

## 基于零样本学习的高光谱图像分类

潘尔婷, 马泳, 黄珺\*, 樊凡, 李嶷, 梅晓光, 马佳义

(武汉大学 电子信息学院, 武汉 430072)

**摘要:** 现有高光谱分类模型的发展大多集中于有监督学习, 大多数方法都在单个高光谱数据立方中进行训练和评估。而高光谱数据集标注困难, 缺少有标签样本, 这限制了分类模型的进一步发展。零样本学习旨在仅用可见类别的信息来预测未知类别, 为高光谱图像的跨数据集训练和测试提供了可能性。受零样本学习的最新进展启发, 本文将其引入高光谱图像分类, 选择了两种不同的视觉特征提取模型, 并引入了在外部语料库上训练的语义特征提取模型中获取的标签语义描述, 然后比较分析了零样本学习的两种映射方式, 包括在语义-视觉特征映射和视觉-语义特征映射。本文在高光谱图像分类任务中复现和微调了五种经典的零样本学习模型。实验结果表明零样本学习在高光谱图像分类任务中具有一定的发展潜力。

**关键词:** 高光谱图像; 图像分类; 特征提取; 零样本学习;

## Hyperspectral Image Classification Based on Zero-shot Learning

PAN Erting, MA Yong, HUANG Jun\*, FAN Fan, LI Hao, MEI Xiaoguang, MA Jiayi

(Wuhan University, Wuhan 430072, China)

**Abstract:** The development of existing hyperspectral classification models mainly focus on the supervised learning and most of methods are trained and evaluated on a single hyperspectral data cube. However, labeling hyperspectral data is laborious and there is a lack of labeled samples. It limits the further development of hyperspectral image classification. Zero-shot learning aims to only use the information of seen categories to predict unseen categories, which provides a possibility of training and testing the hyperspectral classification models across datasets. Inspired by the latest progress of zero-shot learning, we introduce it into hyperspectral image classification. In this work, two different visual feature extraction models and a semantic feature extraction model trained on an external corpus are selected for feature extraction. Then, two feature mapping styles of zero-shot learning including semantic-visual feature mapping and visual-semantic feature mapping are compared and analyzed. Five classic zero-shot learning models are re-implemented and fine-tuned for the hyperspectral image classification task. The experimental results show that zero-shot learning has great potential for development in hyperspectral image classification tasks.

**Key words:** Hyperspectral Image; Image Classification; Feature Extraction; Zero-shot Learning

## 0 引言

高光谱图像相比传统遥感图像具有更高的光谱分辨率, 通常以数据立方体的形式呈现。丰富的空间和光谱信息极大地增强了图像的感知能力, 使高光谱

遥感技术广泛应用于各种领域, 例如环境监测, 资源勘探, 军事侦察和灾难评估等<sup>[1-3]</sup>。

高光谱图像分类是指根据图像中的光谱和空间信息对每个像素点进行类别标记, 是高光谱数据分析

中最为关键的技术之一[4, 5]。与传统高光谱图像分类方法不同,深度学习方法凭借其具有可挖掘高光谱图像深层次特征的优势,已经成为高光谱图像分类的有力工具。但是,由于目前高光谱图像的公开数据集少,有标注样本少并且人工标注成本昂贵,几乎所有现有的高光谱图像分类方法都在单个数据集上进行训练和测试[6-8]。

现有分类模型的发展大多集中于有监督学习,对有标注数据的数量和质量有越来越高的要求。这也激发了小样本学习甚至零样本学习技术的发展[9]。零样本学习是迁移学习的一种,指依据现有类别的信息推断未见过的类别,这里的“零”指的是测试的样本类别在训练集中没有出现过。为了推断未见过的类别信息,零样本学习引入了自然语言处理中的语义信息来生成对类别的高维语义描述,通过在已知类别中建立视觉特征和语义特征之间的联系来实现新的类别的识别。

将零样本学习引入高光谱图像分类中,可以实现跨数据集对分类模型进行训练和测试,有效降低算法对数据集的依赖和标注的压力。本文的工作中结合了两种视觉特征提取算法,引入了源自在外部语料库中训练的自然语言处理模型,在高光谱图像分类任务中复现了五种经典的零样本学习算法。实验结果表明,零样本学习在该任务中有一定的应用潜力,跨数据集进行训练和测试的模式可以成为未来高光谱分类的发展趋势之一。

## 1 算法

零样本学习的实现有三个步骤,首先是特征提取,包括高光谱数据的视觉特征提取和类别的语义特征提取;第二步是特征映射,用于在训练集中建立视觉特征和语义特征之间的联系;第三步是标签推理,即将训练得到的映射关系应用于测试集数据中,推理得到最终标签。

### 1.1 特征提取

如图 1 所示,在特征提取步骤中,图像特征提取模型  $\phi(\cdot)$  将训练集和测试集的数据  $x^{tr}, x^{te}$  编码至视觉特征空间,得到对应的视觉特征  $\phi(x^{tr}), \phi(x^{te})$ 。而类别语义描述模型  $\psi(\cdot)$  则将训练集和测试集中的类别名称  $y^{tr}, y^{te}$  编码为类别语义特征  $\psi(y^{tr}), \psi(y^{te})$ 。

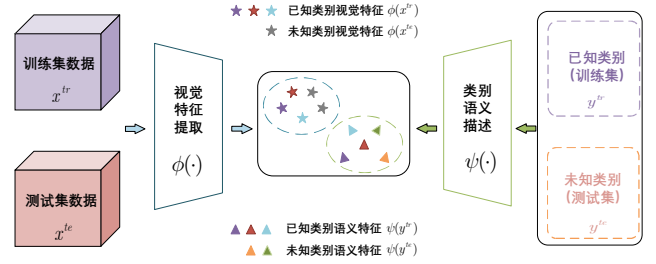


图 1 特征提取图示

Fig.1 Illustration of feature extraction.

图像视觉特征提取旨在将高光谱图像中的目标像素相关的光谱和空间信息编码至视觉特征空间中,为训练集和测试集图像提取鲁棒的高光谱视觉特征。在高光谱图像分类领域已经提出了多种性能优越的分类模型,可以灵活使用分类模型的中间层输出视觉特征。本文中选取了两个不同的视觉特征提取模型进行实验。

类别语义空间的构建可以将训练集和测试集中的类别映射到同一语义特征空间。现有的零样本学习方法中主要使用两种类型的类别语义描述,即属性和词向量。属性描述的是不同类别是否具有某一特性,通常是人工标注的,而高光谱数据中缺少这一信息来源。词向量的转换则是借助于在外部语料库中训练的自然语言处理模型。本文的选择使用词向量作为语义特征描述。该模型通常在一个大型无标注的文本语料库中预训练,通过学习单词之间的语义关联,将每个单词表示为一个固定维度的向量。为了得到高光谱类别的词向量,我们选用了在谷歌新闻数据集上预训练的 word2vec 模型[10],该模型中包含 300 万单词和短语的词向量,维度为 300 维。

### 1.2 特征映射

特征提取阶段得到的视觉特征和语义特征不同维度的,因此需要学习特征映射  $f$  将视觉特征和语义特征关联起来。

$$f = W[\phi(x^{tr}), \psi(y^{tr})] \quad (1)$$

在零样本学习中,根据嵌入空间的选择,特征映射有两种选择,如图 2 所示。一个是在语义-视觉特征映射 ( $V \rightarrow S$ ),指将语义特征投影至视觉特征空间,另一种是视觉-语义特征映射 ( $S \rightarrow V$ ),即将视觉特征投影至语义特征空间。目标函数的设计则是针对如何衡量映射后两部分特征的相似性。

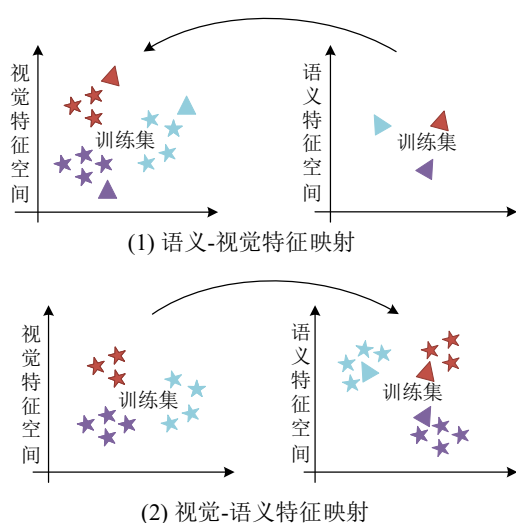


图 2 特征映射图示

Fig.2 Illustration of feature mapping.

语义-视觉特征映射方式的典型方法是 DeVise<sup>[11]</sup> 和 ALE (Attribute Label Embedding)<sup>[12]</sup>, 都选择了语义特征空间作为嵌入空间。DeVise 方法中设计了基于排名损失的双线性相容函数, 而 ALE 是学习的是一种基于铰链排名损失优化的映射。视觉-语义特征映射方式的代表方法有 DEM(Deep Embedding Model)<sup>[13]</sup>和 RN (Relation Network)<sup>[14]</sup>, 这两种方法都设计了一个子网络将视觉特征映射到语义空间。

## 1.3 标签推理

如图 3 所示, 在训练集数据上学习到特征映射后, 标签推理步骤即将该映射应用在测试集数据的视觉特征和类别语义特征中, 再通过度量两种特征的相似性来为测试集数据推理出对应的标签。

最常用的度量学习方法是使用最近邻分类器, 该方法依据两种特征的最近距离来进行分类。上文介绍的典型算法中, DeVise 和 ALE 的优化目标是训练过程中基于边界的排名损失函数。另外还有采用线性回归来获得投影矩阵  $W$  的 SAE (Semantic Auto-Encoder)<sup>[15]</sup>, 通过计算投影后两特征之间的余弦距离, 再使用  $K$  均值聚类来推理标签。

DEM 中的标签推理是通过简单的距离度量来实现的, 在视觉特征空间中使用  $K$  最近邻搜索, 将视觉特征与语义特征的投影进行匹配。与前面方法中的固定度量或浅层学习度量不同, RN 选择通过深度学习与嵌入相结合的非线性相似度量来识别视觉特征和

语义特征是否匹配, 回归得到相关程度的分数, 最大得分为 1, 不匹配则得分为 0。

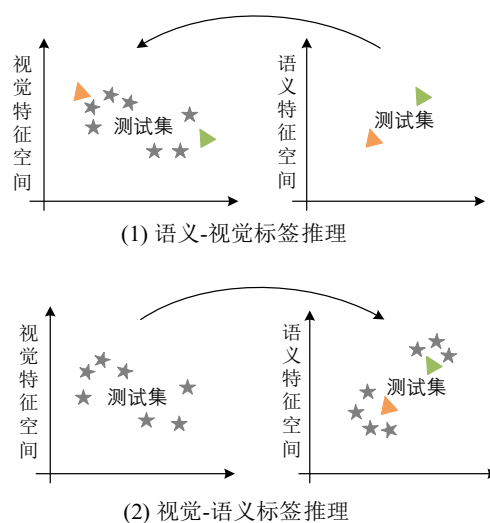


图 3 标签推理图示

Fig.3 Illustration of label prediction.

## 2 实验

### 2.1 实验数据



图 4 Salinas 数据集的伪彩图和参考标签图

Fig.4 The false color image and the corresponding ground-truth map of Salinas dataset.

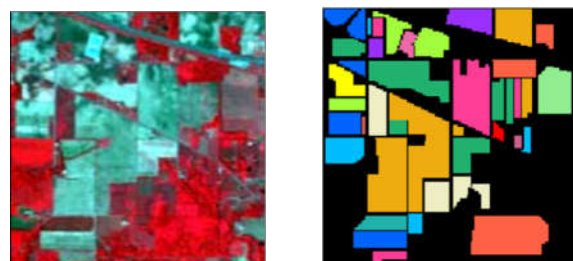


图 5 Indian Pines 数据集的伪彩图和参考标签图

Fig.5 The false color image and the corresponding ground-truth map of Indian Pines dataset.

# 第五届高光谱技术及应用研讨会 会议论文集

表 1 SA 和 IP 数据集的类别名称及有标签样本数量表

Table1 The categories and the corresponding number of labeled samples of Salinas and Indian Pines datasets.

编号	SA		IP	
	名称	数量	名称	数量
1	Brocoli_green_weeds_1	2009	Alfalfa	46
2	Brocoli_gree_weeds_2	3726	Corn-notill	1428
3	Fallow	1976	Corn-mintill	830
4	Fallow_rough_plow	1394	Corn	237
5	Fallow_smooth	2678	Grass-pasture	483
6	Stubble	3959	Grass-trees	730
7	Celery	3579	Grass-pasture-mowed	28
8	Grapes_unstrained	11271	Hay-windrowed	478
9	Soli_vinyard_develop	6203	Oats	20
10	Corn_seneseed_green_weeds	3278	Soybean-notill	972
11	Lettuce_romaine_4wk	1068	Soybean-mintill	2455
12	Lettuce_romaine_5wk	1927	Soybean-clean	593
13	Lettuce_romaine_6wk	916	Wheats	205
14	Lettuce_romaine_7wk	1070	Woods	1265
15	Vinyard_untrained	7268	Building-Grass-Trees-Drives	386
16	Vinyard_vertical_trellis	1807	Stone-Steel-Towers	93

本文选择了 Salinas 数据集(SA)和 Indian Pines 数据集(IP)进行对比实验,其伪彩图和对应的参考标签图如图 4 和图 5 所示。选择这两个公开数据集有以下原因:(1) SA 和 IP 数据集都是由 AVIRIS 传感器采集的,光谱波段范围为 $400 - 2500nm$ ;(2) 两数据集均采集自农作物区域,包含的是不同的农作物类别。各类别名称和对应的有标签样本数目如表 1 所示。另外,Salinas 数据集大小为 $512 \times 217 \times 204$ ,Indian Pines 数据集大小为 $145 \times 145 \times 200$ 。

## 2.2 实验设置

本文实验进行跨数据集的训练和测试,设计了两组实验,一是以 SA 为训练集,IP 为测试集,记为 SA-IP,二是以 IP 为训练集,SA 为测试集,记为 IP-SA。如表 1 所示,IP 中有几个类别有标签样本过少,本文中暂不考虑样本数少于 200 的类别,因此参与训练和测试的是 IP 中的 12 个类别和 SA 中的 16 个类别。

本文实验选择 top-k 的准确率作为性能评价指标。

其他实验设置细节如下:

(1) 视觉特征提取模型:本文中选择了两种视觉特征提取模型进行实验,一个是使用光谱特征的 RNN 模型<sup>[8]</sup>,提取的视觉特征维度为 128 维,一个是使用光谱-空间特征的 SSUN 模型<sup>[16]</sup>,提取的视觉特征维度为 512 维。

(2) 语义空间构建:高光谱数据集中很多类别都是由多个单词组成的,我们加载的是每个单词对应的词向量,然后求平均得到该类别对应的词向量。但不是所有的类别名称都能在语料库中找到对应的单词,因此我们对某些类别的名称进行了修正。比如“Lettuce\_romaine\_4wk”中的“4wk”表示的是生长周期,但是这并不是一个正常的单词,因此我们将该类别修正为“Lettuce\_romain\_4”。

(3) 特征映射:实验中所涉及各个对比算法的相关参数设置如表 2 所示。

表 2 各模型相关参数设置表

Table 2 Parameter configuration of comparative methods

模型	参数
DeVise <sup>[11]</sup>	Margin=0.1; r = 0.0001
WLE <sup>[12]</sup>	r = 0.001
SAE <sup>[15]</sup>	$\lambda = 100000$ ; r = 0.0001
DEM <sup>[13]</sup>	$\lambda = 0.0001$ ; r = 0.0001
RN <sup>[14]</sup>	r = 0.0001

## 2.3 实验结果

各对比算法的实验结果如表 3 所示。SA-IP 组测试集有 12 个类别,IP-SA 组有 16 个类别。首先,由于这两组数据集中包含一些语义相近但光谱差距大的类别,如“Soybean-notill”,“Soybean-mintill”和“Soybean-clean”,还有“Brocoli\_green\_weeds\_1”和“Brocoli\_green\_weeds\_2”,等等。不同类别在语义描述上区分难度大,从而导致整体分类性能上表现不佳。

两种不同特征提取算法的对比表明,由于 SA 和 IP 数据集中大都是农作物类别,包含大量的同质区域,空间特征的利用有较大作用,联合使用光谱和空间特征有更好的分类性能。对比两种不同映射方式的实验结果,可以看出,语义-视觉特征映射(S→V)是更好的表现。在所有零样本学习的模型中,RN 这种使用子网络学习非线性映射的模型在分类精度上有更好的表现。



# 第五届高光谱技术及应用研讨会 会议论文集

表 3 各对比算法的实验结果表

Table 3 Experimental results of comparative methods.

精度	模型	映射 方向	SA-IP		IP-SA	
			RNN	SSUN	RNN	SSUN
Top1	DeVise	V→S	13.46	15.02	10.95	12.57
	WLE	V→S	8.57	10.89	6.82	7.33
	SAE	V→S	10.14	11.03	6.04	8.46
		S→V	11.03	12.09	7.92	7.29
	DEM	S→V	16.65	<b>20.87</b>	10.63	9.05
	RN	S→V	<b>19.87</b>	13.75	<b>13.26</b>	<b>14.11</b>
Top3	DeVise	V→S	28.36	31.21	24.52	26.83
	WLE	V→S	22.49	23.76	16.14	20.04
	SAE	V→S	23.57	26.88	15.37	23.15
		S→V	27.69	25.41	16.94	22.97
	DEM	S→V	28.35	<b>34.02</b>	23.85	26.08
	RN	S→V	<b>31.67</b>	28.36	<b>29.82</b>	<b>37.46</b>

## 3 结 论

针对高光谱图像标注数据困难且缺少有标注数据的难点,本文通过引入零样本学习,实现了跨数据集对高光谱图像分类算法的训练和评估。基于零样本学习的高光谱图像分类方法首先使用特征提取模型分别提取图像的视觉特征和标签的语义描述,然后通过特征映射学习视觉特征和语义特征之间的关联,最后将学习到的映射应用在测试集数据中推理得到标签。本文中使用了两种不同的视觉特征提取模型,引入了在外部语料库中学习的 word2vec 模型,并且复现了五种经典的零样本学习方法。本文使用同一传感器采集的两个数据集进行跨数据集的训练和评估。实验结果表明,视觉-映射的模式能更好地学习到两种特征之间的联系,零样本学习在高光谱图像分类任务中有一定的发展潜力。

## 参考文献:

- [1] 童庆禧, 张兵, 张立福. 中国高光谱遥感的前沿进展[J]. 遥感学报, 2016,20(05):689-707.
- [2] 张兵. 高光谱图像处理与信息提取前沿[J]. 遥感学报, 2016,20(05):1062-1090.
- [3] Li S, Song W, Fang L, et al. Deep Learning for Hyperspectral Image Classification: An Overview[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019,57; 57(9; 9):6690-6709.

- [4] Li C, Ma Y, Mei X, et al. Hyperspectral image classification with robust sparse representation[J]. IEEE Geoscience and Remote Sensing Letters, 2016,13(5):641-645.
- [5] Jiang J, Ma J, Wang Z, et al. Hyperspectral Image Classification in the Presence of Noisy Labels[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019,57(2):851-865.
- [6] Paoletti M E, Haut J M, Plaza J, et al. Deep learning classifiers for hyperspectral imaging: A review[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019,158:279-317.
- [7] Mei X, Pan E, Ma Y, et al. Spectral-Spatial Attention Networks for Hyperspectral Image Classification[J]. Remote Sensing, 2019,11(8):963.
- [8] Mou L, Ghamisi P, Zhu X X. Deep recurrent neural networks for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017,55(7):3639-3655.
- [9] Xian Y, Lampert C H, Schiele B, et al. Zero-Shot Learning—A Comprehensive Evaluation of the Good, the Bad and the Ugly[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019,41(9):2251-2265.
- [10] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality, 2013[C].
- [11] Frome A, Corrado G S, Shlens J, et al. Devise: A deep visual-semantic embedding model, 2013[C].
- [12] 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, 2013,36(3):453-465.
- [13] Zhang L, Xiang T, Gong S. Learning a deep embedding model for zero-shot learning, 2017[C].
- [14] Sung F, Yang Y, Zhang L, et al. Learning to compare: Relation network for few-shot learning, 2018[C].
- [15] Kodirov E, Xiang T, Gong S. Semantic Autoencoder for Zero-Shot Learning, 2017[C].July.
- [16] Xu Y, Zhang L, Du B, et al. Spectral-Spatial Unified Networks for Hyperspectral Image Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018,56(10):5893-5909.

**基金编号:** 国家自然科学基金(61903279); 湖北省自然科学基金(2019CFB162, 2018CFA006)

**作者简介:** 潘尔婷 (1996-), 女, 湖南宁乡人, 硕士, 高光谱图像分析, Email: panerting@whu.edu.cn

**导师简介:** 马泳(1971-), 男, 湖北襄阳人, 博士, 信号与信息处理, Email: mayong@whu.edu.cn