

基于局部保持的遥感场景零样本分类算法

吴晨¹, 王宏伟², 王志强², 袁昱纬³, 刘宇², 程红², 全吉成^{2*}

¹海军航空大学, 山东 烟台 264000;

²空军航空大学, 吉林 长春 130022;

³91977 部队, 北京 102200

摘要 目标域遥感图像特征分布的变化, 导致遥感场景零样本分类性能下降, 针对该问题, 提出一种基于局部保持的遥感场景零样本分类算法。首先, 为减少冗余信息, 采用解析字典学习方法, 将源域中的场景图像特征和类别语义词向量嵌入到同一稀疏编码空间, 并实现两者稀疏系数的强制对齐, 以建立图像特征与词向量之间的关系; 然后, 通过保留图像特征空间中场景图像间的局部近邻关系, 增强场景图像对应稀疏系数的鉴别性, 以有助于对稀疏系数进行聚类分析; 最后, 为适应目标域图像特征分布变化, 采用 k -means 算法对目标域场景图像的稀疏系数进行聚类, 并以初始中心的类别标签作为对应的聚类簇中场景的类别标签。实验分别采用 GoogLeNet 和 VGGNet 图像特征, 以数据集 UCM 作为源域遥感场景集, 对目标域场景集 RSSCN7 进行零样本分类, 获得了最高 50.67% 和 53.29% 的总体分类准确度, 比现有算法各提升了 8.06% 和 9.70%。实验结果表明: 该算法能够适应目标域遥感场景图像特征分布的变化, 显著提升遥感场景零样本分类效果, 具有一定的优越性。

关键词 遥感; 零样本分类; k -means 算法; 解析字典学习; 图像特征

中图分类号 TP753

文献标识码 A

doi: 10.3788/AOS201939.0728001

Zero-Shot Classification for Remote Sensing Scenes Based on Locality Preservation

Wu Chen¹, Wang Hongwei², Wang Zhiqiang², Yuan Yuwei³, Liu Yu²,
Cheng Hong², Quan Jicheng^{2*}

¹University of Naval Aviation, Yantai, Shandong 264000, China;

²Aviation University of Air Force, Changchun, Jilin 130022, China;

³The 91977 of Peoples Liberation Army of China, Beijing 102200, China

Abstract Due to the change of image feature distribution in target domain, the performance of zero-shot classification for remote sensing scenes degrades. To solve this problem, a zero-shot classification algorithm for remote sensing scenes based on locality preservation is proposed. Firstly, in order to reduce redundant information, the analysis dictionary learning method was exploited to embed the image features and word vectors of the source domain into the common sparse coefficient space, and the sparse coefficients were compulsively aligned for establishing the relationship between the image features and word vectors. Then, the discriminability of sparse coefficients of scene images was enhanced by preserving the local neighborhood relationship among scene images, which is helpful for clustering analysis on the sparse coefficients. Finally, in order to adapt to the change of image feature distribution, the k -means algorithm was utilized to cluster the sparse coefficients of scene images, and the class labels of the initial centers were used as the scene class labels. With the UCM remote sensing scene dataset as the source domain, zero-shot classification experiments were carried out on RSSCN7 scene dataset of the target domain via two type image features, i. e., GoogLeNet and VGGNet. The highest overall accuracies of 50.67% and 53.29% are obtained, which outperform the state-of-the-art algorithms by 8.06% and 9.70%, respectively. The experimental results show that this method can adapt to the feature distribution of remote sensing scenes, and significantly improve the zero-shot classification performance with certain advantages.

Key words remote sensing; zero-shot classification; k -means algorithm; analysis dictionary learning; image features

收稿日期: 2019-01-21; 修回日期: 2019-03-03; 录用日期: 2019-03-21

基金项目: 国家自然科学基金青年基金(61301233)

* E-mail: jicheng_quan@126.com

1 引 言

近年来,随着遥感图像空间分辨率不断提升,传统的基于对象的图像分类方法难以满足遥感图像分析需要,因此人们提出了基于场景的遥感图像分类方法^[1-2]。所谓“场景”是指遥感图像中具有主题地物含义的局部图像块,目前其已成为场景分类的基本单元^[3]。然而现有遥感图像场景分类需要标注大量的场景训练样本,而且学习得到的场景识别模型属于监督分类模型,难以扩展识别其他场景类别。针对该问题,Li等^[4]提出一种遥感场景零样本分类算法,该算法通过将零样本分类与遥感场景结合,实现对场景样本标注成本的降低以及对新场景类别的扩展。

作为一种特殊的无监督分类方法,零样本分类(ZSC)算法使用类别名称的语义词向量提供类别之间的距离关系,通过迁移源域中的类别标注样本中的知识,推断识别目标域中类别的样本,且源域与目标域没有交叉重叠的类别。ZSC算法具有降低样本标注成本和灵活扩展识别能力的优势,因此近年来受到广泛关注^[5]。依据学习过程中是否引入未标注的目标域类别样本,ZSC算法分为直推式(T型)和推导式(I型)两种类型。由于没有未标注的目标域类别样本参与学习,I型ZSC算法更符合ZSC算法的理念。Xian等^[5]在细粒度ZSC研究中,Xian等^[5]通过引入隐式变量模型到兼容函数学习过程中,提出了隐式嵌入(LatEm)算法。Wang等^[6]针对映射函数泛化能力不足的问题,通过利用语义映射还原目标域中类别的流形结构,提出了关系知识迁移(RKT)算法。Zhang等^[7]提出联合隐式相似性嵌入(JLSE)算法。Zhang等^[8]提出语义相似性嵌入(SSE)算法。Wang等^[9]提出的双向隐式嵌入(BiDiLEL)算法,将图像特征和语义特征分别映射到公共空间,并在该公共空间中完成零样本分类。Li等^[10]提出的双视觉语义映射(DMaP)算法,利用了语义空间流形和视觉语义映射迁移能力之间的关系。ZSC算法的基本原理是将源域中监督学习的分类模型,迁移到目标域中新类别的样本上。这导致I型ZSC算法存在一定的域偏移问题,降低了分类效果。为减轻域偏移问题,学习时引入未标注的目标域类别样本,提出T型ZSC算法。Kodirov等^[11]提出利用目标域词向量正则化映射函数的无监督域适配算法(UDA)。Guo等^[12]提出基于共享

模型空间学习的T型ZSC算法(SMS)。Fu等^[13]提出基于多视角嵌入的T型ZSC算法(TME)。此外,冀中等^[14-15]对ZSC算法也进行了相关研究。

不过,现有ZSC算法都假定源域和目标域中的遥感场景图像的来源相同,即光照、成像时间等属性未发生变化。但在实际应用中这些条件通常很难满足,这导致由图像特征提取器获得的图像特征的分布发生显著变化,降低遥感场景ZSC效果。针对该问题,本文采用解析字典学习方法分别提取源域图像特征和语义词向量的稀疏系数,并强制使两者的稀疏系数一致,从而实现源域的图像特征与语义词向量对齐。同时,为增强稀疏系数的鉴别性,进而提高聚类效果,将图像特征空间中的局部近邻关系,保持到稀疏编码空间。最后,对目标域场景的稀疏系数进行 k -means聚类,并以初始中心的类别标签,作为对应的聚类簇的场景类别标签。

2 解析字典学习

稀疏化表示是减少冗余信息的一种重要方法^[16]。字典学习(DL)是获取输入样本稀疏系数的典型方法。字典学习方法分为两类,即合成字典学习(SDL)和解析字典学习(ADL)。虽然SDL方法应用广泛,但计算复杂度较高。而ADL通常具有闭式解,编码能力良好,计算效率较高,近年来受到广泛关注^[17-18]。ADL的基本公式为

$$\arg \min_{\mathbf{Z}, \mathbf{\Omega}} \frac{1}{2} \|\mathbf{Z} - \mathbf{\Omega}\mathbf{X}\|_{\text{F}}^2, \text{ s.t. } \mathbf{\Omega} \in \mathbf{\Gamma},$$

$$\|z_i\|_0 \leq T_0, \quad (1)$$

式中, $\mathbf{X}=[x_1, \dots, x_n] \in \mathbb{R}^{m \times n}$ 为 n 个输入样本的特征(m 维)矩阵, $x_i \in \mathbb{R}^m$ 为第 i 个样本特征, $z_i \in \mathbb{R}^m$ 为 x_i 的稀疏系数,其稀疏性采用 L_0 范数及参数 T_0 实现, $\mathbf{Z}=[z_1, \dots, z_n] \in \mathbb{R}^{m \times n}$ 为 \mathbf{X} 的稀疏系数矩阵; $\mathbf{\Omega}$ 为解析字典, $\mathbf{\Gamma}$ 是为避免出现平凡解,对 $\mathbf{\Omega}$ 限制的log-det条件^[19]。ADL的任务是给定输入特征矩阵 \mathbf{X} ,来寻找最佳稀疏系数矩阵 \mathbf{Z} 及解析字典 $\mathbf{\Omega}$ 。

3 具体算法

算法的具体流程如图1所示。

3.1 源域遥感场景的局部保持

现有研究证明,稀疏表示可以减少冗余信息,提升分类效果^[20]。将图像特征和语义词向量分别嵌

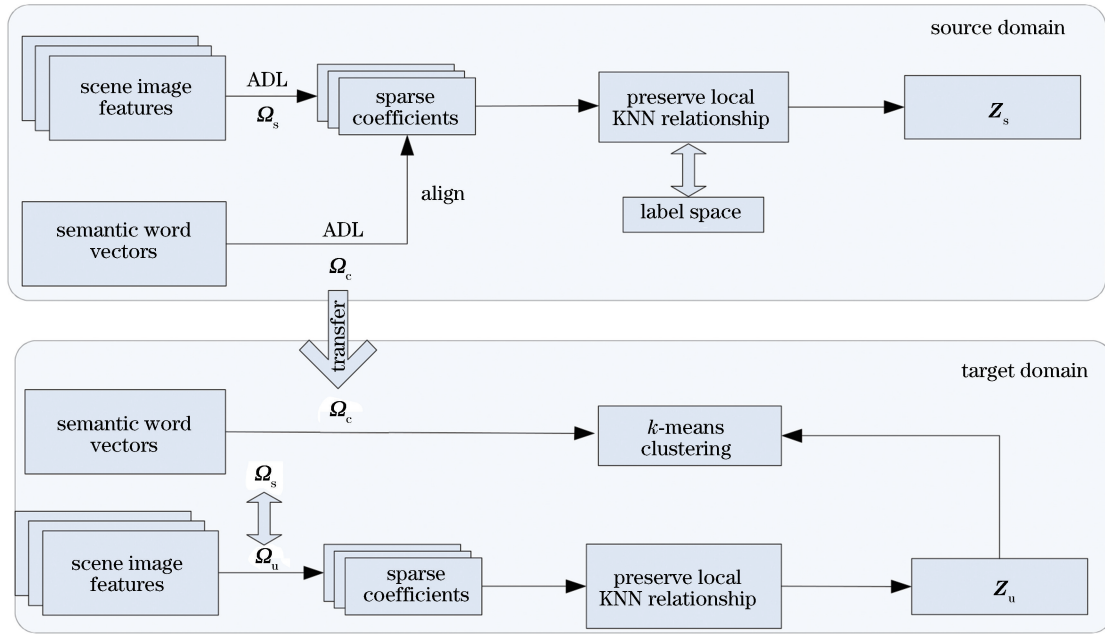


图 1 所提算法的整体框架图

Fig. 1 Whole framework of the proposed algorithm

入到中间的公共空间^[9],是零样本分类研究中常用的方法。本文采用稀疏编码空间作为中间的公共空间,运用稀疏编码效率较高的 ADL 方法,提取图像特征和语义词向量的稀疏编码系数。虽然场景图像特征空间与语义词向量空间的物理意义及表现方式迥异,但由于场景图像特征与类别语义词向量表达的是相同的场景对象,因此嵌入同一稀疏编码空间后,两者的稀疏系数应保持一致,即实现场景图像特征与类别语义词向量的互相对齐。本文通过 ADL 方法将源域场景的图像特征与类别语义词向量嵌入到同一稀疏编码空间,并强制两者的稀疏系数相同,实现图像特征与类别词向量的对齐,从而建立两者之间的联系。

虽然图像特征与语义词向量在稀疏编码空间中进行了对齐,但这种对齐是无监督式的,因此稀疏系数之间的鉴别性不足,即相同类别场景的稀疏系数没有互相靠近,不同类别场景的稀疏系数没有互相远离,使得分类效果下降。已有研究证明,增强鉴别性可以显著提升分类效果^[21]。

近年来,基于流形理论的局部保持算法,在降维、分类等任务中得到广泛关注^[22]。流形理论认为:高维图像特征空间中隐含着反映图像之间局部近邻关系的低维度空间(即流形)。在流形上,不同类别的图像样本分别聚集在不同的位置,彼此之间互相远离,具有良好的鉴别性。通过将图像特征空间中场景样本间的局部近邻关系,传递到稀疏编码

空间,可增强稀疏编码系数间的鉴别性,从而提升分类效果。 K 最近邻(KNN)是一种常用的局部近邻表示,记录了每个样本与其 K 个近邻样本间的距离关系。通过将每个场景图像的 KNN 局部关系保持到稀疏编码空间,可使在图像特征空间流形中靠近或远离的样本,在稀疏编码空间依然靠近或远离,从而增强稀疏编码系数之间的鉴别性。源域中的场景图像具有类别标签,因此采用监督式局部保持。源域遥感场景的局部保持的目标损失函数 ζ_s 可表示为

$$\zeta_s = \min_{\mathbf{Z}_s, \mathbf{\Omega}_s, \mathbf{\Omega}_c} \|\mathbf{Z}_s - \mathbf{\Omega}_s \mathbf{X}_s\|_F^2 + \|\mathbf{Z}_s - \mathbf{\Omega}_c \mathbf{C}_s^{\otimes}\|_F^2 + \lambda \text{tr}[\mathbf{Z}_s \mathbf{L}_s (\mathbf{Z}_s)^T],$$

$$\text{s.t. } \mathbf{\Omega}_c, \mathbf{\Omega}_s \in \mathbf{\Gamma}, \|\mathbf{z}_i^s\|_0 \leq T_0, \forall i, \quad (2)$$

式中:第一项和第二项为稀疏编码项,分别将场景图像特征、语义词向量转换到稀疏编码空间,且强制两者的稀疏系数相同;第三项为局部保持项,将图像特征空间中的 KNN 局部近邻关系保持到稀疏编码空间,以增强稀疏系数的鉴别性; $\lambda \geq 0$ 为局部保持项的重要性系数; $\mathbf{X}_s \in \mathbb{R}^{d_x \times N_s}$ 为源域遥感场景的图像特征矩阵, d_x 为图像特征维数, N_s 为源域中遥感场景的个数; $\mathbf{Z}_s \in \mathbb{R}^{d_z \times N_s}$ 为 \mathbf{X}_s 对应的稀疏编码系数矩阵, \mathbf{z}_i^s 为 \mathbf{Z}_s 的第 i 列向量, d_z 为稀疏编码系数的维数; $\mathbf{C}_s^{\otimes} \in \mathbb{R}^{d_c \times N_s}$ 为源域场景对应的类别词向量组成的矩阵, d_c 为语义词向量的维数; $\mathbf{\Omega}_s \in \mathbb{R}^{d_z \times d_x}$ 为源域场景图像特征的解析字典; $\mathbf{\Omega}_c \in \mathbb{R}^{d_z \times d_c}$ 为语义词向量的解析字典; $\mathbf{L}_s = \mathbf{D}_s - \mathbf{W}_s$ 为拉普拉斯矩阵, \mathbf{W}_s 是反映源域场景之间局部近邻关系的仿射矩

阵,其元素 $W_s(i, j)$ 表示图像特征 x_i^s 和 x_j^s 之间的

KNN 近邻关系,即

$$W_s(i, j) = \begin{cases} \exp(-\|x_i^s - x_j^s\|^2/2), & x_i^s \in N_K(x_j^s) \text{ or } x_j^s \in N_K(x_i^s) \\ & \text{and label}(x_i^s) = \text{label}(x_j^s) \quad ; \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

D_s 为对角矩阵且其对角元素为 W_s 中对应的元素之和,即 $D_s(i, j) = \sum_j W_s(i, j)$, $N_K(x_j^s)$ 为 x_j^s 的 K 个近邻, $\text{label}(x_j^s)$ 为 x_j^s 的类别标签。因此可知, (2) 式中第三项局部保持项的物理意义为 Z_s 的局部近邻关系与 X_s 的局部近邻关系 (W_s) 之间的差异性损失。通过降低这种损失,使 Z_s 与 X_s 的局部近邻关系趋于一致,从而在稀疏编码空间中实现源域场景的局部保持。(2) 式对变量 Z_s 、 Ω_s 和 Ω_c 同时非凸,可以通过固定其他变量方式,逐个迭代求解。由于 Z_s 一般可初始化为样本标签 one-hot 向量,而 Ω_s 和 Ω_c 的初始化较难,因此迭代过程需先更新 Ω_s 和 Ω_c ,最后更新 Z_s ,具体过程如下:

1) 固定 Z_s 、 Ω_c 不变,更新 Ω_s ,此时的目标函数为

$$\arg \min_{\Omega_s} \|Z_s - \Omega_s X_s\|_F^2, \text{ s.t. } \Omega_s \in \Gamma. \quad (4)$$

为避免出现平凡解,引入对解析字典 Ω_s 的 $\log\text{-det}$ 条件 $R(\Omega_s) = \|\Omega_s\|_F^2 - \lg|\det[(\Omega_s)^T \Omega_s]|$,其中 $\lg|\det[(\Omega_s)^T \Omega_s]|$ 用于抑制出现无效解,避免 Ω_s 中的行重复或全 0。因此,更新 Ω_s 的目标函数为

$$\arg \min_{\Omega_s} \|Z_s - \Omega_s X_s\|_F^2 + R(\Omega_s). \quad (5)$$

(5) 式的求解步骤^[19]: 首先,将 $X_s (X_s)^T + I$ 矩阵 Cholesky 分解为 LL^T ; 其次,对 $L^{-1} X_s Z_s$ 进行奇异值分解,分解为 $Q\Sigma R^T$; 最后,得到 $\Omega_s = 0.5R[\Sigma + (\Sigma^2 + 2I)^{1/2}]Q^T L^{-1}$ 。

2) 固定 Ω_s 、 Z_s 不变,更新 Ω_c ,此时的目标函数为

$$\arg \min_{\Omega_c} \|Z_s - \Omega_c C_\otimes^s\|_F^2, \text{ s.t. } \Omega_c \in \Gamma. \quad (6)$$

同理, Ω_c 的更新与 Ω_s 相同。

3) 固定 Ω_s 、 Ω_c 不变,更新 Z_s ,此时的目标函数为

$$\arg \min_{Z_s} \|Z_s - \Omega_s X_s\|_F^2 + \|Z_s - \Omega_c C_\otimes^s\|_F^2 + \lambda \text{tr}[Z_s L_s (Z_s)^T], \text{ s.t. } \|z_i^s\|_0 \leq T_0, \forall i. \quad (7)$$

通过对 Z_s 求导并置 0, 得到 $Z_s = (\Omega_s X_s + \Omega_c C_\otimes^s)(2I + \lambda L_s)^{-1}$, 并对其进行稀疏化处理。

迭代循环步骤 1)~3), 直至循环结束或总的损失函数值下降幅度小于门限值。对于门限值的设置,除了考虑损失下降的幅度值外,还应考虑其占总体损失的比例。由于本文实验过程中,算法的总体损失可快速下降至收敛状态,因此采用固定迭代次数的方式,直至迭代循环结束。

3.2 目标域遥感场景的局部保持

对目标域遥感场景进行局部保持时,需要利用解析字典将图像特征嵌入到稀疏编码空间,提取稀疏编码系数。由于目标域与源域的图像特征分布不同,因此由源域遥感场景学习到的解析字典 Ω_s , 不能直接作为目标域图像特征的解析字典 Ω_u 。两个解析字典虽不相同,但存在一定的相关性,通过引入正则项 $\|\Omega_s - \Omega_u\|_F^2$ 控制偏移量,实现 Ω_s 到 Ω_u 的迁移。由于目标域中的场景图像没有对应的类别标签,因此采用无监督式 KNN 局部关系保持。目标域遥感场景局部保持的目标损失函数 ζ_u 为

$$\zeta_u = \min_{Z_u, \Omega_u} \|Z_u - \Omega_u X_u\|_F^2 + \|\Omega_s - \Omega_u\|_F^2 + \mu \text{tr}[Z_u L_u (Z_u)^T], \quad \|z_i^u\|_0 \leq T_0, \forall i, \quad (8)$$

式中: 第一项为稀疏编码项,将目标域场景的图像特征嵌入到稀疏编码空间,获取稀疏系数; 第二项为解析字典 Ω_u 的正则项,便于稳定求解 Ω_u ; 第三项为局部保持项, $\mu \geq 0$ 为其重要性系数; $X_u \in \mathbb{R}^{d_x \times N_u}$ 为目标域遥感场景的图像特征矩阵, N_u 为目标域遥感场景个数; $Z_u \in \mathbb{R}^{d_z \times N_u}$ 为 X_u 对应的稀疏编码系数矩阵; $\Omega_u \in \mathbb{R}^{d_z \times d_x}$ 为目标域场景图像特征的解析字典; $L_u = D_u - W_u$ 为目标域拉普拉斯矩阵; W_u 是反映目标域场景之间局部近邻关系的仿射矩阵,其元素 $W_u(i, j)$ 表示图像特征 x_i^u 和 x_j^u 之间的无监督式 KNN 近邻关系,即

$$W_u(i, j) = \begin{cases} \exp(-\|x_i^u - x_j^u\|^2/2), & x_i^u \in N_K(x_j^u) \text{ or } x_j^u \in N_K(x_i^u) \\ 0, & \text{otherwise} \end{cases}; \quad (9)$$

D_u 为对角矩阵且其对角元素为 W_u 中对应的元素之和,即 $D_u(i, i) = \sum_j W_u(i, j)$ 。与源域局部保持相似,(8)式第三项局部保持项的物理意义为 Z_u 的局部近邻关系与 X_u 的局部近邻关系(W_u)之间的差异性损失。通过降低这种损失,使 Z_u 与 X_u 的局部近邻关系趋于一致,从而在稀疏编码空间中实现目标域场景的局部保持。(8)式对变量 Z_u 和 Ω_u 同时非凸,难以直接求解,同样通过固定其他变量方式逐个求解,此时 Ω_u 初始化为 Ω_s ,因此可以先求解 Z_u ,具体过程如下:

1) 初始化 Ω_u 为 Ω_s ,

2) 固定 Ω_u 不变,更新 Z_u ,此时的目标函数为

$$\arg \min_{Z_u} \|Z_u - \Omega_u X_u\|_F^2 + \mu \text{tr}[Z_u L_u (Z_u)^T],$$

$$\text{s.t. } \|z_i^u\|_0 \leq T_0, \forall i, \quad (10)$$

通过对 Z_u 求导并置 0,得到 $Z_u = \Omega_u X_u (I + \mu L_u)^{-1}$ 并对其进行稀疏化处理。

3) 固定 Z_u 不变,更新 Ω_u ,此时目标函数为

$$\arg \min_{\Omega_u} \|Z_u - \Omega_u X_u\|_F^2 + \|\Omega_s - \Omega_u\|_F^2, \quad (11)$$

同样对 Ω_u 求导并置 0,得到 $\Omega_u = [Z_u (X_u)^T + \Omega_s] [X_u (X_u)^T + I]^{-1}$ 。

迭代循环步骤 2)~3),直至循环结束或损失函数值下降幅度小于门限值。

3.3 目标域遥感场景的 k -means 聚类

由于成像时间、光照等条件不同,目标域的遥感图像的色调、纹理等属性与源域的遥感图像相比发生明显变化,导致通过图像特征提取器获得的图像特征的分布也发生明显变化。虽然 T 型 ZSC 方法在学习过程中,引入目标域图像特征,在一定程度上减轻因源域与目标域的场景类别不同而导致的域偏移问题,但这些 ZSC 方法均假定目标域与源域的遥感图像的来源相同,图像特征分布未发生变化,因此当目标域遥感图像特征分布变化时,现有 ZSC 方法不能适应这种变化,从而导致分类效果下降。

目标域遥感图像特征的分布虽然发生变化,但是流形分布特点依然存在,即不同类别场景仍聚集在流形上的不同位置。利用该特点,对增强鉴别性后的目标域稀疏系数进行无监督聚类,可以发现场景样本间内在的类别结构关系,从而减轻图像特征分布变化的影响。 k -means 因其简单、高效的特点,成为目前使用最广泛的聚类算法。 k -means 旨在将样本集划分为 k 个聚类簇,使聚类簇中的样本距离自己聚类簇中心的平均距离最小。因此,为适应目

标域图像特征分布变化,本文以稀疏化后的 $\Omega_c C_u$ 为初始中心($C_u \in \mathbb{R}^{d_c \times n_u}$ 为目标域场景类别的类别词向量组成的矩阵, n_u 为目标域场景类别个数),对 Z_u 进行 k -means 聚类,最后以 $\Omega_c C_u$ 的类别标签,作为各对应聚类簇的类别标签,从而实现对目标域场景样本的分类。

3.4 算法流程步骤

基于局部保持的遥感场景零样本分类算法的具体步骤如下。

输入:源域遥感场景的图像特征矩阵 $X_s \in \mathbb{R}^{d_x \times N_s}$,源域场景对应的类别词向量组成的矩阵 $C_s \in \mathbb{R}^{d_c \times N_s}$,目标域遥感场景的图像特征矩阵 $X_u \in \mathbb{R}^{d_x \times N_u}$,目标域遥感场景类别的类别词向量组成的矩阵 $C_u \in \mathbb{R}^{d_c \times n_u}$,最大迭代次数为 Iter_N。

输出:对 X_u 中样本推断类别标签。

步骤 1:初始化 Z_s 为 one-hot 向量矩阵,根据(3)式计算 W_s 、 D_s 以及 $L_s = D_s - W_s$,根据(9)式计算 W_u 、 D_u 以及 $L_u = D_u - W_u$ 。

步骤 2:固定 Z_s 、 Ω_c 不变,根据(5)式更新 Ω_s 。

步骤 3:固定 Ω_s 、 Z_s 不变,根据(6)式更新 Ω_c 。

步骤 4:固定 Ω_s 、 Ω_c 不变,根据(7)式更新 Z_s 。

步骤 5:固定 Ω_u 不变,根据(10)式更新 Z_u 。

步骤 6:固定 Z_u 不变,根据(11)式更新 Ω_u 。

步骤 7:以稀疏化后的 $\Omega_c C_u$ 为初始中心,对 Z_u 进行 k -means 聚类,最后以 $\Omega_c C_u$ 的类别标签,作为 Z_u 各对应聚类簇的类别标签,从而实现对目标域场景样本的分类。

步骤 8:判断是否达到最大循环次数 Iter_N。若是,则终止执行;若否,则循环执行步骤 2~8。

4 实验及结果分析

4.1 数据集及实验设置

实验采用两种公开的遥感场景集,UC-Merced (UCM)数据集^[23]和 RSSCN7 数据集^[24]。UCM 是使用最广泛的遥感场景集,有 21 类场景,每类 100 张图像,共 2100 张遥感场景图像,图像尺寸为 256 pixel×256 pixel,空间分辨率为 0.3 m,其样本如图 2 所示。RSSCN7 共 2800 张遥感场景图像,分为 7 个类别(1 草地、2 河湖、3 工厂、4 场地、5 森林、6 居民区和 7 停车场),每类 400 张图像,图像尺寸为 400 pixel×400 pixel,均来自谷歌地球,其样本如图 3 所示。UCM 与 RSSCN7 来源不同,色调区别明显,满足本文对源域和目标域遥感场景图像的要求。

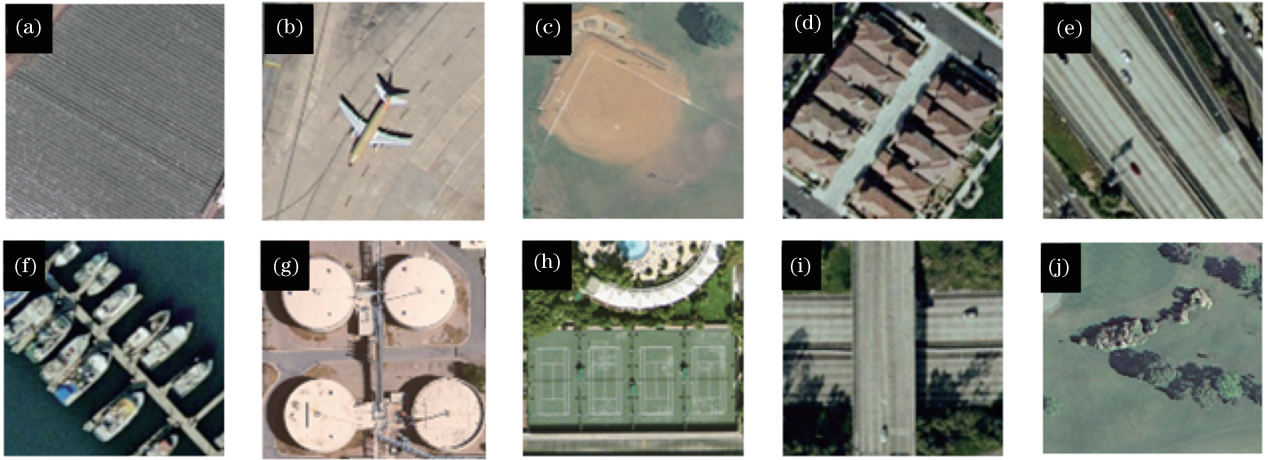


图 2 UCM 数据集若干类的样本。(a)农田;(b)飞机;(c)棒球场;(d)密集住宅;(e)高速公路;(f)海港;(g)储罐;
(h)网球场;(i)立交桥;(j)高尔夫球场

Fig. 2 Images of several UCM classes. (a) Farmland; (b) airplane; (c) baseball diamond; (d) dense residential;
(e) freeway; (f) harbor; (g) storage tanks; (h) tennis court; (i) overpass; (j) golf course

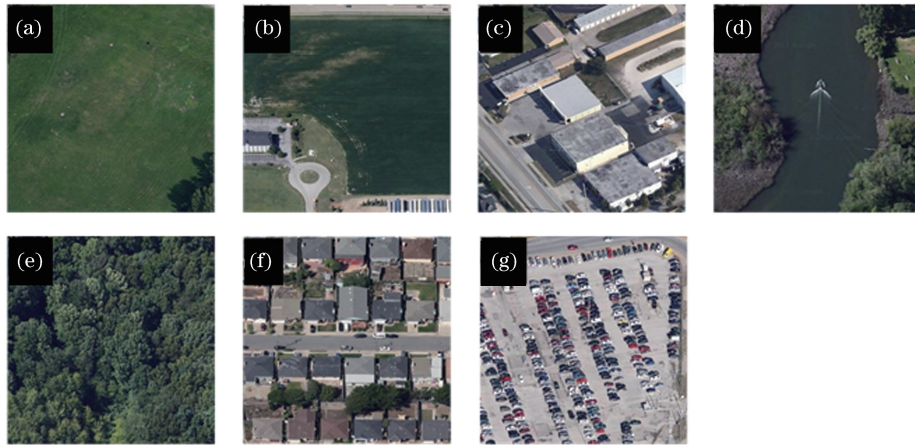


图 3 RSSCN7 数据集各类样本。(a)草地;(b)河湖;(c)工厂;(d)场地;(e)森林;(f)居民区;(g)停车场

Fig. 3 Images of RSSCN7 classes. (a) Grass; (b) river lake; (c) industry; (d) field; (e) forest;
(f) residential; (g) parking

由于深度特征具有更丰富的类别语义信息,实验采用 ImageNet 大规模数据集上预训练的深度卷积神经网络模型 GoogLeNet^[25] 和 VGGNet^[26], 提取全连接层输出(1024 维、4096 维)作为场景的图像特征。采用在维基百科语料上训练的 Word2Vec 词向量作为场景类别的语义词向量,维数为 300。最大迭代次数 Iter_N 设置为 30, KNN 局部近邻关系中设置 $K=5$, 稀疏系数中稀疏的比例设置为 10%。实验以 UCM 作为源域遥感场景集, 对目标域遥感场景集 RSSCN7 进行零样本分类, 并采用总体分类准确度(OA)作为评价指标, OA 值为分类正确的 RSSCN7 场景图像个数占 RSSCN7 场景总数的比例。由于算法初始化及求解过程均未有随机变量引入, 因此得到的 OA 值不需进行结果平均。超参数

λ, μ 的取值范围均为 $\{0, 1\}$, 当 $\lambda=0$ 时, 源域未进行局部保持; 当 $\lambda=1$ 时, 源域进行了局部保持; 当 $\mu=0$ 时, 本文方法属于 I 型 ZSC 方法; 当 $\mu=1$ 时, 本文方法属于 T 型 ZSC 方法。实验采用 64 位 Windows 7 操作系统, 内存 32 GB, 计算平台使用 MATLAB R2016a, CPU 为 Intel (R) Core (TM) i7-6700 3.40 GHz。

4.2 鉴别性分析

本文方法通过采用局部保持, 增强目标域场景的稀疏系数的鉴别性, 提升基于 k -means 聚类的 ZSC 效果。本节旨在分析局部保持项能否增强目标域场景的稀疏系数的鉴别性。图 4 和图 5 分别为 GoogLeNet 和 VGGNet 特征下, \mathbf{X}_u 及不同情形 \mathbf{Z}_u 的 t-SNE 图。可以看出, 目标域遥感场景的图像特

征矩阵 \mathbf{X}_u 的不同类别的场景样本没有互相远离, 相同类别的场景样本没有互相聚集, 即其鉴别性明显不足。这主要因为训练 GoogLeNet 和 VGGNet 深度模型的 ImageNet 图像集不是遥感场景图像, 没有适应遥感图像本身特点, 因而提取的遥感场景 GoogLeNet 和 VGGNet 特征不具有良好的鉴别性。 $\mathbf{Z}_u + \mathbf{I}$ 表示通过本文 I 型方法 ($\mu = 0$) 得到的 \mathbf{X}_u 的稀疏系数 \mathbf{Z}_u 。图 4 和 5 中的 (b)、(c) 为本文 I 型算法计算得到的 \mathbf{Z}_u 的 t-SNE 图, (b) 和 (c) 的主要区别在于: (b) 中的 \mathbf{Z}_u 未进行源域的局部保持 ($\lambda = 0$), 而 (c) 中的 \mathbf{Z}_u 进行了源域的局部保持 ($\lambda = 1$)。图 4 和 5 中的 (b)、(c) 场景图像样本分布的疏密情况与 (a) 基本一致, 不同类别的场景图像没有互相远离, 而且仍有一定的重叠, 此外, 相同类别的样本也没有互相靠近聚集。上述结果说明: I 型的两种情形 ($\lambda = 0, \lambda = 1$) 下 \mathbf{Z}_u 的鉴别性依然不足。 $\mathbf{Z}_u + \mathbf{T}$ 表示通过 T 型方法 ($\mu = 1$) 得到的 \mathbf{X}_u 的稀疏系数 \mathbf{Z}_u 。图 4 和 5 中的 (d)、(e) 为本文 T 型算法计算得到的 \mathbf{Z}_u 的 t-SNE 图, (d) 和 (e) 的主要区别在于: (d) 中的 \mathbf{Z}_u 未进行源域的局部保持 ($\lambda = 0$), 而 (e) 中的 \mathbf{Z}_u 进行了

源域的局部保持 ($\lambda = 1$)。从图 4 和 5 中的 (d)、(e) 可知 \mathbf{Z}_u 场景图像样本的分布情况, 即不同类别的场景样本彼此互相远离, 同类别的场景样本互相靠近聚集, 这与 (a) (b) (c) 有很大不同。由 (d) 和 (e) 的 t-SNE 图可知: T 型的两种情形 ($\lambda = 0, \lambda = 1$) 下 \mathbf{Z}_u 的不同类别场景样本彼此互相远离, 同类场景样本彼此互相靠近, 即 \mathbf{Z}_u 的鉴别性均得到大幅提升, 且明显优于 \mathbf{X}_u 和 $\mathbf{Z}_u + \mathbf{I}$ 。这表明: 目标域的局部保持显著提升了 \mathbf{Z}_u 的鉴别性, 其主要原因是图像特征空间中的场景局部近邻关系隐含了场景分布流形结构, 而场景分布流形结构本身具有同类样本互相靠近、异类样本互相远离的特性。因此, 通过局部保持将这种场景分布流形结构传递到稀疏系数空间, 可使 \mathbf{Z}_u 具有良好的鉴别性。图 6 为本文方法在不同 λ, μ 取值下的 OA 值。可知, 当 $\lambda = 1$ 且 $\mu = 1$ (源域和目标域都进行局部保持) 时, 本文方法在 GoogLeNet 和 VGGNet 两种图像特征上, 均取得最高 OA 值。因此, 源域的局部保持 ($\lambda = 1$) 和目标域的局部保持 ($\mu = 1$) 对提升鉴别性和 OA 值都具有必要性。

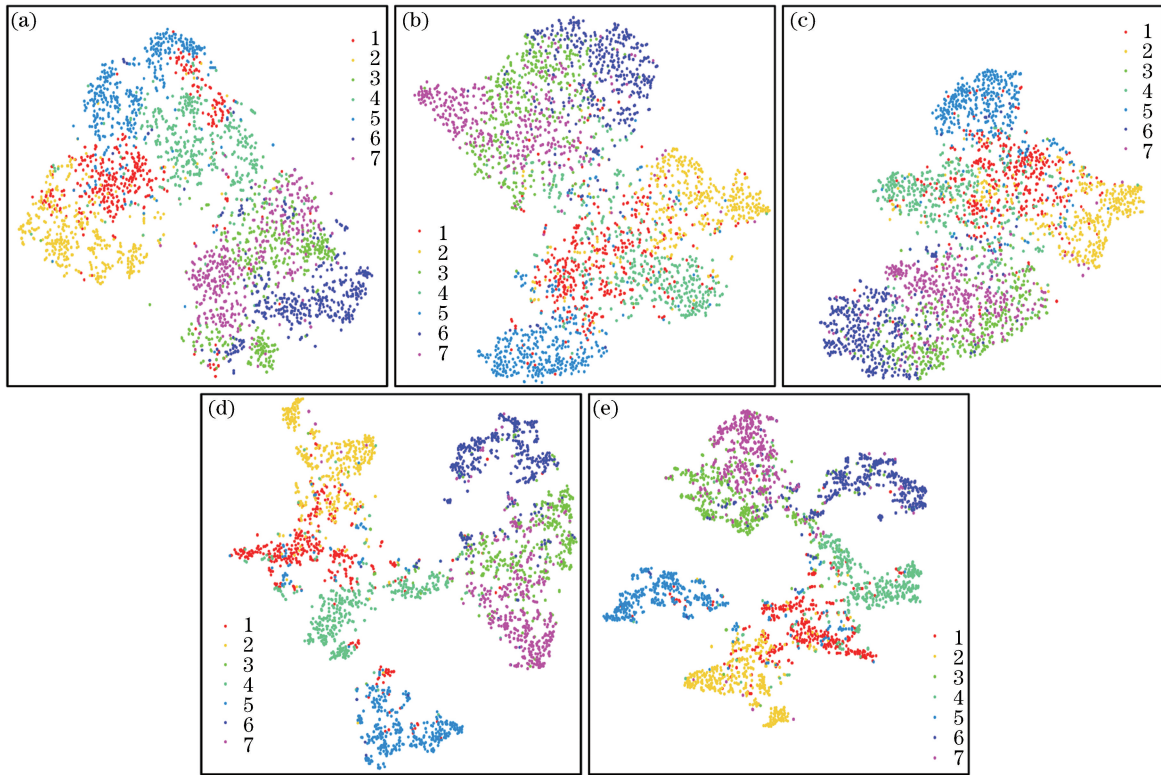


图 4 GoogLeNet 特征下 \mathbf{X}_u 及不同情形 \mathbf{Z}_u 的 t-SNE 图。(a) \mathbf{X}_u ; (b) $\mathbf{Z}_u + \mathbf{I}(\lambda = 0, \mu = 0)$; (c) $\mathbf{Z}_u + \mathbf{I}(\lambda = 1, \mu = 0)$; (d) $\mathbf{Z}_u + \mathbf{T}(\lambda = 0, \mu = 1)$; (e) $\mathbf{Z}_u + \mathbf{T}(\lambda = 1, \mu = 1)$

Fig. 4 The t-SNE graphs of \mathbf{X}_u and different \mathbf{Z}_u on GoogLeNet feature. (a) \mathbf{X}_u ; (b) $\mathbf{Z}_u + \mathbf{I}(\lambda = 0, \mu = 0)$; (c) $\mathbf{Z}_u + \mathbf{I}(\lambda = 1, \mu = 0)$; (d) $\mathbf{Z}_u + \mathbf{T}(\lambda = 0, \mu = 1)$; (e) $\mathbf{Z}_u + \mathbf{T}(\lambda = 1, \mu = 1)$

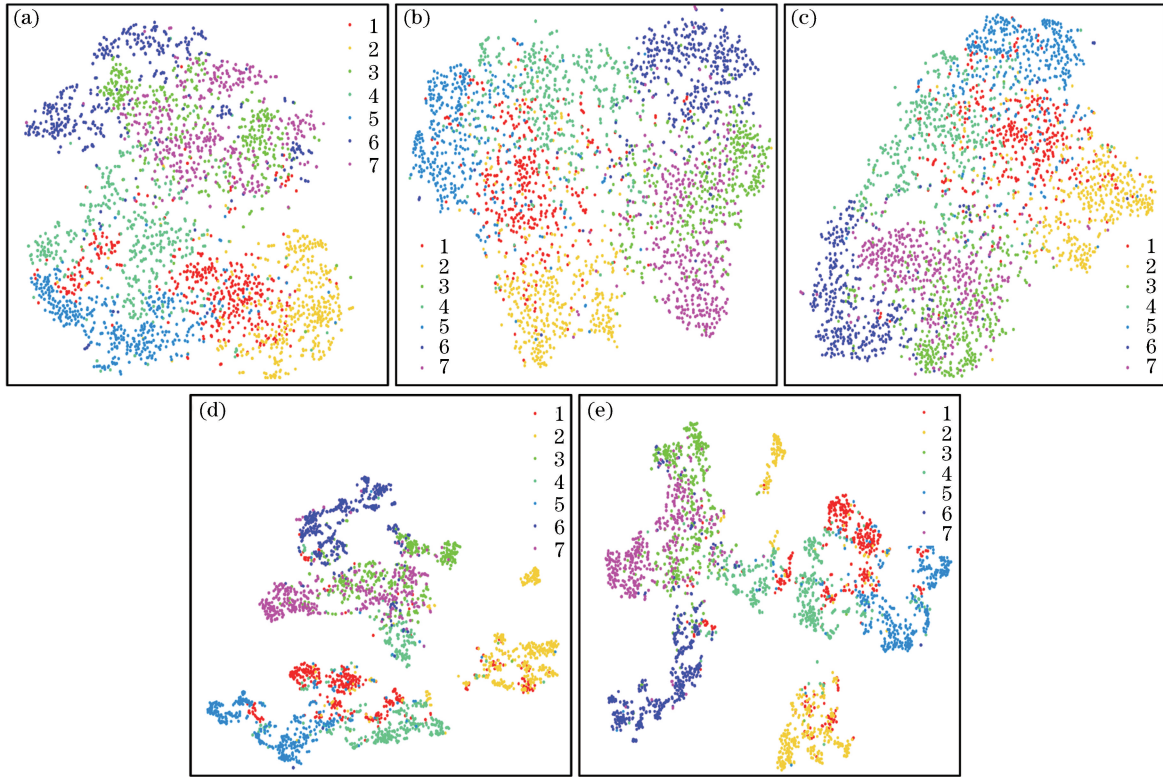


图 5 VGGNet 特征下 X_u 及不同情形 Z_u 的 t-SNE 图。(a) X_u ; (b) $Z_u + I(\lambda=0, \mu=0)$; (c) $Z_u + I(\lambda=1, \mu=0)$; (d) $Z_u + T(\lambda=0, \mu=1)$; (e) $Z_u + T(\lambda=1, \mu=1)$

Fig. 5 The t-SNE graphs of X_u and different Z_u on VGGNet feature. (a) X_u ; (b) $Z_u + I(\lambda=0, \mu=0)$; (c) $Z_u + I(\lambda=1, \mu=0)$; (d) $Z_u + T(\lambda=0, \mu=1)$; (e) $Z_u + T(\lambda=1, \mu=1)$

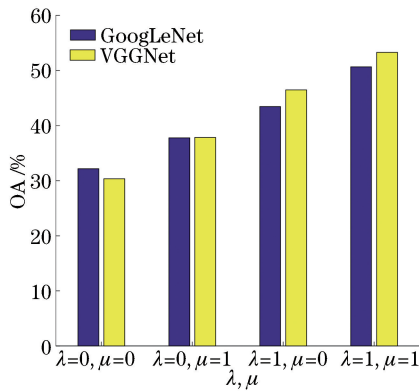


图 6 本文方法在不同 λ, μ 取值下的 OA 值

Fig. 6 OA values of proposed algorithm under different λ and μ

4.3 收敛性分析

图 7 和图 8 分别为 GoogLeNet 和 VGGNet 特征下, 本文算法的总体损失 ζ_{sum} 和 OA 值随迭代次数的变化情况。其中, $\zeta_{sum} = \zeta_s + \zeta_u$, 即源域和目标域损失值之和。从图中看出, 随着迭代次数不断增加, ζ_{sum} 快速下降, 并趋于稳定收敛, OA 值折线逐步提升并趋于稳定。这表明算法具有良好的收敛性, 这主要得益于优化计算过程中的交替求解策略。

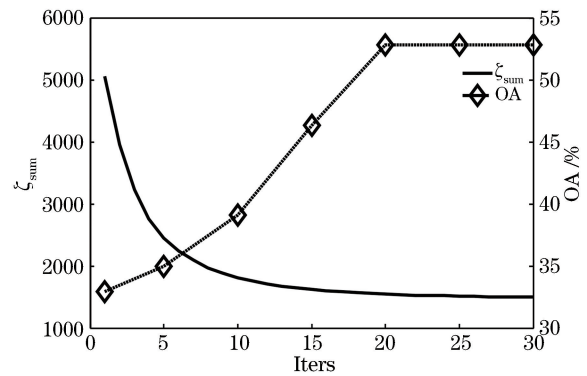


图 7 GoogLeNet 特征上的算法 ζ_{sum} / OA 曲线

Fig. 7 The ζ_{sum} / OA curves of algorithm under GoogLeNet feature

4.4 分类效果分析

4.4.1 总体分类准确度

表 1 是本文及现有典型 ZSC 方法, 以 UCM 为源域遥感场景集, 对目标场景集 RSSCN7 进行分类获取 OA 值。其中, LatEm、BiDiLEL、JLSE、SSE、DMaP、SAE 和 RKT 及本文 RSZSC_I($\mu=0$) 方法属于 I 型 ZSC 方法, 即目标域的场景图像特征未参与训练阶段。UDA、TME、SMS 及本文 RSZSC_T

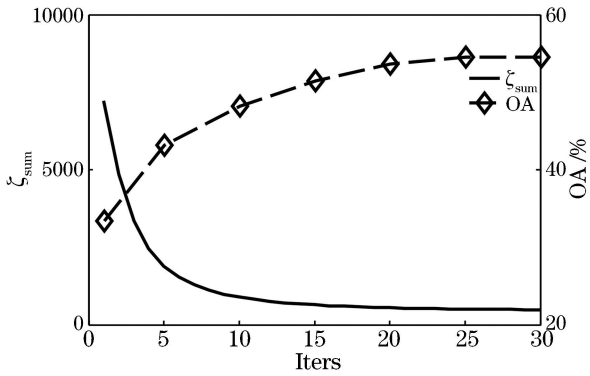


图 8 VGGNet 特征上的算法 ζ_{sum} /OA 曲线

Fig. 8 The ζ_{sum} /OA curves of algorithm under VGGNet feature

($\mu=1$)方法属于 T 型 ZSC 方法,即目标域的场景区图像特征参与训练。RSZSC_I 和 RSZSC_T 均进行了源域局部保持,即 $\lambda=1$ 。从表 1 可知,在 I 型及 T 型两种情形中,本文方法均取得了最高 OA 值。7 种 I 型对比方法中,SSE 方法在 GoogLeNet、VGGNet 特征上分别获得了 39.86% 和 37.36% 的最高 OA 值,但仍分别低于本文 I 型方法 RSZSC_I (GoogLeNet: 43.46%, VGGNet: 46.50%) OA 值 3.6% 和 9.14%。其主要原因是 SSE 方法是将目标域场景类别视为由源域各场景类别混合,基于目标域中相同类别场景的混合比例具有一定的相似性,将图像特征和词向量嵌入到同一语义空间中测量这种相似性,实现目标域场景的零样本分类。因此,SSE 方法中的相似性处理,难以转换为语义空间中场景的鉴别性。而本文 RSZSC_I 方法将源域中场景间局部关系保持到稀疏编码空间,并通过解析字典 Ω_s 和 Ω_c ,将这种鉴别性间接保持到目标域,从而提升了目标域场景稀疏系数的鉴别性。T 型方法中,本文 RSZSC_T 方法在 GoogLeNet、VGGNet 特征上分别获得 50.67% 和 53.29% 的最大 OA 值,优于 UDA (GoogLeNet: 42.61%, VGGNet: 43.59%) OA 值 8.06% 和 9.70%。其主要原因为 UDA 方法利用目标域语义词向量正则化映射函数来减轻域偏移问题,而 UDA 正则化时需要事先利用源域数据对目标域场景的类别标签预测,以作为正则化的目标项,但源域和目标域图像特征分布不同,引入了预测误差,因此降低了目标域样本间鉴别性。

4.4.2 RSSCN7 各类别的分类准确度

图 9 和图 10 分别为 I 型和 T 型 ZSC 方法在 GoogLeNet 和 VGGNet 图像特征上,对目标域各场景类别的分类准确度。可以看出,本文 RSZSC_I 和 RSZSC_T 方法在目标域 7 种场景类别上,均取

表 1 各 ZSC 方法在对 RSSCN7 上的总体分类准确度

Table 1 OA values of ZSC methods on RSSCN7

Method	I/T	OA / %	
		GoogLeNet	VGGNet
LatEm	I	18.79	14.15
BiDiLEL	I	14.51	14.47
JLSE	I	36.61	33.45
SSE	I	39.86	37.36
DMaP	I	34.00	32.07
SAE	I	34.20	35.60
RKT	I	35.47	32.07
RSZSC_I	I	43.46	46.50
UDA	T	42.61	43.59
TME	T	41.96	42.44
SMS	T	37.15	39.18
RSZSC_T	T	50.67	53.29

得了最高的分类准确度,优于典型的 ZSC 方法,证明了本文“增强鉴别性+k-means 聚类”方法能够适应目标域图像特征分布的变化,取得最高的分类准确度。

4.5 计算效率分析

为比较本文与其他 ZSC 方法的计算效率,测试各 ZSC 方法对图像特征 GoogLeNet 的运算耗时,结果如表 2 所示。从表 2 可以看出,在 I 型算法中,DMaP 方法耗时最长,为 396.45 s,其次是 JLSE 方法,耗时为 67.49 s,而本文 RSZSC_I 方法耗时最短,为 11.68 s。在 T 型算法中,TME 耗时最长,为 76.91 s,而本文 RSZSC_T 方法耗时最短,为 14.70 s。这主要因为 ADL 算法的时间复杂度低,且每个更新步骤都有闭式解,使本文方法的运算效率优于典型的 ZSC 方法。

5 结 论

成像时间等条件不同,导致目标域与源域的遥感图像特征分布发生变化,令遥感场景零分类效果下降。针对该问题,提出一种基于局部保持的遥感场景零样本分类算法。该方法的主要特点是:1)利用稀疏编码效率较高的解析字典学习方法,提取图像特征和语义词向量的稀疏系数,减少了冗余信息;2)采用基于流形理论的局部近邻关系保持,显著提升了目标域场景稀疏系数的鉴别性;3)采用 k-means 方法对目标域稀疏系数进行聚类,并推断各

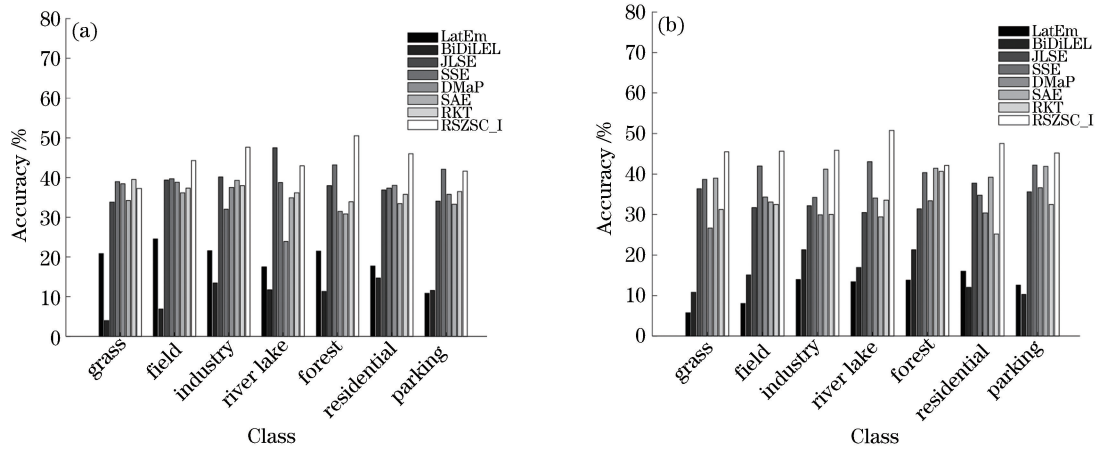


图 9 I 型 ZSC 方法对 RSSCN7 各类别的分类准确度。(a) GoogLeNet; (b) VGGNet

Fig. 9 Accuracies of different classes of RSSCN7 produced by I type ZSC methods. (a) GoogLeNet; (b) VGGNet

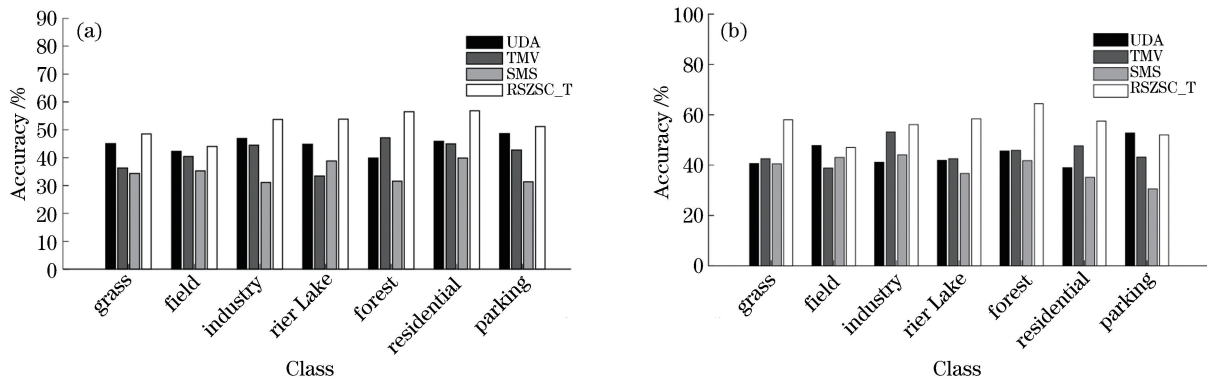


图 10 T 型 ZSC 方法对 RSSCN7 各类别的分类准确度。(a) GoogLeNet; (b) VGGNet

Fig. 10 Accuracies of different classes of RSSCN7 produced by T type ZSC methods. (a) GoogLeNet; (b) VGGNet

表 2 ZSC 算法在 GoogLeNet 特征上的运算耗时

Table 2 Computing time of different ZSC algorithms on GoogLeNet feature

Method	I/T	Time /s
LatEm	I	17.71
BiDiLEL	I	22.59
JLSE	I	67.49
SSE	I	14.35
DMaP	I	396.45
SAE	I	16.21
RKT	I	22.53
RSZSC_I	I	11.68
UDA	T	56.37
TME	T	76.91
SMS	T	49.28
RSZSC_T	T	14.70

场景的类别标签,有效减轻了图像特征分布变化对分类效果的影响。实验结果证明,本文方法能够适应目标域图像特征分布变化,分类效果优于典型的 ZSC 方法,具有一定的优越性和适应性。但该方法仅利用了简单的 KNN 局部近邻关系,下一步将考虑使用更复杂的近邻关系,以进一步提升遥感场景零样本分类效果。

参 考 文 献

[1] Liu D W, Han L, Han X Y. High spatial resolution remote sensing image classification based on deep learning [J]. Acta Optica Sinica, 2016, 36 (4): 0428001.
刘大伟, 韩玲, 韩晓勇. 基于深度学习的高分辨率遥感影像分类研究 [J]. 光学学报, 2016, 36 (4): 0428001.

[2] Chen S Z, Tian Y L. Pyramid of spatial relations for scene-level land use classification [J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(4): 1947-1957.

- [3] Zhang L P, Zhang L F, Du B. Deep learning for remote sensing data: a technical tutorial on the state of the art[J]. *IEEE Geoscience and Remote Sensing Magazine*, 2016, 4(2): 22-40.
- [4] Li A X, Lu Z W, Wang L W, *et al.* Zero-shot scene classification for high spatial resolution remote sensing images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(7): 4157-4167.
- [5] Xian Y Q, Akata Z, Sharma G, *et al.* Latent embeddings for zero-shot classification [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 69-77.
- [6] Wang D H, Li Y, Lin Y T, *et al.* Relational knowledge transfer for zero-shot learning [C]//Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16), February 12-17, 2016, Phoenix, Arizona, USA. California: AAAI Publications, 2016: 2145-2151.
- [7] Zhang Z M, Saligrama V. Zero-shot learning via joint latent similarity embedding [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 6034-6042.
- [8] Zhang Z M, Saligrama V. Zero-shot learning via semantic similarity embedding [C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 4166-4174.
- [9] Wang Q, Chen K. Zero-shot visual recognition via bidirectional latent embedding [J]. *International Journal of Computer Vision*, 2017, 124(3): 356-383.
- [10] Li Y N, Wang D H, Hu H H, *et al.* Zero-shot recognition using dual visual-semantic mapping paths [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 5207-5215.
- [11] Kodirov E, Xiang T, Fu Z Y, *et al.* Unsupervised domain adaptation for zero-shot learning [C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 2452-2460.
- [12] Guo Y C, Ding G G, Jing X M, *et al.* Transductive zero-shot recognition via shared model space learning [C]//Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA. California: AAAI Publications, 2016: 3494-3500.
- [13] Fu Y W, Hospedales T M, Xiang T, *et al.* Transductive multi-view embedding for zero-shot recognition and annotation [M]//Fleet D, Pajdla T, Schiele B, *et al.* *Lecture notes in computer science* 2014. Cham: Springer, 2014, 8690: 584-599.
- [14] Ji Z, Xie Y Z, Pang Y W. Zero-shot learning based on canonical correlation analysis and distance metric learning[J]. *Journal of Tianjin University*, 2017, 50(8): 813-820.
冀中, 谢于中, 庞彦伟. 基于典型相关分析和距离度量学习的零样本学习[J]. *天津大学学报*, 2017, 50(8): 813-820.
- [15] Ji Z, Sun T, Yu Y L. Transductive discriminative dictionary learning approach for zero-shot classification[J]. *Journal of Software*, 2017, 28(11): 2961-2970.
冀中, 孙涛, 于云龙. 一种基于直推判别字典学习的零样本分类方法[J]. *软件学报*, 2017, 28(11): 2961-2970.
- [16] Bao C L, Ji H, Quan Y H, *et al.* Dictionary learning for sparse coding: algorithms and convergence analysis[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(7): 1356-1369.
- [17] Yang M, Chang H Y, Luo W X. Discriminative analysis-synthesis dictionary learning for image classification[J]. *Neurocomputing*, 2017, 219: 404-411.
- [18] Wang J J, Guo Y Q, Guo J, *et al.* Synthesis linear classifier based analysis dictionary learning for pattern classification[J]. *Neurocomputing*, 2017, 238: 103-113.
- [19] Ravishankar S, Bresler Y. Sparsifying transform learning with efficient optimal updates and convergence guarantees[J]. *IEEE Transactions on Signal Processing*, 2015, 63(9): 2389-2404.
- [20] Kreutz-Delgado K, Murray J F, Rao B D, *et al.* Dictionary learning algorithms for sparse representation[J]. *Neural Computation*, 2003, 15(2): 349-396.
- [21] Fernando B, Fromont E, Muselet D, *et al.* Discriminative feature fusion for image classification [C]//2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE, 2012: 3434-3441.
- [22] Tenenbaum J B. A global geometric framework for nonlinear dimensionality reduction [J]. *Science*, 2000, 290(5500): 2319-2323.
- [23] 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.
- [24] Zou Q, Ni L H, Zhang T, *et al.* Deep learning based

- feature selection for remote sensing scene classification [J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(11): 2321-2325.
- [25] Szegedy C, Liu W, Jia Y Q, *et al.* Going deeper with convolutions [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 7298594.
- [26] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2019-01-01]. <https://arxiv.org/abs/1409.1556>.