https://doi.org/10.3799/dqkx.2021.218

集成空间变换结构与深度残差网络的遥感影像场景分类方法

孟亦菲[1],郑贵洲[1*],冀炜臻[2]

1. 中国地质大学地理与信息工程学院,湖北武汉 430078
2. 江西理工大学土木与测绘工程学院,江西赣州 341000

摘 要: 针对传统高分辨率遥感影像的场景分类效率较低,以及卷积神经网络在遥感影像场景分类上由于空间不变性而导致的分类精度不高的问题,提出了一种结合空间变换网络和迁移学习的高分辨率遥感影像场景分类算法.首先,利用ImageNet数据集训练深度残差网络ResNet101得到预训练模型,通过知识迁移提高模型目标探测效率;之后在模型中嵌入空间变换结构,使模型能够主动在空间上变换特征映射,提高模型的鲁棒性;最后,在模型中添加Dropout层减小模型出现过拟合的概率.本方法在AID和NWPU-RESISC45两种不同规模的高分遥感影像数据集上进行了验证,在只有20%训练样本的情况下仍达到了94.30%和93.63%的分类精度.实验结果表明本次改进模型具有更好的特征提取能力,针对易误分类场景的分类结果更优.

关键词: 深度学习;残差网络;空间变换网络;迁移学习;场景分类;遥感.

中图分类号: P237 **文章编号:** 1000-2383(2023)09-3526-13 **收稿日期:** 2021-07-07

Remote Sensing Image Scene Classification Method Integrating Spatial Transformation Structure and Depth Residual Network

Meng Yifei[1], Zheng Guizhou[1*], Ji Weizhen[2]

1. *School of Geography and Information Engineering, China University of Geosciences, Wuhan* 430078, *China*
2. *School of Architectural and Surveying and Mapping Engineering, Jiangxi University of Science and Technology, Ganzhou* 341000, *China*

Abstract: In order to solve the problem that the remote sensing image with small sample set can easily lead to the over-fitting of the training model and the low classification accuracy caused by the spatial invariance of convolution neural network in remote sensing image scene classification, a high-resolution remote sensing image scene classification algorithm based on spatial transformation network and transfer learning is proposed. Firstly, the ImageNet dataset is used to train the deep residual network ResNet101 to obtain the pre-training model, and the training efficiency of the model is improved through knowledge transfer. Then, the spatial transformation structure is embedded in the model, so that the model can actively transform the feature mapping in space and improve the robustness of the model. Finally, the Dropout layer is added to the model to reduce the probability of over-fitting of the model. This method is verified on two high-score remote sensing image data sets of AID and NWPU-

基金项目: 国家自然科学基金重点项目(No. 42130309);山西省大同经济技术开发区城市地质调查项目(No. 2022030115).

作者简介: 孟亦菲(1998—),女,硕士,主要研究方向为深度学习、遥感场景分类. ORCID: 0000-0002-5699-7837. E-mail: cugmyf@cug.edu.cn

* **通讯作者:** 郑贵洲,ORCID: 0000-0002-2890-6395. E-mail: zhenggz@cug.edu.cn

引用格式: 孟亦菲,郑贵洲,冀炜臻,2023.集成空间变换结构与深度残差网络的遥感影像场景分类方法.地球科学,48(9):3526-3538.

Citation: Meng Yifei, Zheng Guizhou, Ji Weizhen, 2023. Remote Sensing Image Scene Classification Method Integrating Spatial Transformation Structure and Depth Residual Network. *Earth Science*, 48(9): 3526-3538.

RESISC45, and the classification accuracy of 94.30% and 93.63% is achieved in the case of only 20% training samples. The experimental results show that the improved model has better feature extraction ability and better classification results for misclassification scenarios.

Key words: deep learning; residual network; spatial transformation networks; transfer learning; scene classification; remote sensing.

0　引言

随着对地观测技术的快速发展,遥感影像在分辨率方面已有很大突破,现存高分卫星(例如IKONOS、Quick Bird)的遥感影像可达到米级甚至是亚米级的分辨率(李冠东等,2019).高空间分辨率遥感影像反映的空间纹理特征和语义信息更为丰富(张康等,2018),且已被广泛应用于城市规划、灾害监测、军事情报、成矿预测和环境监测等方面(李德仁等,2017;余姝辰等,2019;徐永洋等,2020;左仁广等,2021).但是,高分辨率遥感影像的类内多样性和类间相似性较高,单纯利用传统的基于像素和面向对象的遥感影像分类方法已不能对其进行准确地识别与分类.如何精确、高效、显著地解译遥感影像数据是遥感影像研究工作的关键.

近十年间,研究者们在遥感影像场景分类方面做了大量的工作,特别是在场景特征表征方面(van de Sande *et al.*, 2010;Cheng *et al.*, 2013).早期的遥感场景分类方法主要是基于底层特征和中层特征的方法.基于底层特征的算法是通过结合影像颜色、形状和纹理等特征描述并使用相应的分类器分类(Oliva and Torralba, 2001; Yang and Newsam, 2013).典型的方法包括尺度不变特征变换(scale-invariant feature transform, 简称SIFT)、定向梯度直方图(histogram of gradient, 简称HOG)(Dalal and Triggs, 2005; Yang and Newsam, 2008)等.但这类方法在描述复杂场景时存在局限性,难以得到理想的分类精度.基于中层特征的分类方法是通过字典学习,利用视觉词典构建中层语义特征提供更完整的图像描述,常用的方法包括基于词袋模型、特征编码和主题模型的场景分类等(Wallraven *et al.*, 2003; Perronnin *et al.*, 2010; Văduva *et al.*, 2013).虽然这种方法能够在一定程度上提高对遥感影像语义信息的表达能力,但受词袋特征单词模糊和冗余问题的限制,模型泛化能力不足,难以提取遥感图像复杂场景的语义特征信息(余东行等,2020).

深度学习的出现使实现遥感影像的高层特征提取成为了可能.通常,高层特征能够同时包含语义表征和抽象表征,克服了浅层学习模型以及人工提取特征所带来的一系列工作量大或效率低等问题.深度学习方法经过多年的发展在计算机视觉任务方面表现出色,其中卷积神经网络作为图像领域特征提取的典型深度学习模型受到了广泛关注,如AlexNet(Krizhevsky *et al.*, 2012)、CaffeNet(Jia *et al.*, 2014)、VGGNet(Simonyan and Zisserman, 2014)、GoogLeNet(Szegedy *et al.*, 2015)和ResNet(He *et al.*, 2016).然而,地表异质性等因素的影响使得这些模型的特征提取机制无法精细耦合地物类间的划分边界,同时,如何在有限的人工标注数据下训练有效的深度神经网络模型也是遥感影像场景分类的关键问题.近期许多工作(Donahue *et al.*, 2014; Oquab *et al.*, 2014)已经证明,在像ImageNet这样的大型数据集上预先训练、网络学习到的权重可以转移到有限训练数据的识别任务中,可以在一定程度上解决遥感影像数据不足的问题.同时,因对输入数据空间变换具有不变性,卷积神经网络识别一些存在变形的复杂影像数据的能力有限.空间变换网络(spatial transformation network, 简称STN)(Jaderberg *et al.*, 2015)概念的提出,在一定程度上缓解了遥感影像降质和变形对场景识别的影响.王瑞琛(2018)结合传统残差网络结构与双塔结构的特点提出残差双塔网络,并在双塔结构中的一条支路加入空间变换结构,学习两个支路输入图像之间变换参数.基于此,本文综合空间变换网络和迁移学习的优势,提出一种基于改进残差网络(ResNet)的遥感影像场景分类方法.该方法将ResNet101模型在ImageNet数据集训练得到的卷积层迁移至原模型上进行训练,并将空间变化结构嵌入到预训练模型中,增强模型对输入数据的鲁棒性和分类能力;同时在模型中添加Dropout层解决模型过拟合问题,提高模型训练效率.

1 研究方法

1.1 ResNet101模型

ResNet模型的初衷是为了解决直接将网络组合形成深层网络时存在的梯度弥散和精度下降问题，它允许网络尽可能的加深.ResNet引入了一个全新的结构——残差模块（residual block）.ResNet的残差模块结构如图1所示.其在反向传播时的梯度计算公式如下：

$$\frac{\partial H(x)}{\partial x}=\frac{\partial F(x)}{\partial x}+1, \tag{1}$$

式中，x为某一段神经网络的输入；$H(x)$表示期望输出；$F(x)=H(x)-x$，表示学习的目标.

ResNet模型通过引入残差模块解决了网络深度增加引起的退化问题，是最常用的卷积神经网络之一（程国轩等，2018）.He *et al.*（2016）的研究表明，ResNets（包括ResNet-18、ResNet-34、ResNet-50、ResNet-101、ResNet-152）在ImageNet数据集中比其他CNN模型在图像分类方面表现更好，这表明ResNets能够很好地提取图像特征.

在综合考虑网络的特征提取能力以及总计算量等因素基础上，本文采用ResNet101模型进行影像特征的提取与分类.ResNet101的Backbone部分如图2所示.图2包括ResNet101各阶段详细构成和Bottleneck（He *et al.*，2016）详细构成两部分.其中CONV（convolution Layer）表示卷积层，MAX-POOL（max pooling layer）表示最大池化层，BTNK表示残差模块Bottleneck部分.ResNet共经过5个阶段，第1个阶段对输入数据进行预处理，后4个阶段都由Bottleneck组成，分别包括3、4、23、3个Bottleneck. Bottleneck结构能够在很大程度上降低网络的参数量和计算复杂度（Srinivas *et al.*，2021），该结构存在两种情况，第1种情况为输入通道数与输出通道数不同（Bottleneck1），其涉及到的4个参数C、W、$C1$、S分别代表输入通道数、宽度、输出通道数和步长.第2种情况为输入通道数与输出通道数相同（Bottleneck2），涉及到2个参数为C和W，即输入通道数和宽度.输入影像经过ResNet101的5个阶段，其输出经过全局平均池化与全连接层得到特征向量，最后通过Softmax分类器进行类别划分.

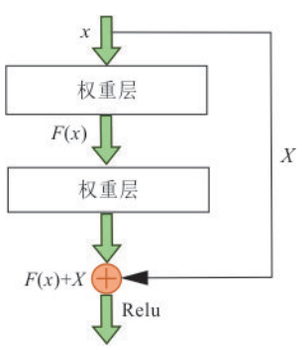

图1 残差学习模块

Fig.1 Residual learning module

1.2 基于改进的ResNet101的高分遥感影像场景分类

本文提出的改进ResNet模型主要包括3个部分，分别为：模型迁移学习、空间变换结构最优输入获取、特征提取与分类，分类流程如图3所示.首先，利用ImageNet数据集对ResNet101训练，保留预训练模型中卷积部分的参数，调整最后的全连接层，获得预训练模型；然后，在模型卷积结构前嵌入空间变换结构，提取输入遥感影像即原始输入I的关注区域，通过映射变换获取模型最优输入I'；最后，利用预训练的ResNet101模型对最优输入I'特征提取，通过Softmax分类器得到最终分类结果.改进后的模型表示为SF-ResNet101.

1.2.1 迁移学习 深度学习的大多数任务通常假设在训练和测试时所采用数据服从相同的输入特征分布，但在现实应用中，实现这个假设十分困难.迁移学习是指利用已有方法和数据去帮助解决具有相似性或相近性任务的方法（Pan and Yang，2010；Weiss *et al.*，2016）.通过迁移学习将网络在大规模数据集中学习到的知识转移到少量的遥感影像数据中，然后设置一个较小的学习率，对模型进行微调即可实现对影像特征学习效果的改进.迁移学习的使用促进了CNN技术在少量标记样本数据集的应用.因此，在高质量遥感图像标签数据匮乏的情况下，为了节省训练时间和计算能力，本文选择采用迁移学习来进行遥感影像场景分类.

迁移学习有两种策略：微调（finetuning）、冻结与训练（freeze and train）.微调是指利用基础数据集得到预训练网络，并在目标数据集上对所有层进行训练.冻结与训练是指冻结部分层的权重，对其余层进行微调（Han *et al.*，2021）.迁移学习策略的选择通常取决于目标数据集的大小，以及目标数据集与基础数据集之间的相似度.本文使用大规模自然数据集ImageNet做为ResNet101网络的源域，考虑到实验采用的aerial images datasets（下文简称AID）数据集和NWPU-RESISC45数据集与基础数据集相

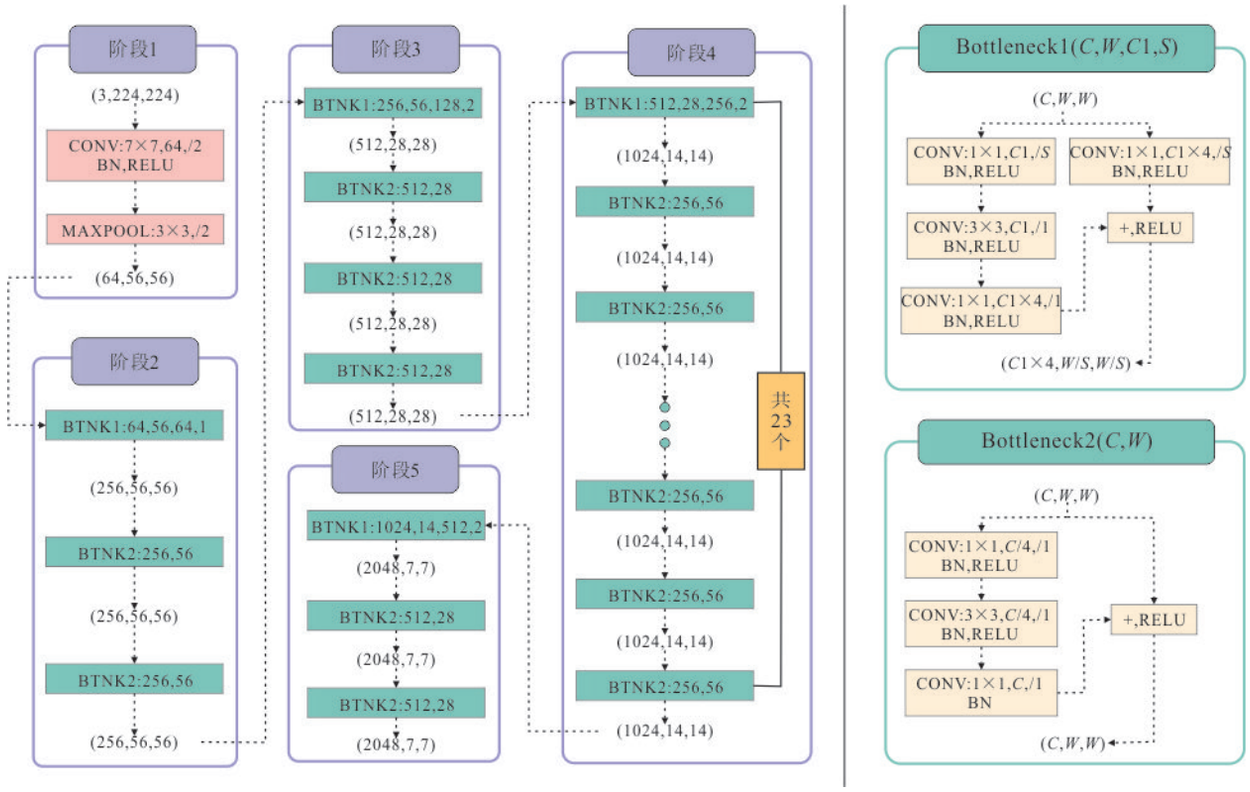

图 2　ResNet101 的 Backbone 部分

Fig.2　The Backbone part of ResNet101

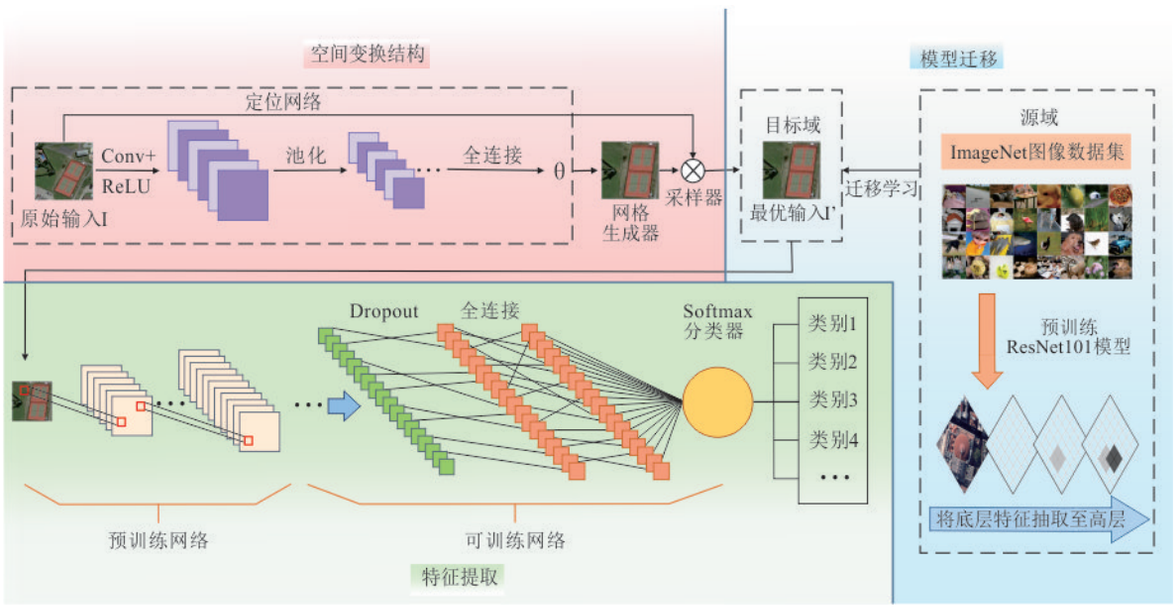

图 3　遥感影像场景分类流程

Fig.3　Flow chart of scene classification of remote sensing image

比数据量小,且场景差距略大;因此,本文采取冻结与训练策略.由于深度学习模型最初的几层网络用于捕获影像中曲线、边缘等基本特征,本文对预训练好的 ResNet101 的前 10 层网络参数冻结迁移,保留其泛性特征,对后面网络层进行训练微调.

1.2.2　空间变换网络　本文将空间变换结构(Jaderberg *et al.*, 2015)嵌入到 ResNet101 网络中,使模型能够针对输入图像自发地预测并调整空间变换

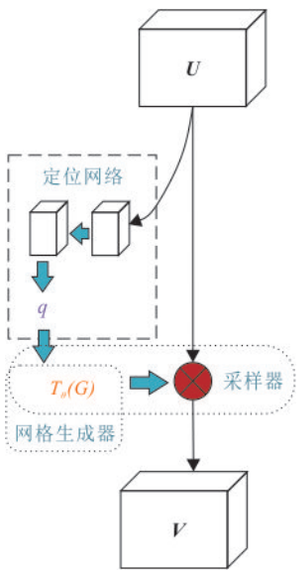

图 4 空间变换结构

Fig.4 Spatial transformation structure

参数，具有对输入数据保持空间不变的能力，不需要额外的训练监督、人工数据增强（如旋转、平移、缩放、倾斜、裁剪）或数据归一化技术．空间变换结构有 3 个组成部分，分别是定位网络、网格生成器和采样器，如图 4 所示．

（1）定位网络．输入的特征图维度为 $U \in R^{H \times W \times C}$，其中 H、W、C 分别为特征图的高度、宽度和通道数量．输出一个变换参数 θ，θ 的大小由特定的变换来决定，在二维仿射变换时，θ 是一个 2×3 的六维向量．

（2）网格生成器．网格生成器利用定位网络层训练输出的变换参数 θ 构造采样网络，得到变换前后像素坐标点的映射关系 $\tau_\theta(G_i)$．假设网格生成器输入图像的像素点坐标为 $\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix}$，输出图像的像素点坐标为 $\begin{pmatrix} x_i^t \\ y_i^t \end{pmatrix}$，其中 $\tau_\theta(G_i)$ 表示一种二维映射函数，其仿射变换表达式如公式（2）所示：

$$\begin{pmatrix} x_i^s \\ y_i^s \end{pmatrix} = \tau_\theta(G_i) = P_\theta \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^t \\ y_i^t \\ 1 \end{pmatrix}. \quad (2)$$

（3）采样器．采样器，通过某种插值方法来确定图像中某一个像素点的灰度值，提供像素采集的功能．当采用双线性插值方法时，公式如下：

$$V_i^c = \sum_n^H \sum_m^W U_{nm}^c \max\left(0, 1 - \left| x_i^s - m \right|\right) \max\left(0, 1 - \left| y_i^s - n \right|\right), \quad (3)$$

式中，V_i^c 为输出特征图上第 c 个通道某一点的灰度值，U_{nm}^c 为输入特征图上第 c 个通道点 (n, m) 的灰度值．且公式（3）对 V_i^c 和 $\left(x_i^s, y_i^s\right)$ 可以求导，即空间变换结构在网络中通过不断训练来修正参数，在提高模型对数据集的识别率和分类精度上有很大帮助．为了使空间变换结构直接作用于输入，本文将空间变换结构嵌入到改进模型的第一部分，使输入影像首先经过空间变换结构，不断调整变换参数 θ 对影像进行映射变换以得到最优输入．最后，将最优输入用于后续的特征提取和分类任务．

1.2.3 Dropout 算法 ResNet101 模型包含的参数数量达 8×10^7，这种含有大量参数的模型在面对小样本数据集训练时易产生过拟合现象．Dropout 能够有效缓解过拟合，在一定程度上具有正则化的效果（He and Chen，2019）．Dropout 在训练过程中随机删除隐藏层神经元，以在不同批量上训练不同的神经网络架构．图 5 比较了不同连接机制下的神经元，在图 5a 中每一层的神经元都与另一层的任意神经元相关联，图 5b 中的一些神经元则被随机删除，以降低计算成本．此外 Dropout 层还可以削弱神经元之间的耦合，减少过拟合，从而有利于数据高层特征的提取．本文改进模型通过在全连通层和输出层之间增加 Dropout 层来提高收敛速度，解决了模型训练时由于迭代次数增加导致的过拟合问题．

1.3 交叉熵损失函数

深度学习中利用损失函数来估量模型的预测值与真实值的不一致程度，从而度量模型的性能，提供模型的优化方向．损失函数分很多种，因此，选择合适的损失函数可以使模型达到更好的训练效果．在机器学习中，训练数据的分布是固定的，因此最小化的模型数据分布与训练数据之间的相对熵

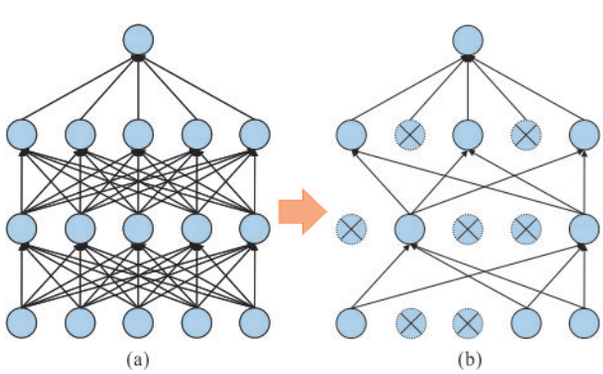

图 5 Dropout 原理示意

Fig.5 Schematic diagram of Dropout principle

(KL 散度)等价于最小化交叉熵,为了使学到的模型更贴合真实数据分布,本文采用交叉熵损失函数(Berman *et al.*,2018)配合 Softmax 层输出类别概率,交叉熵损失函数表达式如下:

$$L(y,\hat{y})=-\sum_{i=1}^{M}y^{(i)}\log_{10}\hat{y}^{(i)}+(1-y^{(i)})\log_{10}(1-\hat{y}^{(i)}),\tag{4}$$

式中,y 和 $\hat{y}$ 分别为图像真实类别和模型预测类别,M 表示样本数量.利用交叉熵损失函数得到的预测类别与真实类别之间的损失值判断预测结果的好坏,从而更有效地训练网络模型.

2　实验设计与结果分析

2.1　实验数据集

为验证 SF-ResNet101 模型的有效性,实验采用两个公开的大规模遥感影像数据集:AID 数据集和 NWPU-RESISC45 数据集.AID 数据集(Xia *et al.*,2017)是由华中科技大学和武汉大学从 GoogleEarth 影像中收集获得的大规模高分遥感场景数据集,该数据集于 2017 年发布.AID 数据集共有 10 000 张图像,包含河流、草地、机场等 30 类遥感场景,每个类别包含 220～420 张不等数量的影像,每张像素尺寸为 600×600,图像空间分辨率为 0.5～8.0 m.图 6 为 AID 数据集部分场景示例.

NWPU-RESISC45 数据集(Cheng *et al.*,2017)于 2017 年由西北工业大学从 GoogleEarth 影像中收集并创建,相较于 AID 数据集,数据量更大、类别更多.其包含海滩、丛林、教堂等 45 类遥感场景,每个类别有 700 张影像,整体共计 31 500 张,每张图像像素尺寸为 256×256,空间分辨率为 0.2～30.0 m.图 7 为 NWPU-RESISC45 数据集部分场景示例.

2.2　实验设置

实验操作系统为 Windows 10 系统,实验基于 Pytorch 框架,处理器为 12 核的 AMD Ryzen 9 3900X,主频为 3.79 GHz,内存为 32 GB,GPU 为 NVIDIA 公司的 RTX2080Ti,显存为 12 GB,Python 版本为 3.7,CUDA 版本为 9.0,cudnn 版本为 8.0.

在实验数据集训练比率选取方面,为了选择合适的训练比率使模型训练效率最大化,本文设置了 4 组不同的训练比率探究对实验结果的影响.如表 1 所示,两组数据集随着训练样本数量的增加,分类准确率均相应提升.考虑到训练比率设置过大会增加模型的训练时长,并且当数据集包含数据量较大时,设置较小的训练比率通常可以满足模型的完整训练需求(Zhang *et al.*,2019).因此,对于 AID 数据集,随机选取每类场景数据总量的 20%、50% 作为训练数据,余下的 80%、50% 作为测试数据.对于

图 6　AID 数据集部分场景示例

Fig.6　Example images of AID dataset

图 7 NWPU-RESISC45 数据集部分场景示例

Fig.7 Example images of NWPU-RESISC45 dataset

表 1 不同训练比率设置下 SF-ResNet101 模型测试集精度对比

Table 1 Accuracy comparison of SF-RESNET 101 model test sets under different training ratio Settings

训练比率	test_acc (%)			
	10%	20%	50%	80%
AID	91.65	**94.30**	**96.52**	96.81
NWPU	**91.66**	**93.63**	93.75	93.77

NWPU-RESISC45 数据集，随机选取每类场景数据总量的 10%、20% 作为训练数据，余下的 90%、80% 作为测试数据. 为了便于及时调整模型超参数，分别从两个数据集的训练数据中划分 20% 作为验证数据集，通过多次验证确定网络最优结构.

实验以批（batch）为单位将数据集输入到模型中，批处理大小设为 16，优化器使用随机梯度下降法（stochastic gradient descent，简称 SGD），初始化学习率为 1×10^{-5}，权重衰减为 5×10^{-3}，动量为 0.9，共训练 60 个 epoch. 训练阶段和测试阶段均设置相同的超参数.

2.3 实验结果与分析

2.3.1 不同模型上的训练情况对比 图 8 为 ResNet101 和 SF-ResNet101 在 AID 数据集上的训练情况，其中横轴表示 epochs（训练步数），图 8a 和图 8b 的纵轴分别表示损失值和精度值，蓝色和绿色实线分别表示 ResNet101 的 train_loss（训练集损失）和 test_loss（测试集损失）、train_acc（训练集精度）和 test_acc（测试集精度），黄色和红色虚线分别表示 SF-ResNet101 的 train_loss 和 test_loss、train_acc 和 test_acc. 在图 8a 中，随着训练步数的增加，train_loss 值不断收敛并在大约第 27 个 epochs 之后趋于稳定，test_loss 值在收敛过程中存在少量波动现象，最终在大约第 25 个 epochs 之后趋于稳定. 在图 8b 中，ResNet101 的 test_acc 在训练初期存在较明显的波动情况，在训练步数达到 25 时，train_acc 和 test_acc 走势一致并趋于稳定，达到最优分类精度. 综合来看，SF-ResNet101 中损失值和精度值的训练情况相对于原 ResNet101 模型收敛过程中波动情况有明显改善，收敛速度更快. 同时对比图 8b 中绿色实线和红色虚线可以看出 SF-ResNet101 模型趋于稳定后的 test_acc 相对于原 ResNet101 模型有明显提升. 由此可知，SF-ResNet101 模型能够更好地提取高分遥感场景高阶特征，具有优秀的场景描述能力和类别区分能力. 同时 Dropout 层的加入能够有效减少模型过拟合现象，进一步提高训练速度.

图 9 为 ResNet101 和 SF-ResNet101 在 NWPU-RESISC45 数据集上的训练情况. 对比图 9a 和 9b 可以看出，SF-ResNet101 模型其 loss 和 accuracy 值曲线波动较小，收敛更快，最优 test_acc

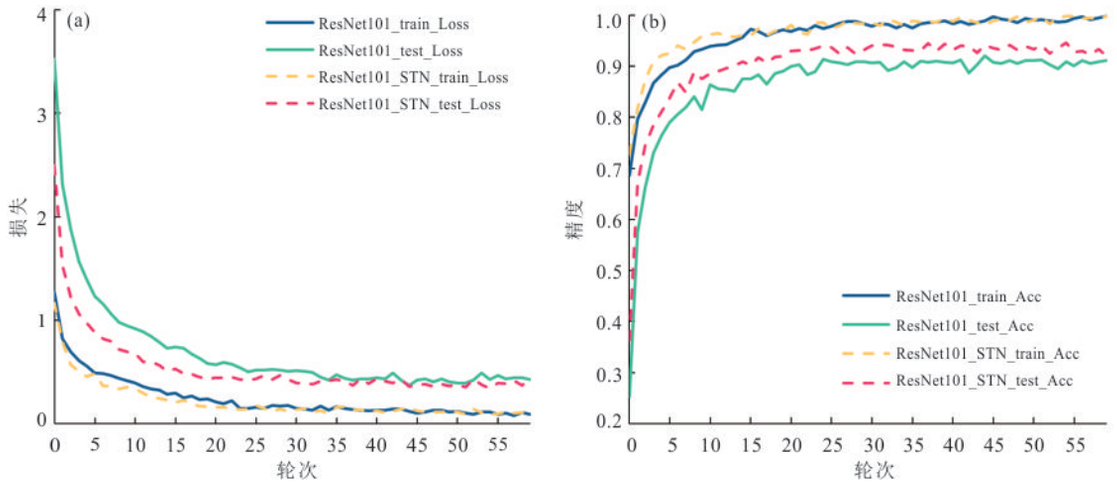

图 8　ResNet101 和 SF-ResNet101 在 AID 数据集上的训练情况

Fig.8　ResNet101 and improved ResNet101 training on AID datasets

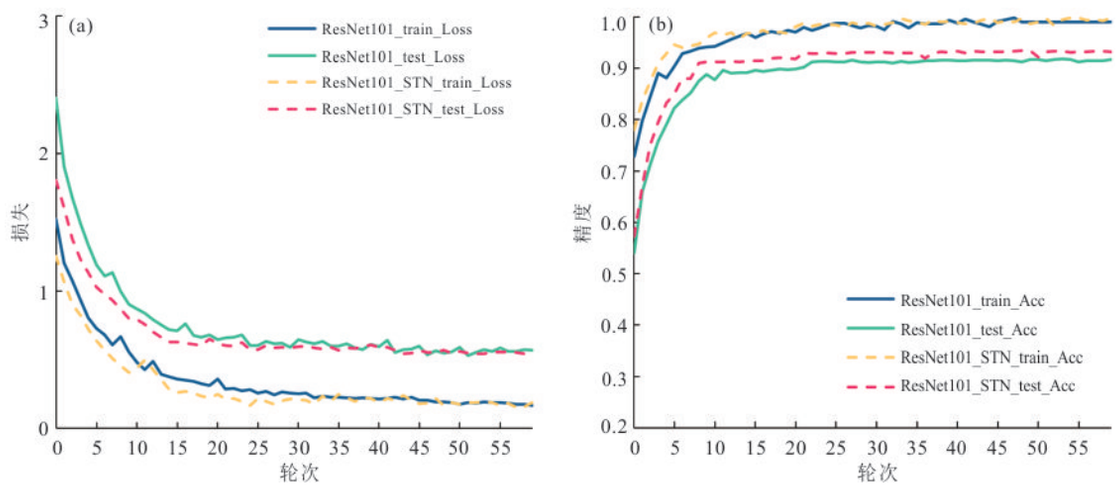

图 9　ResNet101 和 SF-ResNet101 在 NWPU-RESISC45 数据集上的训练情况

Fig.9　ResNet101 and improved ResNet101 training on NWPU-RESISC45 datasets

值也有明显提升. train_acc 和 test_acc 均在 20 个 epochs 趋于稳定. 由图中虚线和实线的对比可得出, SF-ResNet101 模型训练更加稳定, 证明预训练的 ResNet101 模型能够有效地迁移到遥感影像数据中, 提高模型的分类效率.

2.3.2　不同 Dropout 率对改进 ResNet 模型的影响　为了探究不同 Dropout 率设置下模型的分类能力, 本文设计了以下实验, 如图 10 和图 11 所示. 其中蓝色、绿色、黄色和红色实线分别代表 Dropout 率设置为 0、0.1、0.2、0.4 时损失值和精度值随训练步数的变化情况. 首先在 AID 数据集上, 可以看出 Dropout 层的加入有效提高了模型分类能力, 训练曲线稳定性、模型的收敛速度和训练准确率均有提升. 当 Dropout 率设置为 0.1 时模型的分类能力最优.

在 NWPU-RESISC45 数据集上(图 11)同样可以体现出 Dropout 层对模型的优化效果, 当 Dropout 率设置为 0.2 时模型表现最好, 损失值和精度值的曲线波动均为最小, 分类准确率达到了最优. 然而, 当 Dropout 率设置为 0.4 时, AID 数据集和 NWPU-RESISC45 数据集在收敛初期均出现波动较大的情况, 这可能是由于 Dropout 率设置太高, 从而使模型存在一定程度的欠拟合现象.

表 2 统计了两种数据集在不同 Dropout 率设置下的测试集精度. 横向对比可以看出, AID 数据集在 Dropout 率设置为 0.1 时分类准确率最高, NWPU-RESISC45 数据集在 Dropout

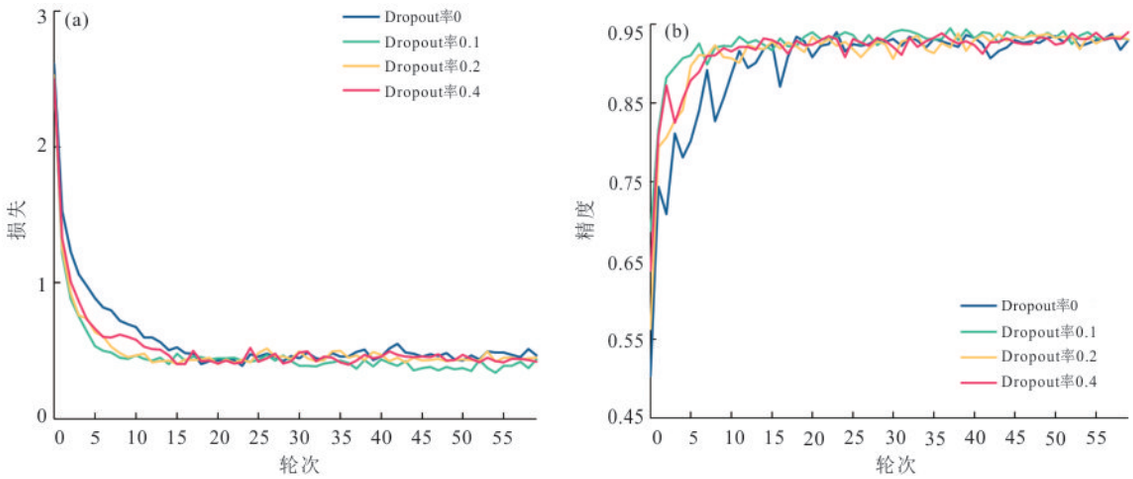

图 10 不同 Dropout 率在 AID 数据集上的训练情况

Fig.10 Different Dropout rate training on AID datasets

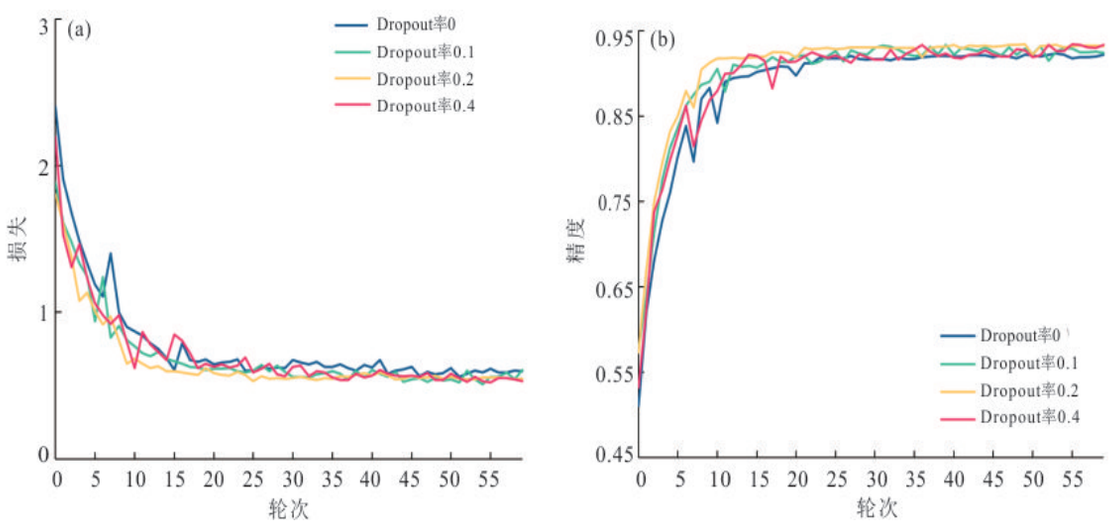

图 11 不同 Dropout 率在 NWPU-RESISC45 数据集上的训练情况

Fig.11 Different Dropout rate training on NWPU-RESISC45 datasets

表 2 不同 Dropout 率测试集精度对比

Table 2 Comparsion of test accuracy of different Dropout rates

Dropout 率	test_acc (%)		
	0.1	0.2	0.4
AID	**94.30**	94.24	94.02
NWPU	93.47	**93.63**	93.60

率设置为 0.2 时分类准确率最高. 因此, 两个数据集在不同模型上的训练情况对比实验中笔者分别将 Dropout 率设置为 0.1 和 0.2.

2.3.3 各模型性能对比 为了评估改进模型的性能, 本文选取了现阶段针对遥感影像场景分类的几种代表性网络模型进行对比, 同时添加了原 ResNet101 模型和嵌入 STN 结构的 ResNet101 模型与 SF-ResNet101 模型的实验对比, 实验数据集训练比率与 2.2 节一致. 各类方法均进行了 10 次随机试验, 并对试验结果进行平均和标准差的计算. 表 3 与表 4 分别为 SF-ResNet101 模型与其他网络模型在 AID 数据集和 NWPU-RESISC45 数据集上的总体精度对比, 可以看到 SF-ResNet101 的分类精度优于大多数典型网络, 取得了现阶段较好的精度. 表 3 中 AID 数据集在训练比率为 50% 的情况下 D-CNNs 模型的训练效果最优, 这可能是由于通过嵌入度量学习约束 D-CNNs 模型判别组合特征距离, 在训练比率较大的情况下能够更好地解决类内多样性与类间相似性影响. 与原 ResNet101 网络模型相比,

表 3　各网络模型在 AID 数据集上的分类精度

Table 3　Classification accuracy of different models on AID dataset

模型	总体精度(%)	
	20% 训练比率	50% 训练比率
GoogleNet(Xia *et al.*,2017)	83.44±0.40	89.36±0.55
VGG-VD16+MSCP+MRA(He *et al.*,2018)	92.21±0.17	96.56±0.18
CNN-CapsNet(Zhang *et al.*,2019)	93.79±0.13	96.32±0.12
D-CNNs(Cheng *et al.*,2018)	90.82±0.16	**96.89±0.10**
ResNet101	92.32±0.23	95.49±0.38
ResNet101+STN	93.58±0.22	95.89±0.27
本文方法	**94.30±0.29**	96.52±0.10

表 4　各网络模型在 NWPU-RESISC45 数据集上的分类精度

Table 4　Classification accuracy of different models on NWPU-RESISC45 dataset

模型	总体精度(%)	
	10% 训练比率	20% 训练比率
Fine-tuned VGGNet-16(Cheng *et al.*,2017)	87.15±0.45	90.36±0.18
VGG-VD16+MSCP+MRA(He *et al.*,2018)	88.07±0.18	90.81±0.13
CNN-CapsNet(Zhang *et al.*,2019)	89.03±0.21	92.60±0.11
D-CNNs(Cheng *et al.*,2018)	89.22±0.50	91.89±0.22
ResNet101	89.27±0.21	91.81±0.19
ResNet101+STN	90.72±0.23	92.47±0.28
本文方法	**91.66±0.15**	**93.63±0.22**

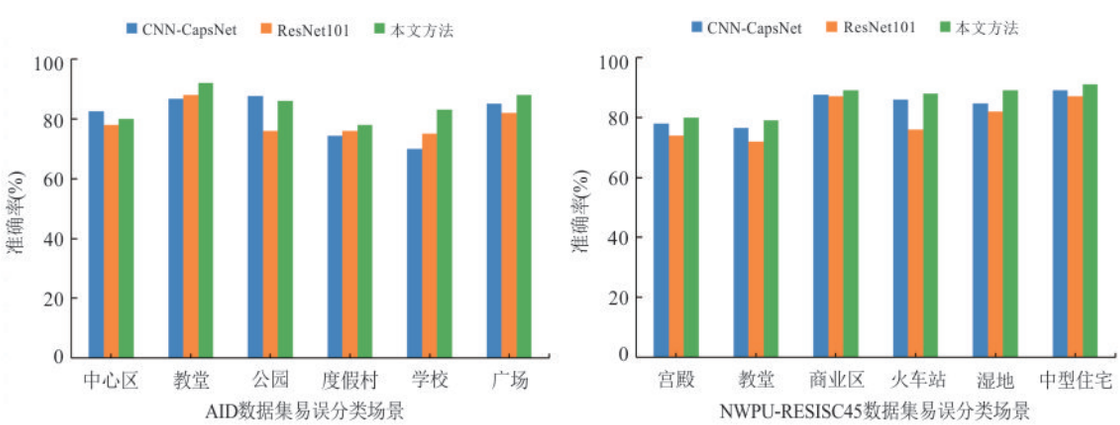

图 12　各网络模型易误分类场景性能对比

Fig.12　Comparison of performance for different models on easily misclassified scene images

SF-ResNet101 的分类准确率提升了 2%～3% 左右,与 ResNet101+STN 方法相比也有 1% 左右的精度提升,这是由于在原始的 ResNet101 网络中引入了迁移学习,使得改进模型能够利用大型自然数据集 ImageNet 的知识,通过 ResNet 模型中的 Bottleneck 残差结构降低维度,提高模型的分类效率.同时,ResNet101+STN 方法较原 ResNet101 网络分类能力也有明显提高,这是由于加入的空间变换结构对输入影像进行仿射变换,增强模型的鲁棒性,从而能够有效提取遥感影像的高层特征.此外,在模型全连接层和输出层之间添加 Dropout 随机失活机制有效解决了过拟合问题,使模型的整体分类性能有所提升.

本文选取了两类数据集分类结果中准确率不到 90% 的场景,分别比较这些场景在不同模型中的

分类准确率并绘制图表,如图12所示.可以直观地看到SF-ResNet101模型在大部分易误分类场景中具有最优的分类精度,分类效果良好.说明本文方法针对易误分类场景同样具有一定的优势.

3 结论

本文进行了基于深度学习的高分遥感影像场景分类的实验研究.联合空间变换网络和迁移学习的优势,提出了一种基于改进残差网络的高分辨率遥感影像场景分类算法.通过在网络中嵌入空间变换结构增强模型的鲁棒性,提高模型特征提取能力,利用迁移学习使模型充分利用在已标注好的大型数据集中训练得到的知识,改进对遥感影像场景识别的学习效果.在AID和NWPU-RESISC45两个高分遥感影像数据集上的对比实验表明,本文提出的SF-ResNet101模型相较于原ResNet101模型收敛更快,训练过程更为稳定,且分类准确率也有明显提升.对比其他遥感影像分类常用网络模型,本文方法具有更好的分类效果.证明了所提出的改进模型的稳定性与高效性.

本文也存在一定的不足之处,在少部分场景上的分类表现较为一般,这可能是由于空间变换结构对于部分包含复杂地物目标场景的特征提取能力较差,仍有待进一步的提高.同时,SF-ResNet101模型参数计算量较大,降低了分类速度,也增加了参数冗余的可能.因此,降低卷积神经网络模型复杂度,进一步完善迁移学习机制和空间变换结构是今后研究的重点.

致谢:感谢匿名审稿专家提出的有益建议!

References

Berman, M., Triki, A. R., Blaschko, M. B., 2018. The Lovasz-Softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-Over-Union Measure in Neural Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City. https://doi.org/10.1109/CVPR.2018.00464

Cheng, G., Guo, L., Zhao, T. Y., et al., 2013. Automatic Landslide Detection from Remote-Sensing Imagery Using a Scene Classification Method Based on BoVW and pLSA. *International Journal of Remote Sensing*, 34(1): 45−59. https://doi. org/10.1080/01431161.2012.705443

Cheng, G., Han, J. W., Lu, X. Q., 2017. Remote Sensing Image Scene Classification: Benchmark and State of the Art. *Proceedings of the IEEE*, 105(10): 1865−1883. https://doi.org/10.1109/JPROC.2017.2675998

Cheng, G., Yang, C. Y., Yao, X. W., et al., 2018. When Deep Learning Meets Metric Learning: Remote Sensing Image Scene Classification via Learning Discriminative CNNS. *IEEE Transactions on Geoscience and Remote Sensing*, 56(5): 2811−2821. https://doi. org/10.1109/TGRS.2017.2783902

Cheng, G. X., Niu, R. Q., Zhang, K. X., et al., 2018. Opencast Mining Area Recognition in High-Resolution Remote Sensing Images Using Convolutional Neural Networks. *Earth Science*, 43(S2): 256−262 (in Chinese with English abstract).

Dalal, N., Triggs, B., 2005. Histograms of Oriented Gradients for Human Detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR′05), San Diego. https://doi.org/10.1109/CVPR.2005.177

Donahue, J., Jia, Y. Q., Vinyals, O., et al., 2014. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. Proceedings of the 31st International Conference on Machine Learning, Beijing. https://doi.org/10.5555/3044805.3044879

Han, X., Zhang, Z. Y., Ding, N., et al., 2021. Pre-Trained Models: Past, Present and Future. *AI Open*, 2: 225−250. https://doi.org/10.1016/j.aiopen.2021.08.002

He, K. M., Zhang, X. Y., Ren, S. Q., et al., 2016. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas. https://doi. org/10.1109/CVPR.2016.90

He, N. J., Fang, L. Y., Li, S. T., et al., 2018. Remote Sensing Scene Classification Using Multilayer Stacked Covariance Pooling. *IEEE Transactions on Geoscience and Remote Sensing*, 56(12): 6899−6910. https://doi.org/10.1109/TGRS.2018.2845668

He, X., Chen, Y. S., 2019. Optimized Input for CNN-Based Hyperspectral Image Classification Using Spatial Transformer Network. *IEEE Geoscience and Remote Sensing Letters*, 16(12): 1884−1888. https://doi. org/10.1109/LGRS.2019.2911322

Jaderberg, M., Simonyan, K., Zisserman, A., et al., 2015. Spatial Transformer Networks. *arXiv*, 1506.02025. https://arxiv.org/abs/1506.02025

Jia, Y. Q., Shelhamer, E., Donahue, J., et al., 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. Proceedings of the 22nd ACM international conference

on Multimedia, Orlando. https://doi. org/10.1145/2647868.2654889

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe. https://doi.org/10.5555/2999134.2999257

Li, D. R., Wang, M., Shen, X., et al., 2017. From Earth Observation Satellite to Earth Observation Brain. *Geomatics and Information Science of Wuhan University*, 42(2): 143－149 (in Chinese with English abstract).

Li, G. D., Zhang, C. J., Wang, M. K., et al., 2019. Transfer Learning Using Convolutional Neural Network for Scene Classification within High Resolution Remote Sensing Image. *Science of Surveying and Mapping*, 44(4): 116－123, 174 (in Chinese with English abstract).

Oliva, A., Torralba, A., 2001. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision*, 42(3): 145－175. https://doi.org/10.1023/A: 1011139631724

Oquab, M., Bottou, L., Laptev, I., et al., 2014. Learning and Transferring Mid-Level Image Representations Using Convolutional Neural Networks. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus. https://doi.org/10.1109/CVPR.2014.222

Pan, S. J., Yang, Q., 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10): 1345－1359. https://doi. org/10.1109/TKDE.2009.191

Perronnin, F., Sánchez, J., Mensink, T., 2010. Improving the Fisher Kernel for Large-Scale Image Classification. European Conference on Computer Vision, Berlin. https://doi.org/10.1007/978-3-642-15561-1_11

Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv*, 1409.1556. https://arxiv.org/abs/1409.1556

Srinivas, A., Lin, T. Y., Parmar, N., et al., 2021. Bottleneck Transformers for Visual Recognition. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville. https://doi. org/10.1109/CVPR46437.2021.01625

Szegedy, C., Liu, W., Jia, Y. Q., et al., 2015. Going Deeper with Convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston. https://doi.org/10.1109/CVPR.2015.7298594

Văduva, C., Gavăt, I., Datcu, M., 2013. Latent Dirichlet Allocation for Spatial Analysis of Satellite Images. *IEEE Transactions on Geoscience and Remote Sensing*, 51(5): 2770－2786. https://doi. org/10.1109/TGRS.2012.2219314

van de Sande, K., Gevers, T., Snoek, C., 2010. Evaluating Color Descriptors for Object and Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9): 1582－1596. https://doi.org/10.1109/TPAMI.2009.154

Wallraven, C., Caputo, B., Graf, A., 2003. Recognition with Local Features: The Kernel Recipe. Proceedings Ninth IEEE International Conference on Computer Vision, Nice. https://doi.org/10.1109/ICCV.2003.1238351

Wang, R. C., 2018. Feature Learning and Patch Matching of Multispectral Images Based on Deep Neural Networks (Dissertation). Beijing University of Posts and Telecommunications, Beijing (in Chinese with English abstract).

Weiss, K., Khoshgoftaar, T. M., Wang, D. D., 2016. A Survey of Transfer Learning. *Journal of Big Data*, 3(1): 1－40. https://doi.org/10.1186/s40537-016-0043-6

Xia, G. S., Hu, J. W., Hu, F., et al., 2017. AID: A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(7): 3965－3981. https://doi.org/10.1109/TGRS.2017.2685945

Xu, Y. Y., Li, Z. X., Xie, Z., et al., 2020. Prediction of Copper Mineralization Based on Semi-Supervised Neural Network. *Earth Science*, 45(12): 4563－4573 (in Chinese with English abstract).

Yang, Y., Newsam, S., 2008. Comparing SIFT Descriptors and Gabor Texture Features for Classification of Remote Sensed Imagery. 2008 15th IEEE International Conference on Image Processing, San Diego. https://doi.org/10.1109/ICIP.2008.4712139

Yang, Y., Newsam, S., 2013. Geographic Image Retrieval Using Local Invariant Features. *IEEE Transactions on Geoscience and Remote Sensing*, 51(2): 818－832. https://doi.org/10.1109/TGRS.2012.2205158

Yu, D. H., Zhang, B. M., Zhao, C., et al., 2020. Scene Classification of Remote Sensing Image Using Ensemble Convolutional Neural Network. *Journal of Remote Sensing*, 24(6): 717－727 (in Chinese with English abstract).

Yu, S. C., Yu, D. Q., Wang, L. C., et al., 2019. Remote Sensing Study of Dongting Lake Beach Changes before and after Operation of Three Gorges Reservoir. *Earth Science*, 44(12): 4275－4283 (in Chinese with English abstract).

Zhang, K., Hei, B. Q., Li, S. Y., et al., 2018. Complex Scene Classification of Remote Sensing Images Based

on CNN. *Remote Sensing for Land & Resources*, 30(4): 49－55 (in Chinese with English abstract).

Zhang, W., Tang, P., Zhao, L. J., 2019. Remote Sensing Image Scene Classification Using CNN-CapsNet. *Remote Sensing*, 11(5): 494. https://doi. org/10.3390/rs11050494

Zuo, R. G., Peng, Y., Li, T., et al., 2021. Challenges of Geological Prospecting Big Data Mining and Integration Using Deep Learning Algorithms. *Earth Science*, 46(1): 350－358 (in Chinese with English abstract).

附中文参考文献

程国轩, 牛瑞卿, 张凯翔, 等, 2018. 基于卷积神经网络的高分遥感影像露天采矿场识别. 地球科学, 43(S2): 256－262.

李德仁, 王密, 沈欣, 等, 2017. 从对地观测卫星到对地观测脑. 武汉大学学报(信息科学版), 42(2): 143－149.

李冠东, 张春菊, 王铭恺, 等, 2019. 卷积神经网络迁移的高分影像场景分类学习. 测绘科学, 44(4): 116－123, 174.

王瑞琛, 2018. 基于深层神经网络的异源图像特征学习及块匹配(硕士学位论文). 北京: 北京邮电大学.

徐永洋, 李孜轩, 谢忠, 等, 2020. 基于半监督神经网络的铜矿预测方法. 地球科学, 45(12): 4563－4573.

余东行, 张保明, 赵传, 等, 2020. 联合卷积神经网络与集成学习的遥感影像场景分类. 遥感学报, 24(6): 717－727.

余姝辰, 余德清, 王伦澈, 等, 2019. 三峡水库运行前后洞庭湖洲滩面积变化遥感认识. 地球科学, 44(12): 4275－4283.

张康, 黑保琴, 李盛阳, 等, 2018. 基于CNN模型的遥感图像复杂场景分类. 国土资源遥感, 30(4): 49－55.

左仁广, 彭勇, 李童, 等, 2021. 基于深度学习的地质找矿大数据挖掘与集成的挑战. 地球科学, 46(1): 350－358.