

基于 ResNet 双注意力机制的遥感图像场景分类^①



乔星星^{1,2,3,4}, 施文灶^{1,2,3,4}, 刘芫汐^{1,2,3,4}, 林耀辉^{1,2,3,4}, 何代毅^{1,2,3,4}, 王磊^{1,2,3,4},
温鹏宇^{1,2,3,4}, 孙雯婷^{1,2,3,4}

¹(福建师范大学 光电与信息工程学院, 福州 350007)

²(福建师范大学 福建省光电传感应用工程技术研究中心, 福州 350007)

³(福建师范大学 医学光电科学与技术教育部重点实验室, 福州 350007)

⁴(福建师范大学 福建省光子技术重点实验室, 福州 350007)

通讯作者: 施文灶, E-mail: swz@fjnu.edu.cn

摘要: 针对基于传统机器学习遥感图像场景分类无法快速有效提取图像特征造成分类结果不准确的问题, 提出一种基于注意力残差网络的遥感图像场景分类的方法, 以残差网络为基准模型, 在通道和空间两个维度上建立注意力模块, 实验过程中对参数进行合理有效的设置, 调整网络层数优化模型, 达到对 UC Merced Land-Use 数据集的有效分类. 实验结果表明, 与基于卷积神经网络结构的遥感图像场景分类方法相比, 该方法达到了 98.1% 的准确率.

关键词: 遥感图像; 场景分类; 卷积神经网络; 残差网络; 注意力模块

引用格式: 乔星星, 施文灶, 刘芫汐, 林耀辉, 何代毅, 王磊, 温鹏宇, 孙雯婷. 基于 ResNet 双注意力机制的遥感图像场景分类. 计算机系统应用, 2021, 30(8): 243-248. <http://www.c-s-a.org.cn/1003-3254/8059.html>

Remote Sensing Image Scene Classification Based on ResNet and Dual Attention Mechanism

QIAO Xing-Xing^{1,2,3,4}, SHI Wen-Zao^{1,2,3,4}, LIU Yuan-Xi^{1,2,3,4}, LIN Yao-Hui^{1,2,3,4}, HE Dai-Yi^{1,2,3,4}, WANG Lei^{1,2,3,4},
WEN Peng-Yu^{1,2,3,4}, SUN Wen-Ting^{1,2,3,4}

¹(College of Photonic and Electronic Engineering, Fujian Normal University, Fuzhou 350007, China)

²(Fujian Provincial Engineering Technology Research Center of Photoelectric Sensing Application, Fujian Normal University, Fuzhou 350007, China)

³(Key Laboratory of Optoelectronic Science and Technology for Medicine (Ministry of Education), Fujian Normal University, Fuzhou 350007, China)

⁴(Fujian Provincial Key Laboratory for Photonics Technology, Fujian Normal University, Fuzhou 350007, China)

Abstract: To deal with the inaccurate classification caused by a failure of quick and effective extraction of image features in the remote sensing image scene classification based on existing machine learning methods, we propose a remote sensing image scene classification method based on residual attention network. With the residual network as the benchmark model, attention modules are created in the dimensions of channel and space. For effective classification of the UC Merced Land-Use dataset, parameters are set reasonably and the model that optimizes the number of network layers is fine-tuned. The results show that the accuracy of our method reaches 98.1% compared with that based on the convolution neural network.

Key words: remote sensing image; scene classification; Convolutional Neural Network (CNN); Residue Network (ResNet); attention module

① 基金项目: 国家自然科学基金青年基金 (41701491); 福建省自然科学基金面上项目 (2017J01464, 2018J01619)

Foundation item: Young Scientists Fund of National Natural Science Foundation of China (41701491); General Program of Natural Science Foundation of Fujian Province (2017J01464, 2018J01619)

收稿时间: 2020-11-27; 修改时间: 2020-12-28; 采用时间: 2021-01-13; csa 在线出版时间: 2021-07-31

1 引言

如今遥感技术日渐发达,遥感图像有丰富的地表信息及更高的空间分辨率^[1],遥感图像场景分类成为一个研究热点.图像场景分类是遥感图像数据处理的重要内容,也是处理图像检测,图片识别问题的重要手段,在计算机领域内已经是极其重要且富有研究价值的方向,在城市规划,土地利用等方面也得到了广泛利用.通过人工对图像进行分类虽然可以得到准确率较高的结果,但随着图片种类及数量的剧增,仅靠传统的人工分类难以满足分类领域的要求,深度卷积神经网络(Convolutional Neural Network, CNN)方法已经开始应用于遥感图像场景分类.面对媒体平台的海量遥感图像,如何通过神经网络模型^[2]来提取数据集的深层次特征已成为当前遥感图像场景分类的关键性问题.

遥感图像场景主要包括森林,河流,田地等自然景象和高速公路,飞机场,建筑物等人文景象,图像体现类内有明显相同特征,类间无明显相同特征的特点^[3].场景是对遥感图像数据的高级语义分析,对遥感数据集进行多场景分类的目的是依照图片里的信息对数据进行语义分类,其关键就在于对图像特征的提取.由于数据源特征信息越来越丰富,遥感图像的场景分类研究也从最初的基于颜色^[4]、密度^[5]、变换域纹理^[6]等低级特征到中级特征再到现在的深度学习,其中,中级特征是在低层特征的基础上对其进行编码学习得到,主要有语义的属性、对象和局部语义3类^[7,8].传统的分类方法对基本特征无明显差异的复杂数据难以表现出好的效果,相比之下基于卷积神经网络^[9]的研究方法对数据源信息进行更深度的挖掘,具有更高的精度.

在2012年 AlexNet^[10]成绩超过传统方法后, VGG^[11], InceptionNet^[12], ResNet^[13]等卷积神经网络模型也相继提出,基于卷积神经网络模型的研究已成为图像分类领域的主流趋势. ResNet 模型在遥感图像场景分类的任务中,通过加深网络结构,加强了提取特征的有效性,从而决定了数据集分类的精度. ResNet 分类网络在特征提取方面已经得到了大量的应用,较于其他的卷积神经网络,残差网络中除了第一层外,每一层的输入都是来源于上一层,数据在经过上一层的卷积或者池化操作时,就对其进行了一次滤波器的处理,输入向量在通过池化操作后尺寸变得更小,从而减少网络的参数数量,抑制过拟合的发生,残差神经网络在某种程度上有效处理了深层网络的梯度消失或爆炸的现

象.考虑到注意力机制可以在卷积神经网络中能够学习特定的特征并抑制不需要的特征,有注意力残差学习模块的网络效果更好.基于卷积神经网络的方法通过自主学习方式^[14]抽象出具体的语义特征,因算法的不断改进也显现出强大的层次化特征学习和良好的迁移性^[15].在大量的数据集训练中能表现出比传统方法更好的性能.

本文采用 Python 语言,基于 PyTorch 框架及 ResNet 模型,对 UC Merced Land-Use (UCM) 遥感图像数据集进行场景分类,并将改进算法与其他网络模型方法进行实验对比.

2 方法

2.1 网络结构及原理

本文采用的基础卷积神经网络是 ResNet 模型,该模型将残差学习的思想加入在传统卷积神经网络中,很好地解决了网络深度较深时梯度消失和精度降低的问题,网络层数尽可能地增多,随之模型的分类能力也会得到大幅度的提高.相较于 VGG 网络, ResNet 网络结合了全局平均池化操作,大大降低了参数量.本文的网络架构设计如图1所示,把注意力模块(Convolutional Block Attention Module, CBAM)加在残差网络中,网络输入是包含单特征地物目标的遥感图像,将特征图送入残差网络之后,通过一系列残差与注意力结合的模块(ResBlock+CBAM)以及残差模块进行特征学习提取,后经全连接层输出数据并完成分类.

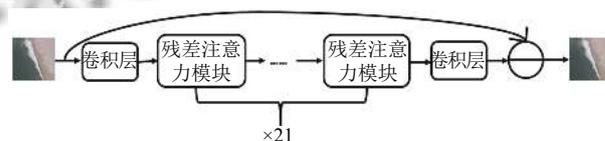


图1 模型结构示意图

基于注意力机制的模块对遥感图像场景分类精度的提高有极大的帮助.由于注意力机制擅长专注挖掘数据的特定信息,实验中融合注意力模块和 ResNet 网络,在给出了特征图后,实现两个层面的注意力模块操作,一个是通道层面的,另一个是空间层面的,然后输出特征图,注意力机制会保留特征图的原始尺寸,而且可以灵活更新通道层面的特征权重,还可以学习到空间层面上各地物目标像素的关联性,能够有效地从背景复杂的遥感图像中提取到特定的特征.

2.2 残差模块

残差模块堆叠形成残差网络, 残差模块的结构如图2所示.

对于某个特征图, 依次经过两个卷积层, 学习并提取到图像特征, 其中第一个卷积层是加了 ReLU 激活函数的, 将最初的特征图和卷积运算后得到的特征图融合, 再经过一个激活函数, 便是最后的结果. 其过程可用式 (1) 表示:

$$y = F(x, \{w_i\}) + x \quad (1)$$

其中, x 是指输入, y 是输出特征图, $F(x, \{w_i\})$ 是网络内卷积操作, $\{w_i\}$ 是指训练参数. 残差模块的主要原理是可以越过几层网络后作用在后面的层, 其能够从某一层激活后, 立马接到其他层或比较深的层, 在网络训练中进行反向传播时, 梯度容易消失, 而残差模块可以很好地改善这个问题.

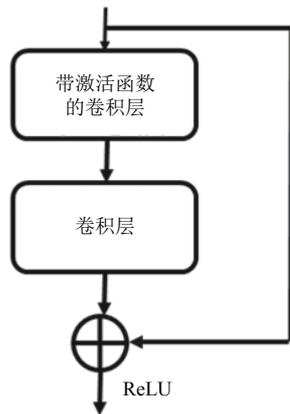


图2 残差模块结构图

2.3 注意力模块

在 CBAM 中, 注意力模块通常分两个, 一个是通道注意力模块, 另一个是空间注意力模块, 分别如图3、图4所示.

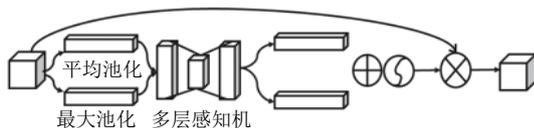


图3 通道注意力模块结构

通道注意力主要是压缩了输入的特征图空间层面的内容, 在特征图的空间内容上分别进行不同的池化操作, 得到两个相应且不一样的表示符来形容空间信息, 然后分别将他们传递到共享网络来自适应调整不

同通道的权重, 最后将得到的特征相加合并, 通过归一化函数把所得的结果进行处理, 获得权重参数, 输出特征向量, 之后与原特征进行乘积运算则可生成新的特征. 其过程可用式 (2) 表示:

$$M_C(x) = \sigma(MLP(AvgPool(x)) + MLP(MaxPool(x))) \otimes x \\ = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \otimes x \quad (2)$$

式中, x 指输入信息, W_1 和 W_0 分别表示共享网络的权重系数, \otimes 指的是像素分别进行乘积运算.

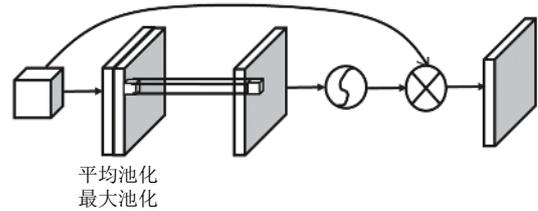


图4 空间注意力模块结构

空间注意力模块是对通道注意力的补充, 它的目的是挖掘最有意义的内容信息, 输入特征图后, 根据通道轴进行平均池化和最大池化操作, 然后将他们再沿着通道方向连接起来生成一个有效的特征表示符, 最后经过卷积运算, 挖掘到各个位置信息之间的内在联系, 然后利用归一化函数处理结果, 输出最终注意力图, 将其与输入特征图做乘积运算来更新特征图的权重.

2.4 注意力残差模块

通过注意力模块, 可以使网络专注于图像的特定区域信息, 从而更高效地完成特定的特征提取与重建. 由于一个标准卷积层具有一定的通道数, 每个通道都自适应学习某一特定特征, 通过注意力模块能够在所有卷积运算中专注对特定特征的提取并忽略无关特征的存在. 通过对特征的有效学习后各通道会更新出权重, 令其与输送进去的特征图进行乘积运算, 使网络模型专注于特定的特征. 近年来注意力机制被广泛使用在深度学习任务中, 利用注意力机制的原理能够处理一部分卷积操作存在的问题, 增强卷积的学习能力.

本文融合了注意力模块与残差模块组建了注意力残差模块, 形成了更深层次的模块, 从而使网络模型的分类能力有了新的突破. 如图5所示, 特征图先进行两次卷积特征提取, 接着经过双注意力机制模块, 将得到的结果与输入图直接相加, 通过 ReLU 激活函数输出. 其过程可表示为式 (3):

$$y = M_s(M_c(F(x, \{w_i\}))) + x \quad (3)$$



图5 注意力残差模块结构

3 实验与结果分析

3.1 数据集

本文实验使用的是 UC Merced Land-Use 数据集, 数据集中的图像是美国地质勘探局从城市区域图像收集的大型图像中手动提取的, 分布在全国的各个城市地区. 图像像素大小为 256×256, 场景图像有 21 类, 实验中数据增强后每类有 250 张, 共 5250 张. 实验中将数据集按训练集 60%, 验证集 40% 配置. 图 6 显示了每类图像的一个示例.

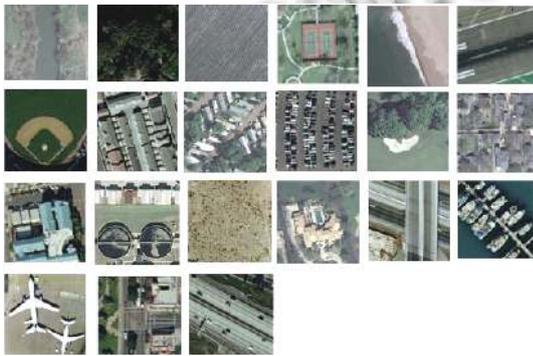


图6 数据集各类示例

卷积神经网络模型具有很强的学习及表达能力, 在数据量不是很充足时, 一般利用数据增强来增加训练样本的数量, 使模型更具鲁棒性, 降低过拟合效应产生. 在图像分类任务中, 通常对现有图像进行拉伸, 翻转, 裁剪, 色彩变换得到更多实验数据. 通过对数据的处理, 可以提高模型的泛化能力, 提取特征能力及学习能力更强. 本实验采用数据增强方法扩大训练样本, 达到增加数据量的目的.

3.2 实验结果与对比

在实验中, 优化器选择 Adam, 初始学习率取值为 0.01, 批尺寸为 16, 一共训练 3000 次, 最终在训练集和验证集上的准确率分别达到了 96.2553% 和 98.1061%. 实验环境配置如表 1 所示.

初始学习率为 0.01, 实验中学习率的值过大, 也许会错过最佳的学习率取值, 准确率下降; 如果取值过小, 会导致收敛很难进行, 调整后的学习率提高了模型的

优化效率, 模型收敛到局部极小值的速度会加快, 也会得到更好的分类精度.

表 1 实验环境配置

参数	配置
操作系统	Windows 10
GPU	GeForce GTX 1080Ti
开发平台	JetBrains PyCharm Community Edition
开发语言	Python 3.7
深度学习框架	PyTorch 1.7

训练次数为 3000 次, 当训练次数太少时, 模型不能充分学习特征, 导致准确率较低; 训练次数过多时, 测试精度先增大后减小, 需要减少迭代次数, 否则容易产生过拟合现象.

本实验批尺寸设置为 16, 小的批尺寸引入的随机性更大, 提取到的图像特征不具有代表性, 难以达到收敛; 大的批尺寸数据集计算的梯度下降方向会更加准确, 但每次纠正参数的时间会增加.

本文实验评估了基于 ResNet 双注意力机制方法的分类能力. 图 7 显示了总体精度, 该方法在训练集和验证集上都取得了较好的分类精度. 图 8 显示了混淆矩阵, 第 i 行和第 j 列的元素表示第 i 类中被分类为第 j 类的测试样本类. 具有相似视觉特征的类可被分类为不正确的类别, 不正确的分类显著减少.

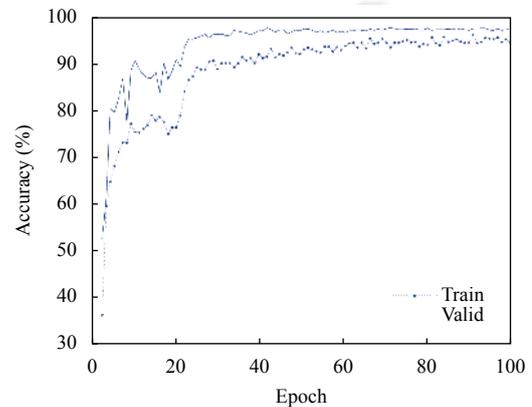


图7 本文方法在 UCM 数据集上的分类精度

Kappa 系数用来检测预测和实际结果是否一致, 也能用来评价分类模型, 是基于混淆矩阵计算的. Kappa 系数的意义就是用来评估图像分类模型的性能, 在遥感领域里通常拿来衡量分类性能的好坏, 若两个结果相差较大, 那么 Kappa 系数就会很小.

对于遥感图像场景分类问题, 所谓一致性就是模型预测结果和实际分类结果是否一致. Kappa 最后求

出的结果在-1~1间,但一般 Kappa 落于 0~1 内,当 $Kappa \geq 0.75$ 时,说明实际类别和预测类别达到了相对较高的一致程度,而本实验 Kappa 系数达到 0.97,进一步说明了该模型的分性能较优越。

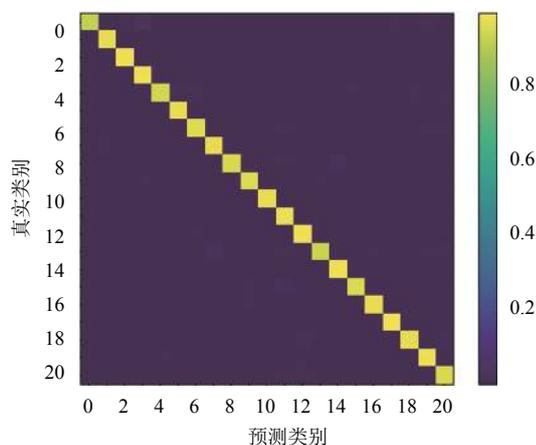


图8 混淆矩阵

ResNet 允许网络尽可能的加深,从实用的角度看,将 5 种结构的 ResNet 模型分别进行实验对比,如表 2 所示,发现 ResNet101 可以取得更高的分类精度,时间上虽然略多于 ResNet18、34、50 三种结构,但为了保证模型在分类上的有效性,故将 CBAM 方法建立在 ResNet101 的基础上完成。

表2 ResNet 各类模型性能对比

模型	验证集时间(s)	准确率(%)
ResNet18	38.25	97.15
ResNet34	39.30	97.72
ResNet50	39.47	97.93
ResNet101	42.49	98.10
ResNet152	43.93	98.01

(1) 机器学习方法与卷积神经网络方法对比

K 近邻算法 (KNN) 和支持向量机算法 (SVM) 在 UCM 数据集上的分类精度^[16]与卷积神经网络模型做出比较,各类算法实验数据总结如表 3 所示,相比传统的分类方法,利用不同的卷积神经网络模型展开实验,可以得到更高的分类精度.在研究与实践中证明,通过机器学习算法对图像特征进行提取和描述时,也易丢掉某些有用的内容,无法保证获取令人满意的结果.通过卷积神经网络对遥感图像进行多场景分类任务,将图像作为输入,通过训练模型来对遥感图像特征进行学习并提取,然后根据特征信息进行分类,很大程度上节省了特征学习和提取的时间,提高了识别的效率。

(2) 卷积神经网络基础模型与本文方法对比

实验模型与卷积神经网络基准模型进行分类对比,由于注意力机制可以在 CNN 模型中能够专注于特定特征的学习,忽略不重要的背景内容,所以本文在 CNN 模型中应用注意力机制使得网络专注于学习图像中具体场景的结构而略去背景图的内容细节,从而提高图像分类的性能.加入注意力机制模块的残差神经网络在遥感图像场景分类中达到了 98.1% 的精度,相比于表 3 中的其他模型^[17],进一步证明了所提方法的优越性。

表3 各类方法与实验结果

方法	准确率(%)
KNN	90.24
SVM	93.81
CaffeNet	94.52
GoogleNet	94.04
VGG-F	95.67
VGG-S	95.42
VGG-M	94.76
VGG-16	95.57
VGG-19	94.29
ResNet	95.75
ResNet&VGG-S	95.17
ResNet&VGG-S&VGG-16	95.90
Ours (ResNet+attention)	98.10

3.3 分析

实验的主要研究内容是基于注意力机制的图像分类算法研究,在 ResNet 网络中加入了注意力模块 CBAM.将该模型用于场景分类,通过在公开数据集 UC Merced Land-Use 上训练,与残差网络进行对比得知,随着注意力模块的加入,网络分类的性能也逐步提升,以此可以表明本算法引入的注意力模块可以提升网络分类的性能。

本文所提的注意力残差网络在 ResNet 上融合 CBAM 注意力模型,其性能已经比复杂的深度神经网络要好,这说明注意力机制可以很好的应用于遥感图像分类领域,简单有效.进一步来说,本文所提出的双注意力网络,由于其设计的通道注意力模块可以对遥感图像条纹强度进行模拟,空间注意力模块可以微调特征图空间每个位置的像素值,使得效果在客观的评价指标上具有更好的性能.通过大量实验,在分类的评价指标上,所提算法都有明显的提升。

训练时,发现验证集的准确率高于训练集,一个可能原因是数据集切分不均匀,这可能造成训练集的内部方差大于验证集,导致训练集的误差更大.对此重新分布训练集和验证集的比例,对数据集进行扩充,更多

的数据往往会得到更多的分布信息,但结果依然是验证集准确率大于训练集的准确率.另一方面训练的精度是每个批次产生的,而验证的精度一般是训练一轮后产生的,验证时的模型是训练一个个批次之后的,有一定的滞后性;一般训练得差不多的模型用来验证,精度要高一点.

4 结语

实验的主要研究内容是基于 ResNet 双注意力机制的遥感图像场景分类.由于注意力机制可以在 CNN 模型中自适应地激励需要的特征而抑制不需要的特征,所以本实验 CNN 模型中应用注意力机制使得网络专注于标志性图像特征的结构而对背景图的内容进行抑制,从而提高数据集的分类性能.实验表明,本文所提出的模型能够在 UC Merced Land-Use 数据集取得较好的分类结果.但如何提高模型普适性,也是下一步的研究方向.

参考文献

- 1 张立福,彭明媛,孙雪剑,等.遥感数据融合研究进展与文献定量分析(1992–2018).遥感学报,2019,23(4):603–619. [doi: 10.11834/jrs.20199073]
- 2 黄立威,江碧涛,吕守业,等.基于深度学习的推荐系统研究综述.计算机学报,2018,41(7):1619–1647. [doi: 10.11897/SP.J.1016.2018.01619]
- 3 何霞,汤一平,王丽冉,等.基于 Faster RCNNH 的多任务分层图像检索技术.计算机科学,2019,46(3):303–313. [doi: 10.11896/j.issn.1002-137X.2019.03.045]
- 4 辛晨.基于深度学习的图像分类及应用研究[硕士学位论文].北京:中国科学院大学(中国科学院遥感与数字地球研究所),2017.
- 5 Liu CA, Chen ZX, Shao Y, *et al.* Research advances of SAR remote sensing for agriculture applications: A review. Journal of Integrative Agriculture, 2019, 18(3): 506–525. [doi: 10.1016/S2095-3119(18)62016-7]
- 6 Wang WX, Wang MF, Li HX, *et al.* Pavement crack image acquisition methods and crack extraction algorithms: A review. Journal of Traffic and Transportation Engineering (English Edition), 2019, 6(6): 535–556. [doi: 10.1016/j.jtte.2019.10.001]
- 7 Zhang QS, Zhu SC. Visual interpretability for deep learning: A survey. Frontiers of Information Technology & Electronic Engineering, 2018, 19(1): 27–39.
- 8 Gong XY, Su H, Xu D, *et al.* An overview of contour detection approaches. International Journal of Automation and Computing, 2018, 15(6): 656–672. [doi: 10.1007/s11633-018-1117-z]
- 9 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.计算机工程与应用,2019,55(12):20–36. [doi: 10.3778/j.issn.1002-8331.1903-0031]
- 10 Zhang YD, Govindaraj VV, Tang CS, *et al.* High performance multiple sclerosis classification by data augmentation and AlexNet Transfer Learning Model. Journal of Medical Imaging and Health Informatics, 2019, 9(2): 2012–2021.
- 11 王相龙,胡钊政,穆孟超,等.基于 VGG 深度卷积神经网络和空间分布的道路裂纹种类识别.交通信息与安全,2019,37(6):95–102.
- 12 赵冰,李平,代明睿,等.基于深度学习的铁路图像场景分类优化研究.计算机系统应用,2019,28(6):228–234. [doi: 10.15888/j.cnki.csa.006937]
- 13 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778.
- 14 田萱,王亮,丁琪.基于深度学习的图像语义分割方法综述.软件学报,2019,30(2):440–468. [doi: 10.13328/j.cnki.jos.005659]
- 15 朱祺琪.面向高分辨率遥感影像场景语义理解的概率主题模型研究[博士学位论文].武汉:武汉大学,2018.
- 16 张馨月.基于 DCNN 的高分辨率遥感图像场景分类[硕士学位论文].长春:吉林大学,2019.
- 17 Han W, Feng RY, Wang LZ, *et al.* A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 145: 23–43. [doi: 10.1016/j.isprsjprs.2017.11.004]