

# 深度学习的半监督遥感图像检索

张洪群<sup>1</sup>, 刘雪莹<sup>1,2</sup>, 杨森<sup>1,2</sup>, 李宇<sup>1</sup>

1. 中国科学院遥感与数字地球研究所, 北京 100094;

2. 中国科学院大学, 北京 100049

**摘要:** 遥感图像数据的海量性、多样性和复杂性等特点对遥感图像检索的速度和精度提出了更高的要求, 其中特征提取是影响遥感图像检索效果的关键。本文方法首先对遥感图像进行预处理, 然后基于稀疏自动编码的方法在大量未标注的遥感图像上进行特征学习得到特征字典, 基于卷积神经网络的思想, 使用训练出来的特征字典对遥感图像进行卷积和池化得到每幅图像的特征图; 接下来使用特征图训练Softmax分类器; 最后对待检索图像分类, 在同一类别中计算特征间的距离, 进而实现遥感图像的检索。实验结果表明, 该方法能够有效提高遥感图像检索的速度和准确度。

**关键词:** 遥感图像检索, 深度学习, 稀疏自动编码, 卷积神经网络, Softmax分类器

**中图分类号:** TP751 **文献标志码:** A

**引用格式:** 张洪群, 刘雪莹, 杨森, 李宇. 2017. 深度学习的半监督遥感图像检索. 遥感学报, 21(3): 406–414

Zhang H Q, Liu X Y, Yang S and Li Y. 2017. Retrieval of remote sensing images based on semisupervised deep learning. Journal of Remote Sensing, 21(3): 406–414 [DOI:10.11834/jrs.20176105]

## 1 引言

基于内容的遥感图像检索CBRSIR(Content-Based Remote Sensing Image Retrieval), 是基于内容的图像检索技术在遥感图像上的应用(李德仁和宁晓刚, 2006)。CBRSIR基于遥感图像本身的特征, 由系统代替用户进行遥感图像特征的描述、提取和分析, 是目前遥感图像检索技术研究的重要方向, 是数字地球等重大项目中解决信息检索难题的一项关键技术。

遥感图像特征提取是CBRSIR的关键环节, 它的好坏直接影响到系统的运行效率和检索效果。传统的特征提取主要依据遥感图像的底层视觉特征, 包括颜色、纹理、形状和空间信息等(刘米娜, 2013); Sheikholeslami等人(1999)用小波分析的方法提取图像的纹理特征, 用于地学图像检索; Kitamoto和Takagi(1994)用可变椭圆分解法按形状和空间分布将台风云图分解为不同层次的一

系列椭圆, 提出了一种基于层次属性关系图的云图检索方法; Zhu等人(2000)用Gabor滤波法提取图像的纹理特征, 用于航空遥感图像的检索; 陆丽珍等人(2004)提出了一种加权颜色和纹理特征距离的方法用于高分辨率卫星和航空遥感图像的检索。然而, 遥感图像存在多样性、复杂性和海量性等特点, 无论是使用单一特征还是综合特征, 图像底层特征和上层理解之间总是存在较大的差异, 即“语义鸿沟”(semantic gap), 这种现象是由于底层特征不能完全反映或者匹配查询意图所造成的。

为了解决“语义鸿沟”的问题, 人们尝试对图像底层特征进行学习分类以获得图像语义特征。Mojsilovic和Rogowitz(2001)通过多层聚类分析寻找图像底层视觉特征和图像语义之间的关系, 进行图像的分类和检索; Zhao等人(2014)等提出了一种基于多尺度的中心圆结构的BOVW模型, 并证明了这种方法在高分辨率遥感影像的分类上是有效的。然而这些获取图像语义特征的方法都具有明

收稿日期: 2016-04-21; 修订日期: 2016-09-18; 优先数字出版日期: 2016-09-25

基金项目: 国家自然科学基金(编号: 61501460)

第一作者简介: 张洪群(1971—), 男, 正高级工程师, 研究方向为遥感卫星地面系统相关技术。E-mail: zhanghq@radi.ac.cn

通信作者简介: 李宇(1986—), 女, 助理研究员, 研究方向为遥感图像处理。E-mail: liyu@radi.ac.cn

显的人工痕迹, 且严重依赖分类或聚类算法, 在实际应用中适应性较差。

近年来, 深度学习的方法(Hinton和Salakhutdinov, 2006; Lecun 等, 2015)在传统的语音识别、目标识别和探测、自然语言处理, 以及其他的一些应用领域都取得了很大的成功(Hu 等, 2015a)。深度学习思想的发展为图像特征的获取提供了新思路。在遥感图像大数据的背景下, 通过对大规模遥感图像数据进行训练, 更加复杂强大的深度学习模型能从根本上揭示海量遥感图像中潜在的复杂而又丰富的数据信息, 得到大量更具代表性的特征信息。深度学习是一种从遥感数据中提取表示的好方法, 用于CBRSIR中前景广阔。Mnih和Hintion(2010)等首次将深度学习运用于遥感领域, 采用深度信念网络DBN(Deep Belief Network)模型对机载遥感图像中的道路进行检测; Luus等人(2015)等采用深度卷积神经网络DCNN(Deep Convolutional Neural Network)进行土地利用分类。然而这些方法大多数采用监督分类, 需要提前对图像进行标注, 对海量遥感图像进行标注是十分繁琐与低效的。

本文提出了一种基于深度学习的半监督遥感图像检索方法。该方法采用稀疏自动编码(sparse autoencoder)算法对海量遥感图像进行无监督学习, 训练时无需大量标签, 避免了对海量遥感图像的标注工作, 大大提高方法的自动化程度。同时, 方法将卷积神经网络CNN(Convolutional Neural Network)共享权值的思想引入到特征提取过程中, 对遥感图像进行卷积处理, 并进行池化以减少特征维度, 提取出有效的特征对图像进行表示。另外, 借助有监督的Softmax分类器将遥感图像分类, 在缩小检索范围的同时提高了检索速度。

## 2 模型方法

本文的模型方法是一种结合无监督方法和有监督方法的半监督检索方法, 检索处理流程如图1所示: 对遥感图像预处理完成后, 首先, 采用无监督的特征学习方法稀疏自动编码进行遥感图像的特征提取; 然后, 对图像进行卷积处理并建立图像特征库, 接下来采用有监督的Softmax分类器进行图像分类; 最后, 采用欧氏距离作为度量准则, 衡量图像之间的相似度, 实现遥感图像检索。

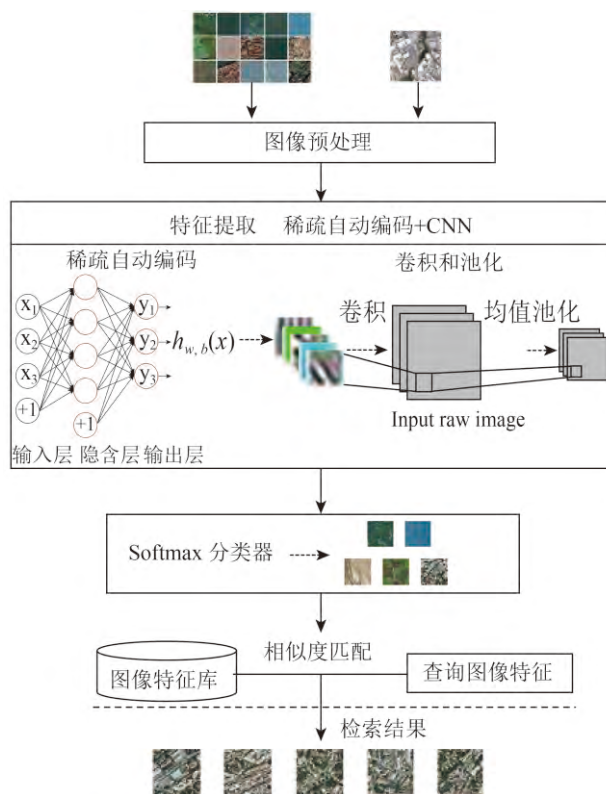


图1 实验流程示意图

Fig. 1 Flow chart of the experiment

### 2.1 图像预处理

图像预处理的主要作用是去除图片噪声, 去除相关性, 加强边缘等, 以提高后续特征提取的有效性。常见的图像预处理方法有白化, 高通滤波, 低通滤波, 高斯去噪等(黄劲 等, 2013)。本文采用ZCA白化(Zero-phase Component Analysis Whitening)的方法进行图像预处理, 它能够保持与原始数据相同的维度, 使得白化后的数据尽可能接近原始的数据, 同时去掉数据之间的相关联系(段宝彬 等, 2015)。

### 2.2 特征提取

#### 2.2.1 稀疏自动编码

自动编码算法(Auto-Encoder)是深度学习的一种典型的算法, 是由Bourlard和Kamp(1988)首次提出的, 该种算法通过无监督学习进行逐层训练, 特征训练时无需有标签的数据集, 并且能够很好地表示数据空间的固有几何结构(鲁珂 等, 2005)。稀疏自动编码(Bengio 等, 2006)是在自动编码的基础上加上稀疏性限制, 即稀疏模式的自动编码。稀疏自动编码是无监督的, 适合在大量的无标注

遥感图像上进行特征字典的提取。它的损失函数表达式中用到的标注的样本值(通常所说的 $y$ 值)等于本身输入的值 $x$ 。稀疏自动编码的典型网络结构如图2所示,包括输入层、隐藏层和输出层,其中,+1为偏置项。稀疏自动编码中的隐藏层神经元的个数是影响方法性能的一个重要参数(Coates等,2011),本文增加了隐藏层神经元的个数,使其个数超过了输入节点数。无稀疏约束时网络的损失函数 $J(W,b)$ 表达式如式(1):

$$J(W,b) = \left[ \frac{1}{m} \sum_{i=1}^m \left( \frac{1}{2} \|h_{W,b}(x^{(i)}) - y^{(i)}\|^2 \right) \right] + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_l} \sum_{j=1}^{s_{l+1}} (W_{ji}^{(l)})^2 \quad (1)$$

式中, $m$ 为输入样本个数; $W, b$ 为网络参数; $n_l$ 为自编码神经网络层数; $s_l$ 是第 $L$ 层的节点数目; $\lambda$ 是规则化系数; $h_{W,b}(x^{(i)})$ 是第 $i$ 组样本神经网络输出层的输出值。

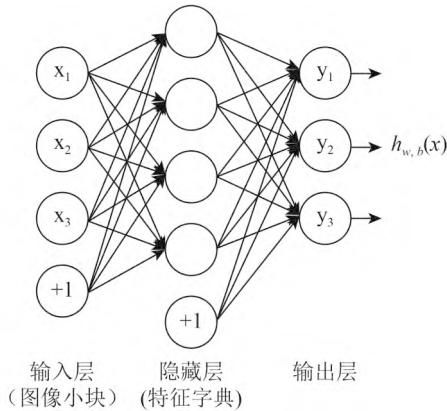


图2 稀疏自动编码示意图

Fig. 2 The diagram of Sparse Autoencoder

稀疏自动编码对网络的隐藏层的输出进行约束,使隐藏层节点输出的平均值尽量为0,进而使大部分的隐藏层节点处于非激活状态。此时的损失函数表达式为

$$J_{sparse}(W,b) = J(W,b) + \beta \sum_{j=1}^{s_2} KL(\rho \parallel \hat{\rho}) \quad (2)$$

式中, $\rho$ 为稀疏性参数(本文为0.05), $\beta$ 为参数,后项为KL距离,表达式如下:

$$KL(\rho \parallel \hat{\rho}) = \rho \log \frac{\rho}{\hat{\rho}} + (1-\rho) \log \frac{1-\rho}{1-\hat{\rho}} \quad (3)$$

用KL距离来测量两个分布间的差异,KL距离是一个以 $\rho$ 为均值和一个以 $\hat{\rho}$ 为均值的两个伯努利随机变量之间的相对熵,隐藏层节点输出平均值 $\hat{\rho}$ 求法如下:

$$\hat{\rho} = \frac{1}{m} \sum_{i=1}^m [a_j^2(x^i)] \quad (4)$$

式中, $a_j^2$ 表示在给定输入为 $x^i$ 的情况下,隐藏层神经元 $j$ 的激活程度。对于激活函数的选取,在隐藏层采用的是sigmoid函数,在输出层采用的是线性函数。

本文中稀疏自动编码的输入为在遥感图像上提取的图像小块。图像小块的大小是经验值,图像小块过小不足以描述图像的特征(如边缘等);图像小块过大会使训练过程中出现过拟合,本文采用的图像小块大小为 $8 \times 8$ 像素。隐藏层结点设置为400个,训练得到400个特征构成特征字典,如下面图3所示。

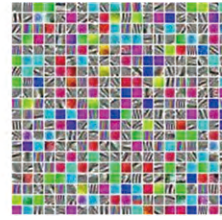


图3 稀疏自动编码特征字典

Fig. 3 The feature dictionary of Sparse Autoencoder

### 2.2.2 卷积和池化

卷积神经网络由加拿大多伦多大学教授LeCun(1989)和他的同事们一起提出。CNN是一种包含卷积层的深度神经网络模型,它的权值共享的网络结构类似于生物神经网络,降低了网络模型的复杂度,减少了权值的数量。目前CNN是人工神经网络的一种,具备很强的类别泛化能力,已经成为当前语音分析和图像识别领域的研究热点。

本文采用CNN的思想,模拟人大脑皮层中视觉皮层不同位置只对局部区域有响应这一特点,每个隐藏层的节点只与一部分连续的输入点连接。如图4所示,将稀疏自动编码学习到的特征字典用于大量的训练图像和测试图像的卷积,得到每幅遥感图像的特征层1。

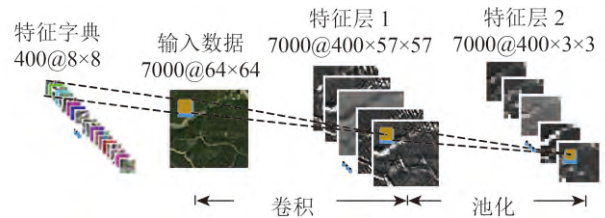


图4 遥感图像卷积和池化过程

Fig. 4 The procedure of convolution and pooling in the remote sensing images

通过卷积获取图像的特征规模十分庞大, 用这些特征进行图像的分类, 很容易出现过拟合的现象, 并且会产生很大的计算量。本文采用池化(Pooling)的方法来处理这一问题。池化是对卷积得到的结果进行统计计算, 解决了卷积神经网络输出维数过大的问题, 大大减少了需要训练的参数, 从而避免了有监督特征分类器在训练过程中产生过拟合现象(Hu 等, 2015b), 同时保持空间不变性。常用的池化方法有最大值池化和均值池化两种, 均值池化能减少邻域大小受限而造成的估计方差增大产生的误差, 更多的保留图像的背景信息, 用于图像检索, 因此本文采用的是均值池化的方法。池化操作后, 输出的特征图分辨率降低, 但能较好的保持高分辨率特征图描述的特征效果(尹宝才等, 2015), 得到如图4所示的特征层2。

### 2.3 Softmax分类器

分类器的选择会对检索结果产生影响, 王杰等人(2014)对不同分类器进行对比后发现, Softmax分类器结合稀疏自编码算法学习到的特征提取网络的分类精度优于其他分类器与特征提取算法的组合。

Softmax分类器用于解决多分类问题, 是一种有监督学习算法(Bishop, 2007), 需要提供训练样本和测试样本, 以及相应的分类标签。

在本文的Softmax分类器中, 有5个不同的类别标签( $k=5$ , 对应的地物类别分别为水体、城市、农田、裸地、植被5种), 对于训练集 $\{(\mathbf{x}^{(1)}, y^{(1)}), (\mathbf{x}^{(2)}, y^{(2)}), \dots, (\mathbf{x}^{(m)}, y^{(m)})\}$ , 有 $y^{(i)} \in \{1, 2, \dots, k\}$ 。对于给定的输入值 $\mathbf{x}^{(i)}$ , 使用假设函数对每一个类别 $j$ 估计出概率值 $p(y^{(i)}=j|\mathbf{x}^{(i)}; \theta)$ 。因此我们的假设函数要输出一个 $k$ 维的向量来表示这 $k$ 个估计值的概率。假设函数 $h_{\theta}(\mathbf{x})$ 形式如下:

$$h_{\theta}(\mathbf{x}) = \begin{bmatrix} p(y^{(i)}=1|\mathbf{x}^{(i)}; \theta) \\ p(y^{(i)}=2|\mathbf{x}^{(i)}; \theta) \\ \vdots \\ p(y^{(i)}=k|\mathbf{x}^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T \mathbf{x}^{(i)}}} \begin{bmatrix} e^{\theta_1^T \mathbf{x}^{(i)}} \\ e^{\theta_2^T \mathbf{x}^{(i)}} \\ \vdots \\ e^{\theta_k^T \mathbf{x}^{(i)}} \end{bmatrix} \quad (5)$$

式中,  $\theta_1, \theta_2, \dots, \theta_k$ 是模型的参数, 等式右边对概率分布进行归一化, 使得所有概率之和为1。使用 $\theta$ 符号来表示全部的模型参数, 在Softmax分类器进行回归计算时, 将 $\theta$ 用一个 $k \times (n+1)$ 的矩阵来表示, 如下所示:

$$\theta = \begin{bmatrix} \theta_1^T \\ \theta_2^T \\ \vdots \\ \theta_k^T \end{bmatrix} \quad (6)$$

Softmax分类器代价函数是logistic回归代价函数的一般形式, 对类别标记的 $k$ 个可能值进行了累加。

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)}=j\} \log \frac{e^{\theta_j^T \mathbf{x}^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T \mathbf{x}^{(i)}}} \right] \quad (7)$$

式中,  $1\{\cdot\}$ 是示性函数,  $1\{\text{值为真的表达式}\}=1$ ,  $1\{\text{值为假的表达式}\}=0$ 。

### 2.4 相似性匹配

对于每一幅待检索的图像, 计算它池化后的特征图, 然后运用Softmax分类器进行分类得到此图像类别, 通过计算待检索图像的特征图与图像库中跟它同一类别的图像的特征图之间的欧氏距离, 实现遥感图像检索。

用 $D(x, y)$ 表示两幅图像 $x$ 与 $y$ 之间的距离,  $\mathbf{x}_i$ 和 $\mathbf{y}_i$ 为图像的特征向量, Euclidean距离公式如下:

$$D(x, y) = \left( \sum_{i=1}^n |\mathbf{x}_i - \mathbf{y}_i|^2 \right)^{1/2} \quad (8)$$

按照特征距离由小到大进行排序, 依次返回检索结果。

## 3 实验与分析

### 3.1 实验数据与评价准则

实验采用的计算机配置是Intel(R) Core(TM) i5-3470处理器, 12 G内存, 使用的遥感数据来源为Pleiades和SPOT 5卫星的全色波段数据。Pleiades全色波段分辨率为0.5 m, SPOT 5全色波段分辨率为2.5 m, 尝试使用不同分辨率遥感图像进行检索, 比较图像的分辨率是否影响检索结果, 探讨本研究使用的方法是否适用于不同分辨率的遥感图像的检索。

从上述两种卫星图像中截取7000幅大小为 $64 \times 64$ 的图像组成图像库, 其中训练图像为3000幅, 测试图像为4000幅。按照图像所包含的主要地物类型, 将图像块划分为5种地物类型: 水体、植被、城市、农田和裸地, 样例图像和类型划分如图5所示。



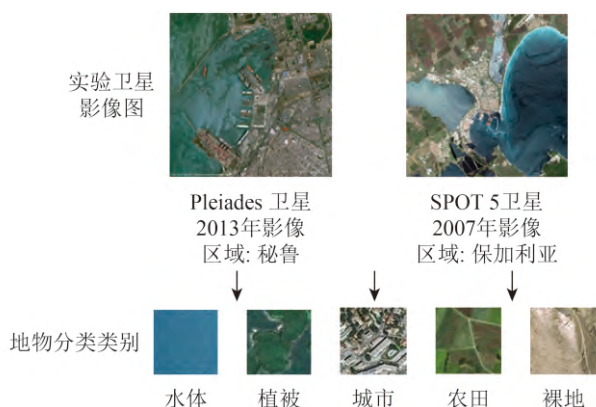


图5 实验卫星影像和地物分类类型

Fig. 5 Experimental satellite images and the types of terrestrial object

分类准确度是对所有测试图像进行统计得到的, 而检索准确度只是选取了20次检索结果的平均值, 因此分类准确度更适合用来评价神经网络的好坏。本文采用分类准确度来衡量神经网络优劣, 公式如下:

$$\text{分类准确度} = \frac{\text{测试图像库中分类正确的图像数}}{\text{测试图像库总图像数}} \quad (9)$$

同时, 本文采用检索准确度来对检索结果进行评价, 公式如下:

$$\text{检索准确度} = \frac{\text{检索结果中返回的相关图像数}}{\text{检索结果中的总图像数}} \quad (10)$$

## 3.2 实验结果与分析

### 3.2.1 检索速度

为了提高检索速度, 本文在设计模型时进行了两方面的考虑:

一是缩小每幅图像的特征数。在7000个图像

块上随机采样时, 得到140000幅大小为 $8 \times 8$ 的小图片, 用小图片训练稀疏自动编码网络, 隐藏节点选取400个, 即得到的特征字典大小为 $400 \times 8 \times 8$ , 整个过程共迭代了400次, 用时31 min。

然后用特征字典对7000幅 $64 \times 64$ 的图像进行卷积, 最终每一幅图像学习到的特征数为:

$$400 \times (64 - 8 + 1) \times (64 - 8 + 1) = 1299600$$

为了防止因卷积过后的系统参数过多导致分类器过拟合, 本文采用大小为 $19 \times 19$ 的方块对卷积后图像进行均值池化, 经过池化以后每一幅图像学习到的特征数为:

$$400 \times \text{floor}(64/19) \times \text{floor}(64/19) = 3600$$

经过池化, 每一幅图像的特征数减少, 从而缩减了计算量, 提高了运行速度。卷积和池化的过程用时61 min。

二是先进行分类再检索。为了提高检索速度和准确度, 首先对图像集进行分类, 总的分类时间为0.35 min(包括使用训练数据训练Softmax分类器和使用测试数据进行分类并统计分类准确度的时间)。

图6中的实验(a)和实验(b)输入为同一幅遥感图像, 当返回图像个数为20个时, 不分类直接进行检索和分类后进行类内检索这两种情况下的检索结果相差较大。进行类内检索有效的减少了相似性高的地物的干扰。在MATLAB中对检索速度进行测试, 分别取10次检索时间的平均值, 不进行分类的检索平均用时为2.404996 s, 进行分类时检索平均用时为1.88439 s(包括检索图像分类时间), 检索速度提高了27.6%, 这对于大数据的遥感图像的检索具有重要的意义。



图6 不进行分类和进行分类检索结果对比图(其中红色方框代表错误的检索结果)

Fig. 6 The comparison of retrieval results without and with classification (The red squares represent wrong results)

3.2.2 检索准确度

(1)隐藏层神经元个数对神经网络的影响。由于分类准确度直接影响检索准确度，这里采用分类准确度来度量网络参数的优劣。针对不同个数的隐藏层神经元进行实验，对应的分类准确度如表1。

表 1 SA不同隐藏层结点数对神经网络的影响  
Table 1 The influences on the result on neural networks made by the number of neuron in the hidden layer of SA

隐藏层结点数	分类准确度/(%)
100	91.850
200	92.300
300	92.400
400	92.425
500	92.075
600	91.975

根据实验的结果，隐藏层结点数在一定的范围内(小于400)，分类精度随着隐藏层的神经元数的增加而提高，超过这个范围，隐藏层神经元数的增多，反而导致分类精度下降。本文最终采用400个隐藏层神经元。

(2)池化区域的大小对神经网络的影响。为了评估池化区域大小对结果的影响，用不同的池化区域进行试验，得到的分类准确度如表2。

表 2 不同池化区域对神经网络的影响  
Table 2 The influences on the results of neural networks made by different regions of pooling

Pooling区域大小	分类准确度/(%)
5	90.775
10	91.950
15	91.475
16	92.000
17	91.950
18	92.275
19	92.425
20	90.250
21	90.425

当池化区域过小和过大时，都会导致结果准确度下降，取准确度最高的时候，即池化区域大小为19×19像素，进行检索实验。

(3)数据集大小对检索准确度的影响。为了验证数据集大小对检索结果的影响，分别采用了1913幅和7000幅64×64大小的遥感图像集进行实

验，得出Softmax分类器的分类准确度分别为82.839%和92.425%。分类完成后对图像进行类内检索，检索结果显示在Matlab的GUI界面中，如图7所示，当待检索图像选自Pleiades卫星的图像时，返回的检索结果以Pleiades图像为主(前10个返回图像均为Pleiades图像)，同时也有Spot图像，只是排在稍后的位置(第11、13、14和15为Spot图像)。这说明本检索方法首先返回同一卫星的图像，同时不同卫星的同类图像也可以检索出来。

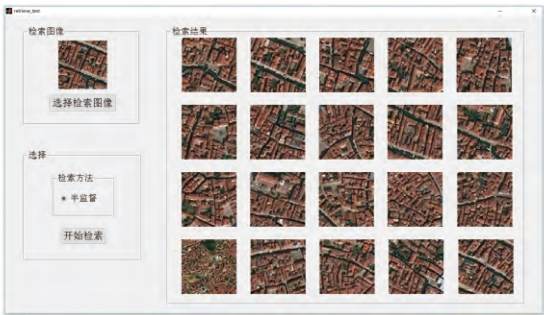


图 7 遥感图像集检索结果示例

Fig. 7 An example of the result of remote sensing image retrieval

在较小遥感数据集和较大遥感数据集上分别进行检索实验，统计结果。对五类地物分别进行检索，每次检索分别返回20、40、60、80和100幅图像，每类地物取20次检索结果平均值，总的检索结果取五类地物平均值。当返回图像数量为100时，检索准确度分别为88.6%和90.6%。其中遥感图像中分地物检索准确度如下面表3和表4所示。

表 3 较小数据集分地物检索准确度表  
Table 3 The accuracy of retrieval on small data set

返回图像个数	水体	植被	城市	农田	裸地	整体
20	0.900	1.000	0.968	0.690	0.968	0.912
40	0.900	0.996	0.968	0.679	0.968	0.909
60	0.900	0.996	0.966	0.670	0.966	0.906
80	0.899	0.993	0.963	0.661	0.963	0.898
100	0.863	0.988	0.962	0.654	0.962	0.886

表 4 较大数据集分地物检索准确度表  
Table 4 The accuracy of retrieval on large data set

返回图像个数	水体	植被	城市	农田	裸地	整体
20	0.973	0.995	0.995	0.625	0.998	0.917
40	0.948	0.984	0.998	0.599	0.998	0.905
60	0.936	0.973	0.998	0.629	0.997	0.906
80	0.927	0.963	0.998	0.656	0.996	0.908
100	0.921	0.948	0.998	0.670	0.994	0.906

从表中可以看出, 水体、植被、城市和裸地检索准确度相对较高, 农田的检索准确度相对较低。随着图像检索数目的增加, 检索准确度下降并不明显。

扩大数据集后, 不同地物和整体的检索准确度都得到了提升。从整体来看, 对于较大数据集, 返回图像数目的增加并没有使检索准确度大幅下降, 检索效果随着返回图像数量的增加趋于稳定。

(4) 底层特征对比实验。采用基于颜色特征和纹理特征这两个底层特征的遥感图像检索方法并进行对比, 图8展示了分别使用图像的颜色直方图(刘鹏宇, 2004)和纹理特征(Haralick和Shangmugam, 1973)进行检索和本文方法进行检索得到的检索准确度。其中, 颜色特征提取时首先将图像由RGB空间转化为HSV空间并进行非等间隔量化(其中 $H$ 空间量化成16级,  $S$ 空间量化成4级,  $V$ 空间量化成4级), 然后将3个颜色分量表示成一维矢量( $L=H \times Q_s \times Q_v + S \times Q_v + V$ ;  $Q_s=4$ ,  $Q_v=4$ ), 接着计算 $L$ 直方图作为颜色特征, 根据颜色特征进行检索; 纹理特征的提取基于灰度共生矩阵, 求能量、熵、惯性矩、相关性这4个参数的均值和标准差作为最终8维纹理特征, 进行遥感图像检索。

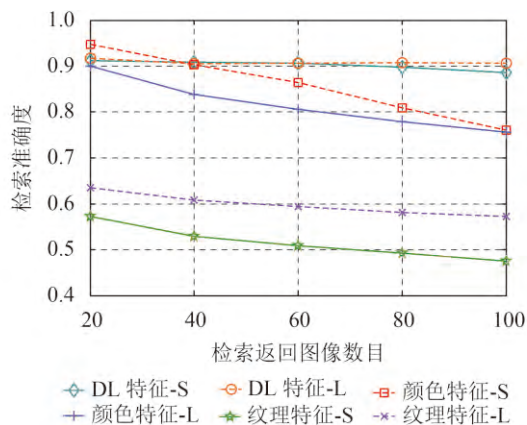


图8 不同方法检索准确度对比

Fig. 8 The comparison of different methods

(图例中S代表小数据集, L代表大数据集)

(In the legend, S represents the small data set, and L represents the large data set)

从图中可以看出, 当使用颜色特征进行检索时, 检索准确度随着返回图像的数目的增加而下降的幅度较大, 而且随着数据集的增大, 检索准确度

下降明显; 纹理特征在较大遥感数据集上的检索结果优于较小遥感数据集, 但是整体检索准确度较差; 使用本文的方法进行检索时, 当返回图像数目增加时, 检索准确度下降较为缓慢, 并且随着遥感数据集的增大, 检索准确度上升, 整体检索效果较好。

对检索结果进一步分析, 图9和10展示了各类地物在不同方法下的检索结果。

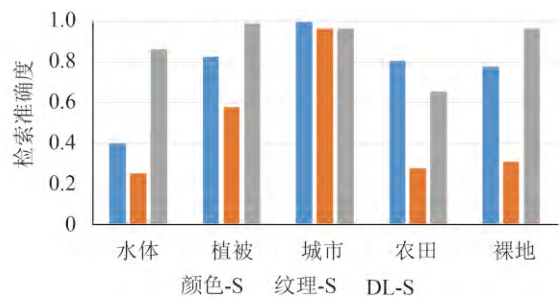


图9 较小遥感数据集上不同地物检索结果图

Fig. 9 The retrieval result of objects in remote sensing images on small data set

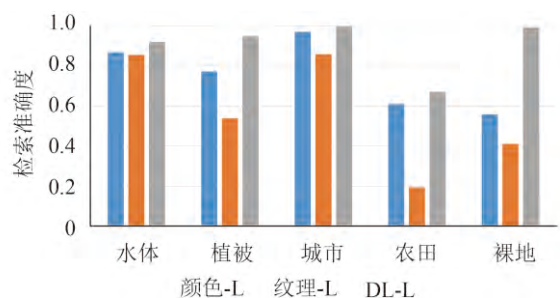


图10 较大遥感数据集上不同地物检索结果图

Fig. 10 The retrieval result of objects in remote sensing images on large data set

从图中看出, 不同的方法对城市检索效果都较好, 深度学习方法对水体、裸地和植被的检索效果都优于其他两类方法。较小数据集的水体样本不充足导致基于颜色和基于纹理的方法检索结果都很差, 但深度学习的方法还是能够较为准确的提取水体特征。由于部分农田和植被, 部分水体和裸地颜色相近, 水体和裸地纹理特征很难区分, 导致基于颜色和基于纹理的检索方法出现较大误差。

(5) 高层特征对比实验。实验使用基于卷积神经网络(CNN)的深度学习框架Caffe(Jia 等, 2014), 在本文的遥感数据集上进行特征训练和分



类, 训练集和测试集图像数量与本文方法相同。网络结构为两个卷积及池化层, 一个全连接层和一个Softmax分类层, 与本文的神经网络的分类结果进行对比。结果如表5所示。

表 5 本文方法与CNN方法结果对比  
Table 5 The comparison between CNN and our algorithm

神经网络	分类准确度	/(%)
CNN	91.601	
本文半监督方法	92.425	

根据表5结果, 本文的半监督的神经网络优于有监督的CNN网络, 在遥感图像集上取得了更好的效果。

## 4 结 论

本文提出了一种基于深度学习的半监督遥感图像检索方法。首先采用无监督的稀疏自动编码算法进行特征提取, 然后进行卷积和池化, 再使用Softmax分类器进行有监督分类。这种方法的优点是可以在大量无标签遥感图像上进行特征学习, 只需要少量具有代表性的有标签数据进行训练, 减少了对图像进行标注的烦琐工作, 适用于大数据遥感图像的检索。并且, 传统的遥感图像检索方法一般需要提取图像的颜色、纹理、形状等特征, 而本文直接在图像像素上进行特征提取, 再基于特征进行遥感图像的检索。为了探索本方法对大数据遥感图像的实用性, 对数据集进行扩大, 并对参数进行调整, 取得了两方面的成果:

一是检索速度的提高。采用先分类再进行类内检索的方法, 检索速度从不分类的2.404996 s改进到了1.88439 s, 提高了27.6%。

二是检索准确度的提高。扩大数据集后, 当返回图像数为100时, 检索准确度从原来的88.6%提高到了90.6%, 优于进行对比的基于颜色特征和基于纹理特征进行遥感图像检索的方法。而且在返回图像数目增加的同时, 检索准确度下降速度缓慢并趋于稳定。同时对稀疏自编码网络隐藏层神经元个数和pooling区域大小等神经网络参数进行优化和调整, 取得了较好的检索结果。

## 参考文献(References)

Bengio Y, Lamblin P, Popovici D and Larochelle H. 2006. Greedy lay-

er-wise training of deep networks//Proceedings of the 19th International Conference on Neural Information Processing Systems. Cambridge: MIT Press: 153-160

Bishop C M. 2007. Pattern Recognition and Machine Learning. New York: Springer

Bourlard H and Kamp Y. 1988. Auto-association by multilayer perceptrons and singular value decomposition. Biological Cybernetics, 59(4/5): 291-294 [DOI: 10.1007/BF00332918]

Coates A, Lee H and Ng A Y. 2011. An analysis of single-layer networks in unsupervised feature learning//Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. Fort Lauderdale, FL, USA: JMLR: 215-223

Duan B B, Han L X and Xie J. 2015. Fuzzy C-means clustering algorithm based on stacked sparse autoencoders. Computer Engineering and Applications, 51(4): 154-157 (段宝彬, 韩立新, 谢进. 2015. 基于堆叠稀疏自编码的模糊C-均值聚类算法. 计算机工程与应用, 51(4): 154-157) [DOI: 10.3778/j.issn.1002-8331.1402-0149]

Haralick R M and Shangmugam K. 1973. Texture features for image classification. IEEE Transaction on Systems, SMC-3(6): 768-780

Hinton G E and Salakhutdinov R R. 2006. Reducing the dimensionality of data with neural networks. Science, 313(5786): 504-507 [DOI: 10.1126/science.1127647]

Hu F, Xia G S, Hu J W and Zhang L P. 2015a. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sensing, 7(11): 14680-14707 [DOI: 10.3390/rs71114680]

Hu F, Xia G S, Hu J W and Zhang L P. 2015b. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sensing, 7(11): 14680-14707 [DOI: 10.3390/rs71114680]

Huang J, Sun Y and Xu H R. 2013. The application of Sparse coding on image retrieval. Digital Technology and Application(11): 76-77,81 (黄劲, 孙洋, 徐浩然. 2013. 稀疏编码(Sparse coding)在图像检索中的应用. 数字技术与应用(11): 76-77,81)

Jia Y Q, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S and Darrell T. 2014. Caffe: convolutional architecture for fast feature embedding. Eprint Arxiv: 1048.5093

Kitamoto A and Takagi M. 1994. Similarity retrieval of satellite cloud imagery based on optimization principle. IEICE Technical Report Pattern Recognition and Understanding, 94(294): 15-22

LeCun Y. 1989. Generalization and network design strategies//Pfeifer R, Schreier Z, Fogelman F and Steels L, eds. Connectionism in Perspective. Zurich, Switzerland: Elsevier: 143-155

LeCun Y, Bengio Y and Hinton G. 2015. Deep learning. Nature, 521(7553): 436-444 [DOI: 10.1038/nature14539]

Li D R and Ning X G. 2006. A new image decomposition method for content-based remote sensing image retrieval. Geomatics and Information Science of Wuhan University, 31(8): 659-662 (李德仁, 宁晓刚. 2006. 一种新的基于内容遥感图像检索的图像分块策略. 武汉大学学报(信息科学版), 31(8): 659-662)

Liu M N. 2013. Remote Sensing Image Retrieval Based on Color and Texture. Xian: Xi'an Technological University (刘米娜. 2013. 基



于颜色和纹理特征的遥感图像检索. 西安: 西安工业大学)

- Liu P Y. 2004. Study on the Algorithms for Content-based Image Feature Extraction. Changchun: Jilin University (刘鹏宇. 2004. 基于内容的图像特征提取算法的研究. 长春: 吉林大学)
- Lu K, Zhao J D, Ye Y L and Zeng J Z. 2005. Algorithm for semi-supervised learning in image retrieval. Journal of UEST of China, 34(5): 669–671 (鲁珂, 赵继东, 叶娅兰, 曾家智. 2005. 一种用于图像检索的新型半监督学习算法. 电子科技大学学报, 34(5): 669–671)
- Lu L Z, Liu R Y and Liu N. 2004. Remote sensing image retrieval using color and texture fused features. Journal of Image and Graphics, 9(3): 328–333 (陆丽珍, 刘仁义, 刘南. 2004. 一种融合颜色和纹理特征的遥感图像检索方法. 中国图象图形学报, 9(3): 328–333) [DOI: 10.11834/jig.20040361]
- Luus F P S, Salmon B P, Van den Bergh F and Maharaj B T J. 2015. Multiview deep learning for land-use classification. IEEE Geoscience and Remote Sensing Letters, 12(12): 2448–2452 [DOI: 10.1109/LGRS.2015.2483680]
- Mnih V and Hinton G E. 2010. Learning to detect roads in high-resolution aerial images//Proceedings of the 11th European Conference on Computer Vision. Berlin Heidelberg: Springer: 210–223 [DOI: 10.1007/978-3-642-15567-3\_16]
- Mojsilovic A and Rogowitz B. 2001. Capturing image semantics with low-level descriptors//Proceedings of the 2001 International Conference on Image Processing. Thessaloniki: IEEE: 18–21 [DOI: 10.1109/ICIP.2001.958942]
- Sheikholeslami G, Zhang A D and Bian L. 1999. A multi-resolution content-based retrieval approach for geographic images. GeoInformatica, 3(2): 109–139 [DOI: 10.1023/A:1009859912970]
- Wang J, Jia Y H and Zhao X. 2014. Tobacco leaf maturity classification based on sparse auto-encoder. Tobacco Science & Technology, 46(9):17–19 (王杰, 贾育衡, 赵昕. 2014. 基于稀疏自编码器的烟叶成熟度分类. 烟草科技, 46(9):17–19)
- Yin B C, Wang W T and Wang L C. 2015. Review of deep learning. Journal of Beijing University of Technology, 41(1): 48–59 (尹宝才, 王文通, 王立春. 2015. 深度学习研究综述. 北京工业大学学报, 41(1): 48–59) [DOI: 10.11936/bjtxb2014100026]
- Zhao L J, Tang P and Huo L Z. 2014. Land-use scene classification using a concentric circle-structured multiscale bag-of-visual-words model. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 7(12): 4620–4631 [DOI: 10.1109/JSTARS.2014.2339842]
- Zhu B, Ramsey M and Chen H. 2000. Creating a large-scale content-based airphoto image digital library. IEEE Transactions on Image Processing, 9(1): 163–167 [DOI: 10.1109/83.817609]

## Retrieval of remote sensing images based on semisupervised deep learning

ZHANG Hongqun<sup>1</sup>, LIU Xueying<sup>1,2</sup>, YANG Sen<sup>1,2</sup>, LI Yu<sup>1</sup>

1. Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract:** Massive data and diversity characteristics exert higher demands on the retrieval of remote sensing images. The feature extraction algorithm is the most outstanding factor that influences the performance of retrieval methods. Traditional feature extraction methods cause the problem of semantic gap, which occurs when the low-level feature does not perfectly reflect or match the purpose of retrieval. BOVW and multilayer cluster analysis have been proposed to solve semantic gap. However, the usability of these methods is limited given their dependence on classification or cluster algorithms and artificiality. A semi supervised deep-learning method was proposed in this paper. This method combines Sparse Auto encoder (SA) and the principle of Convolutional Neural Networks (CNNs).

The method involves four steps: first, the remote sensing images are pretreated with the ZCA whitening method. Second, the feature dictionary is extracted using SA, an algorithm that deals with nonannotated data. Subsequently, the feature dictionary is utilized in image convolution following the principle of CNNs, which imitates the neural net of organisms and decreases the number of features. Average pooling is conducted after image convolution. Both convolution and pooling are implemented to reduce model complexity, thus calculating distance faster. Third, the soft max classifier categorizes remote sensing images into five classes. Lastly, the remote sensing image retrieval is sorted based on the Euclidean distance between the query image and database in the same category as the query image.

Experimental results based on high-resolution remote sensing images demonstrate that the proposed method is effective and is more accurate than the methods that are based on color and texture features. Image classification before sorting speeds up retrieval by 27.6%. In addition, the semi-supervised deep learning algorithm is stable when the number of returned images increases. Given that the number of neurons in the hidden SA layer and the region size of pooling primarily affect retrieval results, this study conducted several optimization experiments on these two parameters. Moreover, the algorithm in this research performed well when the number of images in the data set and the retrieval accuracy on a larger data set increased, which are meaningful for the retrieval of massive remote sensing images.

The semisupervised deep learning algorithm decreases the time of image annotation, which is an exhausting job. Unlike traditional methods that extracts the features of color, texture, and shape, the method in this study directly extracts the feature dictionary from image elements with good accuracy. Moreover, the efficiency of the proposed method is guaranteed by convolution and pooling, which reduce the feature dimension. Furthermore, our experiment proves that this algorithm performs well in retrieving high-resolution remote sensing images.

**Key words:** remote sensing image retrieval, deep learning, sparse auto encoder, convolutional neural networks, Softmax classifier

**Supported by** National Natural Science Foundation of China (No.61501460)