# 基于词向量一致性融合的遥感场景零样本分类方法

吴晨[1]，于光[2]*，张凤晶[2]，刘宇[2]，袁昱纬[3]，全吉成[2]

[1] 海军航空大学，山东 烟台 264001；

[2]空军航空大学，吉林 长春 130022；

[3]91977 部队，北京 102200

**摘要** 遥感场景类别的语义词向量与图像特征原型的距离结构不一致问题，严重影响遥感场景零样本分类效果。针对该问题，利用不同词向量间一致性，提出一种基于解析字典学习的语义词向量融合方法，以提升遥感场景零样本分类效果。首先，采用解析字典学习方法，提取场景类别的不同词向量的公共稀疏系数，并作为融合后的语义词向量；然后，同样采用解析字典学习方法，将场景类别的图像特征原型嵌入到融合后的词向量空间，与融合后的词向量进行结构对齐，降低距离结构的不一致性；最后，通过联合优化获得未知类的图像特征空间类别原型表示，并采用最近邻分类器完成未知类别遥感场景的分类。在 3 种遥感场景数据集和多种语义词向量上进行定量和定性实验。实验结果表明，通过词向量融合可以获得与图像特征原型结构更一致的语义词向量，从而显著提升遥感场景零样本分类的准确度。

**关键词** 遥感；场景分类；零样本分类；结构对齐；词向量融合；解析字典学习

**中图分类号** TP753 **文献标识码** A **doi**：10.3788/AOS201939.0828002

# Zero-Shot Classification Method for Remote-Sensing Scenes Based on Word Vector Consistent Fusion

Wu Chen[1], Yu Guang[2]*, Zhang Fengjing[2], Liu Yu[2], Yuan Yuwei[3], Quan Jicheng[2]

[1] *Naval Aviation University, Yantai, Shandong* 264001, *China*;

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

[3] *The* 91977 *of PLA, Beijing* 102200, *China*

**Abstract** The problem of distance structure difference between the word vectors and visual prototypes of remote-sensing scene classification seriously influences the performance of the zero-shot scene classification. Herein, a fusion method based on analytical dictionary learning is proposed to exploit the consistency among the different kinds of word vectors for the performance improvement of the zero-shot scene classification. Firstly, the common sparse coefficients of different kinds of word vectors of scene classification are extracted by analytical dictionary learning method and acted as the fused word vector. Secondly, the visual prototypes are embedded into and structure-aligned with the fused word vector by analytical dictionary learning method similarly, to reduce the distance structure inconsistency. Finally, the prototypes of the unseen classes in the image feature space are obtained via joint optimization, and the nearest neighbor classifier is used to complete the classification of remote-sensing scenes from the unseen classes. Quantitative and qualitative experiments are also conducted on three remote-sensing scene datasets with the fusion of various word vectors. The experimental results show that the fused word vector is more structure-consistent with the prototypes in the image feature space, and the zero-shot classification accuracies of the remote-sensing scenes can be significantly improved.

**Key words** remote sensing; scenes classification; zero-shot classification; structure alignment; word vector fusion; analytical dictionary learning

**OCIS codes** 280.4788; 100.2960; 100.3008; 100.5010

**收稿日期**：2019-03-15；**修回日期**：2019-03-25；**录用日期**：2019-04-15

**基金项目**：国家自然科学基金(61301233)

* **E-mail**：1471612866@qq.com

## 1 引　言

传统的遥感图像分类方法主要在“像素”和“对象”层面进行，针对的是空间分辨率不高的遥感图像分类任务，然而近年来随着遥感图像空间分辨率不断提升，这些方法越来越难以满足实际需要。场景分类作为高分辨率遥感图像快速分析与信息提取的重要手段，近10年来受到广泛关注[1-2]。这里的“场景”是指具有清晰类别语义的遥感图像块，以其作为遥感图像分类的基本单元，使场景分类能够适应大规模遥感图像快速分析的需要。然而，目前的场景分类方法属于监督分类，无法将识别能力灵活扩展到新类别场景，因此阻碍了遥感场景分类研究的进一步发展。为解决现有场景分类方法的迁移识别能力不足问题，Li等[3]提出了遥感图像零样本场景分类方法，即将场景分类与零样本学习方法结合，提高对新类别场景的迁移识别能力。

零样本分类(ZSC)是一种特殊的无监督分类方法，其基本原理是：以类别名称的语义词向量为桥梁，通过迁移由已知(seen)类别标注样本学习得到识别模型，获得对新的(unseen)类别的识别能力。由于ZSC方法能够在不标注unseen类样本情况下，获得对其的识别能力，因此近年来受到广泛关注[4-11]。为进行细粒度的ZSC，Xian等[4]在兼容函数学习过程中引入隐式变量模型，从而提出隐式嵌入方法(LatEm)。针对映射函数的泛化能力不足问题，Wang等[5]提出的关系知识迁移(RKT)方法通过语义映射方法还原unseen类别的流形结构。Zhang等[6]提出的联合隐式相似性嵌入(JLSE)方法将样本特征和对应的语义嵌入表示作为输入，通过建立两者之间的相似性度，实现对unseen类别样本的ZSC。Zhang等[7]提出的语义相似性嵌入(SSE)方法将源域或目标域数据视为训练类组合，并将其映射到同一语义空间中。Wang等[8]提出的双向隐式嵌入(BiDiLEL)方法利用流形保持原理，将图像特征和语义特征分别映射到第三方的公共空间。Li等[9]提出的双视觉语义映射(DMaP)方法利用语义空间流形和视觉语义映射迁移能力之间的关系，修正了语义词向量。为估计unseen样本特征分布特点以提升ZSC效果，Zhao等[10]提出利用直推式框架(MDP)。除语义词向量外，人工标注的类别属性向量也可用于ZSC研究中，如Lampert等[11]提出的基于类别属性向量表示的零样本分类方法，但是由于类别属性向量的标注成本较大且扩展性较弱，近年来用到ZSC的研究越来越少。语义词向量[12]是采用自然语言训练模型，在大规模文本语料集上，通过无监督学习得到的实体单词高维向量表示。在ZSC中，采用类别名称的语义词向量，提供类别间距离结构关系，来辅助推断图像特征空间unseen类别的原型表示。因此，语义词向量能否反映图像特征空间的类间距离结构关系，是ZSC方法的关键。现有ZSC方法针对均是某一领域内的细粒度类别的分类任务，然而，由于遥感场景类别涉及不同领域，词向量需要反映场景类别间的距离关系。单种语义词向量受训练语料、训练模型限制，难以满足多领域的遥感场景类别的情形。

近几年，随着自然语言处理技术的进步，已能便捷获取不同训练模型(如Word2Vector[13]、Glove[14]等)和不同训练语料(如Wikipedia、Common Crawl等)的语义词向量。这些语义词向量具有一定的一致性，通过融合可获得与图像特征空间场景类别距离结构更一致的语义词向量，从而提升遥感场景ZSC准确度。为利用不同语义词向量间的一致性，本文提出一种基于词向量一致性融合的遥感场景ZSC方法。首先，采用解析字典学习方法，获取各语义词向量的稀疏系数；其次，将各词向量的公共稀疏系数作为融合后的语义词向量表示；然后，再采用解析字典方法，将seen类图像特征原型表示嵌入到融合后的语义词向量空间，与其中的seen类融合语义词向量进行结构对齐，提升模型到unseen类的迁移效果；最后，在图像特征空间以学习得到的unseen类原型表示为中心，采用最近邻分类器对unseen类场景样本进行分类。

## 2 解析字典学习

字典学习方法分为两类，即合成性字典学习(SDL)和解析性字典学习(ADL)。SDL认为输入特征可以由字典和相应稀疏系数重建得到，而ADL则将字典应用到输入特征上，获得特征的稀疏系数。虽然SDL方法应用广泛，但其计算效率不高。而ADL通常具有闭式解，编码能力良好，计算效率较高[15-16]。ADL的基本公式为

$$\arg\min_{\boldsymbol{\Omega},\boldsymbol{Z}} \frac{1}{2}\left\|\boldsymbol{Z}-\boldsymbol{\Omega}\boldsymbol{X}\right\|_{\mathrm{F}}^{2},$$
$$\text{s.t. } \boldsymbol{\Omega}\in\boldsymbol{\Gamma},\ \left\|z_i\right\|_0\leqslant T_0, \tag{1}$$

式中：$\boldsymbol{X}=[\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n]\in\mathbb{R}^{m\times n}$为$n$个输入样本组成的特征矩阵，$\boldsymbol{x}_i\in\mathbb{R}^m$为第$i$个样本；$\boldsymbol{Z}$为$\boldsymbol{X}$的稀疏系数，其样本稀疏性采用$l_0$范数及参数$T_0$实现；

$\boldsymbol{\Omega}$ 为解析字典；$\boldsymbol{\Gamma}$ 是为避免出现平凡解而对 $\boldsymbol{\Omega}$ 的 log-det 限制条件[17]。

## 3 方　法

采用 ADL 方法获得各词向量的稀疏系数，并将公共的稀疏系数作为融合词向量表示，与图像特征空间类别原型结构对齐。首先，由于词向量中存在冗余信息，影响类间距离结构信息表达，需要对其进行稀疏编码处理，以减少冗余信息，突出类间距离结构信息。而解析字典学习方法具有优越的稀疏编码能力，因此本文采用解析字典学习方法，建立稀疏编码项，获取各语义词向量的稀疏系数。其次，为获取不同词向量的一致性，将各词向量的公共稀疏系数作为融合后的语义词向量表示。然后，融合后的词向量空间与场景图像特征空间来源不同，再加上遥感场景类别涉及不同领域（人类生产生活以及自然地貌），导致了两种空间中的场景类间距离存在较大差异，降低了对 unseen 类的迁移效果。因此需要对这种空间差异性进行建模，而 ADL 方法具有较强的稀疏编码能力，能够将场景图像特征嵌入到稀疏的融合后语义词向量空间，从而与其中的 seen 类场景图像特征对齐。最后，通过对 seen 和 unseen 类上的目标函数进行联合迭代计算，获得 unseen 类图像特征原型表示，进而采用最近邻分类器完成对 unseen 样本的分类。本文方法的整体框架如图 1 所示。

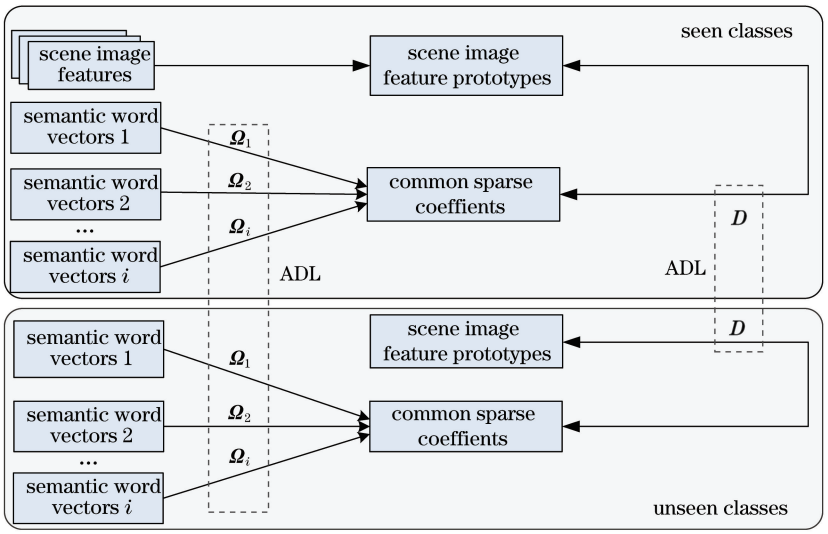


图 1　本文方法的整体框架图

Fig. 1　Whole framework of proposed method

### 3.1　基于解析字典学习的词向量融合方法

词向量融合的目标函数可表示为

$$\begin{cases}\zeta_{\mathrm{s}}=\min\limits_{\boldsymbol{\Omega}_i,\boldsymbol{Z}^{\mathrm{s}},\boldsymbol{D},\boldsymbol{P}^{\mathrm{s}}}\sum\limits_{i=1}^{M}\left\|\boldsymbol{Z}^{\mathrm{s}}-\boldsymbol{\Omega}_i\boldsymbol{C}_i^{\mathrm{s}}\right\|_{\mathrm{F}}^2+\\ \quad\left\|\boldsymbol{Z}^{\mathrm{s}}-\boldsymbol{D}\boldsymbol{P}^{\mathrm{s}}\right\|_{\mathrm{F}}^2+\left\|\boldsymbol{X}^{\mathrm{s}}\boldsymbol{H}_{\mathrm{s}}-\boldsymbol{P}^{\mathrm{s}}\right\|_{\mathrm{F}}^2\\ \zeta_{\mathrm{u}}=\min\limits_{\boldsymbol{\Omega}_i,\boldsymbol{Z}^{\mathrm{u}},\boldsymbol{D},\boldsymbol{P}^{\mathrm{u}}}\sum\limits_{i=1}^{M}\left\|\boldsymbol{Z}^{\mathrm{u}}-\boldsymbol{\Omega}_i\boldsymbol{C}_i^{\mathrm{u}}\right\|_{\mathrm{F}}^2+\left\|\boldsymbol{Z}^{\mathrm{u}}-\boldsymbol{D}\boldsymbol{P}^{\mathrm{u}}\right\|_{\mathrm{F}}^2\end{cases},$$

$$\mathrm{s.t.}\boldsymbol{\Omega}_i\in\boldsymbol{\Gamma},\boldsymbol{D}\in\boldsymbol{T},\left\|\boldsymbol{Z}_i^{\mathrm{s}}\right\|_0\leqslant T_0,\left\|\boldsymbol{Z}_i^{\mathrm{u}}\right\|_0\leqslant T_0,\tag{2}$$

式中：$\zeta_{\mathrm{s}}$、$\zeta_{\mathrm{u}}$ 分别为 seen 和 unseen 类词向量融合的目标函数；$\zeta_{\mathrm{s}}$ 和 $\zeta_{\mathrm{u}}$ 的第一项为稀疏编码项，旨在提取各词向量的一致性稀疏系数作为新的融合词向量表示，第二项为结构对齐项，将融合语义词向量表示与图像特征空间场景类别原型进行结构对齐，$\zeta_{\mathrm{s}}$ 的第三项为 seen 类图像特征原型学习项，旨在学习 seen 类场景的类别原型表示；$\boldsymbol{C}_i^{\mathrm{s}}\in\mathbb{R}^{d_i\times c_{\mathrm{s}}}$ 为 seen 类的第 $i$ 种词向量（共 $M$ 种不同词向量）矩阵，$\boldsymbol{C}_i^{\mathrm{u}}\in\mathbb{R}^{d_i\times c_{\mathrm{u}}}$ 为 unseen 类的第 $i$ 种词向量矩阵，$d_i$ 为第 $i$ 种词向量维度，$c_{\mathrm{s}}$ 为 seen 类数，$c_{\mathrm{u}}$ 为 unseen 类数；$\boldsymbol{\Omega}_i\in\mathbb{R}^{d_i\times d_i}$ 为第 $i$ 种词向量空间对应的解析字典。通过融合不同词向量，获得不同词向量的一致性表示，并与图像特征空间场景类别原型进行对齐，从而计算得到 unseen 类的图像特征原型表示，最后进行最近邻分类。seen 类不同词向量的一致性稀疏系数 $\boldsymbol{Z}^{\mathrm{s}}\in\mathbb{R}^{d\times c_{\mathrm{s}}}$ 为融合后 seen 类别词向量表示，unseen 类不同词向量的一致性稀疏系数 $\boldsymbol{Z}^{\mathrm{u}}\in\mathbb{R}^{d\times c_{\mathrm{u}}}$ 为融合后 unseen 类别词向量表示，其中 $d$ 为融合词向量的维数。$\boldsymbol{P}^{\mathrm{s}}\in\mathbb{R}^{q\times c_{\mathrm{s}}}$ 为 seen 类场景在图像特征空间中的原型，$\boldsymbol{P}^{\mathrm{u}}\in\mathbb{R}^{q\times c_{\mathrm{u}}}$ 为 unseen 类场景在图像特征空间中的原型，$q$ 为图像特征维数；

$\boldsymbol{H}_s\in\mathbb{R}^{N_s\times c_s}$为 seen 样本的类别标签矩阵，其中的行向量表示 seen 样本的类别标签 one-hot 向量。$\boldsymbol{X}^s\in\mathbb{R}^{q\times N_s}$为图像特征空间中 seen 样本的特征矩阵。由于图像特征空间中的类内样本分布结构复杂，简单地以样本均值中心作为类别的原型，没有充分利用 seen 类别样本的信息。因此$\|\boldsymbol{X}^s\boldsymbol{H}_s-\boldsymbol{P}^s\|_F^2$主要作用是通过建立$\boldsymbol{P}^s$与$\boldsymbol{X}^s$间的对应关系，以更灵活地学习$\boldsymbol{P}^s$，而不是仅仅以样本均值作为 seen 类别原型表示。$\boldsymbol{D}\in\mathbb{R}^{d\times q}$为图像特征空间的解析字典，其主要作用是将$\boldsymbol{P}^s$和$\boldsymbol{P}^u$嵌入到融合后的语义词向量空间，与融合后的语义词向量$\boldsymbol{Z}^s$和$\boldsymbol{Z}^u$对齐。

(2)式对目标变量$\boldsymbol{\Omega}_i$、$\boldsymbol{D}$、$\boldsymbol{Z}^s$、$\boldsymbol{Z}^u$、$\boldsymbol{P}^s$和$\boldsymbol{P}^u$同时非凸，难以直接求解，但可采用逐个循环方式进行求解。由于$\boldsymbol{D}$的求解依赖于$\boldsymbol{Z}^s$和$\boldsymbol{Z}^u$，而$\boldsymbol{Z}^s$和$\boldsymbol{Z}^u$一般可初始化为 one-hot 向量矩阵，因此循环求解过程中，最先求解$\boldsymbol{D}$，其次$\boldsymbol{\Omega}_i$，然后$\boldsymbol{Z}^s$和$\boldsymbol{Z}^u$，最后优化$\boldsymbol{P}^s$和$\boldsymbol{P}^u$。而$\boldsymbol{P}^s$初始化为各 seen 类别样本的均值，其中涉及到的 unseen 类原型$\boldsymbol{P}^u$在第一次迭代时未知，因此需要对其赋予初始值，本文采用高斯分布对$\boldsymbol{P}^u$进行随机初始化。具体步骤如下。

1) 固定$\boldsymbol{\Omega}_i$、$\boldsymbol{Z}^s$、$\boldsymbol{Z}^u$、$\boldsymbol{P}^s$、$\boldsymbol{P}^u$，更新$\boldsymbol{D}$

此时的总体目标函数为

$$\min_{\boldsymbol{D}}\|[\boldsymbol{Z}^s,\boldsymbol{Z}^u]-\boldsymbol{D}[\boldsymbol{P}^s,\boldsymbol{P}^u]\|_F^2,\quad \text{s.t. } \boldsymbol{D}\in\boldsymbol{T}。\tag{3}$$

由于$\boldsymbol{D}\in\mathbb{R}^{d\times q}$的行列数不相等，因此需采用正则项$R(\boldsymbol{D})$：

$$R(\boldsymbol{D})=\begin{cases}\|\boldsymbol{D}\|_F^2-\log|\det\boldsymbol{D}^T\boldsymbol{D}|, & d\geqslant q\\ \|\boldsymbol{D}\|_F^2-\log|\det\boldsymbol{D}\boldsymbol{D}^T|, & d<q\end{cases},\tag{4}$$

这里记$[\boldsymbol{Z}^s,\boldsymbol{Z}^u]$为$\boldsymbol{Z}$，记$[\boldsymbol{P}^s,\boldsymbol{P}^u]$为$\boldsymbol{P}$。因此，更新$\boldsymbol{D}$的目标函数为

$$\min_{\boldsymbol{D}}\|\boldsymbol{Z}-\boldsymbol{D}\boldsymbol{P}\|_F^2+\alpha R(\boldsymbol{D}),\tag{5}$$

式中：$\alpha>0$为正则项$R(\boldsymbol{D})$的重要性系数。然而，(5)式仍然对字典$\boldsymbol{D}$难以直接求解，本文采用梯度下降方法进行求解[16]。其中$\|\boldsymbol{Z}-\boldsymbol{D}\boldsymbol{P}\|_F^2$和$R(\boldsymbol{D})$对字典$\boldsymbol{D}$的梯度分别为$\nabla_{\boldsymbol{D}}(\|\boldsymbol{Z}-\boldsymbol{D}\boldsymbol{P}\|_F^2)=2\boldsymbol{D}\boldsymbol{P}\boldsymbol{P}^T-2\boldsymbol{P}\boldsymbol{Z}^T$、$\nabla_{\boldsymbol{D}}[R(\boldsymbol{D})]=-2\boldsymbol{D}^\dagger$。$\boldsymbol{D}^\dagger$为字典$\boldsymbol{D}$的伪逆矩阵。因此，具体的梯度下降公式为

$$\boldsymbol{D}:=\boldsymbol{D}-\eta\times\{\nabla_{\boldsymbol{D}}(\|\boldsymbol{Z}-\boldsymbol{D}\boldsymbol{P}\|_F^2)+\nabla_{\boldsymbol{D}}[R(\boldsymbol{D})]\}=\boldsymbol{D}-2\eta(\boldsymbol{D}\boldsymbol{P}\boldsymbol{P}^T-\boldsymbol{P}\boldsymbol{Z}^T-\boldsymbol{D}^\dagger),\tag{6}$$

式中：超参数$\eta$为梯度下降速率。

2) 固定$\boldsymbol{D}$、$\boldsymbol{Z}^s$、$\boldsymbol{Z}^u$、$\boldsymbol{P}^s$、$\boldsymbol{P}^u$，更新$\boldsymbol{\Omega}_i$

此时的目标函数为

$$\min_{\boldsymbol{\Omega}_i}\sum_{i=1}^{M}\|[\boldsymbol{Z}^s,\boldsymbol{Z}^u]-\boldsymbol{\Omega}_i[\boldsymbol{C}_i^s,\boldsymbol{C}_i^u]\|_F^2,\quad \text{s.t. } \boldsymbol{\Omega}_i\in\boldsymbol{\Gamma},\tag{7}$$

(7)式的求解步骤与(3)式相同。

3) 固定$\boldsymbol{\Omega}_i$、$\boldsymbol{P}^u$、$\boldsymbol{D}$、$\boldsymbol{P}^s$，更新$\boldsymbol{Z}^s$、$\boldsymbol{Z}^u$

此时关于$\boldsymbol{Z}^s$的目标函数为

$$\min_{\boldsymbol{Z}^s}\sum_{i=1}^{M}\|\boldsymbol{Z}^s-\boldsymbol{\Omega}_i\boldsymbol{C}_i^s\|_F^2+\|\boldsymbol{Z}^s-\boldsymbol{D}\boldsymbol{P}^s\|_F^2,\tag{8}$$

对$\boldsymbol{Z}^s$求导并置 0，可得$\boldsymbol{Z}^s=\left(\sum_{i=1}^{M}\boldsymbol{\Omega}_i\boldsymbol{C}_i^s+\boldsymbol{D}\boldsymbol{P}^s\right)/(M+1)$。

此时关于$\boldsymbol{Z}^u$的目标函数为

$$\min_{\boldsymbol{Z}^u}\sum_{i=1}^{M}\|\boldsymbol{Z}^u-\boldsymbol{\Omega}_i\boldsymbol{C}_i^u\|_F^2+\|\boldsymbol{Z}^u-\boldsymbol{D}\boldsymbol{P}^u\|_F^2。\tag{9}$$

同理，对$\boldsymbol{Z}^u$求导并置$\boldsymbol{0}$，可得$\boldsymbol{Z}^u=\left(\sum_{i=1}^{M}\boldsymbol{\Omega}_i\boldsymbol{C}_i^u+\boldsymbol{D}\boldsymbol{P}^u\right)/(M+1)$。按照比例$T_0$保留幅值较大的前若干个元素且其余元素置 0 的方式稀疏化$\boldsymbol{Z}^s$和$\boldsymbol{Z}^u$中的列向量。

4) 固定$\boldsymbol{\Omega}_i$、$\boldsymbol{Z}^s$、$\boldsymbol{Z}^u$、$\boldsymbol{D}$、$\boldsymbol{P}^u$，更新$\boldsymbol{P}^s$

此时的总体目标函数为

$$\min_{\boldsymbol{P}^s}\|\boldsymbol{Z}^s-\boldsymbol{D}\boldsymbol{P}^s\|_F^2+\|\boldsymbol{X}^s\boldsymbol{H}_s-\boldsymbol{P}^s\|_F^2,\tag{10}$$

对$\boldsymbol{P}^s$求导并置$\boldsymbol{0}$，可得$\boldsymbol{P}^s=(\boldsymbol{D}^T\boldsymbol{D}+I)^{-1}(\boldsymbol{D}^T\boldsymbol{Z}^s+\boldsymbol{X}^s\boldsymbol{H}_s)$。

5) 固定$\boldsymbol{\Omega}_i$、$\boldsymbol{Z}^s$、$\boldsymbol{Z}^u$、$\boldsymbol{D}$、$\boldsymbol{P}^s$，更新$\boldsymbol{P}^u$

此时的总体目标函数为

$$\min_{\boldsymbol{P}^u}\|\boldsymbol{Z}^u-\boldsymbol{D}\boldsymbol{P}^u\|_F^2,\tag{11}$$

对$\boldsymbol{P}^u$求导并置 0，可得$\boldsymbol{P}^u=(\boldsymbol{D}^T\boldsymbol{D})^{-1}\boldsymbol{D}^T\boldsymbol{Z}^u$。

迭代循环结束后，在图像特征空间中，以学到的 unseen 类原型表示$\boldsymbol{P}^u$，作为最近邻分类器的中心，对 unseen 类样本进行分类。

### 3.2　本文方法步骤

本文基于词向量一致性融合的遥感场景 ZSC 方法的计算流程如图 2 所示，具体步骤如下：

输入：seen 类场景图像特征$\boldsymbol{X}^s$，$M$种不同的词向量(其中 seen 类词向量$\boldsymbol{C}_i^s$，unseen 类词向量$\boldsymbol{C}_i^u$，$i=1,2,\cdots,M$)，unseen 类场景图像特征矩阵$\boldsymbol{X}^u=[\boldsymbol{x}_1^u,\cdots,\boldsymbol{x}_{N_u}^u]\in\mathbb{R}^{q\times N_u}$，$N_u$为 unseen 类场景图像个数，最大迭代次数 Iter_$N$。

输出：对$\boldsymbol{X}^u$中样本推断类别标签。

步骤 1：初始化$\boldsymbol{Z}^s$和$\boldsymbol{Z}^u$为 one-hot 向量矩阵，初始化$\boldsymbol{P}^s$为各 seen 类别样本的均值，并采用高斯分布对$\boldsymbol{P}^u$进行随机初始化；

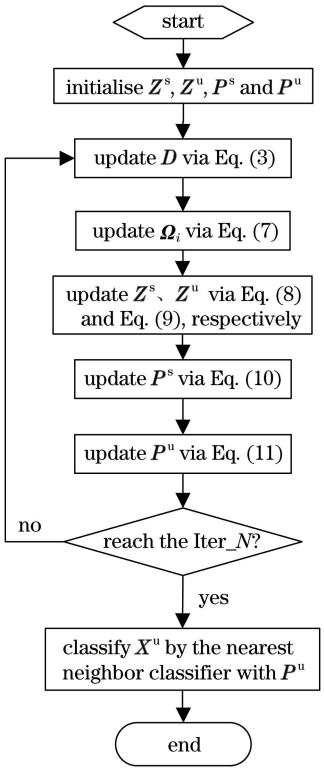


图 2　本文方法运算流程图

Fig. 2　Operational flow chart of proposed method

步骤 2：根据(3)式更新 $\boldsymbol{D}$；

步骤 3：根据(7)式更新 $\boldsymbol{\Omega}_i$；

步骤 4：根据(8)和(9)式分别更新 $\boldsymbol{Z}^s$、$\boldsymbol{Z}^u$；

步骤 5：根据(10)式更新 $\boldsymbol{P}^s=(\boldsymbol{D}^T\boldsymbol{D}+I)^{-1}\times(\boldsymbol{D}^T\boldsymbol{Z}^s+\boldsymbol{X}^s\boldsymbol{H}_s)$；

步骤 6：根据(11)式更新 $\boldsymbol{P}^u=(\boldsymbol{D}^T\boldsymbol{D})^{-1}\boldsymbol{D}^T\boldsymbol{Z}^u$；

步骤 7：判断是否达到最大循环次数 Iter_$N$。若是，则执行步骤 8；若否，则循环执行步骤 2～7；

步骤 8：在图像特征空间中，以 $\boldsymbol{P}^u$ 作为最近邻分类器的中心，推断 unseen 类场景 $\boldsymbol{X}^u$ 的类别标签。

## 4　实验及结果分析

### 4.1　数据集及实验设置

实验采用 3 种遥感场景数据集：UC-Merced (UCM)数据集[18]、航空图像数据集(AID)[19]以及 RSSCN7 数据集[20]。其中，UCM 和 AID 用于定量实验，RSSCN7 用于定性实验，即作为 seen 样本，以测试遥感图像上 unseen 类场景的 ZSC 效果。UCM 有 21 类场景，共 2100 张图像，图像大小为 256 pixel×256 pixel，若干样本如图 3 所示；AID 共有 30 类，共 10000 张场景图像，图像大小为 600 pixel× 600 pixel，若干样本如图 4 所示。RSSCN7 共 2800 张遥感场景图像，分为 7 个类别，图像大小为 400 pixel×400 pixel，其样本如图 5 所示。

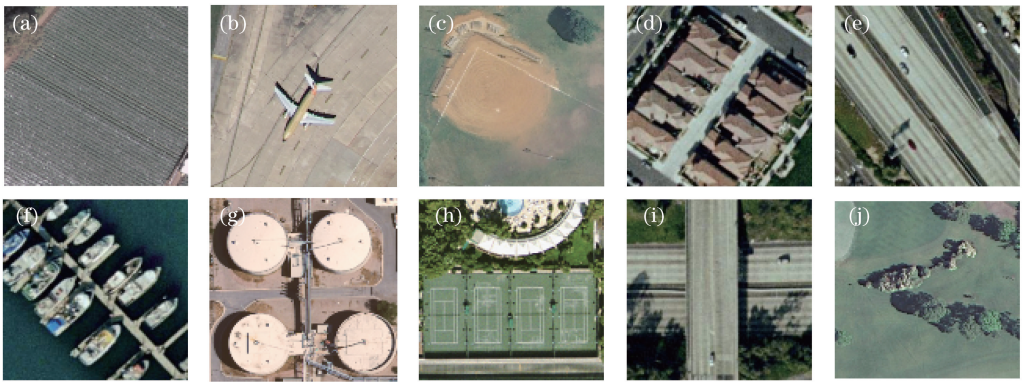


图 3　UCM 数据集若干类的样本。(a)农田；(b)飞机；(c)棒球场；(d)密集住宅；(e)高速公路；(f)海港；(g)储罐；(h)网球场；(i)立交桥；(j)高尔夫球场

Fig. 3　Images of several classes from UCM dataset. (a) Agricultural; (b) airplane; (c) baseball diamond; (d) dense residential; (e) freeway; (f) harbor; (g) storage tanks; (h) tennis court; (i) overpass; (j) golf course

实验采用卷积网络模型 GoogLeNet[21]的全连接层输出作为场景图像特征。词向量融合分为不同训练模型、不同语料词向量融合。其中，不同训练模型的词向量融合实验，涉及 2 种训练模型：Glove (gl)和 Word2Vec (wv)。这两种词向量均在 Wikipedia 语料上训练得到。不同语料词向量融合实验，采用 2 种训练语料：Wikipedia (Wiki)和 Common Crawl(Crawl)，均采用 Glove 模型训练。Iter_$N$ 为 40。定量实验采用总体分类准确度(OA，$x_{OA}$)作为评价指标，$x_{OA}=T_u/N_u$，$N_u$ 为全体 unseen 样本个数，$T_u$ 为正确分类的 unseen 样本个数。UCM 和 AID 分别采用 16/5 和 25/5 的 seen/unseen 类划分。根据实验运行效果，将稀疏比例 $T_0$ 设置为 10%(即稀疏化时保留前 10%最大的元

图 4　AID 数据集若干类的样本。(a)机场;(b)贫瘠地;(c)海滩;(d)桥梁;(e)商业区;(f)运动场;(g)池塘;(h)火车站;(i)体育场;(j)立交桥

Fig. 4　Images of several classes from AID dataset. (a) Airport; (b) bare land; (c) beach; (d) bridge; (e) commercial; (f) playground; (g) pond; (h) railway station; (i) stadium; (j) viaduct

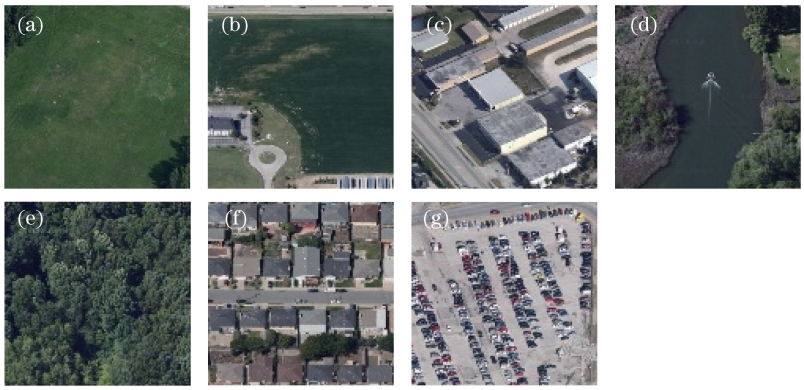

图 5　RSSCN7 数据集类的样本。(a)草地;(b)河湖;(c)工厂;(d)场地;(e)森林;(f)居民区;(g)停车场

Fig. 5　Images of several classes from RSSCN7 dataset. (a) Grass; (b) river laker; (c) industrial; (d) field; (e) forest; (f) residential; (g) parking

素值,其余元素值置 0),超参数 $\alpha$ 取值范围设置为 $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10\}$,超参数 $\eta$ 设置为 0.01。

### 4.2　定量实验结果及分析

在 UCM 和 AID 数据集上进行定量实验,并从结构对齐、超参数取值、融合效果以及与典型 ZSC 方法对比等 4 个方面,分别进行分析。

#### 4.2.1　结构对齐效果分析

ZSC 方法的本质是借助语义词向量提供的类间距离关系,将图像特征空间中类别原型,迁移至 unseen 类,获得 unseen 类的图像特征空间原型表示,最后利用该原型对 unseen 样本进行分类。而本文结构对齐项的实质作用就是降低两种空间类别间距离的不一致性。因此,这里定义语义词向量空间与图像特征空间的类别距离结构差异度为

$$D_{\mathrm{M}}=\left\{\sum_{i,j}\left[d(c_i,c_j)-d(p_i,p_j)\right]^2\right\}^{1/2},\quad(12)$$

式中:$d(c_i,c_j)$ 表示第 $i$、$j$ 类别词向量 $c_i$ 和 $c_j$ 的余弦距离;$d(p_i,p_j)$ 表示第 $i$、$j$ 类别图像特征空间类原型 $p_i$ 和 $p_j$ 的余弦距离。$D_{\mathrm{M}}$ 越大表示两个空间的类间距离结构越不一致,越小则表示越一致。图 6 和 7 分别为不同训练模型、不同训练语料词向量融合前后的 $D_{\mathrm{M}}$ 变化情况。符号⊕表示经

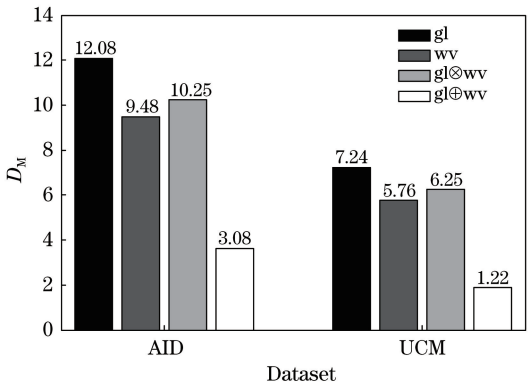


图 6　不同模型词向量融合的结构对齐效果

Fig. 6　Structure alignment performance of word vector fusion with different models

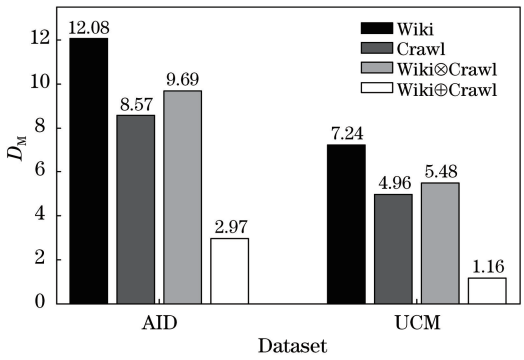


图 7　不同语料词向量融合的结构对齐效果图

Fig. 7　Structure alignment performance of word vector fusion with different corpora

过本文方法($M=2$)融合，符号⊗表示直接串接的词向量。

可以看出，相比未融合的单词向量及直接串接词向量，本文融合方法得到的词向量具有最小的$D_M$值，表明结构对齐效果显著优于直接串接以及未融合的词向量。这主要因为基于 ADL 的结构对齐项能够对融合词向量空间与图像特征空间之间的嵌入关系建模，从而得到与图像特征空间中的类间距离结构更一致的融合词向量。

### 4.2.2　超参数取值分析

本文方法中的超参数 $\alpha$ 取不同值会影响方法的分类效果，为选取最佳的 $\alpha$ 值，分别在不同的 $\alpha$ 取值上进行实验，比较获得的 OA 值，确定最佳超参数，在 UCM 和 AID 数据集上的运行情况如图 8 所示。可以看出，在全体取值范围内，融合词向量下的 OA 值均高于未融合词向量的 OA 值。

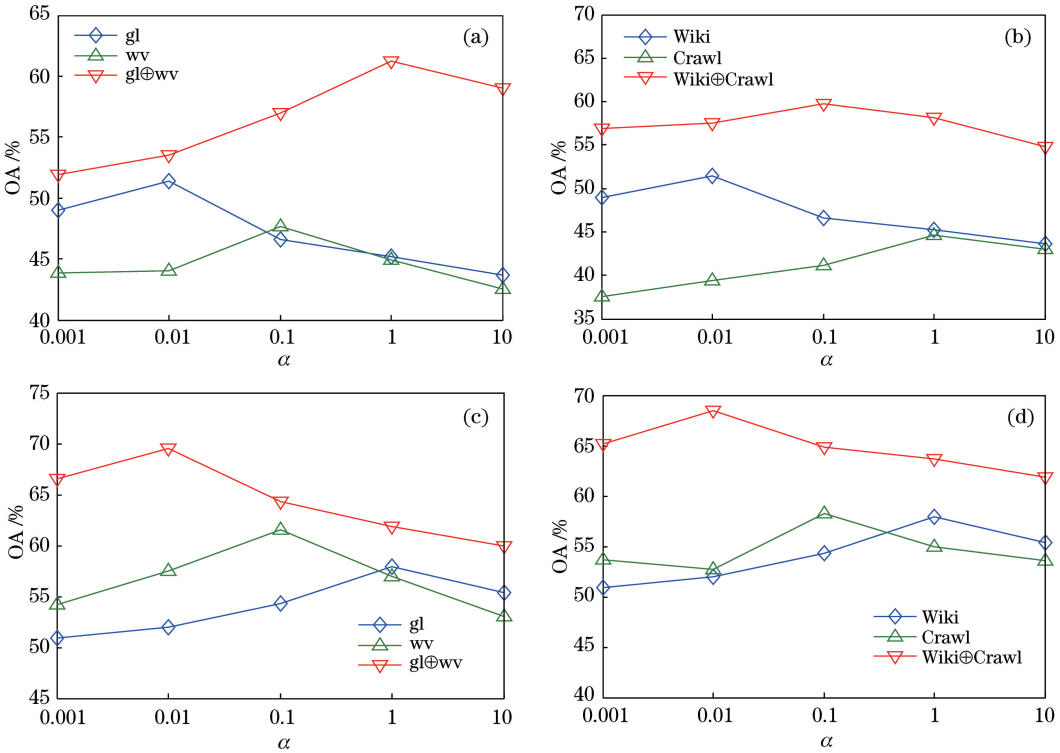


图 8　UCM 和 AID 数据集上本文方法在不同 $\alpha$ 值上的 OA 值。(a) UCM 上不同训练模型词向量融合；(b) UCM 上不同训练语料词向量融合；(c) AID 上不同训练模型词向量融合；(d) AID 上不同训练语料词向量融合

Fig. 8　OA values of proposed method for different $\alpha$ on UCM and AID datasets. (a) Fusion of word vectors from different training models on UCM dataset; (b) fusion of word vectors from different training corpora on UCM dataset; (c) fusion of word vectors from different training models on AID dataset; (d) fusion of word vectors from different training corpora on AID dataset

### 4.2.3　词向量融合效果分析

表 1 为在 UCM 和 AID 上，不同训练模型词向量和不同训练语料词向量的融合前后的 OA 值。其中，训练模型 gl 与 wv 的融合词向量，在 UCM 的 OA 为 61.23%，比未融合的 gl、wv 词向量分别提升了 9.84%和 13.58%；在 AID 的 OA 为 69.47%，比未融合的 gl、wv 词向量分别提升了 11.55%和 7.94%。训练语料 Wiki 与 Crawl 的融合词向量，在 UCM 的 OA 为 59.77%，比未融合的 Wiki、Crawl 词向量分别提升了 8.38%和 15.16%；在 AID 的 OA 为 68.49%，比未融合的 Wiki、Crawl 词向量分别提升了 10.57%和 10.20%。可以看出，融合后的

词向量在2种数据集上的OA值均得到显著提升。图9为在UCM和AID上，不同训练模型词向量和不同训练语料词向量的融合前后的各unseen类的分类准确度。可以看到，融合后的各unseen类的ZSC分类准确度比融合前均有明显提升。结果表明，本文方法能够适应不同unseen类的情形，通过融合不同语义词向量，利用它们间的一致性，显著提升OA值及各场景类别的分类准确度。

表1　不同训练模型词向量和不同训练语料词向量融合前后的OA

Table 1　OA values of different training models and different training corpora before and after fusion of word vectors　%

| Dataset | Fusion of word vectors from different models | | | Fusion of word vectors from different corpora | | |
|---|---|---|---|---|---|---|
| | gl | wv | gl⊕wv | Wiki | Crawl | Wiki⊕Crawl |
| UCM | 51.39 | 47.65 | **61.23** | 51.39 | 42.86/44.61 | **59.77** |
| AID | 57.92 | 61.53 | **69.47** | 57.92 | 56.16/58.29 | **68.49** |

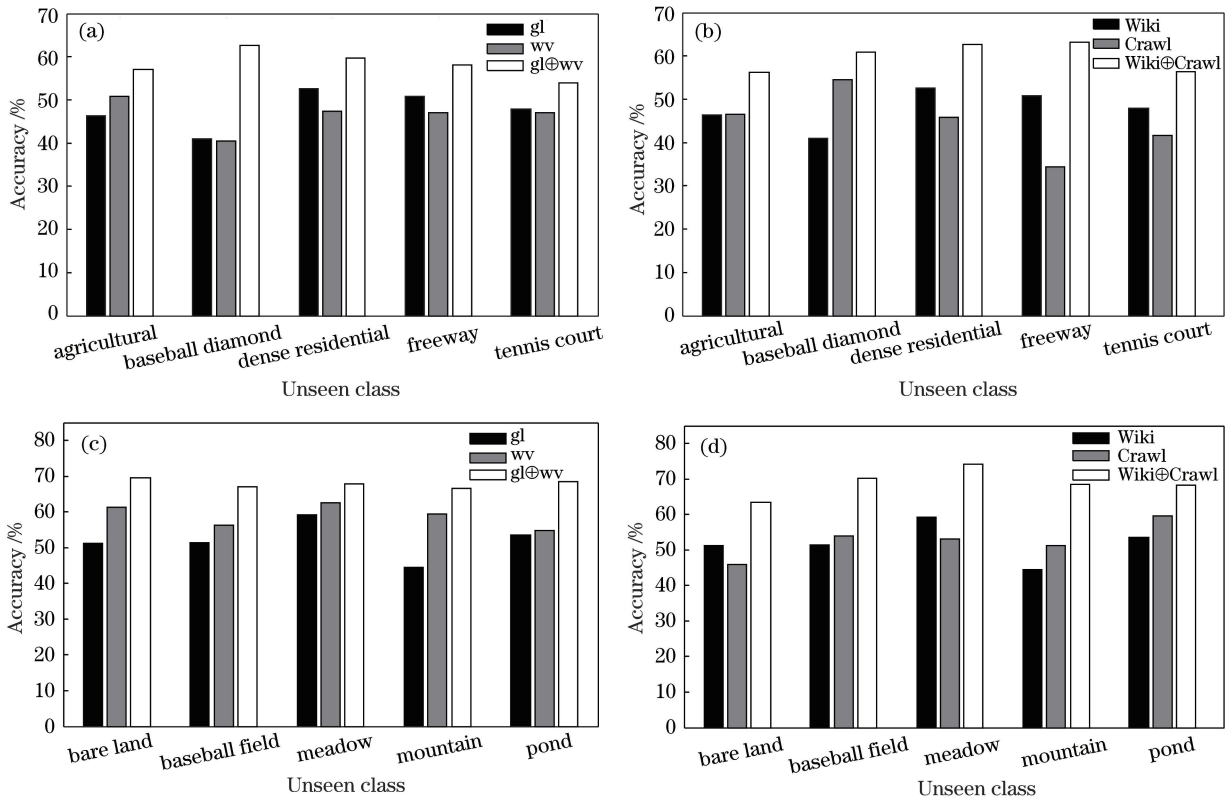


图9　不同训练模型词向量和不同训练语料词向量的各unseen类融合效果。(a) UCM上不同训练模型词向量融合；(b) UCM上不同训练语料词向量融合；(c) AID上不同训练模型词向量融合；(d) AID上不同训练语料词向量融合

Fig. 9　Fusion performance of different training models and different training corpora on unseen classes. (a) Fusion of word vectors from different training models on UCM dataset; (b) fusion of word vectors from different training corpora on UCM dataset; (c) fusion of word vectors from different training models on AID dataset; (d) fusion of word vectors from different training corpora on AID dataset

#### 4.2.4　与典型ZSC方法比较

通过与6种典型ZSC方法进行对比，验证本文方法是否具有更优的ZSC效果。表2中涉及3种语义词向量融合，其中S1为Glove模型在Common Crawl语料上训练的词向量，S2为Word2Vector模型在Wikipedia语料上训练的词向量，S3为Glove模型在Wikipedia语料上训练的词向量。“＋”符号在对比典型ZSC方法中代表词向量串接操作。相比典型ZSC方法，本文方法在数据集UCM和AID上均获得了最高OA值。其中S1＋S2＋S3融合词向量在UCM和AID上获得了最高分类OA值68.56%和76.85%，显著优于对比的典型ZSC方法。在UCM上，S1＋S2＋S3的OA值分别超过S1＋S2、S2＋S3的OA值7.40%、7.33%，

而超过未参与融合的 S3 的 OA 值17.17%；在 AID 上，S1＋S2＋S3 的 OA 值分别超过 S1＋S2、S2＋S3 的 OA 值 6.41%、7.38%，而超过未参与融合的 S3 的 OA 值 18.93%。典型 ZSC 方法中 RKT 表现较好，但在不同词向量下的 OA 值仍低于本文方法，主要原因是：1）RKT 方法没有考虑语义词向量空间与场景图像特征空间的类间距离结构差异，而本文方法通过结构对齐项有效减轻了这种距离结构差异性，提升了到 unseen 类的迁移效果；2）与其他典型 ZSC 方法相似，RKT 方法仅针对单一语义词向量情形，没有考虑多词向量的融合问题，而本文方法基于 ADL 融合不同词向量，通过利用不同词向量之间的一致性，有效提升了 ZSC 效果。由于目前可获取的词向量种类有限，未来随着词向量种类越来越多，可以采用本文方法进行更多种词向量的融合，比如定义 S4 为 World2Vector 模型在 Common Crawl 语料上训练的词向量，由于本文方法对词向量种类没有限制，可对 S1＋S2＋S3＋S4 进行融合，从而获得更高的 ZSC 准确度 OA 值。

表 2　本文方法及对比方法 OA 值

Table 2　OA values of proposed method and relative methods　　%

| Method | UCM | | | | | | AID | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | S1 | S2 | S3 | S1＋S2 | S2＋S3 | S1＋S2＋S3 | S1 | S2 | S3 | S1＋S2 | S2＋S3 | S1＋S2＋S3 |
| LatEm[4] | 18.80 | 20.40 | 19.80 | 33.00 | 23.00 | 20.80 | 15.90 | 22.65 | 23.81 | 18.71 | 28.17 | 21.62 |
| RKT[5] | 40.00 | 39.80 | 44.60 | 40.20 | 43.60 | 43.60 | 48.92 | 48.03 | 48.15 | 48.92 | 50.13 | 53.25 |
| DMaP[9] | 38.20 | 39.60 | 41.60 | 40.80 | 42.00 | 40.20 | 39.24 | 43.44 | 38.54 | 46.67 | 45.22 | 44.97 |
| BiDiLEL[8] | 28.51 | 33.48 | 39.20 | 40.40 | 40.00 | 41.00 | 32.91 | 42.55 | 32.40 | 47.85 | 50.44 | 49.63 |
| JLSE[6] | 37.25 | 34.21 | 45.68 | 37.66 | 34.88 | 38.03 | 36.11 | 34.97 | 42.30 | 35.99 | 43.50 | 45.54 |
| SSE[7] | 38.36 | 39.48 | 37.91 | 38.72 | 39.19 | 38.23 | 38.24 | 37.16 | 39.56 | 34.53 | 40.92 | 43.56 |
| **Proposed** | 44.61 | **47.65** | **51.39** | **61.16** | **61.23** | **68.56** | **58.29** | **61.53** | **57.92** | **70.44** | **69.47** | **76.85** |

图 10 为本文方法及对比典型 ZSC 方法在 UCM 和 AID 数据集上 S1＋S2＋S3 的各个 unseen 类别的分类准确度。可以看出，本文方法在各个 unseen 类别上的分类准确度均优于对比的 ZSC 方法，尤其是优于 LatEm 方法。由此可知，本文方法不仅在 OA 值上优于对比 ZSC 方法，而且在每个 unseen 场景类别上的准确度上同样优于对比方法，进一步证明本文方法的实际效果。

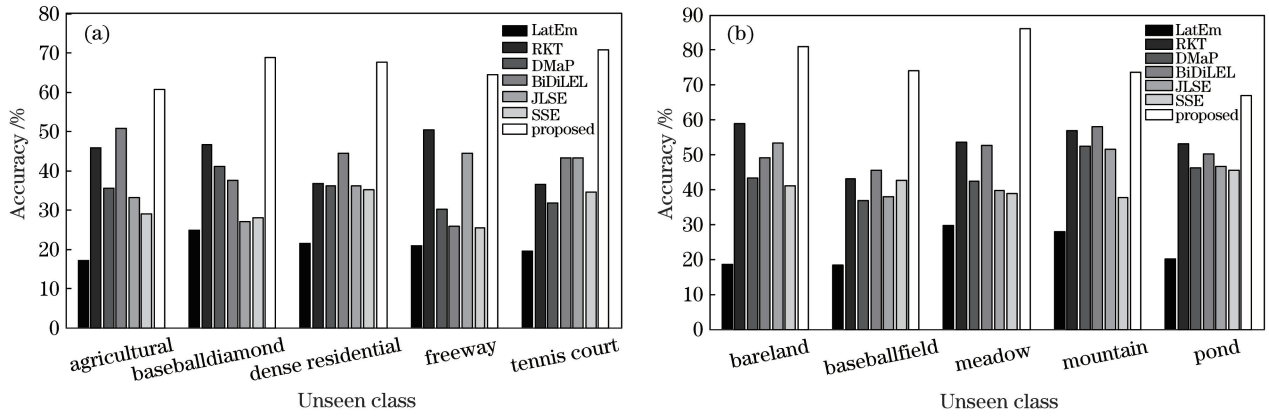


图 10　本文方法及对比方法的各 unseen 类 S1＋S2＋S3 词向量融合效果。(a) UCM 数据集；(b) AID 数据集

Fig. 10　Fusion performance of S1＋S2＋S3 word vectors on unseen classes by proposed method and relative methods. (a) UCM dataset; (b) AID dataset

4.2.5　计算效率分析

为比较本文方法与其他 ZSC 方法的计算效率，测试各 ZSC 方法在 AID 数据集上对 S1 词向量上的计算耗时，结果如表 3 所示。可以看出：DMaP 方法耗时最长，为 409.26 s，其次是 JLSE 方法，耗时为 70.20 s，而本文方法耗时最短，为 17.90 s。这主要因为 ADL 算法的时间复杂度低，使本文方法的运算效率优于对比的典型 ZSC 方法。

表 3　各 ZSC 算法在 AID 数据集上对 S1 词向量上的运算耗时

Table 3　Computing time of different ZSC algorithms on AID dataset with S1 word vector

| Method | Time/s |
|---|---|
| LatEm[4] | 21.66 |
| RKT[5] | 24.24 |
| DMaP[9] | 409.26 |
| BiDiLEL[8] | 28.81 |
| JLSE[6] | 70.20 |
| SSE[7] | 19.74 |
| **Proposed** | **17.90** |

### 4.3　定性实验结果及分析

为定性分析本文方法的实际遥感场景 ZSC 效果，以 RSSCN7 数据集作为 seen 类样本，对 2 幅高分辨率遥感图像 I 和 II（空间分辨率均为 0.3 m）进行 ZSC 分类。unseen 类选择为 ocean、airport 和 runway。S1＋S2＋S3 得到的词向量，用于定性实验。步骤为：首先，用单类别支持向量机（SVM）判断遥感场景样本是否属于 seen 类；然后，对不属于 seen 类的样本，视为 unseen 类样本，采用本文方法进行 ZSC。遥感图像 I 的尺寸为 17920 pixel×10752 pixel，场景尺寸设定为 256 pixel×256 pixel。本文及对比方法在遥感图像 I 上的 ZSC 效果，如图 11 所示。可以看出，本文方法对 unseen 类场景的分类效果优于对比方法，其中 airport 类的场景分类效果更明显。本文及对比的典型 ZSC 方法对于 ocean 类场景均具有良好的识别效果，但对于 airport 类场景的识别效果差异较大。其中对 airport 类场景识别效果最差的方法是 LatEm，可以看出，LatEm 将 airport 类场景误分为 ocean 类场景，其余方法的识别效果优于 LatEm，但是均不如本文方法。

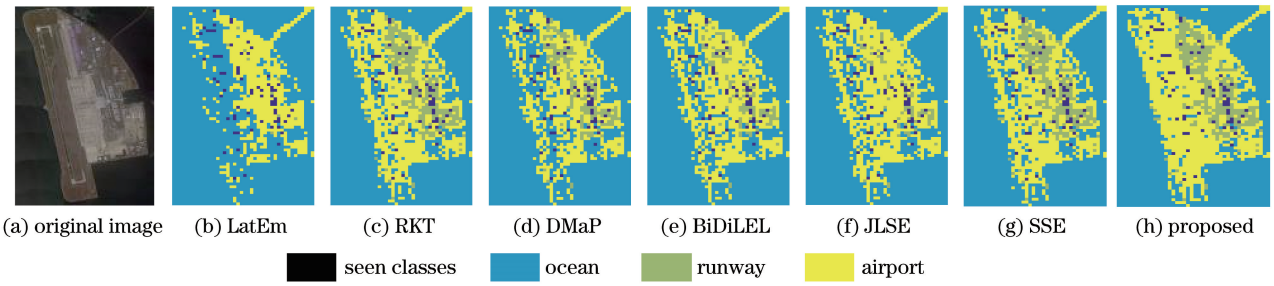


图 11　测试遥感图像 I 的场景 ZSC 效果图

Fig. 11　Scene ZSC results of test remote-sensing image I

遥感图像 II 的尺寸为 25344 pixel×29952 pixel，场景尺寸设定为 256 pixel×256 pixel。本文及对比方法在遥感图像 II 上的 ZSC 效果，如图 12 所示。可以看出，本文词向量融合方法的 ZSC 效果，总体优于对比典型 ZSC 方法，其中 ocean 类场景的分类效果尤其明显。遥感图像 II 的场景组成较遥感图像 I 更复杂，尤其是陆地场景的地物组成种类繁多。与遥感图像 I 的 ZSC 效果不同，不同方法对 ocean 类场景的识别效果差异较大。其中 LatEm 方法将许多陆地场景误分为 ocean 类场景，RKT 等方法对 ocean 类场景识别出现了部分误分现象，只有本文及 SSE 方法对 ocean 类场景的识别效果最佳，但是 SSE 方法对 airport 类的识别效果不如本文方法，因此整体来说本文方法的 ZSC 效果最佳。综合上述定性实验结果可知，本文通过词向量融合的方法能够获得优于对比典型 ZSC 方法的 ZSC 效果。总体而言，本文方法对 ocean 和 airport 类别场景的识别效果优于 runway 的识别效果，主要原因是 ocean 和 airport 类别场景构成比 runway 场景更简单（仅单一的海水和机场水泥地面），而 runway 场景类别组成复杂（包括草地、水泥地面和标识符等）。因此，本文方法对构成简单的场景效果优于构成复杂的场景。

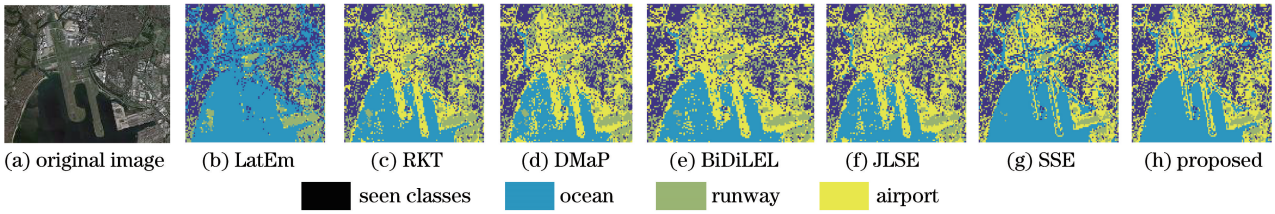


图 12　测试遥感图像 II 的场景 ZSC 效果图

Fig. 12　Scene ZSC results of test remote-sensing image II

## 5 结　论

针对遥感场景类别的语义词向量与图像特征原型的距离结构不一致问题，提出了面向遥感场景ZSC的词向量融合方法，通过定量和定性实验，验证了该方法在不同训练模型、不同训练语料词向量融合的有效性。该方法有以下特点：1)为利用不同词向量一致性，利用解析字典学习方法提取各词向量的公共稀疏编码系数，并作为融合后的词向量；2)为降低结构差异性，将遥感场景图像特征类原型嵌入到融合词向量空间中与其进行对齐。实验结果表明：与典型ZSC方法相比，本文方法在缩小距离结构差异、提升总体分类准确度方面都有更优表现，本文方法能够有效利用不同词向量的一致性，显著提升遥感场景ZSC效果。

## 参　考　文　献


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

[2] 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.

[3] 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.

[4] 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 (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 69-77.

[5] Wang D, Li Y N, Lin Y T, *et al*. Relational knowledge transfer for zero-shot learning[C]//Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA. California: AAAI, 2016: 2145-2151.

[6] Zhang Z M, Saligrama V. Zero-shot learning via joint latent similarity embedding[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 6034-6042.

[7] 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, USA. New York: IEEE, 2015: 4166-4174.

[8] Wang Q, Chen K. Zero-shot visual recognition via bidirectional latent embedding[J]. International Journal of Computer Vision, 2017, 124(3): 356-383.

[9] 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 (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 3279-3287.

[10] Zhao B, Wu B T, Wu T F, *et al*. Zero-shot learning posed as a missing data problem[C]//2017 IEEE International Conference on Computer Vision Workshops (ICCVW), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2616-2622.

[11] Lampert C H, Nickisch H, Harmeling S. Attribute-based classification for zero-shot visual object categorization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(3): 453-465.

[12] Socher R, Ganjoo M, Manning C D, *et al*. Zero-shot learning through cross-modal transfer[C]//26th International Conference on Neural Information Processing Systems, December 5-10, 2013, Lake Tahoe, Nevada. [S.l.: s.n.], 2013: 935-943.

[13] Mikolov T, Chen K, Corrado G, *et al*. Efficient estimation of word representations in vector space[J/OL]. (2013-09-07)[2019-03-01]. https://arxiv.org/abs/1301.3781.

[14] Pennington J, Socher R, Manning C. Glove: global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, Doha, Qatar. [S.l.: s.n.], 2014: 1532-1543.

[15] Yang M, Chang H Y, Luo W X. Discriminative analysis-synthesis dictionary learning for image classification[J]. Neurocomputing, 2017, 219: 404-411.

[16] 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.

[17] 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.

[18] 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-GIS′ 10, November 2-5, 2010, San Jose, California, USA. New York: IEEE, 2010: 270-279.

[19] 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.

[20] 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.

[21] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J/OL]. (2015-04-10)[2019-03-01]. https://arxiv.org/abs/1409.1556.