加深反卷积过程以获得更精细的分割结果。Ghiasi等^[5]提出了基于拉普拉斯金字塔的多分辨率重建结构,使用高分辨率特征映射和乘法门控的跳跃连接,提高分割边界精度。Zhao等^[6]以残差网络为基础提出了金字塔场景分割网络,得到了更好的上下文信息来提升分割效果。第三类研究方法:针对目标自身复杂性带来的困难,Dai等^[7]在CNN中引入了学习空间几何形变的能力,更好地解决了具有空间形变的图像识别任务。Zhang等^[8]提出比例自适应调节感受野的卷积方式来实现场景识别任务。第四类研究方法:特征融合方法,刘帆等^[9]在网络前端将彩色图像和深度图像融合再进行特征提取,李庆武等^[10]针对双目视觉图像分别进行处理并将结果融合,徐岩等^[11]分别使用不同大小的卷积核提取图像中不同尺度的特征,再将特征融合来重建无雾图像。

上述研究探索了深度学习在图像分割中的应用,但仍然存在一些不足。首先,上述文献以大型数据集为研究基础,小样本数据集应用研究少。其次,目前深度网络分割模型研究多考虑使用融合方法,

但是针对边缘特征和位置的保持、提升分割精度仍然存在不足。为此,本文提出融合特征和决策的卷积-反卷积(CD-FFD)图像分割模型,所提模型新颖性主要体现在三个方面:1)模型包括不同分支,从不同来源的图像中获取特征,如对于多光谱遥感影像,一个分支接收多光谱通道图像输入,另一个分支接收全色通道图像输入或高程信息等,构建并联的神经网络从不同源的输入提取特征,获取更多的、互补的信息,进而预测像素类别;2)每个分支构建深度卷积-反卷积网络提取特征,在特征层级进行多尺度特征融合;3)在两个分支网络末端使用类别概率图加权融合,在决策层级进行不同源决策的融合,充分融合两个神经网络分支的优势性能,使性能更优的网络分支在融合时发挥更大的作用。

2 CD-FFD 图像分割模型

CD-FFD图像分割系统采用并联结构,由数据预处理模块、输入模块、特征提取和类别预测模块、类别概率图融合模块、输出模块组成,如图1所示。

2.1 数据预处理模块

1) Weizmann Horse 数据集

Weizmann Horse数据集^[12]包含RGB彩色图像、二值标记图像、灰度图像各328幅,其中前200幅图像作为训练集,后128幅作为测试集。其样本量对于深度网络的训练远远不够,进行数据增强,方法为对训练集图像间隔 5° 进行旋转,间隔5 pixel对图像进行向右、向左移动,同时按照不同比例进行放大和缩小,得到RGB图像、灰度图像、二值标记图像各11973幅(包含11773幅训练图像+200幅验证图像)。在Weizmann Horse数据集中,RGB图像与二值标记图像、灰度图像尺寸不一致,为了使样本图像与标签图像对应起来,将所有图像归一化为同一尺寸(500×500)。扩充后数据集图像中的目标(马)具有不同的姿态、位置,光照、色彩、纹理、背景等信息也复杂各异,分割出目标具有挑战性。

2) vaihigen 数据集

vaihigen数据集^[13]包含33幅IRRG(红外、红、绿三通道)图像和数字表面模型(DSM)信息,图像尺寸不一,其中IRRG图像的三通道内容分别对应近红外、红色和绿色波段,DSM为单通道图像,对应于DSM高度,图像空间分辨率为9 cm。vaihigen数据集中仅有16幅图像包含标记图像,将其中11幅图像(1、3、5、7、13、17、21、23、26、32、37)作为训练集,5幅图像(11、15、28、30、34)作为测试集。训练

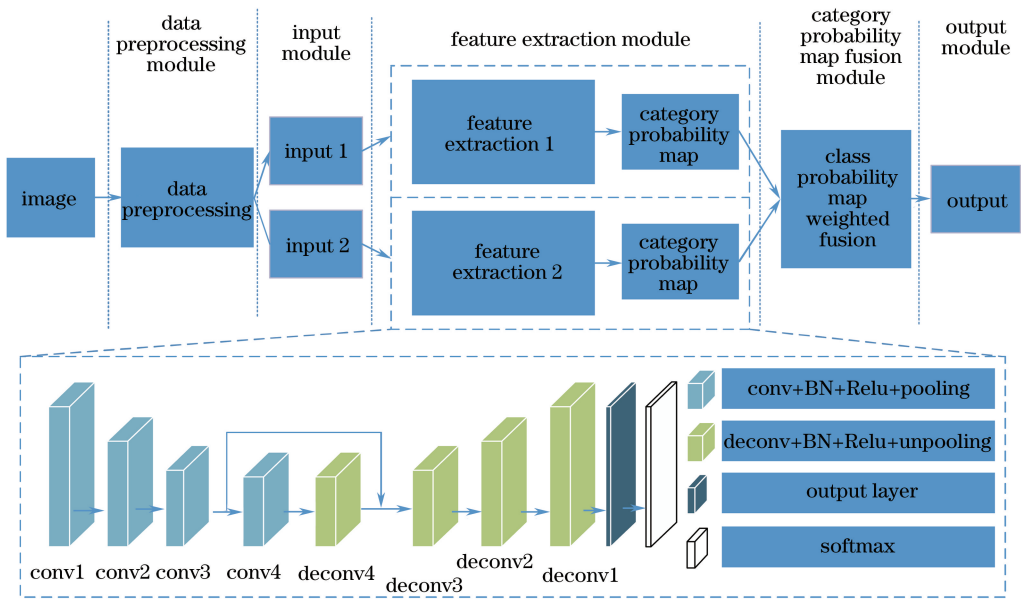


图1 融合特征和决策的卷积-反卷积网络图像分割模型

Fig. 1 Convolution-deconvolution image segmentation model for fusion features and decision

集图像按照不同比例进行放大和缩小,并按 90° 、 180° 、 270° 方向进行旋转。将训练图像按步长进行裁剪,裁剪后每个图像块尺寸为 500×500 ,得到 IRRG 图像、DSM 图像、二值标记图像各 34712 幅。遥感图像中建筑物尺寸、方向、周围环境各异,目标边缘模糊等,分割出目标具有挑战性。

2.2 输入模块

CD-FFD 图像分割系统采用并联结构,具有两个神经网络分支,分别采用不同源信息输入。在 Weizmann Horse 数据集实验中采用 RGB 图像和灰度图像,vaihigen 数据集实验中采用 IRRG 图像和 DSM 图像,分别按照 500×500 的尺寸对各图像进行裁剪作为网络输入。

2.3 特征提取和类别预测模块

CD-FFD 模型包含两个结构相同的特征提取和类别预测模块,分别用于对不同类型的输入信息进行学习和预测,并得到两组权重系数。特征提取和类别预测模块采用卷积-反卷积结构,卷积阶段通过深度 CNN 进行特征提取,反卷积阶段根据特征进行重建,通过 Softmax 层将具有相似特征的像素以高的概率赋予所属类别。卷积和反卷积阶段通过多级尺度特征融合,在全局和局部细节方面为模型提供更多的信息,从而得到更高的分割完整度。

特征提取和类别预测模块由卷积、下采样(池化)、反卷积、上采样(反池化)、输出层及 Softmax 层构成,卷积阶段和反卷积阶段完全对称,呈镜像关系。由于应用中样本数目和目标类别数目相对

PASCAL VOC 等数据集较少,因此选择了有限的深度,包括 10 层卷积层(conv)和 4 层池化层(pooling),反卷积阶段同样包括 10 层反卷积层(deconv)和 4 层反池化层(unpooling)。在每层卷积层和反卷积之后进行数据归一化(BN)和激活操作(ReLu)。网络的输入尺寸设置为 300×300 。

卷积阶段采用 3×3 的卷积核进行特征提取,通过多个 3×3 的堆叠卷积层可以获得和大尺寸卷积层相同大小的感受野,且参数更少。池化层选择最大池化,保留更多的纹理信息,并将特征图进行 2 倍下采样。反卷积阶段使用 3×3 大小的核进行反卷积操作来恢复特征图,连接反池化层将特征图进行 2 倍上采样。反卷积末端连接输出层得到两张特征图,输入 Softmax 层获得两张与输入图像具有相同大小的概率图。

在卷积网络学习过程中,不同深度的卷积层学习到的特征其层次是不同的,浅层的卷积层学习到的是局部特征,随着卷积层深度增加,感受野也随之增大,学习到的特征包含的全局信息更多,一般来说卷积第三阶段(conv3)后得到的分割结果较好,这是由于太过浅层学习的特征对局部变化比较敏感,抗噪性能不佳,而更高层学到的特征对于局部变化不敏感,但缺点是梯度消失使恢复的边界模糊,仅利用这样的特征只能得到一个笼统的预测,很难获得准确的目标边界。因此,将浅层的特征与深层的特征融合有助于解决深层特征梯度消失的问题,以便获得更精细的分割结果。具体到 CD-FFD 模型,

conv1、conv2 阶段层次浅,学习到的特征多为局部特征;conv3、conv4 阶段则包含较多的全局信息,conv4 特征由于卷积层数较深,目标特征则更为突出且背景干扰少,但 conv3 与 conv4 相比包含更多的细节信息,因此将 conv3 与 conv4 的特征融合以提升模型的细节处理能力。在反卷积阶段将 conv4 提取的特征图上采样得到的特征图和 conv3 得到的特征图融合,加强边缘等细节恢复能力。

在 CD-FFD 网络中池化层导致特征分辨率减小,因此在反卷积阶段使用反池化层扩大特征图的尺寸,并且利用卷积阶段池化操作时保留的特征位置信息,实现特征上采样的精细化,提高特征恢复的准确性。反池化操作扩充的位置使用 0 填充得到稀疏的特征图,再对其进行反卷积操作,将稀疏的特征图变得致密,获得更为精细的预测结果图,送入类别概率图融合模块。

2.4 类别概率图融合模块

目标分割任务实际上是预测每个像素的类别,既要较好地表征目标全局特征,又要较多地保留目标边缘特征,因此对细节特征的要求非常高,不同源的数据提供的信息虽有冗余,但又互补,深度模型有利于从不同数据中提取更多的信息,通过融合互补为分割、分类等任务性能的提升提供可能。所提模型通过两个神经网络分支从不同数据中提取特征,进而得到两组类别概率预测图,分别对应于不同数

据得到的类别预测结果,在此进行决策级的融合,对两个网络分支的结果加权融合,进一步提升分割性能。用 X_1 和 X_2 分别表示两个分支网络输出的类别概率图,融合的输出结果表示为

$$Y_i = \lambda_1 X_{1i} + \lambda_2 X_{2i}, \quad (1)$$

式中 i 表示类别编号, λ_1 和 λ_2 分别表示融合时两个分支网络的加权系数,不同的场景在不同的分支网络上表现出不同的预测效果,因此需对不同的场景调节加权系数来获取最优的分割结果。

2.5 输出

类别概率图融合模块对两个分支网络获得的两组概率图采用加权融合进行决策级的融合,得到两张融合后的结果概率图,分别代表了最终的每个像素属于目标和背景的概率,输出模块将两张新的概率图基于最大概率进行类别决策,属于目标的像素点标记为 1,非目标像素点为 0。

2.6 CD-FFD 模型数据处理流程

图 2(a)和(b)分别为 CD-FFD 模型和每一分支网络的数据处理流程,且 CD-FFD 两个分支网络处理流程相同。Weizmann Horse 实验采用 RGB 图像和灰度图像作为输入,得到两个神经网络分支分别用 RGB-Net 和 GRAY-Net 表示。vaihigen 实验采用 IRRG 图像和 DSM 图像作为输入,得到两个神经网络分支分别用 IRRG-Net 和 DSM-Net 表示。

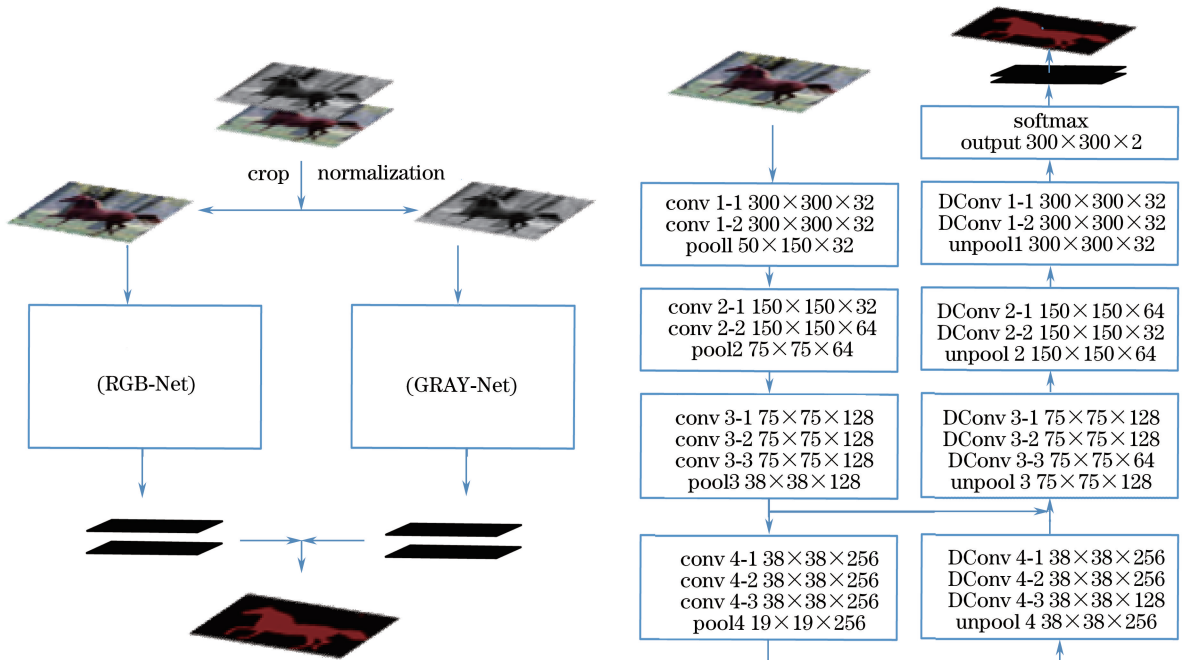


图 2 数据处理流程。(a) CD-FFD 数据处理流程;(b) CD-FFD 每一分支网络数据处理流程

Fig. 2 Flow charts of data processing. (a) Data processing of CD-FFD; (b) each branch network data processing of CD-FFD

2.7 网络训练

由于分割任务较 VGG16 在 ImageNet 上的目标类别和训练样本少很多,本文提出的网络模型卷积层数、卷积核数量均大大减少,因此不使用 VGG16 的预训练模型初始化权重,而是重新训练网络。基于迁移学习的思想训练模型,训练包括两步:1) 在 Weizmann Horse 实验中先使用 PASCAL VOC 2012 数据集训练各分支网络,vaihigen 数据集实验中先使用 Aerial Image 数据集训练各分支网络;2) 使用不同类型的输入信息对上一步训练好的网络模型进行微调得到两个具有不同权重系数的分支网络。

在网络训练时,Weizmann Horse 实验参数设定:批量输入大小为 6;使用随机梯度下降训练模型,学习率为 0.01,每 10000 次迭代将学习率降低 0.1;动量为 0.95;训练损失最大迭代次数为 100000。vaihigen 数据集实验参数设定:批量输入大小为 6;使用随机梯度下降训练模型,学习率为 0.01,每 20000 次迭代将学习率降低 0.1;动量为 0.95;训练损失最大迭代次数为 100000。

2.8 评价方式

采用目标完整度 (COM)^[14]、全局精度 (Global acc) 和交并比 (IOU) 三个指标定量评价分割结果。Global acc 表示每个像素点分类的正确率,COM 度量分割图像目标区域匹配标记图像目标区域的比例,IOU 是一个总体度量,和 COM 相比,IOU 还考虑了目标被错分为背景的情况。它们被定义为

$$P_{\text{globalacc}} = \frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{TP}} + N_{\text{TN}} + N_{\text{FN}} + N_{\text{FP}}}, \quad (2)$$

$$P_{\text{COM}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}, \quad (3)$$

$$P_{\text{IOU}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}} + N_{\text{FN}}}, \quad (4)$$

式中 N_{TP} 表示目标正确分类的像素数目, N_{TN} 表示背景正确分类的像素数目, N_{FN} 表示目标分为背景的像素数目, N_{FP} 表示背景分为目标的像素数目。

3 实验结果和分析

实验包括两部分,第一部分是 CD-FFD 模型在 Weizmann Horse 和 vaihigen 数据集的实验结果和分析,第二部分是 CD-FFD 与现有算法在 Weizmann Horse 和 vaihigen 数据集上的性能评价。实验在配备 64 位 Ubuntu 系统的工作站上完成,硬件配置为 Intel(R) Xeon(R) CPU E5-2690

v32.60GHz 处理器、256 GB 内存和 4 TB 硬盘。使用一块 NVIDIA Tesla K40c 12 GB 显存 GPU 进行加速。CD-FFD 模型的训练使用 Caffe^[15] 平台,代码在 MATLAB 上编写完成。

3.1 CD-FFD 实验结果和分析

3.1.1 Weizmann Horse 数据集

实验采用 RGB 图像和灰度图像分别训练两个神经网络分支,得到两个具有不同权重系数的分支网络,分别用 RGB-Net 和 GRAY-Net 表示。图 3 所示为采用 RGB-Net、GRAY-Net 以及 CD-FFD 模型的分割结果,图 3(a) 为 RGB 图像,图 3(b) 为灰度图像,图 3(c) 为 RGB-Net 分割结果,图 3(d) 为 GRAY-Net 分割结果,图 3(e) 为 CD-FFD 分割结果,图 3(f) 为标签图像。



图 3 CD-FFD 模型分割结果。(a) RGB 图像;(b) 灰度图像;(c) RGB-Net 分割结果;(d) GRAY-Net 分割结果;(e) CD-FFD 分割结果;(f) 标签图像

Fig. 3 Segmentation results of CD-FFD model. (a) RGB image; (b) gray image; (c) segmentation result of RGB-Net; (d) segmentation result of GRAY-Net; (e) segmentation result of CD-FFD; (f) ground-truth

由图 3 可知,灰度图像与 RGB 图像在提取的特征上存在明显差异,采用这些特征得到的分割结果对于不同的图像也呈现了明显差异。CD-FFD 模型得到的分割结果明显优于 RGB-Net 和 GRAY-Net,其目标完整度高,目标和背景的错分少。可见,考虑 RGB 图像和灰度图像的多尺度特

征信息并融合了各自的决策结果的 CD-FFD 模型具有明显的优势。

为了充分验证 CD-FFD 模型的分割性能,分别在 Weizmann Horse 数据集实验验证图像(样本量 200,目标形态复杂,包含各种角度样本)和 Weizmann Horse 测试集(样本量 128,目标形态简单,样本角度单一)上进行了测试,评价结果分别如表 1 和表 2 所示。

表 1 CD-FFD Weizmann Horse 200 幅验证图像评价结果

Type	Average COM	Average global acc	Average IOU
RGB-Net	0.9183	0.9542	0.7815
GRAY-Net	0.9169	0.9407	0.7327
CD-FFD	0.9329	0.9656	0.8188

表 2 CD-FFD Weizmann horse 128 幅测试图像评价结果

Type	Average COM	Average global acc	Average IOU
RGB-Net	0.9279	0.9672	0.8834
GRAY-Net	0.9360	0.9648	0.8753
CD-FFD	0.9407	0.9708	0.8949

由表 1、表 2 可知,在测试集与 Weizmann Horse 数据集上,CD-FFD 模型同样在 COM、Global acc 和 IOU 三个指标上都取得了最好的结果,表明在多尺度的特征级和分割的决策级融合不同源的信息可以有效提高分割精度。

3.1.2 vaihigen 数据集

vaihigen 数据集实验中采用 IRRG 图像和 DSM 图像分别训练两个神经网络分支,DSM 图像包含成像对象的高度信息,有利于区分高低不同的目标,可为建筑物提取提供有效的信息,来源于两种输入的分支网络分别用 IRRG-Net 和 DSM-Net 表示。图 4 所示采用 IRRG-Net、DSM-Net 以及 CD-FFD 模型的分割结果。其中图 4(a)为原始 IRRG 图像,图 4(b)为原始 DSM 图像,图 4(c)为 IRRG-Net 分割结果,图 4(d)为 DSM-Net 分割结果,图 4(e)为 CD-FFD 分割结果,图 4(f)为标签图像。

由图 4 可知,不同的图像目标类型和特征上的差异导致有的图像根据 IRRG 数据得到的分割结果优于 DSM,有的则相反,在目标和背景的错分方面也表现各异;而 CD-FFD 模型由于在特征和决策阶段的融合使最终的分割结果在视觉上明显优于

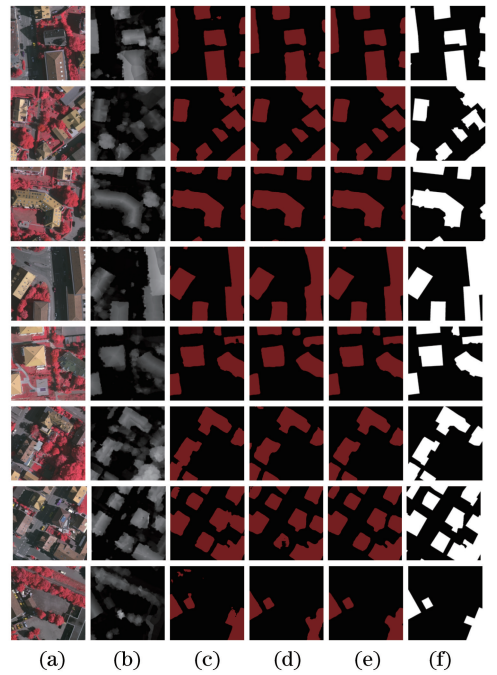


图 4 CD-FFD 模型分割结果。(a) IRRG 图像;(b) DSM 图像;(c) IRRG-Net 分割结果;(d) DSM-Net 分割结果;(e) CD-FFD 分割结果;(f) 标签图像

Fig. 4 Segmentation results of CD-FFD model. (a) IRRG image; (b) DSM image; (c) segmentation result of IRRG-Net; (d) segmentation result of DSM-Net; (e) segmentation result of CD-FFD; (f) ground-truth

IRRG-Net 和 DSM-Net;目标完整度高,目标和背景的错分少,对于各幅图像在三个评价指标上也均优于 IRRG-Net 和 DSM-Net。为了充分验证 CD-FFD 模型的分割性能,在 vaihigen 测试集上进行了测试,评价结果如表 3 所示。

表 3 CD-FFD Vaihigen 5 幅测试图像评价结果

Table 3 Evaluation results of 5 test images from vaihigen dataset

Type	Average COM	Average global acc	Average IOU
IRRG-Net	0.9316	0.9668	0.8698
DSM-Net	0.9221	0.9610	0.8420
CD-FFD	0.9496	0.9748	0.8966

3.2 CD-FFD 模型和其他模型比较分析

3.2.1 Weizmann Horse 数据集

在 Weizmann Horse 数据集^[12]上对 CD-FFD 模型和现有的研究成果进行了比较。由于现有文献对实验结果均采用 Global acc 和 IOU 两个指标进行评价,为了统一也选用 Global acc 和 IOU 两个评价指标。CD-FFD 模型与现有研究成果对比如表 4

所示。

表4 Weizmann Horse 数据集 128 幅测试图像
现有成果对比

Table 4 Comparison among existing results of 128 test
images from Weizmann Horse dataset

Number	Method	Average global acc	Average IOU
1	Ref. [16]	94.6	80.1
2	Ref. [17]	95.8	84.0
3	Ref. [18]	95.7	84.0
4	Ref. [19]	94.9	79.9
5	CD-FFD	97.2	89.5
6	CD-FFD+CRF	97.6	90.1

表4将CD-FFD模型在Weizmann Horse数据集上的分割结果与现有的研究成果进行对比,其中文献[16]、[17]是基于传统的机器学习方法,文献[18]、[19]是基于深度学习的方法。文献[16]方法基于监督学习,使用核结构支持向量机(SVM)从一组训练样本中区别地学习,将高级别对象相似性信息(通过对象特征描述符)与多个低级别分割线索进行组合以生成最适合的分割图像。文献[17]方法为级联算法,由于相似对象倾向于共享相似的局部形状,因此将图像块与示例图像进行局部形状匹配来获取更细致的分割结果。文献[18]方法即条件随机场(CRF)学习方法,利用深度CNN获取学习的特征,并使用结构化支持向量机(SSVM)学习CRF参数来进行物体分割。文献[19]方法利用CNN提取的局部通用特征和递归神经网络(RNN)检索远端依赖关系的能力来提高分割性能。表4第5行是CD-FFD模型在Weizmann Horse数据集上的分割结果,第6行是CD-FFD模型使用CRF做后处理的情况时得到的结果。表4结果表明,CD-FFD模型与文献[16]~[19]方法相比,均获得了较好的结果。与CRF学习方法、文献[19]方法相比,CD-FFD模型包含两个结构相同的特征提取和类别预测模块,用于对不同类型的输入信息进行学习和预测,有利于从不同数据中提取更多的信息,通过融合互补为分割、分类等任务性能的提升提供可能,性能更优。与文献[17]方法相比,该模型的平均IOU提高了5.5%,使用CRF进行后处理,其平均IOU则提高了6.1%。

3.2.2 vaihigen 数据集

为了与现有的研究成果比较,在vaihigen数据集测试集上采用Global acc进行评价,结果对比如表5所示。

表5 Vaihigen 数据集 5 幅测试图像现有成果对比
Table 5 Comparison among existing results of 5 test
images from vaihigen dataset

Number	Method	Average global acc
1	SegNet ^[20]	0.9078
2	CNN+RF ^[21]	0.9423
3	CNN+RF+CRF ^[21]	0.9430
4	Ref. [22]	0.9450
5	CD-FFD	0.9748

比较方法中SegNet是具有对称结构的编码-解码深度卷积网络,编码阶段使用卷积网络提取特征,解码阶段使用卷积网络来进行特征上采样。CNN+RF方法针对原图及手工提取特征分别使用CNN、随机森林(RF)方法进行类别预测,再将输出结果进行融合获取最终分割结果,其中CNN包含4层卷积层和两个全连接层。CNN+RF+CRF是在CNN+RF方法基础上再结合CRF进行平滑处理获取分割结果。文献[22]方法在SegNet网络的末端引入多核卷积层来进行多尺度预测,并使用残差网络将原图及手工提取进行融合特征。上述研究方法均未考虑神经网络内部不同深度卷积层之间特征提取层次不同,CD-FFD模型将浅层的特征与深层的特征进行融合,有助于解决深层特征梯度消失的问题,获得更精细的分割结果;在考虑信息融合时,CNN+RF采用乘法运算融合,预测结果受分割性能差的一方影响较大,文献[22]方法直接对特征进行融合,而CD-FFD采取决策级融合方式,在两个分支网络末端使用类别概率图加权融合,充分融合两个神经网络分支的优势性能,使性能更优的网络分支在融合时发挥更大的作用。与上述方法相比,CD-FFD模型取得了最高的分割精度,与文献[22]方法相比,其平均Global acc提高了2.98%。

4 CD-FFD 模型其他图像实验

在CD-FFD实验中,分别对训练集进行了数据增强,以及缩放、旋转等操作,除了能扩充训练集样本数量使得模型训练充分之外,旋转等操作还可以增加目标角度变化范围,使得模型可以应对具有多种角度变化,以及形态多样的目标的复杂图像分割。以Weizmann Horse数据集训练得到的CD-FFD模型为例进行说明。

图5所示为CD-FFD模型对于目标形态复杂的图像的分割结果,主要考查CD-FFD模型对训练数据集以外图像的分割性能。这些图像来源于网络,和Weizmann Horse数据集中的图像在目标尺度、

位置、姿态、数目上均差别巨大,这给使用目标形态简单、每幅图只有一个目标的小样本数据集训练得到的模型带来了巨大的挑战。图 5(a)为原始 RGB 图像,图 5(b)为 CD-FFD 分割结果。

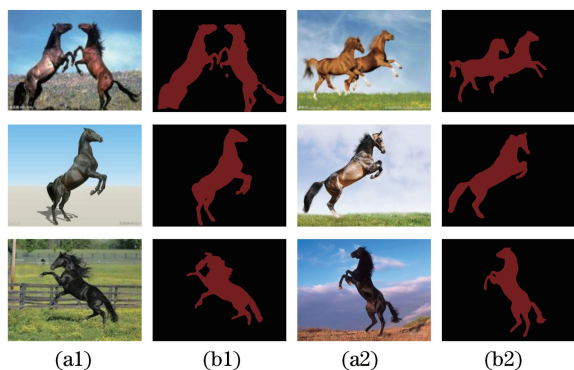


图 5 CD-FFD 模型对其他图像的分割结果。

(a)原始 RGB 图像;(b) CD-FFD 分割结果

Fig. 5 Segmentation results of other images by CD-FFD model. (a) Original RGB images; (b) segmentation results of CD-FFD

由图 5 可知,CD-FFD 模型依然可以获得很好的分割结果,表明该模型具有比较广泛的应用范围。

5 结 论

提出 CD-FFD 图像分割模型,该模型从不同分支、不同来源的图像中获取特征,获取更多互补的信息。每个分支构建深度卷积-反卷积网络提取特征,分别在特征层级和决策层级进行融合,充分融合两个神经网络分支的优势性能,提升算法性能。在 Weizmann Horse、vaihigen 数据集上获得了有竞争力的性能,分割精度分别提高了 6.1% 和 2.98%。另外,通过数据增强方法,模型获得了更强的稳健性,并且在随机选取网络上的其他图像上依然可以获得很好的分割结果,也体现了该模型具有比较广泛的应用范围。未来的研究内容主要是解决高分辨率遥感图像识别中存在的目标遮挡问题。

参 考 文 献

- [1] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 3431-3440.
- [2] Liu W, Rabinovich A, Berg A C. Parsenet: looking wider to see better[J]. arXiv, 2015: 1506.04579.
- [3] Wang P Q, Chen P F, Yuan Y, *et al.* Understanding convolution for semantic segmentation [C] // IEEE Winter Conference on Applications of Computer Vision (WACV), 2018: 1451-1460.
- [4] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation [C] // IEEE International Conference on Computer Vision (ICCV), 2015: 1520-1528.
- [5] Ghiasi G, Fowlkes C C. Laplacian pyramid reconstruction and refinement for semantic segmentation [C] // European Conference on Computer Vision. Springer, 2016: 519-534.
- [6] Zhao H S, Shi J P, Qi X J, *et al.* Pyramid scene parsing network [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6230-6239.
- [7] Dai J F, Qi H Z, Xiong Y W, *et al.* Deformable convolutional networks [C] // IEEE International Conference on Computer Vision (ICCV), 2017: 764-773.
- [8] Zhang R, Tang S, Zhang Y D, *et al.* Scale-adaptive convolutions for scene parsing [C] // IEEE International Conference on Computer Vision (ICCV), 2017: 2050-2058.
- [9] Liu F, Liu P Y, Zhang J N, *et al.* Joint detection of RGB-D images based on double flow convolutional neural network [J]. Laser & Optoelectronics Progress, 2018, 55(2): 021503.
- [10] Li Q W, Zhou Y Q, Ma Y P, *et al.* Salient object detection method based on binocular vision [J]. Acta Optica Sinica, 2018, 38(3): 0315002.
- [11] Xu Y, Sun M S. Convolution neural network image defogging based on multi-feature fusion [J]. Laser & Optoelectronics Progress, 2018, 55(3): 031012.
- [12] Borenstein E, Ullman S. Combined top-down/bottom-up segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(12): 2109-2125.
- [13] Gerke M. Use of the stair vision library within the ISPRS 2D semantic labeling benchmark (vaihigen) [J]. Technical Report, University of Twente, 2014.

- [14] Heipke C, Mayer H, Wiedemann C, *et al.* Evaluation of automatic road extraction [J]. International Archives of Photogrammetry and Remote Sensing, 1997, 32(3 SECT 4W2): 151-160.
- [15] Jia Y, Shelhamer E, Donahue J, *et al.* Caffe: convolutional architecture for fast feature embedding [C] // Proceedings of the 22nd ACM International Conference on Multimedia. ACM, 2014: 675-678.
- [16] Bertelli L, Yu T, Vu D, *et al.* Kernelized structural SVM learning for supervised object segmentation [C] // Computer Vision and Pattern Recognition, 2011: 2153-2160.
- [17] Yang J, Price B, Cohen S, *et al.* Patchcut: data-driven object segmentation via local shape transfer [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1770-1778.
- [18] Liu F Y, Lin G S, Shen C H. CRF learning with CNN features for image segmentation [J]. Pattern Recognition, 2015, 48(10): 2983-2992.
- [19] Visin F, Ciccone M, Romero A, *et al.* Reseg: a recurrent neural network-based model for semantic segmentation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2016: 41-48.
- [20] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [21] Paisitkriangkrai S, Sherrah J, Janney P, *et al.* Effective semantic pixel labelling with convolutional networks and conditional random fields [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2015: 36-43.
- [22] Audebert N, Le Saux B, Lefèvre S. Semantic segmentation of earth observation data using multimodal and multi-scale deep networks [C] // Asian Conference on Computer Vision. Cham: Springer, 2016: 180-196.