

# 基于显著性特征和 DCNN 的高分遥感影像场景分类

吕欢欢\*, 刘涛\*\*, 张辉, 彭国峰, 张峻通

辽宁工程技术大学软件学院, 辽宁 葫芦岛 125105

**摘要** 高分遥感影像的场景分类是解译遥感影像信息的重要工作之一。为了准确提取出目标信息, 针对高分遥感影像场景分类中存在的背景复杂、目标多样、目标信息与背景信息难以区分等问题, 提出了一种基于显著性特征和深度卷积神经网络(DCNN)的高分遥感影像场景分类方法。首先, 利用  $K$ -means 聚类与超像素分割算法得到影像的颜色空间分布与颜色对比图, 融合不同对比图, 以得到显著图。然后, 通过对数变换增强显著图中的特征, 采用自适应阈值分割方法提高目标的区分度并划分出目标区域和背景区域, 以提取出感兴趣区域。最后, 构建了一种用于提取深层语义特征的 DCNN 模型, 并将得到的特征输入网络模型中进行训练和分类。实验结果表明, 本方法能有效区分主要目标信息与背景信息, 减少无关信息的干扰, 在 UC-Merced 数据集和 WHU-RS 数据集上的分类精度分别为 96.10% 和 95.84%。

**关键词** 大气光学; 高分遥感影像; 场景分类; 显著性检测; 卷积神经网络; 深层语义特征

中图分类号 P237

文献标志码 A

doi: 10.3788/LOP202158.2001002

## High-Resolution Remote Sensing Scene Classification Based on Salient Features and DCNN

Lü Huanhuan\*, Liu Tao\*\*, Zhang Hui, Peng Guofeng, Zhang Juntong

College of Software, Liaoning Technical University, Huludao, Liaoning 125105, China

**Abstract** Scene classification of high-resolution remote sensing image is one of the important tasks in interpreting remote sensing image information. In order to extract the target information accurately, we propose a high-resolution remote sensing image scene classification method based on salient features combined with deep convolutional neural network (DCNN) to solve the problems of complex background, diverse targets, and difficult to distinguish between target information and background information in the classification of high-scoring remote sensing image scenes. First,  $K$ -means clustering algorithm and super-pixel segmentation algorithm are used to generate the color spatial distribution map and color contrast map of the image, and the different contrast maps are fused to get the saliency map. Then, the features in the saliency map are enhanced through logarithmic transformation, and the adaptive threshold segmentation method is used to improve the discrimination of the target and divide the target area and the background area, and extract the area of interest. Finally, a DCNN model is constructed to extract deep semantic features and classification, and the obtained features are input into the network model for training and classification. Experimental results show that the method can effectively distinguish the main target information from the background information and reduce the interference of irrelevant information. The classification accuracy of the method on the UC-Merced data set and WHU-RS data set are 96.10% and 95.84%, respectively.

**Key words** atmospheric optics; high-resolution remote sensing image; scene classification; saliency detection; convolutional neural network; deep semantic feature

**OCIS codes** 010.0280; 100.4996; 070.5010; 100.2960

收稿日期: 2020-11-25; 修回日期: 2020-12-24; 录用日期: 2021-01-11

基金项目: 国家自然科学基金(41871379, 61540056)、辽宁省自然科学基金指导计划(20180550450)

通信作者: \*lvhh2010@126.com; \*\*85578981@qq.com

# 1 引言

相比遥感影像,高分遥感影像包含了丰富的空间细节信息,是对地精细观测的重要数据来源,能清晰表达地物目标的空间结构和纹理特征,更好地反映地物目标的尺寸、形状及相邻关系,使对地物目标的精确解译成为可能,已被广泛应用于土地利用、土地覆盖确定、自然灾害检测、城市规划、地理空间目标检测等领域<sup>[1-3]</sup>。但随着分辨率的提高,高分辨率影像场景数据量的大幅增加以及背景复杂性和目标多样性的不断增强给分类任务带来一定困难。此外,遥感场景是由语义目标基于某种空间分布构成的具有深层语义信息的复杂影像区域,在进行场景分类时,无关信息容易与目标区域混淆,导致分类精度降低。因此,如何有效提取高分遥感影像的目标信息和深层语义特征,从而实现场景的自动识别和分类对于高分影像的应用具有重要意义。

早期的场景分类方法主要基于纹理、颜色特征等底层语义特征和基于中层语义特征<sup>[4]</sup>,如 Lowe<sup>[5]</sup>利用尺度不变特征变换(SIFT)特征对旋转、亮度变化和尺度缩放的不变性对局部特征进行描述,提出了一种高效的区域检测方法。Oliva 等<sup>[6]</sup>提出了一种基于全局低层次特征(Gist)的方法,无需对场景进行目标切割,直接将整个场景作为对象进行计算,计算速度较快,但精度有限。这些基于低层特征的方法简单且易于实现,但分类精度较低。基于中层特征的方法通过对局部特征进行编码,构建能表达整张影像的中层特征,其中,最具代表性的就是词袋模型(BoVW),如 Yang 等<sup>[7]</sup>利用视觉词典的共现结合 BoVW 模型,提出了一种非线性空间金字塔匹配核(SPCK++)方法,相比 BoVW 模型,该方法的分类精度更高,但复杂度也较高。Zhao 等<sup>[8]</sup>将概率模型潜在狄利克雷分配(LDA)应用于场景中,用 LDA 作为分类器的一部分(P-LDA 模型)并用 LDA 作为主要特征提取器(F-LDA 模型),提高了 LDA 的分类精度。上述方法面对复杂场景类型和多尺度的高分影像时可能丢失局部特征,导致分类精度不高。近年来,基于深度学习的方法逐渐在图像分类领域中得到了广泛的应用<sup>[9]</sup>。如马永杰等<sup>[10]</sup>提出了一种基于稠密连接网络进化的卷积神经网络(D-ECNN)图像分类方法,对深度网络结构与参数进行自适应优化,在车辆数据集上的准确率可达到 95%。何小飞等<sup>[11]</sup>利用显著性采样提取显著信息块,利用卷积神经网络(CNN)提取高层次特征并用

支持向量机(SVM)进行分类。刘芳等<sup>[12]</sup>基于离散余弦变换(DCT)与 CNN 提出了一种 DCT-CNN 方法,对陆地场景地貌图像主要目标信息的深度特征进行分类,取得了较高的分类精度。李冠东等<sup>[13]</sup>基于迁移学习方法在 Inception-v3 预训练模型中提取特征并用全连接(FC)层进行分类输出,一定程度上提高了模型的泛化能力。任永梅等<sup>[14]</sup>提出了一种基于三维(3D)CNN 的点云图像船舶分类方法,通过设计的 6 层 3D CNN 提取体素网格图像的高水平特征,以捕捉结构信息,提高了对船舶数据集的分类精度。

综上所述,当前 CNN 对遥感影像的场景分类已有初步研究但其分类精度仍有待进一步提升,特别是当场景复杂度增加且信息量增大时,分类效果会急剧下降。为了进一步提升高分遥感影像场景的分类精度并充分利用影像的显著性特征和深层语义特征,本文提出了一种基于显著性特征和深度 CNN(DCNN)的高分遥感影像场景分类方法。该方法利用影像的显著性特征提取感兴趣区域(ROI),通过构建 DCNN 模型提取影像的深层语义特征并实现分类。

## 2 基本原理

### 2.1 简单线性迭代聚类超像素分割

常见的超像素分割算法有 Graph-based 算法、超像素(Superpixel)算法、分水岭算法和简单线性迭代聚类(SLIC)算法等。对比多种超像素分割算法的运算时间和结构发现,SLIC 算法进行超像素分割的时间效率和储存效率优于其他分割算法,且具有良好的边界依从性,能得到像素块紧凑、超像素数量灵活的分割结果。SLIC 算法<sup>[15]</sup>根据国际照明委员会(CIE)公布的 CIELab(Luminosity、a、b)颜色空间和二维坐标构成的五维向量对图像中的特征进行聚类,可生成紧凑、近似均匀的超像素块,且每个大小不同的超像素块都具有相同的性质,有效减少了影像的冗余信息。若要将图像中  $N$  个像素点分割为  $K$  个超像素区域,则每个区域的大小为  $N/K$ ,相邻种子点  $m$ 、 $n$  的初始距离  $S = \sqrt{N/K}$ 。根据颜色距离与空间距离在每个种子点的邻域内为每个像素点分配类标签,搜索范围为  $2S \times 2S$ ,计算方法可表示为

$$d_c = \sqrt{(l_m - l_n)^2 + (a_m - a_n)^2 + (b_m - b_n)^2}, \quad (1)$$

$$d_s = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2}, \quad (2)$$

$$D' = \sqrt{(d_c/N_c)^2 + (d_s/N_s)^2}, \quad (3)$$

式中,  $d_c$  为颜色距离,  $l$  为亮度,  $a$ 、 $b$  为两个颜色通道的值,  $d_s$  为空间距离,  $(x_m, y_m)$ 、 $(x_n, y_n)$  为种子点的坐标,  $D'$  为距离度量,  $N_c$  为类内最大颜色距离, 一般取为 10,  $N_s = S$  为最大空间距离。一般情况下, 将(1)式~(3)式迭代优化 10 次后, 就能得到比较理想的分割效果。

## 2.2 K-means 聚类

K-means 聚类算法是基于距离相似性的聚类算法, 计算复杂度接近线性运算且可以根据影像特点选择合适的  $K$  值, 通过不断计算和迭代聚类, 优化初始监督学习样本分类不合理的情况。通过对比样本之间的相似性, 将相似样本划分为同一类, 以便对图像进行高效处理。在一个包含  $K$  个类别的数据集中, 随机选取第  $k$  类  $c_k$  的聚类中心  $\mu_k$ , 以当前聚类中心和选定的度量方式分类样本点, 不断计算并更新聚类中心, 样本点与聚类中心的距离可表示为

$$E(c_k) = \sum_{x_s \in c_k} \|x_s - \mu_k\|, \quad (4)$$

式中,  $x_s$  为第  $k$  类中第  $s$  个样本点的位置,  $\mu_k$  为第  $k$  个聚类中心。可以发现, 样本点与聚类中心越近, 表明两者的相似度越高。

## 2.3 卷积神经网络

CNN 是一类包含卷积计算且具有深度学习结构的前馈神经网络, 通过对输入影像进行卷积运算突出影像的细节信息并消除图像噪声。池化层可对数据进行压缩, 减少参数量, 防止数据过拟合, 使图像具有缩放、位移不变性。CNN 通常用多个卷积层和池化层进行训练, 以学习全局特征, 并根据实际需要确定池化层的层数、参数和全连接层的分类器类型<sup>[16]</sup>。

1) 卷积层: 利用卷积层进行图像特征提取时, 每个神经元的输入与前一层的局部感受野相连接, 以提取局部特征。输入图像与卷积层的计算可表示为

$$x_j^l = f\left(\sum_{i \in N_j} x_i^{l-1} \times k_{ij}^l + b_j^l\right), \quad (5)$$

式中,  $x_j^l$  为第  $l$  层的第  $j$  个特征图,  $N_j$  为二维向量的宽度,  $k_{ij}^l$  为第  $l$  层的第  $j$  个卷积核, 其长度为  $i$ , 宽度为  $j$ ,  $b_j^l$  为第  $l$  层特征映射的偏置项,  $f(\cdot)$  为激活函数。

2) 池化层: 对卷积获取的特征进行采样, 即池化操作, 通常分为最大池化和平均池化。采用最大

池化提取区域中的最大值, 可降低特征维度, 避免过拟合。池化函数可表示为

$$x_j^l = f\left(\sum_{i \in N_j} x_i^{l-1} \times k_{ij}^l + b_j^l\right), \quad (6)$$

第  $l$  层池化过程可表示为

$$x_j^l = f[\omega_j^l X_{\text{down}}(x_j^{l-1}) + b_j^l], \quad (7)$$

式中,  $X_{\text{down}}$  为池化函数,  $\omega_j^l$  为该层的采样系数。

3) 全连接层: 全连接层将卷积操作提取的特征整合到一起, 输出为一个值, 可起到分类器的作用。全连接层可由卷积操作实现, 即对前层的全连接操作可以转化为卷积核为  $1 \times 1$  的卷积操作, 对前层的卷积操作可转化为卷积核为  $i \times j$  的全局卷积 ( $i$  和  $j$  为前一层的高和宽) 操作。全连接层中影响模型性能的参数包括全连接层的总层数 (长度)、单个全连接层的神经元数 (宽度) 和激活函数。

4) Dropout 层: Dropout 层是深度学习常用的正则化方法, 若模型参数较多, 而训练样本太少, 训练出的模型容易产生过拟合现象, 具体表现为模型在训练数据上的损失函数较小, 预测概率高, 而在测试数据上的损失函数较大, 预测概率低。在前向传播的每个训练批次中, 通过忽略一般的特征检测器 (使一般的隐藏节点值为 0), 即让某个神经元的激活值以一定概率停止工作, 从而减少模型对某些局部特征的过分依赖, 明显减少过拟合现象。

5) Softmax 分类器: 在深度学习中, 常用 Softmax 函数进行分类, 该函数可将一些输出的神经元映射到  $(0, 1)$  之间, 使分类的概率和为 1。Softmax 函数可表示为

$$S_q = \frac{\exp(V_q)}{\sum_c \exp(V_q)}, \quad (8)$$

式中,  $V_q$  为分类器前一层的输出,  $q$  为类别,  $C$  为类别总数,  $S_q$  为当前元素的指数与所有元素指数和的比值, 该函数可将多个类别的输出值转化为相应类型的概率。

## 3 分类原理

本方法的具体流程如图 1 所示, 主要包括高分遥感影像的显著性检测、ROI 提取和基于 DCNN 的场景分类三个阶段。具体步骤: 1) 基于超像素分割与 K-means 聚类提取影像的显著性特征, 并生成显著图; 2) 对显著图进行增强与自适应阈值分割, 得到显著图的二值化图像, 并提取 ROI; 3) 构建一种

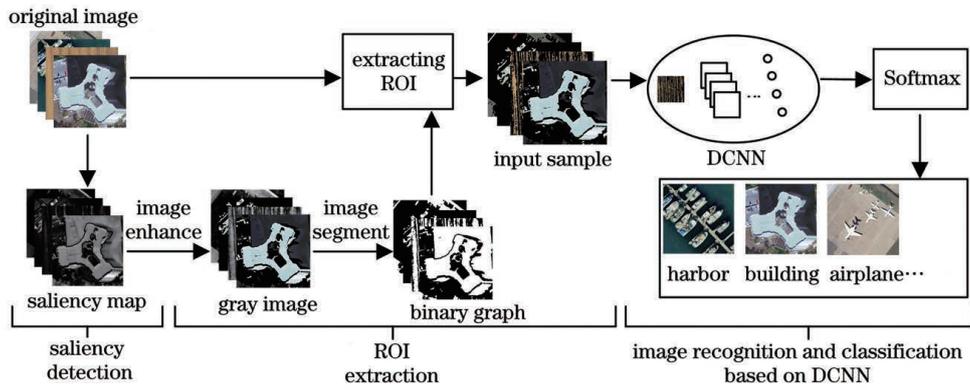


图 1 本方法的流程图

Fig. 1 Flow chart of our method

DCNN 模型对深层语义特征进行提取和分类,并将训练样本输入网络模型中进行训练,然后在测试阶段用模型对样本进行分类。

### 3.1 显著性特征的提取与融合

#### 1) 颜色对比图的构建

颜色对比度可以表达像素点的显著性差异,为了提高计算效率,将超像素分割后的像素块作为基本单位,基于颜色的局部对比度,以像素点之间的欧氏距离为权重,通过计算像素块与影像中心的空间距离,得到颜色对比图。针对高分遥感影像具有的数据量大和细节信息多等特点,加强对超像素尺寸和紧凑性的控制,计算相邻种子点  $m$ 、 $n$  所在区域的 CIELab 颜色空间差异,  $m$  所在区域的颜色对比值  $C_m$  可表示为

$$C_m = \sum_{n=1}^N \| \mathbf{v}_m - \mathbf{v}_n \| W(\mathbf{w}_m, \mathbf{w}_n), \quad (9)$$

$$\| \mathbf{v}_m - \mathbf{v}_n \| =$$

$$\sqrt{(l_m - l_n)^2 + (a_m - a_n)^2 + (b_m - b_n)^2}, \quad (10)$$

$$W(\mathbf{w}_m, \mathbf{w}_n) = 1 / \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2}, \quad (11)$$

式中,  $\mathbf{v}_m$  和  $\mathbf{v}_n$  为区域  $m$  和  $n$  的 Lab 颜色,可表示为  $(l, a, b)$ ,  $\mathbf{w}_m$ 、 $\mathbf{w}_n$  为区域  $m$ 、 $n$  的中心坐标,且

$$\sum_{n=1}^N (\mathbf{w}_m, \mathbf{w}_n) = 1.$$

不同超像素个数  $N$  对分类结果的影响较大,为了得到更好的分割效果,用 SLIC 算法对随机选取的 1000 张数据集图像进行聚类,超像素个数分别为 50、100、200、300、500、800 时算法的边缘召回率 (ER) 和欠分割率 (UR) 如图 2 所示。可以发现,SLIC 算法的分割效果随超像素数的增加逐渐变好,欠分割率逐渐降低,边缘召回率逐渐提高。当超像素数为 50~200 时,分割效果提升明显;当超像素数

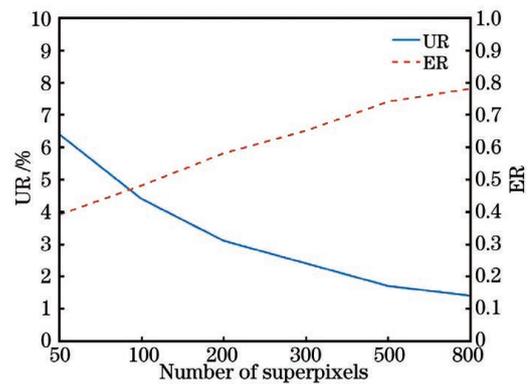


图 2 不同超像素个数的分割结果

Fig. 2 Segmentation results with different number of superpixels

为 200~500 时,分割效果提升缓慢;当超像素数大于 500 时,分割效果提升特别缓慢。此外,SLIC 算法的分割时间随超像素数的增加呈线性增长趋势,综合分割效果与时间效率,将超像素数设置为 500。

#### 2) 颜色空间分布图的构建

图像的空间分布可以有效区别显著性区域与背景区域,因此,采用 K-means 聚类将具有相同颜色特征的图像分为一类,通过计算每个类的类内颜色空间方差与类间颜色空间方差,得到颜色的空间分布情况。颜色空间分布越大,表明显著性越低,目标为背景的概率较大;反之,则表明目标为显著性区域的概率较大。第  $t$  个颜色分量的颜色对比值  $V_t$  可表示为

$$V_t = \left[ \exp(-D_t^{\text{in}}) \times \exp(-D_t^{\text{out}}) + \frac{1 - \exp(-D_t^c)}{2\sigma^2} \right], \quad (12)$$

$$D_t^{\text{in}} = \frac{\sum_{z_i \in m_t} \| \mathbf{z}_i - \boldsymbol{\mu}_t \|^2}{n_t}, \quad (13)$$

$$D_t^{\text{out}} = \sum_{i=1}^M D_t^{\text{in}} = \sum_{i=1}^M \frac{\sum_{z_i \in m_i} \|z_i - \mu_i\|^2}{n_i}, \quad (14)$$

$$D_t^c = \frac{\sum_{z_i \in m_i} \|c_i - c_i^c\|^2}{n_i}, \quad (15)$$

式中,  $D_t^{\text{in}}$  为类内空间方差, 即图像中每个类的颜色位置到该类中心位置的距离;  $D_t^{\text{out}}$  为类间空间方差, 即类内颜色位置与其他类中心位置的偏移程度;  $D_t^c$  为表示类内颜色相似性的类内颜色方差,  $\mu_i$  为聚类中心,  $z_i$  为像素点的坐标,  $m_i$  为该颜色分量像素的集合,  $c_i$  为像素点的  $l$ 、 $a$ 、 $b$  分量,  $c_i^c$  为颜色分量对

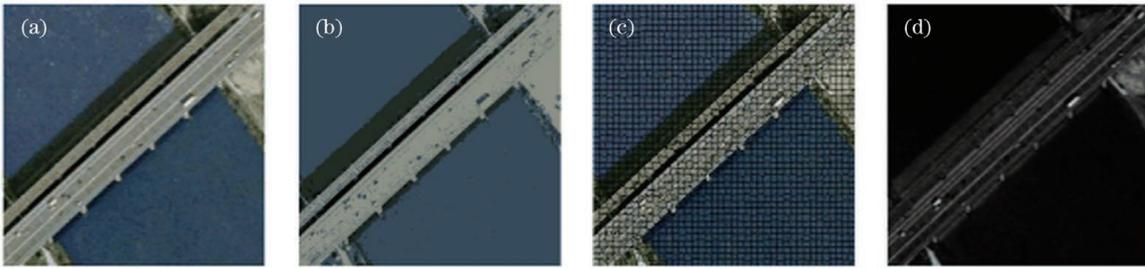


图 3 显著图的提取结果。(a)原始图像;(b) $K$ -means 聚类;(c)超像素分割;(d)融合结果

Fig. 3 Extraction result of saliency map. (a) Original image; (b)  $K$ -means clustering; (c) superpixel segmentation; (d) fusion result

### 3.2 感兴趣区域的提取

为了更好地区分显著图中的目标物体, 需提取显著图的 ROI, 具体步骤如下。

1) 显著图的增强: 显著图中灰度较高的部分比较集中, 而目标往往处于低灰度区域, 对数变换能扩展影像中范围较窄的灰度区域, 因此, 可通过压缩高灰度区域达到强调图像低灰度区域的目的, 具体操作可表示为

$$g(x, y) = c \log [1 + vr(x, y)], \quad (17)$$

式中,  $g(x, y)$  为坐标点  $(x, y)$  处增强后的灰度值,  $c$  为常数项,  $r(x, y)$  为原始图像在  $(x, y)$  处的灰度值,  $v$  为灰度的扩展程度, 当  $v$  在  $[10, 50]$  之间时, 对影像的灰度增强效果最好。

#### 2) 基于自适应阈值的显著图分割

基于自适应阈值的图像分割方法可将图像分为背景 (Background) 和目标 (Object), 并通过设定的阈值使两类像素的区分度达到最大, 得到较为准确的分割效果, 可表示为

$$S(x, y) = \begin{cases} 0, & S(x, y) < T \\ 1, & S(x, y) > T \end{cases}, \quad (18)$$

式中,  $T$  为最佳分割阈值,  $S(x, y)$  为像素点  $(x, y)$  的像素值。

应通道的颜色均值,  $n_i$  为集合中像素的个数, 参数  $\sigma$  为 5。

#### 3) 显著图的生成

结合图像的空间分布与颜色分布特征, 融合颜色对比图与颜色空间分布图像, 得到最终的显著性区域。  $m$  点的显著值  $S_m$  可表示为颜色对比值  $C_m$  与对应区域颜色空间分布值  $V_m$  的积, 即

$$S_m = C_m \times V_m. \quad (16)$$

图 3 为融合后的显著图, 可以发现, 融合后的显著图可以有效提取图像的显著性区域, 具有较好的检测效果。

### 3) 感兴趣区域的提取

为了分离目标区域和背景区域, 根据二值化结果中各个像素点的取值, 设置原始影像上各个对应像素点的值, 从而在原始图像上构建 ROI。对于二值化图像上像素值为 0 的点, 对原始图像上对应的像素点赋值为 0, 即二值化图像上的非 0 部分就是所求的目标区域。此外, 数据集中交叉路口、河流和立交桥的目标区域像素值为 0, 因此, 可将这些图像中像素值为 1 的部分作为被切割对象, 使原始图像上对应像素点的值为 0, 如图 4 所示。

### 3.3 基于深度卷积神经网络的影像分类

#### 1) 网络结构

本方法构建的 DCNN 模型如图 5 所示, 共计 8 层, 包括 5 个卷积层 (Conv) 与池化层 (maxpool)、3 个全连接层 (FC), 此外, 还包括 1 个 Softmax 函数。其中, 卷积层与池化层交替排列设置, 相比单层次的结构, 该网络能提取到图像更本质的特征。

特征提取的具体步骤: 输入尺寸为  $256 \times 256 \times 3$  (3 为图像的通道数) 的图像, 第 1 层用尺寸为  $3 \times 3$ 、步长为 1 的 3 通道卷积核进行滤波操作, 输出图像的尺寸为  $256 \times 256 \times 16$ , 再经过  $2 \times 2$  滤波器进行池化操作后, 得到尺寸为  $128 \times 128 \times 16$  的特征

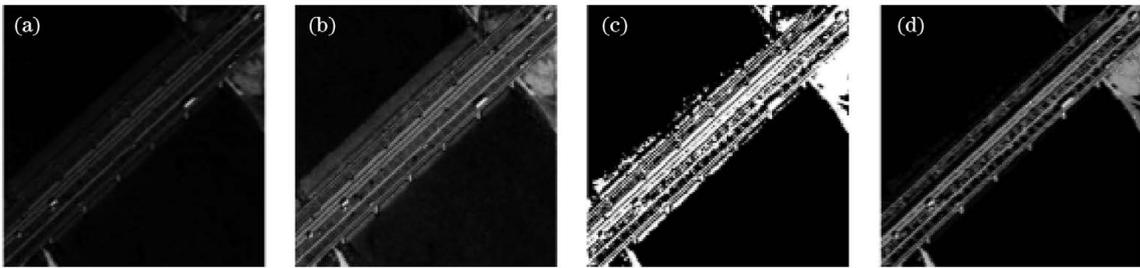


图 4 ROI 的提取结果。(a)显著图;(b)灰度增强图;(c)二值化图;(d)ROI

Fig. 4 Extraction result of the ROI. (a) Saliency map; (b) gray enhancement map; (c) binarization map; (d) ROI

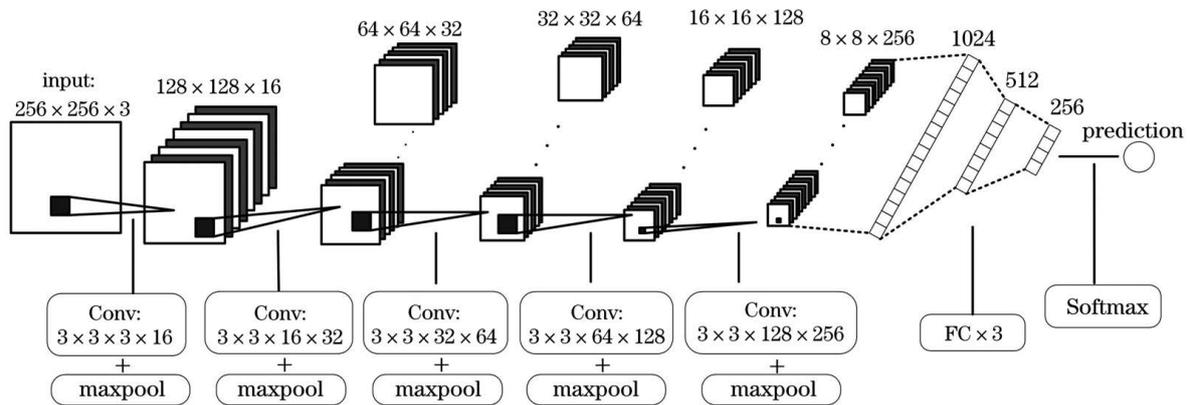


图 5 DCNN 的结构

Fig. 5 Structure of the DCNN

图。第 2 层用尺寸为  $3 \times 3$ 、步长为 1 的 16 通道卷积核对第 1 层的输出结果进行滤波操作,得到尺寸为  $128 \times 128 \times 16$  的特征图,再经过池化操作得到尺寸为  $64 \times 64 \times 32$  的特征图。第 3 次、第 4 次和第 5 次卷积和池化操作后,输出特征图的尺寸分别为  $32 \times 32 \times 64$ 、 $16 \times 16 \times 128$  和  $8 \times 8 \times 256$ 。将生成的特征图输入全连接层中,经过第 1 个全连接层后输出 1 个 1024 维向量;经过第 2 个全连接层后输出 1 个 512 维向量;经过第 3 个全连接层后输出 1 个 256 维向量,即特征图转换成了  $1 \times 256$  维的向量。最后,将特征图输入 Softmax 分类器,得到最终的预测结果。

为了提高网络的非线性映射能力,避免梯度消失

现象,在每个卷积层、池化层和全连接层后引入线性整流单元(ReLU)作为网络的激活函数,可表示为

$$f_{\text{ReLU}}(z) = \max(0, z), \quad (19)$$

式中,  $z$  为输入数据,当输入值非负时,返回原值;否则,返回 0。为避免过拟合问题,在每个全连接层后加入 Dropout 函数,以减少模型对某些局部特征的过分依赖。

## 2) 数据增广

深度学习方法需要利用大量数据学习和训练模型,为了得到更好的训练结果和泛化能力,需要对数据集进行扩充,实验分别对原始图像进行随机水平翻转、垂直翻转和亮度调节,以达到扩充训练集的目的,图像的扩充结果如图 6 所示。



图 6 数据的扩充结果。(a)原始图像;(b)水平翻转;(c)垂直翻转;(d)亮度调节

Fig. 6 Expanded results of the data. (a) Original image; (b) horizontal flip; (c) vertical flip; (d) brightness adjustment

### 3) 模型参数调整

为了增加网络训练的准确性和高效性,获得最优的训练结果,对 CNN 中的几个重要参数进行调整,以降低网络训练的难度。具体参数:Batch size 为 100 张,学习率为 0.0001,衰减系数为 0.00001,卷积核大小为  $3 \times 3$ 。通过不同的迭代次数,判断模型的分类精度和损失,其中,epoch 为梯度下降的超参数,定义了学习算法在整个训练数据集中的迭代次数,可影响模型的泛化能力。不同数据集下模型

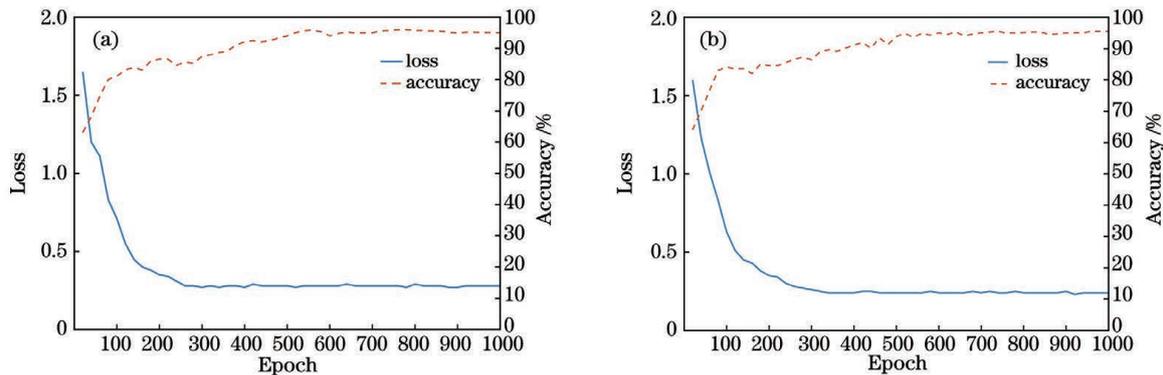


图 7 不同模型的损失函数和分类精度。(a)UC-Merced 数据集;(b)WHU-RS 数据集

Fig. 7 Loss function and classification accuracy of different models. (a) UC-Merced data set; (b) WHU-RS data set

## 4 实验结果与分析

### 4.1 实验数据集

1) UC-Merced 数据集<sup>[17]</sup>是由 UC-Merced 计算机视觉实验室在 2010 年发布的数据集,选自美国 USGS 国家城市地图中不同地区的航空遥感影像,如图 8(a)所示。该数据集包含 2100 张遥感图像,共 21 类场景,每类 100 张图像,分辨率为 0.3 m,每张图像的尺寸为 256 pixel $\times$ 256 pixel。

2) WHU-RS 数据集<sup>[18]</sup>是武汉大学在 2011 年发布的数据集,如图 8(b)所示。该数据集包含地球遥感影像中不同地区的 19 个场景类别,每个类别包括 50 张尺寸为 600 pixel $\times$ 600 pixel 的图像,总计 950 张。

### 4.2 实验设置

在 UC-Merced 数据集和 WHU-RS 数据集中均随机选取 80% 的图像作为训练样本,剩余图像作为测试样本,并对比本方法和其他方法的分类精度。

#### 1) 实验环境

实验在显卡为 NVIDIA GeForce GTX1080、Inter core i7-7700K CPU@4.00 GHz、内存为 64 G 的设备中完成,代码由 Matlab 和 Python 语言编写,DCNN 框架由 TensorFlow 实现。

的精度和损失如图 7 所示,可以发现,随着训练次数的增加,模型的损失值逐渐下降并趋于稳定,这表明模型预测结果与实际结果的误差不断降低,直到达到最小值。此外,测试集的分类精度不断提高,这表明模型的泛化能力较好。当 UC-Merced 数据集和 WHU-RS 数据集的迭代次数分别超过 600 次和 500 次后,训练样本的损失函数和测试样本的精度均趋于稳定,表明模型达到了较优的水平。

#### 2) 评价指标

混淆矩阵(Confusion matrix)是一种衡量分类效果的可视化工具,采用矩阵的格式对比分类结果和真实数据,可表示为

$$C = \begin{bmatrix} m_{11} & \cdots & m_{1c} \\ \vdots & & \vdots \\ m_{c1} & \cdots & m_{cc} \end{bmatrix}, \quad (20)$$

式中, $c$  为类别总数, $m_{ij}$  为样本数据中实际属于第  $i$  类但被识别为第  $j$  类的样本数量,矩阵对角线上的数值越大,表明分类精度就越高。

总体分类精度(OA)是一个概率统计值,定义为被正确分类的样本数量和总分类样本数量的比值,可以从总体上评价分类方法对整个数据集的分类性能,也可以单独计算每个类别的总体分类精度,即某个类别中被识别正确的个数占该类别数量的比例。各类别的平均分类精度(AA)指对应类别分类精度的平均值。

### 4.3 对比方法

用本方法和已有相关分类方法在数据集上进行对比实验,以验证本方法的有效性和准确性。对比方法包括何小飞等<sup>[11]</sup>结合显著性样本块和 CNN 的(Saliency + Multi-CNN)方法、许凤晖等<sup>[19]</sup>基于多尺度特征与 DCNN 模型的(MS-DCNN)场景分类

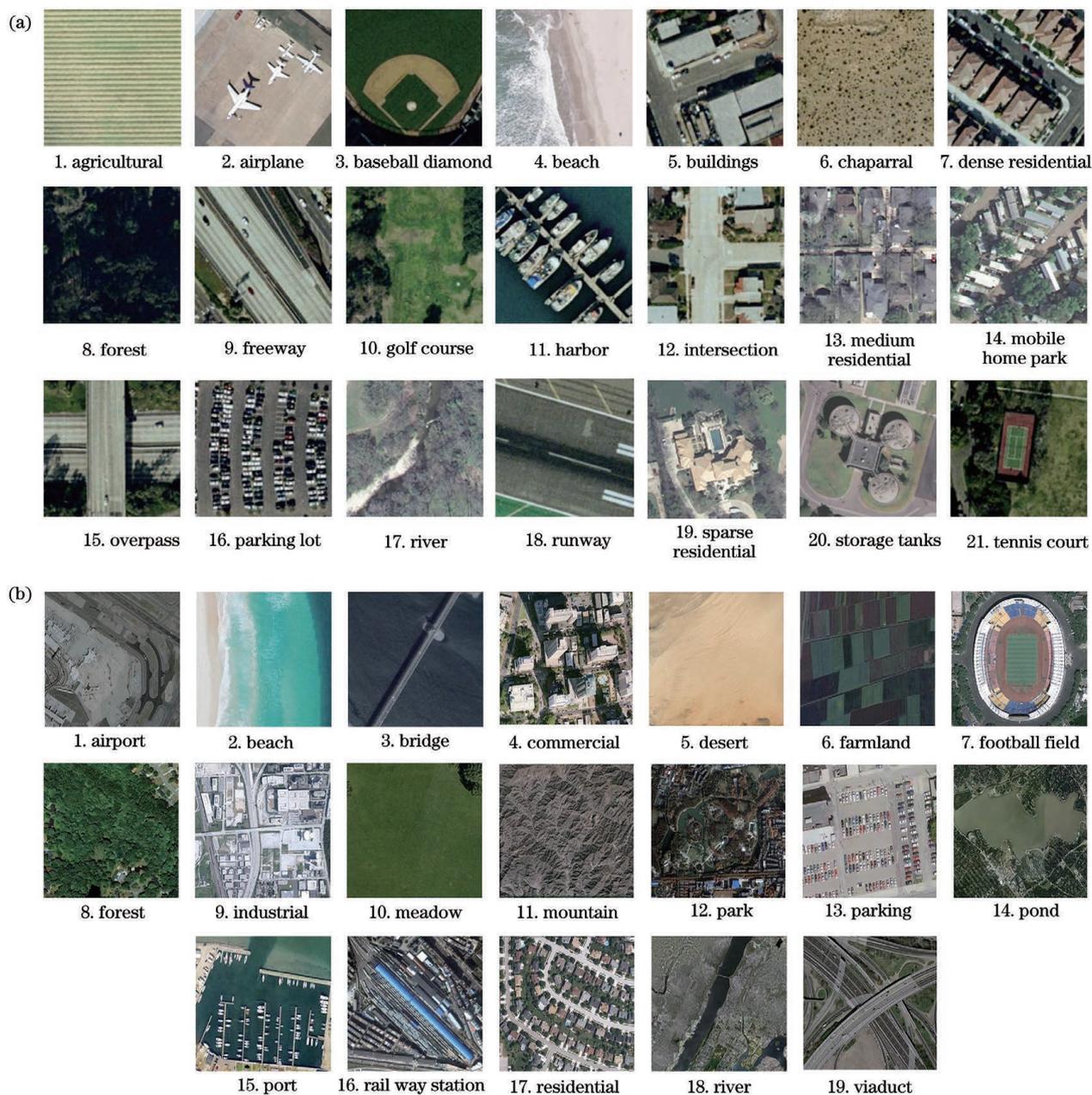


图 8 不同数据集中的图像。(a)UC-Merced 数据集;(b)WHU-RS 数据集

Fig. 8 Images in different data sets. (a) UC-Merced data set; (b) WHU-RS data set

方法以及郑卓等<sup>[20]</sup>提出的联合多尺度 CNN (JMCNN)分类方法。其中,MS-DCNN 方法利用非下采样 Contourlet 变换对遥感图像进行多尺度分解,并采用多核支持向量机(MKSVM)实现遥感图像的场景区分,取得了较高的分类精度。JMCNN 方法利用多通道特征提取和多尺度特征融合增强特征表达,并利用 Softmax 分类器进行分类,计算速度和分类精度均较高。

#### 4.4 UC-Merced 数据集的实验结果

本方法对原始高分遥感图像(飞机、棒球场、海港、河流、储罐和稀疏住宅)的分类结果如图 9 所示,

可以发现,本方法提取的图像能够保留原始场景中显著目标的重要特征,减少背景信息对分类结果的干扰。

本方法对 UC-Merced 数据集中各类别的分类精度如表 1 所示,图 10 为对应的混淆矩阵。可以发现,本方法在 UC-Merced 数据集中的平均分类精度为 96.10%,在农田、棒球场、沙滩、灌木丛、森林、高速公路、高尔夫球场、海港、交叉路口、移动房区、立交桥、停车场、河流、跑道和稀疏住宅这 15 个类型的场景中均取得了较好的分类精度。其中,12 类场景的分类精度达到了 100%,原因是这些场景往往有

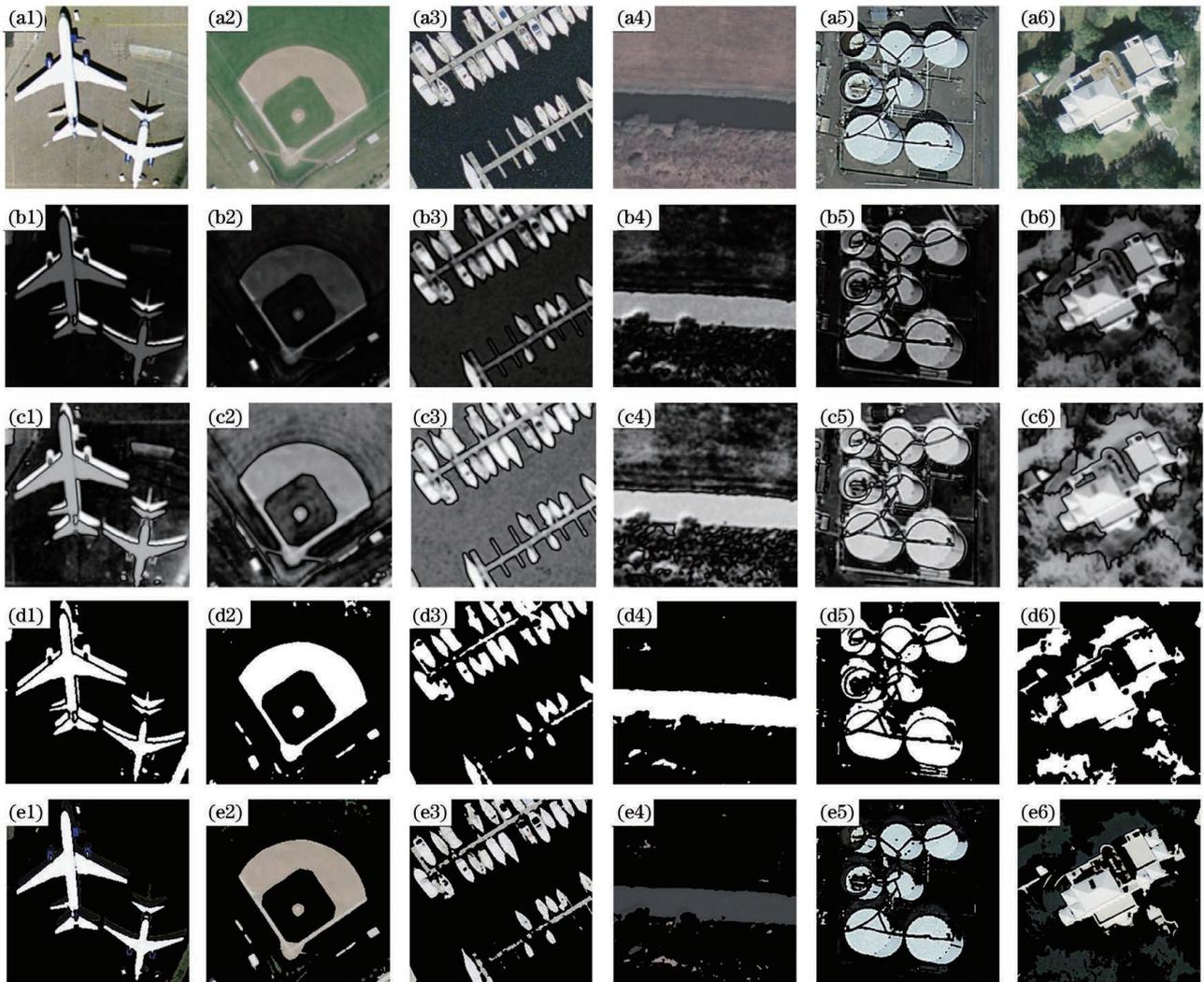


图 9 UC-Merced 数据集的实验结果。(a)原始图像;(b)显著图;(c)灰度增强图;(d)二值化图;(e)ROI

Fig. 9 Experimental results of the UC-Merced data set. (a) Original image; (b) saliency map; (c) gray enhancement map; (d) binarization map; (e) ROI

表 1 UC-Merced 数据集的分类精度

Table 1 Classification accuracy of the UC-Merced data set

unit: %

| No. | Scene             | Accuracy | No. | Scene              | Accuracy | No. | Scene              | Accuracy |
|-----|-------------------|----------|-----|--------------------|----------|-----|--------------------|----------|
| 1   | agricultural      | 96       | 8   | forest             | 95       | 15  | overpass           | 100      |
| 2   | airplane          | 93       | 9   | freeway            | 100      | 16  | parking lot        | 100      |
| 3   | baseball diamond  | 100      | 10  | golf course        | 100      | 17  | river              | 100      |
| 4   | beach             | 100      | 11  | harbor             | 100      | 18  | runway             | 96       |
| 5   | buildings         | 84       | 12  | intersection       | 100      | 19  | sparse residential | 100      |
| 6   | chaparral         | 100      | 13  | medium residential | 84       | 20  | storage tanks      | 88       |
| 7   | dense residential | 92       | 14  | mobile home park   | 100      | 21  | tennis court       | 90       |

Average accuracy

96.10

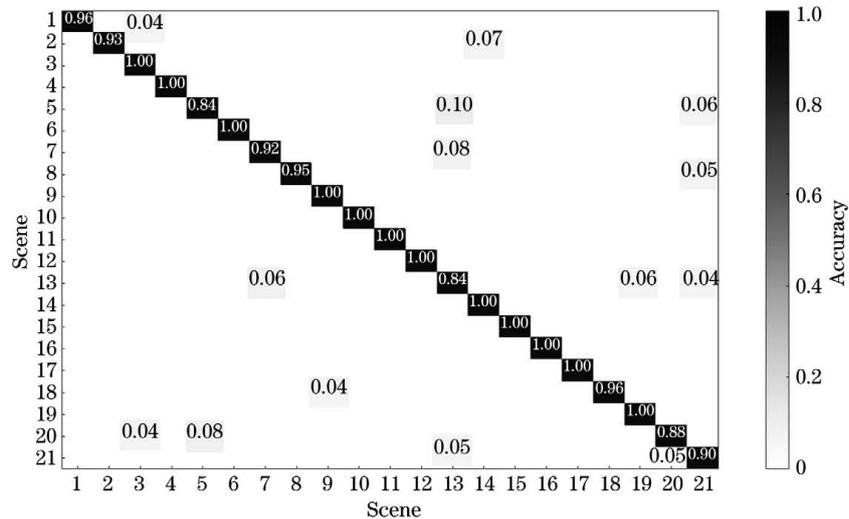


图 10 本方法在 UC-Merced 数据集上的混淆矩阵

Fig. 10 Confusion matrix of our method on the UC-Merced data set

较明显的特征,而本方法能提取出这些具有鉴别性的显著性特征和深层语义特征。在目标的纹理、颜色等特征具有一定相似性的场景中,本方法的分类精度有所降低,飞机、建筑、密集型住宅、中密度住宅、储罐、网球场这 6 类场景的平均分类精度仅能达到 88.5%。原因是这些场景的相似度较高,显著性检测的结果都是建筑物,区别仅在于建筑物的大小和密集程度。此外,储罐由于大小、颜色等类内差异较大,且高分影像均为俯视角拍摄,易与棒球场和颜色、形状相似的建筑物混淆,分类精度低于 90%,但这些易混淆场景类别的分类精度仍维持在较高水准。这表明本方法构建的深度网络模型能对深层次的语义特征进行提取和处理,对于相似度较高的场景仍能进行有效分类。

不同方法在 UC-Merced 数据集上的分类精度如表 2 所示,可以发现,在 UC-Merced 数据集上, Saliency + Multi-CNN 方法的分类精度达到了 92.86%,虽然该方法利用样本块标签获得不同层的特征表达,但没有考虑背景和目标之间的关系,降低了模型的区分能力,导致分类结果的提升效果有限。

表 2 不同方法在 UC-Merced 数据集上的分类精度

Table 2 Classification accuracies of different methods on the UC-Merced data set

| Method               | Accuracy / % | Time / s |
|----------------------|--------------|----------|
| Saliency + Multi-CNN | 92.86        | 2.27     |
| MS-DCNN              | 91.34        | 5.90     |
| JMCNN                | 88.30        | 0.81     |
| Our method           | 96.10        | 1.95     |

MS-DCNN 方法综合了高频子带与低频子带的优势,提高了分类精度,但没有考虑影像的关键区域,而 ROI 包含了更多的影像信息,分类过程计算量大,每张图像的处理时间长。JMCNN 方法的多通道特征提取器由三个单通道浅层卷积网络构成,相比深层次的 CNN 模型,对影像深层特征的提取不足,分类精度较低,但降低了计算复杂度,分类效率较高。本方法的分类精度可达到 96.10%,且处理每张图像的时间仅为 1.95 s,这表明本方法能提取出影像的关键特征,并通过对背景区域和目标区域的分割提高了模型对复杂场景的区分能力,得到具有鉴别性的深层语义特征,能很好地解析和识别不同复杂程度的场景。

#### 4.5 WHU-RS 数据集的实验结果

本方法对 WHU-RS 数据集中原始高分遥感影像(桥、足球场、工业区和高架桥)的分类结果如图 11 所示,可以发现,本方法在背景复杂的场景中能提取出清晰的目标区域影像,从而实现高效的地物目标识别与分类。

本方法对 WHU-RS 数据集各类别的分类精度与总体分类精度如表 3 所示,对应的混淆矩阵如图 12 所示。可以发现,在 WHU-RS 数据集中,本方法对 10 类场景的分类精度达到 100%,沙滩、停车场和港口与 UC-Merced 数据集的分类精度一致,均为 100%,这表明本方法对这几类场景有着较好的泛化能力。对山、沙漠等纹理差异性小的场景分类精度为 100%,对复杂度较高的场景也能维持较高的分类水平。对于商业区、工业区和住宅区这些显著性区域相似度高的场景,本方法的平均精度仍



图 11 WHU-RS 数据集中的实验结果。(a)原始图像;(b)显著图;(c)灰度增强图;(d)二值化图;(e)ROI

Fig. 11 Experimental results in the WHU-RS data set. (a) Original image; (b) saliency map; (c) gray enhancement map; (d) binarization map; (e) ROI

表 3 WHU-RS 数据集的分类精度

Table 3 Classification accuracy of the WHU-RS data set

unit: %

| No. | Scene      | Accuracy | No. | Scene          | Accuracy | No. | Scene            | Accuracy |
|-----|------------|----------|-----|----------------|----------|-----|------------------|----------|
| 1   | airport    | 100      | 8   | football field | 100      | 15  | pond             | 100      |
| 2   | beach      | 100      | 9   | forest         | 92       | 16  | port             | 100      |
| 3   | bridge     | 100      | 10  | industrial     | 93       | 17  | rail way station | 93       |
| 4   | commercial | 91       | 11  | meadow         | 86       | 18  | residential      | 91       |
| 5   | desert     | 100      | 12  | mountain       | 100      | 19  | river            | 92       |
| 6   | farmland   | 91       | 13  | park           | 100      |     |                  |          |
| 7   | viaduct    | 92       | 14  | parking        | 100      |     |                  |          |

Average accuracy

95.84

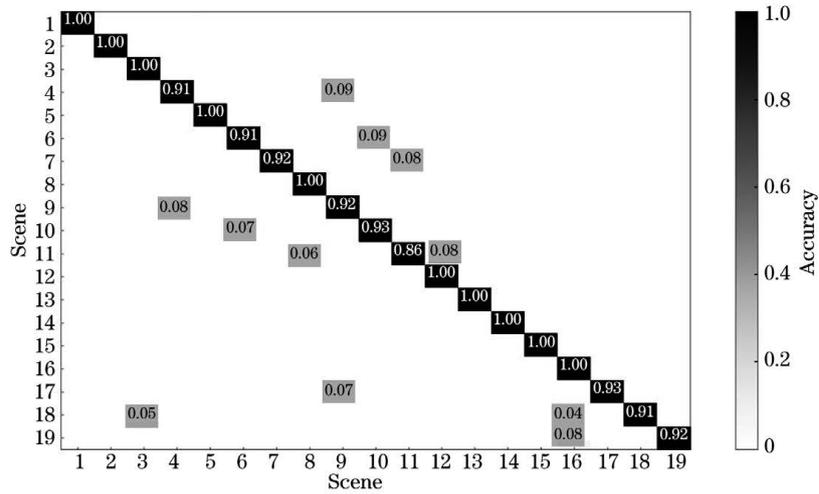


图 12 本方法在 WHU-RS 数据集上的混淆矩阵

Fig. 12 Confusion matrix of our method on the WHU-RS data set

达到了 91.67%。此外,本方法对绝大多数场景的分类精度高于 90%,仅有草地类别的平均精度为 86%,原因是草地的颜色和纹理单一,数据增广前后的差别不大,模型训练效果不明显。此外,草地的颜色特征与足球场和山的相似度较高,而本方法基于颜色对比度与空间分布的显著性检测区分效果较差。

不同方法在 WHU-RS 数据集上的分类结果如表 4 所示,可以发现,本方法在该数据集上的分类精度达到了 95.84%,高于其他三种对比方法。原因是 WHU-RS 数据集中包含了更多的地物信息,遥感图像的复杂度更高,分辨率也更高,导致本方法处理每张图像的时间有所增加,但仍维持在较高水平。这表明本方法能有效提取高分辨率遥感影像的显著性特征与深层语义特征,通过 ROI 与背景区域的分离实现了高效准确的场景分类。

表 4 不同方法在 WHU-RS 数据集上的分类精度

Table 4 Classification accuracies of different methods on the WHU-RS data set

| Method               | Accuracy /% | Time /s |
|----------------------|-------------|---------|
| Saliency + Multi-CNN | 91.80       | 3.13    |
| MS-DCNN              | 90.05       | 7.33    |
| JMCNN                | 87.63       | 0.98    |
| Our method           | 95.84       | 2.32    |

## 5 结 论

针对高分遥感影像的分类问题上,提出了一种基于显著性特征和 DCNN 的高分遥感影像场景分

类方法。该方法充分利用了影像的显著性特征和深层语义特征,通过对目标区域和背景区域的分割提取 ROI,从而提高模型对不同复杂程度场景的区分能力。利用构建的 DCNN 对具有鉴别性的深层特征进行提取和分类,提高了分类精度。但本方法仍存在许多不足,如分类过程需要人工构建神经网络模型,如何根据不同高分影像的特点选择最优的网络层数和改进网络结构还需进行深入研究。此外,如何提高本方法在纹理、颜色相似度较高场景中的分类精度也是需要进一步完善的问题。

## 参 考 文 献

[1] Li D R, Tong Q X, Li R X, et al. Current issues in high-resolution earth observation technology [J]. Scientia Sinica (Terrae), 2012, 42(6): 805-813. 李德仁, 童庆禧, 李荣兴, 等. 高分辨率对地观测的若干前沿科学问题[J]. 中国科学: 地球科学, 2012, 42(6): 805-813.

[2] Fu Q, Guo C, Luo W L. Land use change detection based on GF-1 satellite remote sensing images [J]. Laser & Optoelectronics Progress, 2020, 57(16): 162802. 付青, 郭晨, 罗文浪. 基于高分一号卫星遥感影像的土地利用变化检测[J]. 激光与光电子学进展, 2020, 57(16): 162802.

[3] Zhang X N, Zhong X, Zhu R F, et al. Scene classification of remote sensing images based on integrated convolutional neural networks [J]. Acta Optica Sinica, 2018, 38(11): 1128001. 张晓男, 钟兴, 朱瑞飞, 等. 基于集成卷积神经网络的遥感影像场景分类[J]. 光学学报, 2018, 38(11): 1128001.

[4] Zhao L J, Tang P, Huo L Z, et al. Review of the

- bag-of-visual-words models in image scene classification [J]. *Journal of Image and Graphics*, 2014, 19(3): 333-343.
- 赵理君, 唐娉, 霍连志, 等. 图像场景分类中视觉词包模型方法综述 [J]. *中国图象图形学报*, 2014, 19(3): 333-343.
- [5] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [6] Oliva A, Torralba A. Modeling the shape of the scene: a holistic representation of the spatial envelope [J]. *International Journal of Computer Vision*, 2001, 42(3): 145-175.
- [7] Yang Y, Newsam S. Spatial pyramid co-occurrence for image classification [C] // 2011 International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain. New York: IEEE Press, 2011: 1465-1472.
- [8] Zhao B, Zhong Y F, Zhang L P. Scene classification via latent Dirichlet allocation using a hybrid generative/discriminative strategy for high spatial resolution remote sensing imagery [J]. *Remote Sensing Letters*, 2013, 4(12): 1204-1213.
- [9] Fang X, Wang G H, Yang H C, et al. High resolution remote sensing image classification combining with mean-shift segmentation and fully convolution neural network [J]. *Laser & Optoelectronics Progress*, 2018, 55(2): 022802.
- 方旭, 王光辉, 杨化超, 等. 结合均值漂移分割与全卷积神经网络的高分辨率遥感影像分类 [J]. *激光与光电子学进展*, 2018, 55(2): 022802.
- [10] Ma Y J, Liu P P. Convolutional neural network based on DenseNet evolution for image classification algorithm [J]. *Laser & Optoelectronics Progress*, 2020, 57(24): 241001.
- 马永杰, 刘培培. 基于 DenseNet 进化的卷积神经网络图像分类算法 [J]. *激光与光电子学进展*, 2020, 57(24): 241001.
- [11] He X F, Zou Z R, Tao C, et al. Combined saliency with multi-convolutional neural network for high resolution remote sensing scene classification [J]. *Acta Geodaetica et Cartographica Sinica*, 2016, 45(9): 1073-1080.
- 何小飞, 邹峥嵘, 陶超, 等. 联合显著性和多层卷积神经网络的高分辨率遥感影像分类 [J]. *测绘学报*, 2016, 45(9): 1073-1080.
- [12] Liu F, Lu L X, Huang G W, et al. Landform image classification based on discrete cosine transformation and deep network [J]. *Acta Optica Sinica*, 2018, 38(6): 0620001.
- 刘芳, 路丽霞, 黄光伟, 等. 基于离散余弦变换和深度网络的地貌图像分类 [J]. *光学学报*, 2018, 38(6): 0620001.
- [13] Li G D, Zhang C J, Wang M K, et al. Transfer learning using convolutional neural network for scene classification within high resolution remote sensing image [J]. *Science of Surveying and Mapping*, 2019, 44(4): 116-123, 174.
- 李冠东, 张春菊, 王铭恺, 等. 卷积神经网络迁移的高分辨率影像场景分类学习 [J]. *测绘科学*, 2019, 44(4): 116-123, 174.
- [14] Ren Y M, Yang J, Guo Z Q, et al. Ship classification method for point cloud images based on three dimensional convolutional neural network [J]. *Laser & Optoelectronics Progress*, 2020, 57(16): 161022.
- 任永梅, 杨杰, 郭志强, 等. 基于三维卷积神经网络的点云图像船舶分类方法 [J]. *激光与光电子学进展*, 2020, 57(16): 161022.
- [15] Cui L L, Xu J L, Xu G, et al. Image saliency detection method based on a pair of feature maps [J]. *Journal of Image and Graphics*, 2018, 23(4): 583-594.
- 崔玲玲, 许金兰, 徐岗, 等. 融合双特征图信息的图像显著性检测方法 [J]. *中国图象图形学报*, 2018, 23(4): 583-594.
- [16] Meng Q X, Wu X. Scene classification of high-resolution remote sensing image based on deep convolution neural network [J]. *Bulletin of Surveying and Mapping*, 2019(7): 17-22.
- 孟庆祥, 吴玄. 基于深度卷积神经网络的高分辨率遥感影像场景分类 [J]. *测绘通报*, 2019(7): 17-22.
- [17] Zhao B, Zhong Y F, Xia G S, et al. Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(4): 2108-2123.
- [18] Dai D X, Yang W. Satellite image classification via two-layer sparse coding with biased image representation [J]. *IEEE Geoscience and Remote Sensing Letters*, 2011, 8(1): 173-176.
- [19] Xu S H, Mu X D, Zhao P, et al. Scene classification of remote sensing image based on multi-scale feature and deep neural network [J]. *Acta Geodaetica et Cartographica Sinica*, 2016, 45(7): 834-840.
- 许凤晖, 慕晓冬, 赵鹏, 等. 利用多尺度特征与深度神经网络对遥感影像进行场景分类 [J]. *测绘学报*, 2016, 45(7): 834-840.
- [20] Zheng Z, Fang F, Liu Y Y, et al. Joint multi-scale convolution neural network for scene classification of high resolution remote sensing imagery [J]. *Acta Geodaetica et Cartographica Sinica*, 2018, 47(5): 620-630.
- 郑卓, 方芳, 刘袁缘, 等. 高分辨率遥感影像场景的多尺度神经网络分类法 [J]. *测绘学报*, 2018, 47(5): 620-630.