

基于深度对抗域适应的高分辨率遥感影像跨域分类

滕文秀^{1**}, 王妮^{2,3*}, 陈泰生^{2,3}, 王本林^{2,3,4}, 陈梦琳^{2,3}, 施慧慧³

¹南京林业大学林学院, 江苏 南京 210037;

²滁州学院地理信息与旅游学院, 安徽 滁州 239000;

³安徽省地理信息智能感知与服务工程实验室, 安徽 滁州 239000;

⁴河海大学地球科学与工程学院, 江苏 南京 210098

摘要 提出一种基于深度对抗域适应的高分辨率遥感影像跨域分类方法。利用深度卷积神经网络 VGG16(Visual Geometry Group)学习场景影像的深度特征,然后利用对抗学习方法最小化源域和目标域特征分布差异。利用 RSI-CB256(Remote Sensing Image Classification Benchmark)、NWPU-RESISC45(Northwestern Polytechnical University Remote Sensing Image Scene Classification)和 AID(Aerial Image data set)数据集构建源域数据集,并将 UC-Merced(University of California, Merced)和 WHU-RS 19(Wuhan University Remote Sensing)两个数据集作为目标域数据集进行实验,实验结果表明,所提方法在目标域数据集没有标签的情况下,能够提高模型对目标域数据集的泛化能力。

关键词 遥感; 场景分类; 无监督域适应; 卷积神经网络; 生成对抗网络

中图分类号 TP753

文献标识码 A

doi: 10.3788/LOP56.112801

Deep Adversarial Domain Adaptation Method for Cross-Domain Classification in High-Resolution Remote Sensing Images

Teng Wenxiu^{1**}, Wang Ni^{2,3*}, Chen Taisheng^{2,3}, Wang Benlin^{2,3,4},
Chen Menglin^{2,3}, Shi Huihui³

¹College of Forest, Nanjing Forestry University, Nanjing, Jiangsu 210037, China;

²School of Geographic Information and Tourism, Chuzhou University, Chuzhou, Anhui 239000, China;

³Anhui Engineering Laboratory of Geographical Information Intelligent Sensor and Service,
Chuzhou, Anhui 239000, China;

⁴School of Earth Sciences and Engineering, Hohai University, Nanjing, Jiangsu 210098, China

Abstract In this study, a deep adversarial domain adaptation method is proposed for cross-domain classification in high-resolution remote sensing images. A deep convolutional neural network VGG16 is used to learn the deep features of scene images. The adversarial learning method is used to minimize the difference of feature distribution between source and target domains. RSI-CB256 (Remote Sensing Image Classification Benchmark), NWPU-RESISC45 (Northwestern Polytechnical University Remote Sensing Image Scene Classification) and AID (Aerial Image data set) are used as source domain datasets, and UC-Merced (University of California, Merced) and WHU-RS 19 (Wuhan University Remote Sensing) are used as target domain datasets. The experimental results denote that the proposed method can improve the generalization ability of the model for target domain dataset without labels.

Key words remote sensing; scene classification; unsupervised domain adaptation; convolutional neural network; generative adversarial networks

OCIS codes 280.4788; 100.4996; 100.2960

收稿日期: 2018-12-17; 修回日期: 2018-12-18; 录用日期: 2018-12-25

* E-mail: wnstrive@163.com; ** E-mail: wenxiu_teng@163.com

1 引言

随着全球对地观测系统的不断发展和完善,地球空间信息的获取已经进入了“三高”(高空间分辨率、高光谱分辨率、高时间分辨率)和“三多”(多平台、多传感器和多角度)时代^[1],大量的对地观测影像数据已经成为人类认识理解世界的重要信息来源^[2]。高分辨率遥感影像分类是遥感影像智能解译的一项重要任务。遥感影像空间分辨率的逐步提升,改变了遥感对地观测的尺度,小尺度的观测使遥感影像能够提供复杂的地表结构信息,为地表更高层次的语义信息提取提供了数据源^[3]。然而,遥感影像的像素或者超像素仅包含地物底层信息,不能对影像高层语义内容进行描述,因此,传统的“面向像素”或“面向对象”的分类方法无法满足遥感影像高层次内容解译的需求^[4]。在此背景下,“面向场景”的遥感影像分类引起了广泛关注,并已成为遥感领域的一个活跃的研究课题。

近年来,几种“面向场景”的遥感影像分类方法相继提出,如基于颜色、纹理、形状等底层特征的场景分类方法^[5-7],基于视觉词袋模型和概率主题模型的中层特征场景分类方法^[8-11]以及基于深度特征的场景分类方法等。基于深度卷积神经网络^[12-13]遥感影像场景分类研究在2015年前后相继出现,并在几个基准的遥感数据集上取得了先进结果^[14],其研究主要依据两条思路展开:一是将基于大规模自然影像数据训练得到的深度卷积神经网络模型直接或经过微调后应用于遥感影像场景分类,Penatti等^[15]首次利用深度卷积神经网络处理高分遥感场景分类问题,将大规模自然影像数据集上预训练的深度神经网络直接迁移至遥感场景数据集,利用神经网络中间层的输出特征作为场景全局特征表达,取得了突出的分类效果;龚希等^[16]利用包含场景局部细节信息的卷积层特征和包含场景全局信息的全连接层特征表达遥感影像场景。二是基于公开的遥感影像场景数据集设计并训练一个新的卷积神经网络^[17],Basu等^[18]基于两个小尺度遥感影像场景数据集测试了不同深度的卷积神经网络分类性能;Zhang等^[19]提出一种参数共享的卷积神经网络集成框架,并利用基于UC-Merced数据集训练的4种不同结构的卷积神经网络模型对遥感图像场景进行分类。

上述方法大多针对一个域进行实验,机器学习方法假设训练数据和测试数据独立同分布,当训练数据与测试数据之间的数据分布存在差异时,模型

性能大幅下降^[20]。针对单个领域分类任务设计的监督学习方法存在两个问题:1)不同遥感数据集之间由于遥感平台、传感器、拍摄角度、分辨率、拍摄时间的差异,造成同一类型的地物在不同的域中差异巨大,针对单个领域的分类方法对跨域的数据集不具有稳健性,模型泛化能力较弱;2)监督学习方法需要通过目视解译或地面调查获取大量的标记样本,该方法成本很高,并且人工标注很难跟上当前影像获取的步伐,特别是当任务涉及到非常大的区域或无法进行目视解译时,如树种分类、土地利用分类和叶绿素浓度分类等。从机器学习的角度来看,这种数据漂移问题可以通过迁移学习方法来解决,具体为无监督域适应(UDA,简称域适应),它包括有标签源域和无标签目标域两个数据集,假设源域和目标域类别空间和特征空间一致,仅特征的分布不同,目的是通过域适应方法学习到域不变的特征表示,提高模型的泛化能力。综上所述,针对遥感影像标注数据获取困难、模型泛化能力较弱的问题,如何充分利用过去的知识和恰当的域适应方法,训练出泛化能力较强的模型,是目前遥感领域亟待解决的热点和难点问题^[21]。

近年来,域适应问题的研究大多针对几个数据量较小的办公数据集和手写数字数据集^[22-24],和自然图像相比,遥感影像数据更复杂多变,因此,将自然图像的域适应方法直接用于遥感影像上很难取得较好的分类效果。也有一些关于遥感影像的域适应方法的针对性研究,如文献^[25]中利用预训练模型初始化特征,利用源域数据的交叉熵损失、最大均值差异(MMD)减少源数据和目标数据的分布差异;文献^[26]中通过非对称自适应层处理数据转移问题,并通过联合最小化分布的差异和类别分离学习网络权重,由于缺少丰富类别的源域数据集,这些研究大多只对数据集中的少部分类别进行研究,具有一定的局限性。

针对上述问题,本文提出一种基于深度对抗域适应的高分辨率遥感影像跨域分类方法。利用深度卷积神经网络VGG16学习场景影像的深度特征;利用对抗学习方法最小化源域和目标域特征分布差异,以提高模型对目标域的分类性能;为了验证所提方法的有效性利用RSI-CB256(Remote Sensing Image Classification Benchmark)、NWPU-RESISC45(Northwestern Polytechnical University Remote Sensing Image Scene Classification)和AID(Aerial Image data set)数据集构建源域数据集,利

用 UC-Merced (University of California, Merced) 和 WHU-RS19 (Wuhan University Remote Sensing) 两个通用数据集作为目标域数据集进行实验。

2 原理与方法

2.1 生成对抗网络原理

生成式对抗网络模型 (GANs) 是 Goodfellow 等^[27]提出的一种生成式模型,用于解决许多计算机视觉领域中的相关任务,如图像生成、图像超分辨率、无监督学习、半监督学习以及图像着色等。GANs 的基本思想源自博弈论的二人零和博弈,由一个生成器 (Generator) 和一个判别器 (Discriminator) 组成,其通过对抗学习的方式进行训练,目的是估测数据样本的潜在分布,并生成新的数据样本。

从一个任意的分布采样出随机噪声,并输入到生成器。生成器尽可能拟合真实数据的分布以欺骗判别器;判别器的输入数据包括真实样本和生成器

生成的假样本两个部分,判别器则尽可能区分出两部分样本,生成器和判别器二者通过相互对抗更新迭代,最终达到一个纳什均衡。此时,生成器能够拟合真实数据的分布,生成接近真实数据的样本,判别器无法区分数据来自生成器还是现实世界。GAN 的损失函数可表示为

$$\min_G \max_D [V(G, D)] = E_{x \sim P_{\text{data}}(x)} \log \times [D(x)] + E_{z \sim P_z(z)} \log \{1 - D[G(z)]\}, \quad (1)$$

式中: x 为来自真实数据分布 $P_{\text{data}}(x)$ 的真实图像; z 为来自分布 $P_z(z)$ 的采样噪声; $G(z)$ 表示生成的图像; \log 表示似然代价函数; $E_{\cdot \sim \Omega}[f(\cdot)]$ 表示函数 $f(\cdot)$ 在分布 Ω 下的数学期望。生成器 G 通过使 $D[G(z)]$ 最大化进行学习,而判别器 D 通过使 $D[G(z)]$ 最小化进行训练。根据(1)式给出的损失函数,最优模型可表示为

$$G^* = \min_G \max_D [V(G, D)]. \quad (2)$$

对于给定的最优生成器 G ,最优的判别器为 $\max_D [V(G, D)]$,生成最优判别器的过程为

$$\begin{aligned} V(G, D) &= E_{x \sim P_{\text{data}}(x)} \log [D(x)] + E_{x \sim P_G} \log [1 - D(x)] = \\ &= \int_x \{P_{\text{data}}(x) \cdot \log [D(x)]\} dx + \int_x \{P_G(x) \cdot \log [1 - D(x)]\} dx = \\ &= \int_x \{P_{\text{data}}(x) \cdot \log D(x) + P_G(x) \cdot \log [1 - D(x)]\} dx. \end{aligned} \quad (3)$$

若(3)式取最大值,对于给定的 x ,积分项须最大。在给定数据和 G 的前提下, $P_{\text{data}}(x)$ 与 $P_G(x)$ 都可视为常数,分别用 a 、 b 来表示,则函数 $a \log(D) + b \log(1 - D)$ 关于 D 的最大值为 $a/(a + b)$,因此判别器的最优解为

$$D^*(x) = P_{\text{data}}(x) / [P_{\text{data}}(x) + P_G(x)]. \quad (4)$$

给定最优判别器 D^* 和优化生成器 G ,将 D^* 代入 $V(G, D)$,可得

$$\begin{aligned} \max [V(G, D)] &= V(G, D^*) = E_{x \sim P_{\text{data}}(x)} \log \{P_{\text{data}}(x) / [P_{\text{data}}(x) + P_G(x)]\} + \\ &= E_{x \sim P_G} \log \{P_G(x) / [P_{\text{data}}(x) + P_G(x)]\} = \int_x P_{\text{data}}(x) \cdot \log \{2P_{\text{data}}(x) / [P_{\text{data}}(x) + P_G(x)]\} dx + \\ &= \int_x P_G(x) \cdot \log \{2P_G(x) / [P_{\text{data}}(x) + P_G(x)]\} dx = \\ &= -2 \log 2 + \text{KL}\{P_{\text{data}}(x) \parallel [P_{\text{data}}(x) + P_G(x)]/2\} + \text{KL}\{P_G(x) \parallel [P_{\text{data}}(x) + P_G(x)]/2\} = \\ &= 2 \cdot \text{JS}[P_{\text{data}}(x) \parallel P_G(x)] - 2 \log 2, \end{aligned} \quad (5)$$

式中:KL 表示 KL 散度 (Kullback-Leibler divergence); JS 表示 JS (Jensen-Shannon Divergence) 散度。由(5)式可知,在给定最优判别器 D 的条件下,GANs 的损失函数等价于最小化真实分布 $P_{\text{data}}(x)$ 和 $P_G(x)$ 之间的 JS 散度,当交替更新优化判别器和生成器时,判别器逐渐接近最优,生

成器会使 $P_{\text{data}}(x)$ 和 $P_G(x)$ 之间的 JS 散度越来越小,即生成器 G 最终会使真实数据和生成数据的分布非常接近。从 GANs 的原理可以看出,生成器和判别器相互对抗训练能够对齐两个域的分布,为此将对抗学习方法引入高分辨率遥感影像跨域分类,最小化源域和目标域特征分布的差异。

2.2 本文方法

问题描述: 给定一个具有 n^s 个标记的源域 $X_s = \{(x_i^s, y_i^s)\}_{i=1}^{n^s}$ 和一个具有 n^t 个无标记的目标域 $X_t = \{(x_j^t)\}_{j=1}^{n^t}$, 假设两者的特征空间相同和类别空间均相同, 即 $X_s = X_t$ 且 $Y_s = Y_t$, 但源域和目标域的特征分布不同, 即 $P_s(X_s, Y_s) \neq P_t(X_t, Y_t)$, 目标是通过最小化源域和目标域特征分布 (P_s 和 P_t) 之间的差异指导特征学习, 学习能够正确分类目标域的特征 M_t , 利用对

抗域适应方法提高模型对目标域数据集的泛化能力。所提方法总体框架如图 1 所示。将大型遥感影像场景分类数据集作为源域 X_s , 利用深度卷积神经网络 VGG16 学习源域数据的深度特征 M_s , 并利用源域数据的特征重新初始化一个深度卷积神经网络, 采用对抗学习方法最小化源域和目标域数据特征分布的差异, 进而学习目标域的深度特征 M_t , 最后利用 M_t 和分类器 C 对无标注目标域进行分类。

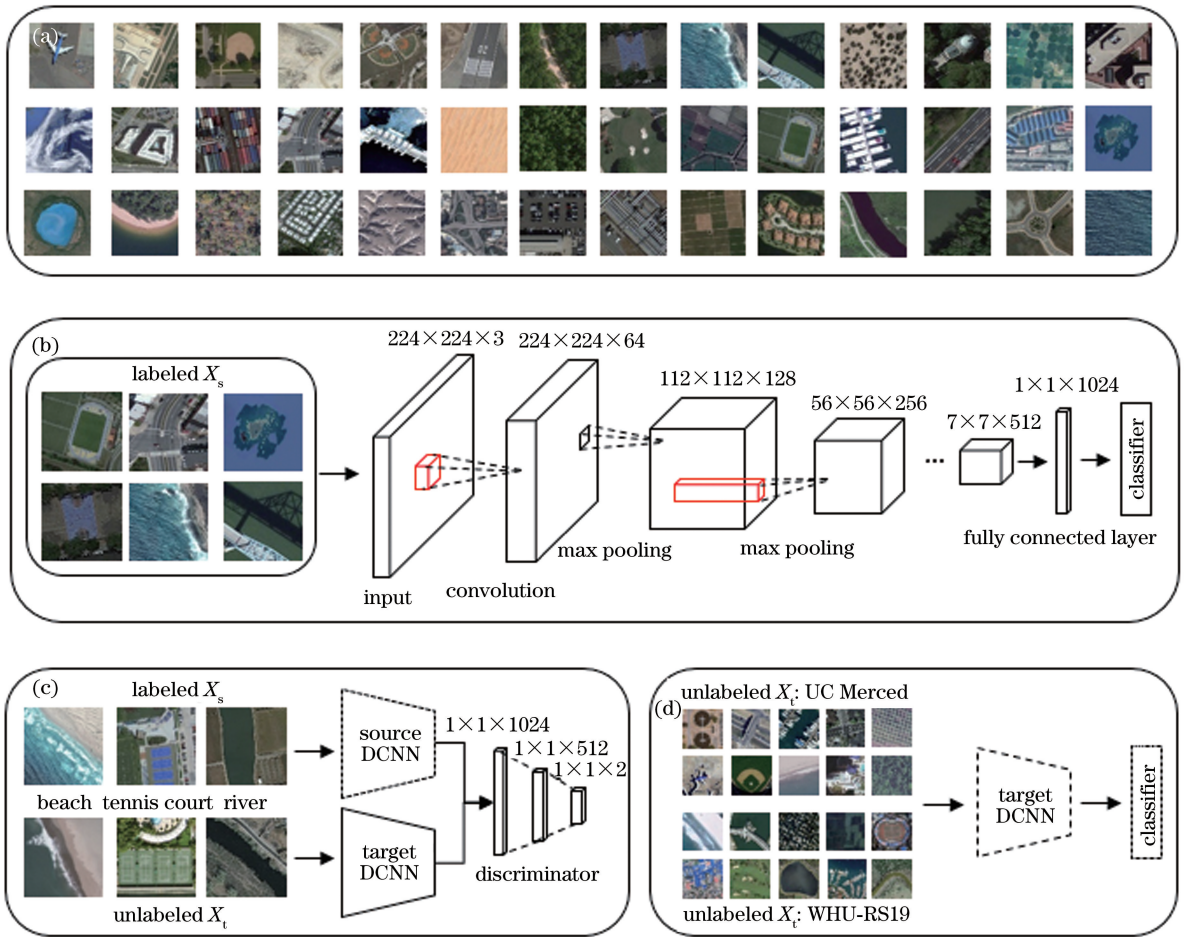


图 1 所提方法框架。(a) 高分辨率遥感影像数据集; (b) 源域深度卷积神经网络; (c) 无监督对抗域适应; (d) 遥感影像场景分类

Fig. 1 Framework of proposed method. (a) High spatial resolution remote sensing image dataset; (b) source deep convolutional neural network; (c) unsupervised adversarial domain adaptation; (d) remote sensing image scene classification

为了充分利用源域知识, 首先利用深度卷积神经网络和源域数据集, 学习场景影像的深度特征表示, 具体的网络结构为 VGG 团队在 ILSVRC 2014 (ImageNet Large Scale Visual Recognition Competition) 年比赛中提出的 16 层网络结构。对于训练损失, 选择标准的监督损失方法, 则损失函数公式为

$$\min_{M_s, C} L_{cls}(X_s, Y_s) = -E_{(x_s, y_s) \sim (x_s, y_s)} \sum_{k=1}^K I_{k=y_s} \{\log C[M_s(x_s)]\}, \quad (6)$$

式中: $I\{\cdot\}$ 表示一个指示函数。采用 Softmax 分类器, $I\{\cdot\}$ 表示一个指示函数, $I\{\text{true}\} = 1$, $I\{\text{false}\} = 0$ 。

基于 2.1 节对生成对抗网络的推导分析, 生成器和判别器相互对抗学习能够对齐两个域分布, 为此将对抗学习方法引入高分辨率遥感影像跨域分类, 最小化源域场景影像和目标域场景影像的特征分布差异, 利用对抗学习方法提高模型对目标域数据集的泛化能力。在对抗学习中使用最大

最小损失作为损失函数训练网络,为了能够利用标准的最小损失函数训练生成器(目标域的深度卷积神经网络),所提方法将生成对抗网络的损失函数分为两个独立的损失函数:一个用于生成器,另一个用于判别器。将判别器的损失函数保持不变,生成器的损失利用反向标签进行修改,修改后的损失函数为

$$\begin{aligned} \min_D L_{\text{adv}_D}(X_s, X_t, M_s, M_t) = & \\ - E_{x_s \sim X_s} \{ \log \{ D[M_s(x_s)] \} \} - & \\ E_{x_t \sim X_t} \{ \log \{ 1 - D[M_t(x_t)] \} \}, & \quad (7) \end{aligned}$$

$$\begin{aligned} \min_{M_t} L_{\text{adv}_M}(X_s, X_t, D) = & \\ - E_{x_t \sim X_t} \{ \log \{ D[M_t(x_t)] \} \}. & \quad (8) \end{aligned}$$

修改后的损失函数可以直接使用反向传播和梯度下降方法优化网络。在域适应阶段固定源域特征,仅学习目标域的特征,模拟生成对抗网络的学习方式,将真实图像的分布固定,把生成器生成的分布匹配到真实图像的分布上。在对抗学习中,先固定目标域网络,优化判别网络,提高判别网络对数据来自源域和目标域的区分能力,然后固定判别网络,优化目标域网络,目标域网络和判别器二者通过相互对抗更新迭代,最小化源域和目标域的特征分布,直至达到一个纳什均衡。为了学习到更多的目标域特征,目标域不共享源域的参数,仅利用源域参数初始化目标域模型,可避免目标数据集数据量较少导致的过拟合问题。

为了优化损失函数,所提方法采用 Adam 优化算法^[28]和反向传播更新模型参数,Adam 优化算法是经典随机梯度下降法(SGD)的扩展,在训练过程中,SGD对所有的权重都保持一个单一的学习速率,而 Adam 算法将动量与 RMSProp(Root Mean Square Prop)优化算法相结合,利用梯度的一阶矩和二阶矩估计自适应选择每个训练参数的学习率,经过偏置校正后,每一次的迭代学习率都需确定范围,使得参数平稳,故可以为不同的参数计算出不同的自适应学习率。该算法对内存需求较小,具体表示为

$$w_t = w_{t-1} - a_t m_{d_w} / \sqrt{\hat{v}_{d_w}} + \epsilon, \quad (9)$$

$$b_t = b_{t-1} - a_t m_{d_b} / \sqrt{\hat{v}_{d_b}} + \epsilon, \quad (10)$$

式中: w_t 、 b_t 、 a_t 分别为第 t 次迭代时的权值、偏置和学习速率; m_{d_w} 和 m_{d_b} 分别为权值与偏置的偏差纠正动量值; $\sqrt{\hat{v}_{d_w}}$ 和 $\sqrt{\hat{v}_{d_b}}$ 分别为权值与偏置的偏差纠正

RMSProp 值; ϵ 为 Adma 算法中的超参数。

3 实验与分析

3.1 实验数据集介绍

3.1.1 源域数据集构建

由于缺少源域数据集,过去的研究只选择验证数据集中的少部分类别验证算法的有效性,具有一定的局限性。为了有效验证所提方法的性能,构建一个遥感影像场景分类数据集作为源域。与 ImageNet 相比,收集带标注的遥感影像数据更困难,然而,现阶段遥感影像数据集的构建已取得了一些成果,如 RSI-CB256(Remote Sensing Image Classification Benchmark)^[29],其利用众包的 OSM(Open Street Map)的 POI(Point of Interest)数据标注全球范围内的大规模遥感影像,根据中国土地利用现状分类要求与 ImageNet 的分层分级机制建立了严格的地物分类体系,该数据集包含 35 类地物,共 36000 幅影像块,每类约为 690 幅影像块,数据来源于 Google Earth 和 Bing Maps,空间分辨率为 0.22~3 m。与其他遥感数据集相比,RSI-CB256 具有高质量、多样性和大尺度等优点,但该数据集的场景类别覆盖不全,和 UC-Merced(University of California, Merced)数据集相比,仅包含其中的 13 类,通过对比现有的遥感影像场景分类数据集发现,NWPU-RESISC45(Northwestern Polytechnical University Remote Sensing Image Scene Classification)^[30]与 RSI-CB256 只有 14 个类别重复,30 个类别不重复,和 RSI-CB256 具有很强的互补性,并且该数据集分辨率从 0.2~30 m,覆盖区域包括全球 100 多个国家和地区。

本文将 NWPU-RESISC45 不重复的 30 个类别添加到 RSI-CB 框架中,其中 RSI-CB256 将不同密度的 3 种居民区合并为 1 种居民区,而大部分数据集都根据不同密度将居民区划分,因此,本文将 RSI-CB256 的居民区数据手动划分至不同密度的居民区。最后利用多传感器多分辨率(0.5~8 m)的 AID(Aerial Image data set)^[31]数据集进一步扩充该数据集,最终构建了多传感器、多分辨率、多地理区域和多类别的源域遥感影像场景分类数据集,该数据集包含 6 个大类,71 个子类,共 69087 张场景影像。具体类别信息如表 1 所示,粗体表示从 NWPU-RESISC45 和 AID 中补充至 RSI-CB256 数据集的类别。

表1 源域遥感影像场景分类数据集类别

Table 1 Dataset category for remote sensing image scene classification of source domain

Large class(six)	Subclass(seventy-one)
Construction land	city_building, container, storage_room, pipeline, town, baseball_diamond, basketball_court, golf_course, tennis_court, ground_track_field, church, commercial_area, industrial_area, mobile_home_park, palace, stadium, thermal_power_station, dense_residential, medium_residential, sparse_residential, Square, Center, Park, Resort, School, Playground
Ultivated land	green_farmland, dry_farm, bare_land, circular_farmland, rectangular_farmland, terrace
Transportation	airplane, airport_runway, avenue, highway, harbor, parkinglot, crossroads, bridge, airport, overpass, railway, railway_station, ship, roundabout
Water area	beach, dam, hirst, lakeshore, river, sea, stream, island, lake, sea_ice, wetland
Woodland	artificial_grassland, sparse_forest, forest, mangrove, sapling, river_protection_forest, shrubwood, chaparral, meadow
Other	desert, snow_mountain, mountain, sandbeach, cloud

3.1.2 目标域数据集

为验证所提方法的有效性,采用 UC-Merced^[8]和 WHU-RS^[32]两个国际通用的场景分类数据集进行实验。

1) UC-Merced: 该场景数据为 Yang 和 Newsam 从美国 USGS (United States Geological Survey) 的一个大型航空正射影像中采集的不同地区的影像,分辨率为 0.3 m,共包含 21 类场景,具体包含:农田、飞机、棒球场、海滩、建筑物、丛林、

森林、高速公路、高尔夫球场、港口、移动房车场、十字路口、高架桥、停车场、密集居住区、中等密集居住区、稀疏居住区、河流、机场跑道、储油区和网球场,影像示例如图 2 所示,影像大小为 256 pixel×256 pixel,每一类场景包含 100 张影像,共 2100 张影像。该数据集类间差距小,类内差距大,是一个非常具有挑战性的数据集,在遥感影像场景分类领域极具有代表性,广泛用于场景分类研究。



图2 UC-Merced数据集示例

Fig. 2 Sample images of UC-Merced dataset

2) WHU-RS: 该数据集为武汉大学测绘遥感信息工程国家重点实验室从 Google Earth 影像中采集的不同分辨率、不同地区的影像,共包含 19 类场景,具体包含:飞机场、海滩、桥梁、商业区、沙漠、牧场、足球场、森林、港口、工业区、草地、山脉、公园、停车场、池塘、火车站、居民区、河流、高架桥,影像示例如图 3 所示,每幅影像大小为 600 pixel×600 pixel,每个类别包含约 50 张影像,共 1005 张影像。该数据集中一些场景类别具有更复杂的空间分布、光照、尺度和方向变化,同一类别的场景样本差异性更大,不同类别的场景更易混淆,和 UC-Merced 数据集相

比更具有挑战性。

3.2 实验设置

本文实验中的源域和目标域的卷积神经网络结构如表 2 所示,判别网络结构如表 3 所示。对于模型的训练,UC-Merced 和 WHU-RS 两个数据集的模型训练参数设置完全一样,训练速率衰减周期设置为 50 epoch, Batch size 设置为 64。对于 Adam 优化算法,初始学习速率为 0.0001,速率衰减因子为 0.9, epsilon 设置为 10^{-8} , VGG16 模型输入图像大小为 224 pixel×224 pixel,因此将场景影像大小缩放到 224 pixel×224 pixel。为了定量分析所提

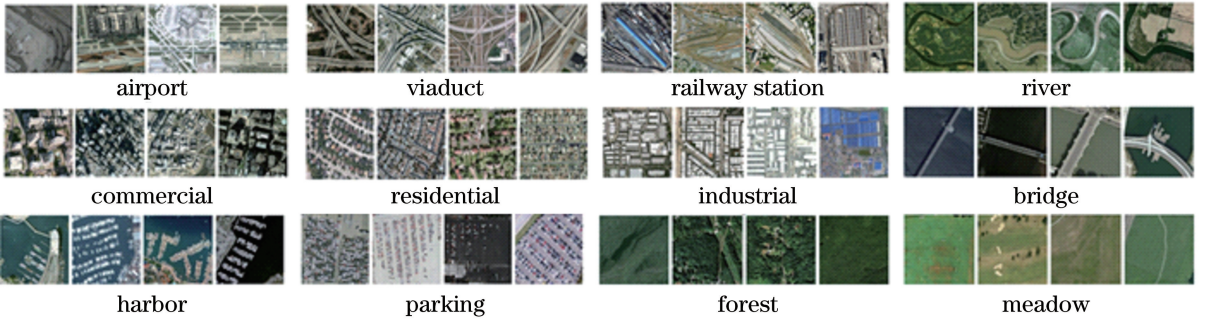


图3 WHU-RS数据集示例

Fig. 3 Sample images of WHU-RS dataset

表2 源域与目标域的深度卷积神经网络结构

Table 2 Deep convolution neural network structure of source domain and target domain

Layer name	Layer type	Output size / (pixel×pixel×pixel)
Source /target DCNN	Input	224×224×3
	Convolution×2	224×224×64
	Map pooling	112×112×64
	Convolution×2	112×112×128
	Map pooling	56×56×128
	Convolution×3	56×56×256
	Map pooling	28×28×256
	Convolution×3	28×28×512
	Map pooling	14×14×512
	Convolution×3	14×14×512
	Map pooling	7×7×512
	Fully connected	1×1×1024
Softmax	1×1×18/1×1×21	

表3 判别网络结构

Table 3 Structure of discriminator

Layer name	Layer type	Output size / (pixel×pixel×pixel)
Discriminator	Fully connected	1×1×1024
	Fully connected	1×1×512
	Fully connected	1×1×2
	Softmax	1×1×2

方法的有效性,利用混淆矩阵、总体精度(OA)和Kappa系数评价分类性能,研究域适应方法对目标域数据集的泛化能力。

3.3 实验结果与分析

1) UC-Merced数据集分类混淆矩阵、OA和Kappa系数如图4所示。其中图4(a)给出了源域数据集上学习到的深度特征直接对目标域数据集分类的结果,OA和Kappa系数分别为70.43%和0.690;图4(b)给出了采用域适应方法后目标域的分类结果,OA和Kappa系数分别为86.71%和

0.857。由于源域遥感数据集和目标域遥感数据集的差异较大,直接将源域数据集和深度卷积神经网络训练的模型迁移至目标数据集上时,所得结果精度较低,总体精度和Kappa系数分别为70.43%和0.690,而采用所提方法后精度明显提高,OA和Kappa系数分别提高了16.28%和0.167。对比两个混淆矩阵可以发现,域适应方法使大部分场景类别的精度提升,密集居住区和中等密集居住区以及建筑物容易产生混淆,移动停车场和建筑物以及密集居住区容易产生混淆,这些场景中都包含建筑物、植被和道路等基本组成单元,区别仅为这些基本单元的空间分布和密度不同,特征非常相似,因此容易产生混淆,但大部分场景类别分类精度较高,表明域适应方法能够有效最小化源域和目标域特征分布,提高目标域数据集的分类性能。

2) WHU-RS数据集分类混淆矩阵、OA和Kappa系数如图5所示。其中图5(a)所示为源域数据集学习到的深度特征直接对目标域数据集分类的结果,OA和Kappa系数分别为87.06%和0.863;图5(b)所示为采用域适应方法后目标域的分类结果,OA和Kappa系数分别为97.41%和0.973。采用所提方法后精度有明显提高,OA和Kappa系数分别提高了10.35%和0.110。对比两个混淆矩阵可以发现,域适应方法使大部分场景类别的精度提升,商业区和居民区混淆问题较严重,虽然WHU-RS数据集和UC-Merced数据集相比,一些场景类别具有更复杂的空间分布、光照、尺度和方向变化,同一类别场景样本差异性更大,但本文方法对该数据集的分类精度较高,说明所提方法能够有效克服源域和目标域由于光照、尺度和方向不同导致的特征分布不同的问题,能够有效最小化源域和目标域特征分布,提高目标域数据集的分类性能。UC-Merced数据集和WHU-RS数据集相比,分类精度较低的

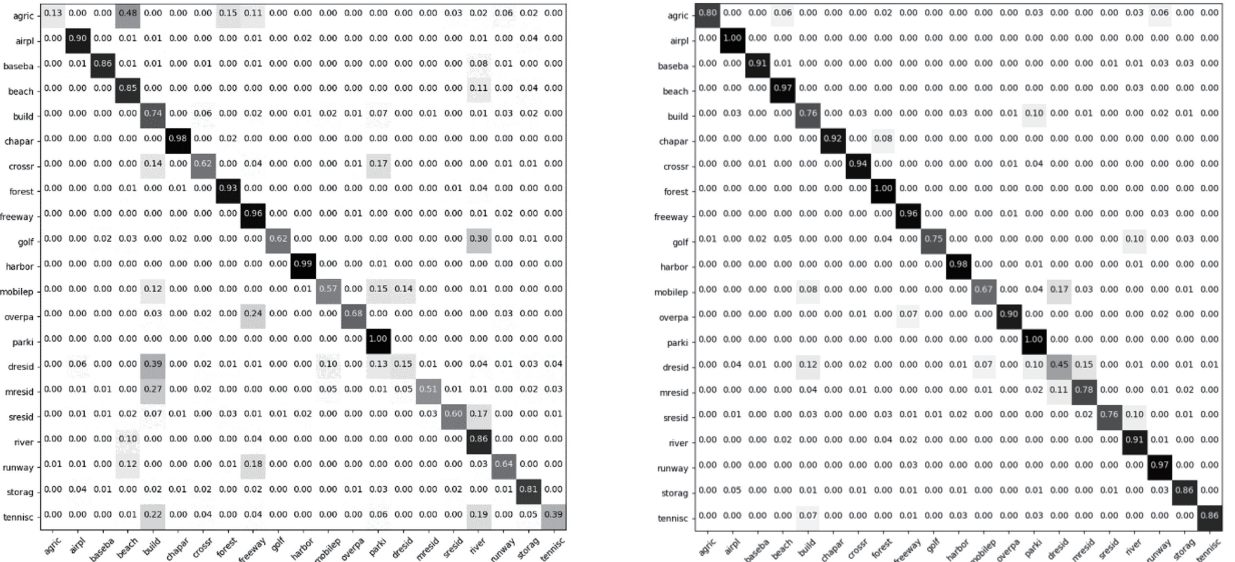


图 4 UC-Merced 数据集的分类混淆矩阵。(a)源域分类精度；(b)域适应分类精度

Fig. 4 Classification confusion matrix of UC-Merced dataset. (a) Classification accuracy of source domain; (b) classification accuracy of domain adaptation

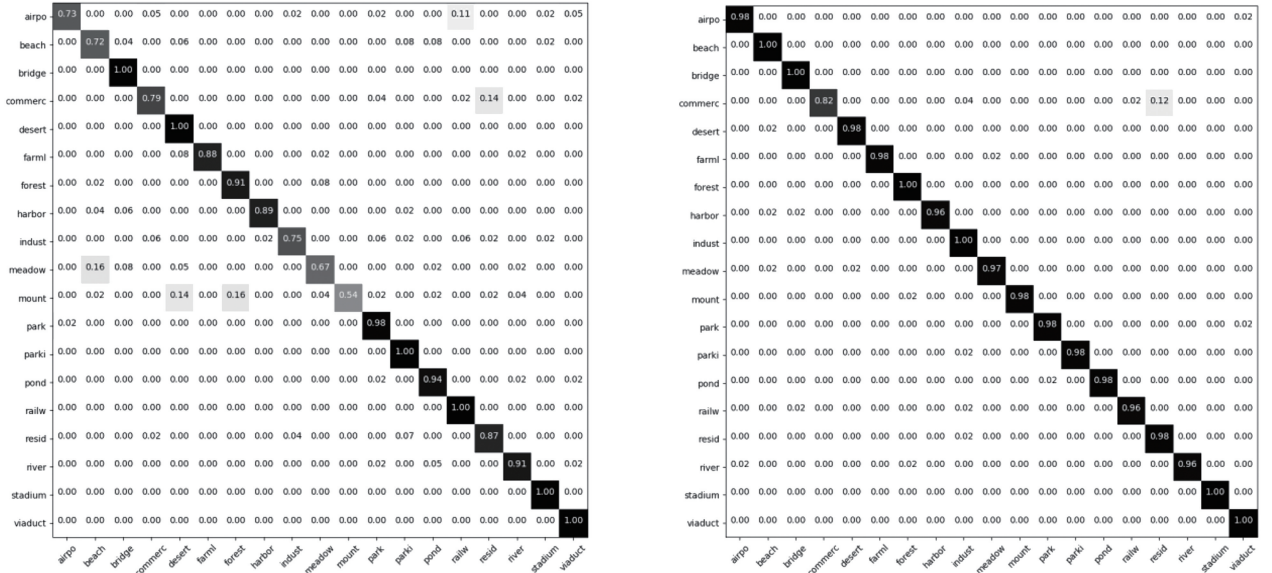


图 5 WHU-RS 数据集的分类混淆矩阵。(a)源域分类精度；(b)域适应分类精度

Fig. 5 Classification confusion matrix of WHU-RS dataset. (a) Classification accuracy of source only; (b) classification accuracy of domain adaptation

主要原因是该数据集中密集居住区、中等密集居住区和建筑物等特征相似程度很高。

综上所述,所提出的基于深度对抗域适应的高分辨率遥感影像跨域分类方法是一种有效的分类方法,在目标域无标签的情况下能够提高模型的泛化能力,有效解决了源域和目标域特征分布不同导致的目标域数据集分类精度较低的问题。

3.4 与其他方法对比结果

将本文方法与其他方法进行实验对比。对比的方法有:1) Source Only,利用源域数据集和深度卷积神经网络学习场景影像的深度特征,直接对目标域数据分类,不使用域适应方法;2) MMD^[22],使用 MMD 损失函数最小化源域和目标域的特征分布差异,使用最后一个全连接层输出计算 MMD 损失;

3) DANN^[23] (Domain-Adversarial Training of Neural Networks), 利用对抗学习方法最小化源域和目标域的特征分布差异, 在域分类器前加入一个梯度反转层, 使用的模型和参数与所提方法一致。各方法的总体精度如表 4 所示。

从表 4 可看出, 和 MMD 方法相比, 所提方法在 UC-Merced 和 WHU-RS 两个数据集上的总体精度分别提高了 4.52% 和 2.58%; 和 DANN 方法相比提高了 5.09% 和 3.58%。综上所述, 所提方法优于其他方法。

表 4 不同算法分类精度

Table 4 Classification accuracy of each algorithm %

Algorithm	Source only	MMD	DANN	Proposed
UC-Merced	70.43	82.19	81.62	86.71
WHU-RS	87.06	94.83	93.83	97.41

4 结 论

提出了一种基于深度对抗域适应的高分辨率遥感影像跨域分类方法。理论分析了对抗学习方法最小化源域与目标域的特征分布差异的原理; 利用深度卷积神经网络 VGG16 学习场景影像的深度特征; 利用对抗学习方法最小化源域和目标域特征分布差异, 利用反向传播和 Adam 算法优化最小损失函数, 提高了模型对目标域数据集的泛化能力。

为了验证所提方法有效性, 构建了一个多传感器、多分辨率、多地理区域、多类别的源域遥感影像场景分类数据集作为源域, 利用 UC-Merced 和 WHU-RS 两个通用数据集作为目标域数据集验证所提方法的有效性。实验结果表明, 采用所提方法, UC-Merced 数据集的 OA 和 Kappa 系数分别提高 16.28% 和 0.167; WHU-RS 数据集的 OA 和 Kappa 系数分别提高 10.35% 和 0.110; 所提方法能够有效减小源域和目标域的特征分布差异, 在目标域数据集没有标签的情况下, 能够提高模型对目标域数据集的泛化能力, 是一种有效的遥感影像跨域分类方法。

参 考 文 献

[1] Li D R, Tong Q X, Li R X, *et al.* Some frontier scientific issues in high-resolution Earth 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] Zhang L P, Huang X. Advanced processing

techniques for remotely sensed imagery[J]. *Journal of Remote Sensing*, 2009, 13(4): 559-569.

张良培, 黄昕. 遥感影像信息处理技术的研究进展[J]. *遥感学报*, 2009, 13(4): 559-569.

- [3] Zhang F. Deep learning for very high resolution remote sensing data analysis [D]. Wuhan: Wuhan University, 2017. 张帆. 面向高分辨率遥感影像分析的深度学习方法研究[D]. 武汉: 武汉大学, 2017.
- [4] Hu F. Feature learning based high resolution remote sensing image scene classification [D]. Wuhan: Wuhan University, 2017. 胡凡. 基于特征学习的高分辨率遥感图像场景分类研究[D]. 武汉: 武汉大学, 2017.
- [5] Xia G S, Yang W, Delon J, *et al.* Structural high-resolution satellite image indexing [C] // ISPRS TC VII Symposium-100 Years ISPRS, July 5-7, 2010, Vienna, Austria. [S. l. : s. n.], 2010: 298-303.
- [6] Risojević V, Momić S, Babić Z. Gabor descriptors for aerial image classification [M] // Dobnikar A, Lotrič U, Šter B. Adaptive and Natural Computing Algorithms. Berlin, Heidelberg: Springer, 2011, 6954: 51-60.
- [7] Yu H, Yang W, Xia G S, *et al.* A color-texture-structure descriptor for high-resolution satellite image classification[J]. *Remote Sensing*, 2016, 8(3): 259.
- [8] 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, November 2-5, 2010, San Jose, California. New York: ACM, 2010: 270-279.
- [9] Hu F, Xia G S, Wang Z F, *et al.* Unsupervised feature learning via spectral clustering of multidimensional patches for remotely sensed scene classification[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(5): 2015-2030.
- [10] Lienou M, Maitre H, Datcu M. Semantic annotation of satellite images using latent dirichlet allocation[J]. *IEEE Geoscience and Remote Sensing Letters*, 2010, 7(1): 28-32.
- [11] Zhong Y F, Zhu Q Q, Zhang L P. Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(11): 6207-6222.

- [12] Yan M, Zhao H D, Li Y H, *et al.* Multi-classification and recognition of hyperspectral remote sensing objects based on convolutional neural network [J]. *Laser & Optoelectronics Progress*, 2019, 56(2): 021702.
闫苗, 赵红东, 李宇海, 等. 基于卷积神经网络的高光谱遥感地物多分类识别[J]. *激光与光电子学进展*, 2019, 56(2): 021702.
- [13] Teng W X, Wen X R, Wang N, *et al.* Tree species classification and mapping based on deep transfer learning with unmanned aerial vehicle high resolution images[J]. *Laser & Optoelectronics Progress*, 2019, 56(7): 072801.
滕文秀, 温小荣, 王妮, 等. 基于深度迁移学习的无人机高分影像树种分类与制图[J]. *激光与光电子学进展*, 2019, 56(7): 072801.
- [14] Qian X L, Li J, Cheng G, *et al.* Evaluation of the effect of feature extraction strategy on the performance of high-resolution remote sensing image scene classification [J]. *Journal of Remote Sensing*, 2018, 22(5): 758-776.
钱晓亮, 李佳, 程焱, 等. 特征提取策略对高分辨率遥感图像场景分类性能影响的评估[J]. *遥感学报*, 2018, 22(5): 758-776.
- [15] Penatti O A B, Nogueira K, dos Santos J A. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 44-51.
- [16] Gong X, Wu L, Xie Z, *et al.* Classification method of high-resolution remote sensing scenes based on fusion of global and local deep features [J]. *Acta Optica Sinica*, 2019, 39(03): 0301002.
龚希, 吴亮, 谢忠, 等. 融合全局和局部深度特征的高分辨率遥感影像场景分类方法[J]. *光学学报*, 2019, 39(3): 0301002.
- [17] Li E Z. Research on remotely sensed image scene deep learning and application [D]. Nanjing: Nanjing University, 2017.
李二珠. 遥感图像场景深度学习与应用研究[D]. 南京: 南京大学, 2017.
- [18] Basu S, Ganguly S, Mukhopadhyay S, *et al.* DeepSat - a learning framework for satellite imagery [EB/OL]. (2015-09-11) [2018-12-01]. <https://arxiv.org/abs/1509.03602>.
- [19] Zhang F, Du B, Zhang L P. Scene classification via a gradient boosting random convolutional network framework [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(3): 1793-1802.
- [20] Xu S H, Mu X D, Zhang X M, *et al.* Unsupervised remote sensing domain adaptation method with adversarial network and auxiliary task [J]. *Acta Geodaetica et Cartographica Sinica*, 2017, 46(12): 1969-1977.
许夙晖, 慕晓冬, 张雄美, 等. 结合对抗网络与辅助任务的遥感影像无监督域适应方法[J]. *测绘学报*, 2017, 46(12): 1969-1977.
- [21] Gong J Y, Ji S P. Photogrammetry and deep learning [J]. *Acta Geodaetica et Cartographica Sinica*, 2018, 47(6): 693-704.
龚健雅, 季顺平. 摄影测量与深度学习[J]. *测绘学报*, 2018, 47(6): 693-704.
- [22] Long M S, Cao Y, Wang J M, *et al.* Learning transferable features with deep adaptation networks [C] // Proceedings of the 32nd International Conference on International Conference on Machine Learning, July 6-11, 2015, Lille, France. Massachusetts: JMLR. org, 2015: 97-105.
- [23] Ganin Y, Ustinova E, Ajakan H, *et al.* Domain-adversarial training of neural networks [J]. *Journal of Machine Learning Research*, 2017, 17(1): 2096-2030.
- [24] Tzeng E, Hoffman J, Saenko K, *et al.* Adversarial discriminative domain adaptation [EB/OL]. (2017-02-17) [2018-12-01]. <https://arxiv.org/abs/1702.05464>.
- [25] Othman E, Bazi Y, Melgani F, *et al.* Domain adaptation network for cross-scene classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(8): 4441-4456.
- [26] Ammour N, Bashmal L, Bazi Y, *et al.* Asymmetric adaptation of deep features for cross-domain classification in remote sensing imagery [J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, 15(4): 597-601.
- [27] Goodfellow I J, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets [C] // Proceedings of the 27th International Conference on Neural Information Processing Systems, December 8-13, 2014, Montreal, Canada. Cambridge: MIT Press, 2014: 2672-2680.
- [28] Kingma D P, Ba J. Adam: a method for stochastic optimization [EB/OL]. (2017-01-30) [2018-12-01]. <https://arxiv.org/abs/1412.6980>.
- [29] Li H, Tao C, Wu Z, *et al.* RSI-CB: a large scale remote sensing image classification benchmark via

- crowdsourced data [EB/OL]. (2017-06-11) [2018-12-01]. <https://arxiv.org/abs/1705.10450>.
- [30] Cheng G, Han J W, Lu X Q. Remote sensing image scene classification: benchmark and state of the art [J]. Proceedings of the IEEE, 2017, 105(10): 1865-1883.
- [31] Xia G S, Hu J W, Hu F, *et al.* AID: a benchmark data set for performance evaluation of aerial scene classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(7): 3965-3981.