doi: 10.6046/gtzyyg.2020.03.03
引用格式: 李宇,肖春姣,张洪群,等. 深度卷积融合条件随机场的遥感图像语义分割[J]. 国土资源遥感,2020,32(3):15-22. (Li Y, Xiao C J, Zhang H Q, et al. Remote sensing image semantic segmentation using deep fusion convolutional networks and conditional random field, 2020, 32(3): 15-22.)

深度卷积融合条件随机场的遥感图像语义分割

李 宇[1], 肖春姣[1], 张洪群[1], 李湘眷[2], 陈 俊[1]
(1. 中国科学院遥感与数字地球研究所,北京 100094; 2. 西安石油大学计算机学院,西安 710065)

摘要: 为了实现高分辨率光学遥感图像的语义分割,提出了一种基于深度卷积融合条件随机场的图像语义分割方法。该方法在全卷积神经网络模型的基础上增加反卷积融合结构结合不同深度的池化层结果,将浅层的细节信息和高层的语义信息融入网络模型,同时将条件随机场的参数推断以迭代层的形式嵌入网络架构,搭建网络模型,在模型训练的正反向传播过程中综合利用遥感图像丰富的细节信息和上下文信息,实现端到端的遥感图像语义分割。在高分辨率遥感图像中进行的实验结果显示: 随着模型中反卷积融合结构结合池化层深度的增加,语义分割处理精度逐渐提高,语义分割结果中的地物轮廓也更清晰、准确; 上下文信息的引入也在一定程度上提高了图像语义分割的精度。实验表明该方法能够较好地保持语义对象内部区域的一致性,有效提高图像语义分割的精度。

关键词: 遥感图像; 语义分割; 全卷积神经网络; 条件随机场
中图法分类号: TP 751 **文献标志码**: A **文章编号**: 1001-070X(2020)03-0015-08

0 引言

图像语义分割是计算机视觉领域的一项关键技术,其目标是对图像的每个像素进行语义标注[1]。图像语义分割将图像分割和目标识别结合起来,同时解决目标边界识别和类别检测问题,能够满足图像中目标精细定位的需求,具有重要的应用价值。随着遥感技术的发展,遥感图像的空间分辨率不断提高,呈现的结构、纹理等细节信息更加清晰,地物边界和空间布局等上下文信息日益丰富,为图像语义分割提供了良好的数据条件; 同时遥感图像信息量巨大、数据复杂等特点也给图像语义分割带来了困难和挑战。

传统的图像语义分割方法由图像分割、特征提取、语义分割模型训练 3 部分组成[2-4],首先利用图像分割方法得到区域块,然后提取区域块的特征(颜色、纹理等),最后通过模型训练建立特征与高层语义之间的联系,实现图像语义分割。传统方法中的特征提取通常需要依靠先验知识进行人工特征选择和设计,不仅设计复杂且对后续的分类模型要求较高,使得该类方法的应用较为局限。同时高分辨率遥感图像的场景复杂且存在阴影、噪声、遮挡等,采用传统方法难以获得理想的分割结果。

近年来,卷积神经网络(convolutional neural network, CNN)因避免了传统人工显式特征提取,能够自动从海量数据中获取特征,受到越来越多研究者们的关注[5-9]。图像语义分割方法也随之有了全新的发展,一系列语义分割方法相继提出: Long 等[10]提出了全卷积神经网络(fully convolutional networks, FCN),用卷积层替换 CNN 网络结构中的全连接层,实现了端到端的像素级分类; Sherrah 等[11]采用无下采样层的 FCN 模型版本对航空遥感图像进行了语义标注; Badrinarayanan 等[12]提出基于编码-解码结构的 SegNet 模型,通过将编码过程中最大池化操作的像素索引传输到解码器中,保留了部分细节信息,在一定程度上改善了分割精度; Ronneberger 等[13]提出了 U-Net 网络模型,通过跳跃连接融合解码信息与对应的编码信息,成功应用到医学影像语义分割中。然而,上述方法由于没有考虑到像素与像素之间的上下文关联,分割结果缺乏语义的空间一致性。为引入上下文信息,进一步优化语义分割效果,条件随机场(conditional random field, CRF)被引入到 CNN 模型框架中。Chen 等[14]

收稿日期: 2019-08-16; **修订日期**: 2019-10-28
基金项目: 国家重点研发计划项目"天空地一体化协同观测、数据整合与应急信息提取技术研究"(编号: 2016YFB0502502)和国家自然科学基金项目"基于语义模型的高分辨率卫星遥感图像人造目标检测方法研究"(编号: 61501460)共同资助。
第一作者: 李 宇(1986-),女,博士,工程师,主要从事遥感图像处理的研究。Email: liyu@radi.ac.cn。
通信作者: 张洪群(1971-),男,正高级工程师,主要从事遥感卫星数据处理技术研究。Email: zhanghq@aircas.ac.cn。

提出了 DeepLab 模型,在 FCN 后接全连接 CRF,对结果进行平滑约束,增加了区域一致性和连续性,克服了定位精度问题; Zheng 等[15] 提出了 CRFasRNN 模型,将 CRF 的学习推理过程嵌入 CNN 架构中,通过反向传播算法对整个深度网络进行端到端的训练,避免了后处理,但其位置信息精度不如 DeepLab 模型。由于 CNN 中的逐层池化操作,细节信息在网络的浅层到深层的传播过程中不断衰减,最终的图像语义分割精度有待提升。

在前人工作基础上,本文提出了一种深度卷积融合条件随机场(deep fusion convolutional networks and conditional random field, DFN – CRF)的遥感图像语义分割方法。首先,在 FCN 框架中加入反卷积融合结构; 同时,将 CRF 的求解过程以迭代层的形式融合到深层网络架构中,搭建 DFN – CRF 模型; 最后,在模型训练的过程中结合图像中丰富的细节信息和上下文信息,实现端到端的图像语义分割。方法不仅利用 FCN 自动提取深度语义特征的优势,提升了模型泛化能力; 而且通过反卷积融合结构引入多尺度特征,将浅层细节信息和深层语义信息结合起来做模型预测,有效地提高模型的处理精度。同时,利用 CRF 引入上下文信息,可以更好地定位地物,进一步提升模型的处理效果。另外,端到端的网络模型缩减了人工预处理和后续处理,模型可以根据数据优化参数,提高模型的契合度。

1　方法概述

本文的遥感图像语义分割处理流程(图 1)包括模型训练和测试 2 个阶段。在模型训练阶段,首先通过反卷积融合结构将高层语义信息和浅层细节信息相结合,同时在深度网络模型中以迭代层的形式融入全连接条件随机场,进而引入上下文信息,构建 DFN – CRF 网络模型; 然后利用随机梯度下降法进行模型参数学习,完成网络模型的训练。在测试阶段,即利用训练阶段得到的模型对待处理的遥感图像中的像素进行标记推断,最终实现端到端的遥感图像语义分割。

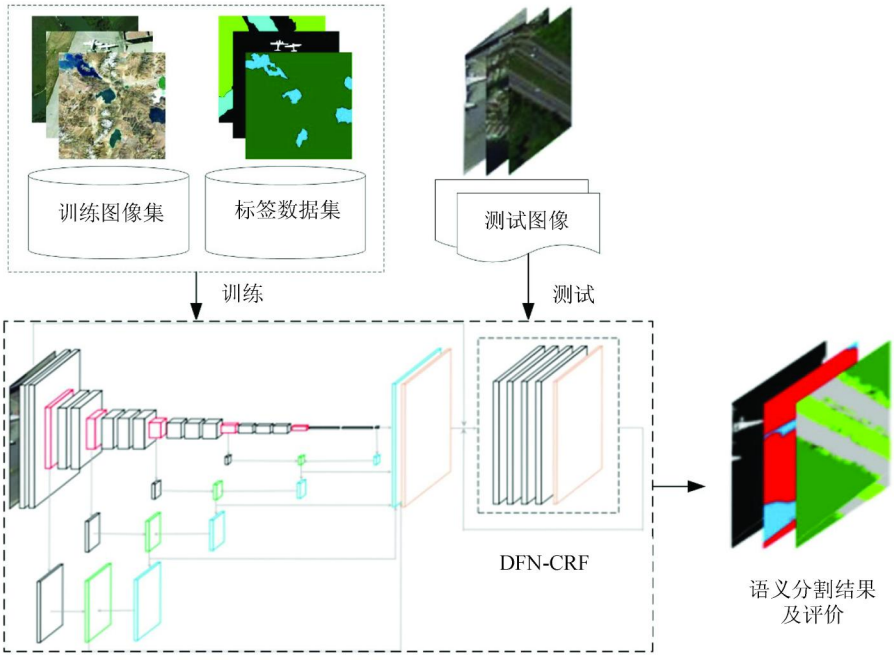


图 1　图像语义分割处理流程

Fig. 1　Flow chart of the image semantic segmentation algorithm

1.1　FCN

FCN 基本结构主要由输入层、卷积层、池化层和输出层组成。

1)输入层。读取数据为 $h \times g \times d$ 的 d 维数组,其中 h 和 g 为数据尺寸,d 为特征或通道数。本文输入图像的尺寸为 $224 \times 224 \times 3$。

2)卷积层。用于提取局部特征,又称特征提取层。通过一组滤波器(卷积核)对输入的数据进行卷积,然后将卷积结果传递给非线性激活函数。本文采用的激活函数为修正线性单元(rectified linear units, ReLU),函数定义为:

$$f(x) = \max(0, x) \quad 。 \tag{1}$$

3)池化层。用于进行非线性降采样,通过对卷积特征进行降维,保留有用信息的同时减少计算复杂度,加快网络训练速度。常用的池化有最大池化和平均池化。本文采用最大池化方法。

4)输出层。通过 Softmax 分类器输出每个像素

点属于各语义类别的概率,并选择最大概率值所对应的类别结果作为最终输出。Softmax 函数的公式定义为:

$$h_i = \frac{\exp(X_i)}{\sum_{j=1}^{k}\exp(X_j)} \quad , \tag{2}$$

式中: X_i 为最后隐藏层的输出; h_i 为类 i 的判别概率; j 为类别,$j=1,2,\cdots,k$。

1.2 反卷积融合结构

借助多层网络架构,CNN 能够从输入图像自动获取到多层特征: 浅层感知域较小,可以得到局部区域的细节信息; 深层感知域大,能够得到更加抽象的语义信息。然而,CNN 中的逐层池化,不仅使得细节信息在网络的浅层到深层的传播过程中不断衰减,也不断缩小了图像尺寸。因此,本文方法采用反卷积融合结构结合浅层细节信息和深层语义信息来实现像素级图像语义标注。

反卷积融合结构主要由卷积层、反卷积层和融合层组成,其中融合层是将来自不同层的结果求和后输出。反卷积操作本质上也是卷积操作,区别在于卷积实现下采样,反卷积操作是完成上采样。以 Caffe 框架中的反卷积操作为例,卷积操作的前向和反向传播过程,前向传播过程为:

$$\boldsymbol{O} = \boldsymbol{A} \times \boldsymbol{B} \quad , \tag{3}$$

式中: $\boldsymbol{A}$ 为卷积核矩阵; $\boldsymbol{B}$ 为图像特征矩阵; $\boldsymbol{O}$ 为输出矩阵。反向传播时,有矩阵微分公式

$$\frac{\partial \boldsymbol{M}x + b}{\partial x} = \boldsymbol{M}^{\mathrm{T}} \quad , \tag{4}$$

式中: $\boldsymbol{M}$ 为任意矩阵; b 为任意常数。由此推出

$$\frac{\partial Loss}{\partial \boldsymbol{B}} = \frac{\partial Loss}{\partial \boldsymbol{O}} \cdot \frac{\partial \boldsymbol{O}}{\partial \boldsymbol{B}} = \boldsymbol{A}^{\mathrm{T}} \frac{\partial Loss}{\partial \boldsymbol{O}} \quad , \tag{5}$$

式中 *Loss* 为损失函数。所以反卷积操作就是正向传播时左乘 $\boldsymbol{A}^{\mathrm{T}}$、反向时左乘 $\boldsymbol{A}$ 的运算。

1.3 CRF 模型

CRF 模型最早是由 Lafferty 等[16]提出的一种判别式概率无向图学习模型。对于遥感图像语义分割而言,假定 I 表示给定的遥感图像,X 表示相应的标签图像,根据 Hammersley - Clifford 定理[17],CRF 的后验概率可近似为:

$$P(X \mid I) = \frac{1}{Z}\exp\{-\sum_{c\in C}\psi_c(X_c)\} \quad , \tag{6}$$

式中: Z 为归一化常数; C 为所有基团 c 的集合; ψ_c 为基团 c 的势函数; $X_c = \{X_i, i \in c\}$, X_i 为像素 i 的类别标签。相应的 Gibbs 能量函数为:

$$E(x) = \sum_{c\in C}\psi_c(x_c) \quad 。 \tag{7}$$

满足最大后验概率的标注向量 $\boldsymbol{x}^*$ 定义为:

$$\boldsymbol{x}^* = \mathrm{argmax}_x p(x \mid I) = \mathrm{argmin}_x E(x) \quad 。 \tag{8}$$

模型的能量函数为:

$$E(x) = \sum_i \psi_1(x_i) + \sum_{i\in I, j\in N_i}\psi_2(x_i, x_j) \quad , \tag{9}$$

式中: $\psi_1(x_i)$ 为一阶势函数,描述单像素点观测信息与其相应标记之间的关系; N_i 为像素 i 的邻域位置集合; $\psi_2(x_i, x_j)$ 为二阶势函数,描述像素之间的关系,鼓励相似的像素分配相同的标签,相差较大的像素分配不同的标签。具体采用的 Gibbs 能量函数为:

$$\psi_1(x_i) = -\ln P(x_i) \quad , \tag{10}$$

$$\psi_2(x_i, x_j) = \mu(x_i, x_j)\sum_{m=1}^{K}\omega^{(m)}k^{(m)}(\boldsymbol{f}_i, \boldsymbol{f}_j) \quad , \tag{11}$$

式中: $\mu(x_i, x_j)$ 为判别函数,若 $x_i \neq x_j$,则 $\mu(x_i, x_j) = 1$,否则为 0; $\omega^{(m)}$ 为各高斯核函数相应的权重,m 为标签类别,$m=1,2,\cdots,K$; $k^{(m)}(\boldsymbol{f}_i, \boldsymbol{f}_j)$ 为高斯核函数; $\boldsymbol{f}_i$ 和 $\boldsymbol{f}_j$ 分别为像素 i 位置和像素 j 位置的特征向量。

采用平均场算法[18]来实现模型参数推断,通过最小化 KL 距离 $D(Q\|P)$,以近似分布 $Q(X)$ 代替精确分布 $P(X)$,即

$$Q(x) = \prod_i Q_i(x_i) \quad , \tag{12}$$

最小化过程的迭代计算公式为:

$$Q_i(x_i = l) = \frac{1}{Z_i}\exp\Big\{-\psi_1(x_i) - \sum_{l'\in L}\mu(l, l')\sum_{m=1}^{K}\omega^{(m)}\sum_{j\neq i}k^{(m)}(f_i, f_j)Q_i(l')\Big\} \quad , \tag{13}$$

式中 Z_i 为归一化常数。

迭代过程可分为 5 个步骤: ①“消息传递” $\sum_{j\neq i}k^{(m)}(\boldsymbol{f}_i, \boldsymbol{f}_j)Q_j(l) \rightarrow \tilde{Q}_i^{(m)}(l)$; ②“滤波结果求加权和” $\sum_m \omega^{(m)}\tilde{Q}_i^{(m)}(l) \rightarrow Q_i'(l)$; ③“兼容性转换” $\sum_{l\in L}\mu(x_i, l)Q_i'(l) \rightarrow \hat{Q}_i(x_i)$; ④“一元项加入” $\psi_1(x_i) - \hat{Q}_i(x_i) \rightarrow \bar{Q}_i(x_i)$; ⑤“类别概率归一化” $\frac{1}{Z_i}\exp(\bar{Q}_i(x_i)) \rightarrow Q_i(x_i)$。

本文方法将 CRF 模型以循环神经网络(recur-

rent neural networks, RNN) 迭代层的形式加入到网络模型中, RNN 迭代层包括 4 个卷积层和 1 个 Softmax 输出层。“消息传递”和“滤波结果求加权和”相当于卷积操作;“兼容性转换”和“一元项加入”都相当于使用 1 ×1 的卷积核对全图进行卷积操作;“类别概率归一化”相当于使用 Softmax 分类器, 然后迭代求解条件随机场模型的参数。

1.4　DFN – CRF 网络结构

本文方法在 FCN 中加入反卷积融合结构, 并以 RNN 迭代层的形式引入将 CRF 模型参数推断过程, 搭建 DFN – CRF 模型, 在模型训练过程中同时利用细节信息和上下文信息。网络模型的网络结构参数如图 2 所示, 各层旁边标注为卷积(或池化、反卷积)核大小以及核个数。输入图像为 256 ×256 的三通道图像, 卷积层 3 ×3 ×64 表示使用 64 个尺寸为 3 ×3 卷积核的卷积层, 池化层为使用尺寸为 2 ×2 卷积核的最大池化层, 反卷积层 4 ×4 ×21 表示使用 21 个尺寸为 4 ×4 卷积核的反卷积层。

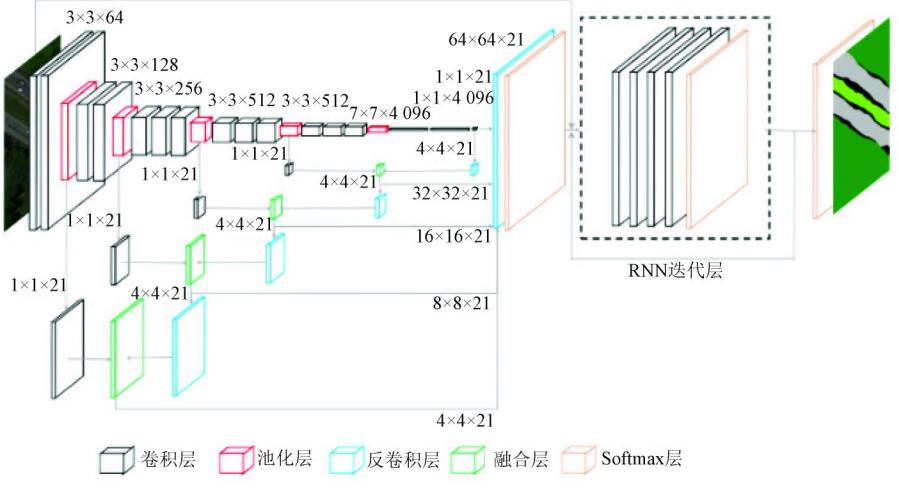


图 2　DFN – CRF 网络模型

Fig. 2　Model map of DFN – CRF

FCN 模型最终的卷积层结果只有最后一层池化层的信息, 仅对这一结果进行反卷积操作的模型记为 DFN – CRF – 5。根据融合层级的不同, 将融合池化层深度到 pool4, pool3, pool2, pool1 的模型, 分别记为 DFN – CRF – 4, DFN – CRF – 3, DFN – CRF – 2, DFN – CRF – 1。以 DFN – CRF – 4 为例, 反卷积融合结构将最后卷积层的结果反卷积到 pool4 层输出结果尺寸, 与 pool4 层卷积后结果通过融合层输出, 再将融合层结果反卷积到输入图像尺寸, 得到最终的像素级图像语义分割结果。模型 DFN – CRF – 3 将 DFN – CRF – 4 中融合层结果反卷积到 pool3 层输出结果尺寸, 与 pool3 层卷积后结果通过融合层输出, 再将融合层结果反卷积到输入图像尺寸, 模型 DFN – CRF – 2 和 DFN – CRF – 1 以此类推。

2　实验与结果

2.1　实验数据与评价准则

模型训练及测试数据来自公开的遥感数据集 NWPU[19], 包含 R, G, B 3 个波段数据。NWPU 数据集在视角、对象姿势、空间分辨率、光照、阴影等方面都有丰富的变化, 具有较高的类内多样性, 能够很好地检验语义分割方法的可靠性。模型泛化能力验证部分的实验数据采用来自 UC Merced Land – Use[20] 数据集的遥感图像。选取 20 种遥感地物语义类别进行实验, 分别为: 飞机、菱形棒球场、灌木丛、圆形农田、密集住宅区、沙漠、森林、高速公路、高尔夫球场、湖泊、草地、中等密集住宅区、山、停车场、矩形农田、河流、海冰、稀疏住宅区、梯田、湿地。模型训练及测试数据集每类 700 张图像; 模型泛化能力验证数据集每类 100 张图像。

实验平台服务器处理器为 Intel Xeon(R) CPU E5 – 2620 2.00GHz, 显卡为 NVIDIA GeForce GTX TITAN X, 采用 Python 2.7.13 与 Matlab R2016b 混合编程。网络模型随机选取数据集中的 12 000 张遥感图像进行训练, 其余作为测试图像。各网络的训练参数设置为: 学习率固定为 1×10^{-14}, 动量设置为 0.99, 权重衰减参数为 0.000 5, 采用随机梯度下降法进行网络训练, 反卷积层权重初始化采用双线性插值法。

本文从定性和定量 2 个方面对图像语义分割结果进行评价。定性评价主要从直观感觉如分割边缘是否明确、轮廓是否清晰等方面进行评估; 定量评价主要采用像素平均精度(per pixel accuracy, PA)、

类别平均精度(mean class accuracy, CA)以及平均像素交叠率(mean intersection over union, MIOU)3个指标[21],具体计算方法为:

$$PA = \frac{\sum_i n_{ii}}{\sum_i \sum_j n_{ij}}, \tag{14}$$

$$CA = \frac{1}{n_c} \sum_i \frac{n_{ii}}{\sum_j n_{ij}}, \tag{15}$$

$$MIOU = \frac{1}{n_c} \sum_i \frac{n_{ii}}{\sum_j n_{ij} + \sum_j n_{ji} - n_{ii}}, \tag{16}$$

式中: n_c 为数据集语义类别数; n_{ij} 为类别 i 被预测为类别 j 的像素总数。

2.2　细节信息影响分析

为了分析细节信息对遥感图像语义分割结果的影响,本文对融合不同深度信息的网络模型进行了对比实验。细节信息的融合尺度由少到多,分别是 DFN – CRF – 5, DFN – CRF – 4, DFN – CRF – 3, DFN – CRF – 2 和 DFN – CRF – 1。各模型迭代 50 000 次的训练时间及一幅图像平均测试时间如表 1 所示。

表 1　各模型训练及测试时间

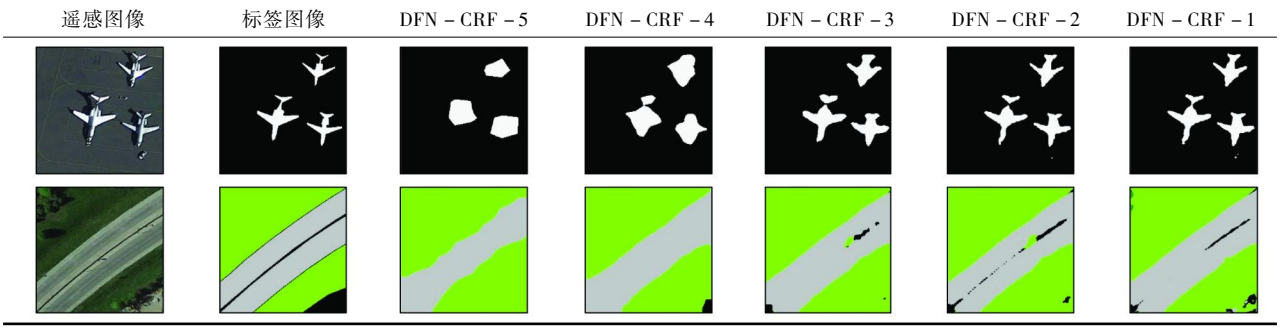
Tab. 1　Training and testing time of each model

模型	训练时间/min	测试时间/s
DFN – CRF – 5	474	0.61
DFN – CRF – 4	5 013	0.69
DFN – CRF – 3	5 237	0.73
DFN – CRF – 2	5 589	0.74
DFN – CRF – 1	5 976	0.77

采用不同层级融合结构的 DFN – CRF 模型进行遥感图像语义分割,结果如表 2 所示。其中,白色为飞机,灰色为公路,绿色为草地,黑色为背景。从表 2 中可以看出,DFN – CRF – 5 的语义分割结果非常粗糙,无法识别出遥感图像中具有固定形状的类别,存在多处语义分割错误(结果图中基本看不出飞机轮廓、高速公路边缘,与标签图像不符)。DFN – CRF – 4 模型通过融合结构引入了 pool4 层尺度的细节信息进行模型训练,相较于 DFN – CRF – 5 模型其语义分割结果明显有所提升(结果图中飞机形状渐显、高速公路的轮廓逐渐清晰)。DFN – CRF – 3 模型在 DFN – CRF – 4 基础上结合 pool3 尺度的细节信息进行语义分割,其结果已可以清楚地识别出

表 2　采用不同层级融合结构的 DFN – CRF 模型语义分割结果对比

Tab. 2　Comparison with the semantic segmentation results of DFN – CRF using different levels of fusion structure



飞机,高速公路间的背景类别被逐渐识别出来,边缘更细化。DFN – CRF – 2 和 DFN – CRF – 1 模型各自又在前一个模型的基础上融合了 pool2 层和 pool1 层尺度的细节信息,从结果中可以看出相较之前模型的语义分割结果,DFN – CRF – 2 和 DFN – CRF – 1 的语义分割结果主要改善了各语义类别的细小部分,使其边缘更加精细、轮廓更清晰、更接近标签图像(如飞机类别的尾部更细化、左下角飞机尾部与机身逐渐连接,高速公路间背景类别被误语义分割为草地类别的部分减少)。

进一步通过定量评价进行结果分析,各模型语义分割精度如表 3 所示。可知,从 DFN – CRF – 5 到 DFN – CRF – 1, PA 值, CA 值和 MIOU 值都逐渐增加,分别提高了 1.10, 4.48 和 5.08 百分点,与定性评价结果一致。

表 3　各模型语义分割结果定量评价

Tab. 3　Training and testing time of each model

模型	*PA*	*CA*	*MIOU*
DFN – CRF – 5	0.908 990	0.855 277	0.768 730
DFN – CRF – 4	0.910 455	0.878 851	0.799 701
DFN – CRF – 3	0.918 773	0.897 060	0.809 149
DFN – CRF – 2	0.918 969	0.898 627	0.816 557
DFN – CRF – 1	0.919 967	0.900 086	0.819 535

综合表 2 和表 3 可以看出,DFN – CRF – 5 模型的语义分割结果精度较低,这是因为 DFN – CRF – 5 模型中没有加入反卷积融合结构,其输出结果仅包含 pool5 层信息,由于池化操作损失了图像中的细节信息; 从 DFN – CRF – 4 到 DFN – CRF – 1,模型语义分割结果的精度逐渐提高,这是由于随着融合结构中引入池化层深度的增加,不同尺度的细节信息加入了模型中进行网络训练,从而使语义分割结果轮廓更清晰、明确,边缘更细化。

2.3　上下文信息影响分析

为分析上下文信息对遥感图像语义分割的优化效果,对上下文信息的不同利用方式进行对比实验。本文方法所提 DFN – CRF – 1 模型融合细节信息和上下文信息进行训练,仅利用细节信息未引入上下文信息的模型记为 DFN – 1,语义分割定量评价结果如表 4 所示。为了进一步分析上下文信息的影响,DFN – CRF – 1 和 DFN – 1 模型中各类地物的 PA 值如图 3 所示。

表 4　DFN – 1 和 DFN – CRF – 1 语义分割结果定量评价

Tab. 4　Quantitative evaluation of semantic segmentation results of DFN – 1 and DFN – CRF – 1

模型	训练时间/min	*PA*	*CA*	*MIOU*
DFN – 1	5 749	0.912 465	0.883 090	0.811 262
DFN – CRF – 1	5 976	0.919 967	0.900 086	0.819 535

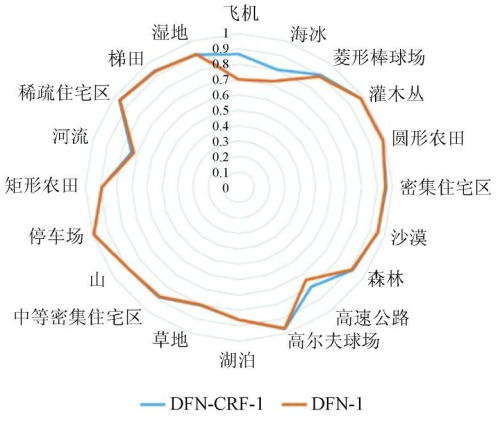


图 3　DFN – 1 和 DFN – CRF – 1 地物类别 PA 对比

Fig. 3　PA contrast diagram of different model's results between DFN – 1 and DFN – CRF – 1

从表 4 中可看出,DFN – 1 模型语义分割结果的定量评价低于 DFN – CRF – 1,这是因为 DFN – 1 模型没有加入上下文信息,缺乏对结构、边缘等空间上下文信息的约束; DFN – CRF – 1 模型相比 DFN – 1 模型在语义分割结果定性和定量评价中都有所提升,因为方法通过全连接条件随机场的高斯二阶势函数引入了图像中丰富的上下文信息,从而增加了对结构、边缘等空间信息的约束。DFN – CRF – 1 较 DFN – 1 模型,PA 值提高了 0.75 百分点左右,CA 值提高了 1.70 百分点左右,MIOU 值提高了 0.83 百分点左右。从图 3 中可以看出,上下文信息的引入对不同地物类别的优化效果不同,对细节信息丰富、结构边缘明显的地物类别(如飞机、高速公路等类别)的影响较为明显。

2.4　对比实验

将本文方法与目前常用的基于 CNN 的 FCN8s, DeepLab, CRFasRNN 等语义分割模型进行了对比实验,训练集和测试集保持一致,实验结果如表 5 所示。可以看出,DeepLab 模型的训练时间最短,但精度最低; 本文方法相比 FCN8s 和 CRFasRNN 模型训练时间虽然略长,但其 MIOU 值分别比其他 3 个模型提高了 3.61 百分点,2.15 百分点和 1.43 百分点左右。说明通过反卷积融合结构和 RNN 迭代层在模型训练的过程中同时融入细节信息和空间上下文信息,有效提高了遥感图像语义分割的精度。

表 5　本文方法与其他方法结果对比

Tab. 5　Comparison between our method and other methods

模型	训练时间/min	*MIOU*
DeepLab	1 692	0.783 401
CRFasRNN	4 853	0.798 024
FCN8s	4 917	0.805 267
DFN – CRF – 1	5 976	0.819 535

2.5　泛化能力验证实验

为了验证 DFN – CRF – 1 模型的泛化能力,本文在 UC Merced Land – Use 数据集上进行了语义分割实验,实验结果如图 4 所示。从图中可以看出 DFN – CRF – 1 模型能够较好地将河流、飞机、高速公路等地物从遥感图像中分割出来,并一定程度上保留其边缘、轮廓,保持了语义对象的内部一致性。总体来看,对数据来源不同于模型训练数据的遥感图像,本文方法也能够较好地实现图像语义分割,模型泛化能力较强。

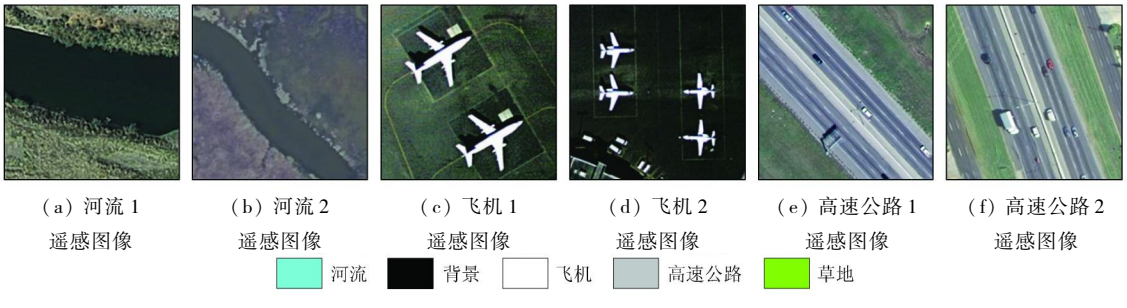


图 4 – 1　UC Merced Land – Use 数据集实验结果

Fig. 4 – 1　Experimental results of UC Merced Land – Use dataset

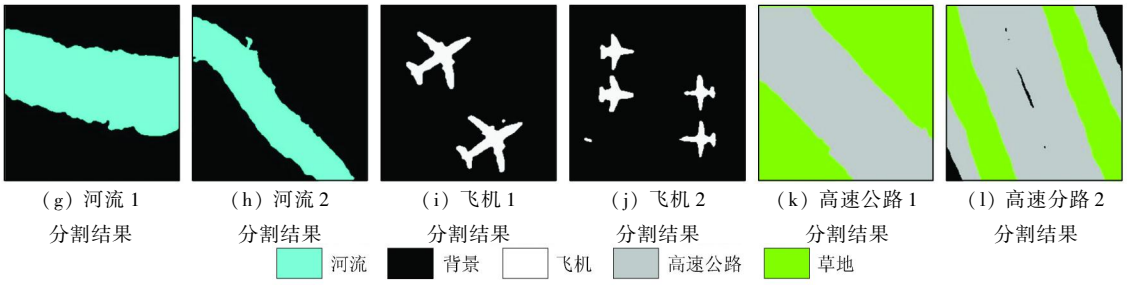


图 4 – 2　UC Merced Land – Use 数据集实验结果

Fig. 4 – 2　Experimental results of UC Merced Land – Use dataset

3　结论

本文提出了一种 DFN – CRF 语义分割方法,在 FCN 基础上加入反卷积融合结构,同时将 CRF 的参数推断过程以 RNN 迭代层的形式嵌入整个网络框架中,构建了端到端的 DFN – CRF 模型。通过对模型的实验与评价,得到如下结论:

1)采用的反卷积融合结构,能够融合多尺度特征,将浅层细节信息和深层语义信息结合起来做模型预测,有效地提高了模型的处理精度。

2)嵌入式的 CRF 模型不仅能够引入上下文信息,更准确地定位地物,进一步提升模型的处理效果,而且端到端的网络处理架构,可以避免后处理,简化训练过程。

3)对比实验表明,随着融合结构结合池化层深度的增加,语义分割结果轮廓更清晰,边缘更细化,处理精度逐渐提高; 空间上下文信息的引入有效提升了语义分割的精度,较好地保持了语义对象内部区域的一致性。

参考文献(References):

[1] 魏云超,赵 耀. 基于 DCNN 的图像语义分割综述[J]. 北京交通大学学报,2016,40(4):82 – 91.
Wei Y C,Zhao Y. A review on image semantic segmentation based on DCNN[J]. Journal of Beijing Jiaotong University,2016,40(4):82 – 91.

[2] 罗 冰. 语义对象分割方法研究[D]. 成都:电子科技大学,2018.
Luo B. Research on segmentation of semantic objects[D]. Chengdu:University of Electronic Science and Technology of China,2018.

[3] 韩 铮,肖志涛. 基于纹元森林和显著性先验的弱监督图像语义分割方法[J]. 电子与信息学报,2018,40(3):610 – 617.
Han Z,Xiao Z T. Weakly supervised semantic segmentation based on semantic texton forest and saliency prior[J]. Journal of Electronics & Information Technology,2018,40(3):610 – 617.

[4] Zhao J,Zhong Y F,Zhang L P. Detail – preserving smoothing classifier based on conditional random fields for high spatial resolution remote sensing imagery[J]. IEEE Transactions on Geoscience & Remote Sensing,2015,53(5):2440 – 2452.

[5] 叶发茂,罗 威,苏燕飞,等. 卷积神经网络特征在遥感图像配准中的应用[J]. 国土资源遥感,2019,31(2):32 – 37. doi:10.6046/gtzyyg.2019.02.05.
Ye F M,Luo W,Su Y F,et al. Application of convolutional neural network feature to remote sensing image registration[J]. Remote Sensing for Land and Resources,2019,31(2):32 – 37. doi:10.6046/gtzyyg.2019.02.05.

[6] Zhao W D,Li S S,Li A,et al. Hyperspectral images classification with convolutional neural network and textural feature using limited training samples[J]. Remote Sensing Letters,2019,10(5):449 – 458.

[7] 张 康,黑保琴,李盛阳,等. 基于 CNN 模型的遥感图像复杂场景分类[J]. 国土资源遥感,2018,30(4):49 – 55. doi:10.6046/gtzyyg.2018.04.08.
Zhang K,Hei B Q,Li S Y,et al. Complex scene classification of remote sensing images based on CNN[J]. Remote Sensing for Land and Resources,2018,30(4):49 – 55. doi:10.6046/gtzyyg.2018.04.08.

[8] Zhang R,Li G Y,Li M L. et al. Fusion of images and point clouds for the semantic segmentation of large – scale 3D scenes based on deep learning[J]. ISPRS Journal of Photogrammetry and Remote Sensing,2018,143:85 – 96.

[9] Kanmffmeyer M,Salberg A B,Jenssen R. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks[C]//Proceedings of Computer Vision and Pattern Recognition Workshops. 2016:680 – 688.

[10] Long J,Shelhamer E,Darrell T. Fully convolutional networks for semantic segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE,2015:3431 – 3440.

[11] Sherrah J. Fully convolutional networks for dense semantic labelling of high – resolution aerial imagery[EB/OL]. (2016 – 06 – 08)[2019 – 08 – 16]. http://arxiv.org/abs/1606.02585v1.

[12] Badrinarayanan V,Handa A,Cipolla R. SegNet:A deep convolutional encoder – decoder architecture for robust semantic pixel – wise labelling[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,309(12):2481 – 2495.

[13] Ronneberger O,Fischer P,Brox T. U – Net:Convolutional networks for biomedical image segmentation[C]//Proceedings of the International Conference on Medical Image Computing and Computer – Assisted Intervention. Berlin:Springer,2015:234 – 241.

[14] Chen L C,Papandreou G,Kokkinos L,et al. DeepLab:semantic

image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 40(4): 834 – 848.

[15] Zheng S, Jayasumana S, Romera – Paredes B, et al. Conditional random fields as recurrent neural networks[C]//IEEE International Conference on Computer Vision. 2015: 1529 – 1537.

[16] Lafferty J, Mccallum A, Pereira F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data[C]// Eighteenth International Conference on Machine Learning. 2001, 3(2): 282 – 289.

[17] Hammersley J M, Clifford P. Markov fields on finite graphs and lattices[EB/OL]. [2012 – 01 – 30]. http://www.statslab.cam.ac.uk/ ~ grg/books/hammfest/hamm – cliff.pdf.

[18] Krahenbuhl P, Koltun V. Efficient inference in fully connected CRFs with gaussian edge potentials[C]//Proceeding NIPS'11 Proceeding of the 24th International Conference on Neural Information Processing Systems Advances in Neural Information Processing Systems. 2011: 109 – 117.

[19] Cheng G, Han J, Lu X Q. Remote sensing image scene classification: benchmark and state of the art[C]//Proceedings of the IEEE. 2017, 105(10): 1865 – 1883.

[20] Yang Y, Newsam S. Bag – of – visual – words and spatial extensions for land – use classification[C]//Proceedings of the 18th Sigspatial International Conference on Advances in Geographic Information Systems. 2010: 270 – 279.

[21] Alberto G G, Sergio O E, Sergiu O, et al. A review on deep learning techniques applied to semantic segmentation[EB/OL]. (2017 – 04 – 22)[2019 – 08 – 16]. http://arxiv.org/abs/1704.06857.

Remote sensing image semantic segmentation using deep fusion convolutional networks and conditional random field

LI Yu[1], XIAO Chunjiao[1], ZHANG Hongqun[1], LI Xiangjuan[2], CHEN Jun[1]

(1. *Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing* 100094, *China*;
2. *School of Computer Science, Xi'an Shiyou University, Xi'an* 710065, *China*)

Abstract: A method for remote sensing image semantic segmentation based on deep fusion convolutional networks and conditional random field is proposed. First, the fully convolutional networks framework with deconvolutional fusion structure is utilized to integrate the pooling – layer results at different levels. The low – level features with rich detail information are combined with the high – level features via deconvolutional fusion module. At the same time, the parameter inference process of conditional random field is embedded in the network architecture by adding recurrent neural networks iteration layers. In addition, the deep fusion convolutional networks and conditional random field model is established. Then, in the model training stage, the abundant detail information and context information in the image are introduced simultaneously to the positive and negative propagation. And lastly, the remote sensing image semantic segmentation is accomplished by the end – to – end framework. Semantic segmentation experiments were performed on the high – resolution optical remote sensing images, and the results show that, with the increase of the depth of deconvolution fusion layer in the model, semantic segmentation results are more refined, and the contour of terrestrial object is more accurate. The introduction of context information also improves the accuracy of image semantic segmentation to a certain extent. It is concluded that the proposed method can better maintain the consistency of internal areas of semantic object and effectively improve the accuracy of semantic segmentation.

Keywords: remote sensing image; semantic segmentation; fully convolutional networks; conditional random field

(责任编辑: 张 仙)