doi: 10.6046/zrzyyg.2022028
引用格式: 李国祥,夏国恩,白丽明,等. 密集特征哈希的遥感场景分类[J]. 自然资源遥感,2023,35(1):66 - 73. (Li G X, Xia G E, Bai L M, et al. Remote sensing image classification based on DenseNet feature hashing[J]. Remote Sensing for Natural Resources, 2023, 35(1): 66 - 73.)

密集特征哈希的遥感场景分类

李国祥[1,2], 夏国恩[2,3], 白丽明[3], 马文斌[1,2]
(1. 广西财经学院教务处,南宁 530003; 2. 广西财税大数据分析工程研究中心,南宁 530003; 3. 广西财经学院工商管理学院,南宁 530003)

摘要: 针对遥感场景的精准分类,提出了密集网络特征哈希的场景分类算法。基于密集网络输出的高层语义特征,经全连接层过渡降维后,激活函数生成归一化的特征向量作为分类层的判断输入,形成端到端的分类网络。训练后的网络作为特征提取器,将测试数据激活层特征映射生成二值哈希码,最后采用支持向量机分类。所提出的算法分别在 UC Merced, WHU 和 NWPU - RESISC45 公开数据集进行了验证,分别与传统局部特征描述子、迁移学习、深度特征编码 3 个层次的多种算法进行了对比,实验结果表明,相比于传统中低层语义特征,分类准确度得到大幅度提高; 相比于深度学习网络的迁移,密集特征映射表达精细,聚集影像核心类别判断要素,更符合遥感影像的特征分布; 相比于深度特征编码算法,特征结构简单,分类精度高,迁移和拓展性强,可以满足不同遥感场景分类要求。

关键词: 迁移学习; 特征编码; 密集网络; 哈希码
中图法分类号: TP 393 **文献标志码**: A **文章编号**: 2097 - 035X(2023)01 - 0066 - 08

0 引言

遥感图像的精准分类是遥感内容分析的重要技术支撑,随着遥感图像分辨率的提高,其所包含的几何、纹理、结构,以及内容场景信息,使得遥感图像类内距离变大,类间距离变小,检索和分类任务面对更大的挑战。近年来,随着深度学习技术和硬件环境的发展,卷积神经网络(convolutional neural network, CNN)广泛应用于图像的检索和分类任务,网络深度也由最初 AlextNet[1] 的 8 层逐步发展到 Resnet[2] 的最多 152 层, DenseNet[3] 的最多 264 层,深度学习的网络结构在不断优化,网络学习能力不断增强。虽然遥感图像承载的信息与普通自然图像不同,但是二者的特征分布具有高度的相似性[4],大量实验证明自然图像所训练的 CNN 在遥感图像处理领域同样具有良好的效果。经过训练的网络各层对原始遥感图像的激活响应,具有优秀的特征表达能力[5],其中通过对卷积层和全连接层特征的编码与池化,可以进一步提升高层语义表征能力,这种深度学习特征正逐渐取代传统的描述算子[6]。

深度学习方法在遥感图像分类中的应用大致可以分为以下 3 种:

1)深度神经网络的迁移学习。主要思想是利用预训练和微调的 CNN 模型迁移至遥感数据集中进行分类。由于预训练模型已经在大规模图像数据集进行了训练,各参数具备了很好的适应性,通过修改 CNN 模型全连接层节点数并进行参数微调,一般都可以实现较好的识别效果。

2)深度特征的编码融合。该类算法将 CNN 作为特征提取器,以网络某一全连接层或卷积层的输出响应作为特征向量,直接输入或编码后输入分类器。按照实现过程,可以分为特征提取和特征编码两个部分。Hu 等[7] 在不同尺度上提取预训练网络的卷积层特征,多个特征连接并编码形成最终图像特征; Cheng 等[8] 将深度学习特征与词袋模型(bag of words, BOW), Fisher 向量(Fisher vector, FV),局部聚合描述符向量(vector of locally aggregated descriptors, VLAD)等编码方法相融合; 王鑫等[9] 训练核函数将全连接层特征映射到高维空间,形成最终

收稿日期: 2022 - 01 - 21; **修订日期**: 2022 - 06 - 06
基金项目: 国家自然科学基金资助项目"网络客户特征分析与流失预测研究"(编号: 71862003)、"大容量图像可逆信息隐藏理论与方法研究"(编号: 62162006)和广西高校中青年教师基础能力提升资助项目"基于特征混合编码的遥感场景分类研究"(编号: 2021KY0650)共同资助。
第一作者: 李国祥(1984 -),男,硕士,副教授,主要研究方向为模式识别、人工智能。Email: masterlgx@163.com。
通信作者: 夏国恩(1977 -),男,博士,教授,主要研究方向为机器学习、商务智能。Email: gandlf007711@163.com。

分类特征；Wang 等[4]将不同的深度学习网络卷积层进行 VLAD 编码降维后，与全连接层特征组成遥感图像特征；He 等[10]使用预训练 CNN 模型将多个卷积层特征集成，计算这些叠加特征的协方差矩阵池化特征；Cheng 等[11]通过优化新的目标函数，引入度量学习正则化项，构成 D－CNN 网络模型，提高针对于遥感图像类内多样性和类间相似性的识别精度；刘异等[12]将 Fisher 编码和 CaffeNet 全连接层特征结合到一起形成新的特征向量。

3）深度特征的多方法集成学习。余东行等[13]利用 Stacking 集成方法，构建基分类器和元分类器，提取的深度学习特征首先利用基分类器预测概率分布，并将其作为元分类器的输入，得到图像的分类结果；李亚飞等[14]在 AlexNet 基础上修改网络结构，去掉部分卷积层和池化层，使用光谱特征向量、纹理特征矩阵、亮度、绿度和湿度等形成混合特征，作为 CNN 的输入；张哲益等[15]将 CNN 与脉冲神经元的稀疏脉冲结合，实现对遥感图像的稀疏表征。

与传统特征描述方法相比，上述深度学习方法优势明显，但也存在一些问题，比如迁移学习模型经过相当长的迭代后，分类精度基本保持稳定，瓶颈效应凸显，提升空间有限。深度特征的编码和多方法集成，学习过程相对复杂，特征存储和计算代价偏高。为此，本文将哈希算法引入密集网络结构，将密集网络输出的高层语义特征映射为二进制的特征表达，构建适应遥感图像的分类模型。

1　密集网络的特征提取分析

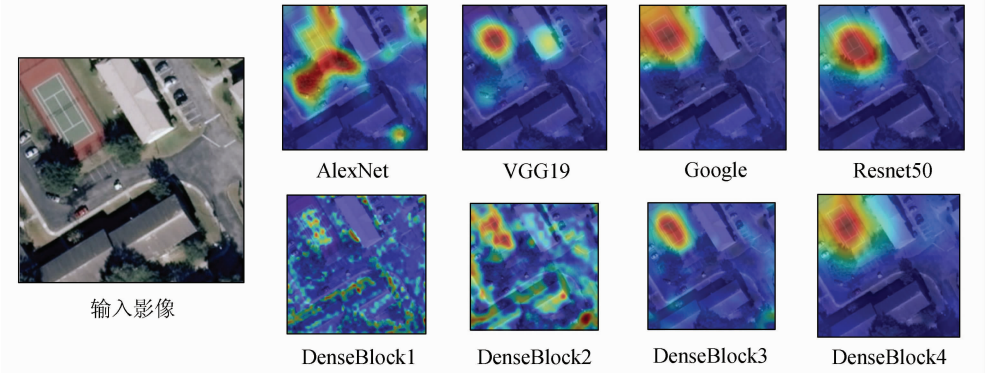
DenseNet 是 2017 年的最优深度模型，在 Resnet 基础上提出的密集连接模式，主要由密集块和过渡层组成。密集块中每一个上层网络节点输出不仅顺序传递至相连接的下游层节点，同时也传递至所有非直接连接的下游节点，这些跨节点的输入与上层连接点的输出串联形成新的输入特征，这样使得任一层的输入包含上游所有结点信息，各层特征得到充分复用，具体可以表示为：

$$\boldsymbol{X}_l = H_l([\boldsymbol{X}_0, \boldsymbol{X}_1, \cdots, \boldsymbol{X}_{l-1}]) \quad , \qquad (1)$$

式中：$\boldsymbol{X}_l$ 为第 l 层特征向量输出；$\boldsymbol{X}_0, \boldsymbol{X}_1, \cdots, \boldsymbol{X}_{l-1}$ 为上游各层特征连接；H_l 为对应层的变换函数，主要包括归一化、ReLU 和卷积 3 种非线性变换的顺序操作。每层特征向量的连接融合可以有效防止过拟合以及层数加深而导致的梯度消失问题。

过渡层主要包括卷积和池化 2 部分。各密集块间的连接对于特征输入尺寸要求不同，过渡层的主要作用就是对上游密集块输出特征的降维池化，在对特征进一步融合的同时保证有效输入下游密集块。DenseNet 中各层和块的连接设计，使得各部分特征有效融合，极大程度保留图像的高层语义特征。

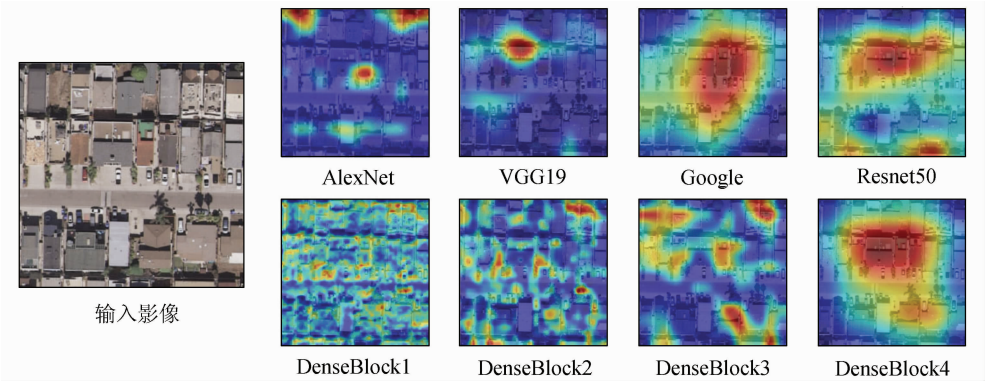
为了直观感受各块特征输出，利用梯度加权激活方法将特征输出可视化[16]，同时与其他主流网络中的最后卷积层特征进行对比。图 1 以网球场（图 1（a））和稠密居住区（图 1（b））2 幅遥感图像作为输入，其中网球场影像直观上容易识别并定位，稠密居住区影像内容复杂，容易引起类别混淆。图中红色代表对类别判断起主要贡献部分，DenseBlock1—DenseBlock 4 分别代表了 DenseNet 第 1—4 密集块的特征输出，可以看出，随着特征在不同密集块中的递进传输，输出特征对于类别判断的焦点逐渐集中在图像核心要素内。对于易于区分的网球场图像，各类特征都能集中在所属类别的主要标的物。对于判断难度较大的稠密居住区图像，AlexNet 和 VGG19 卷积层输出的判断重点对类间区分贡献不明显，Google 特征在关注居住区的同时也将图中的道路纳入其中，容易引起类别的混淆。Resnet50 同样具有不错的高层语义表征能力。



（a）网球场

图 1－1　不同深度特征的可视化

Fig. 1－1　Visualization of different deep features



（b）稠密居住区

图 1－2　不同深度特征的可视化

Fig. 1－2　Visualization of different deep features

2　密集特征的哈希表示

图像的紧凑哈希表示能够有效减少后续检索和分类的计算成本，其核心思想是将图像从原始高维空间映射为低维汉明空间的二进制编码。自深度学习与哈希算法首次结合以来，众多改进的深度哈希算法在图像检索领域获得广泛应用，其中 Lin 等[17]采用激活函数将 AlexNet 全连接层特征映射为哈希编码，方法简单且效果良好。但遥感影像除了具有传统图像特征之外，还具有丰富空间、纹理等语义信息，因此在有限的二进制编码中表现出不同遥感图像深层次的语义信息是关键，为此，本文通过提取 DenseNet 高层语义特征的输出，根据哈希码长设置全连接层结点，对 DenseBlock4 的输出降维，之后利用 sigmoid 函数将密集特征映射为归一化的特征向量，通过目标数据集的训练调参，形成最终遥感场景分类的哈希特征生成网络。测试数据在网络末端激活函数的归一化输出，经二值化即为图像哈希特征。最后使用支持向量机分类。框架如图 2 所示。

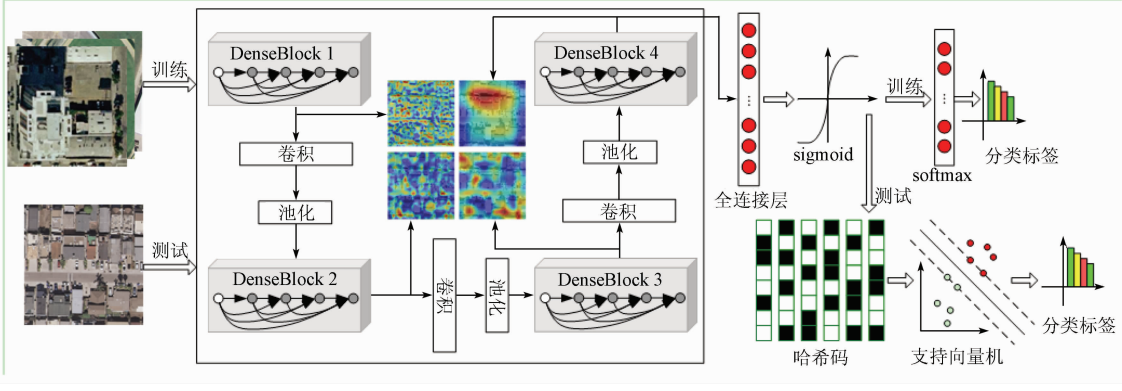


图 2　密集特征哈希算法结构图

Fig. 2　Illustration of dense feature hashing algorithm

3　实验

3.1　实验数据集

为了验证本文方法的有效性，实验在 3 个通用的遥感数据集上进行测试分析。第一个数据集为 UC Merced Land Use Dataset（UCM），每张图像大小为 256 像素 × 256 像素，21 类不同的遥感场景，每一类 100 张，共计 2 100 张。第二个数据集为 WHU－RS19 Data Set（WHU），每个图像尺寸为 600 像素 × 600 像素，整个数据集包含 19 类场景，每一类约为 50 张，总共 1 005 张。第三个数据集是 NWPU－RESISC45（NWPU）数据集，每个图像尺寸为 256 像素 × 256 像素，包含飞机、机场、棒球场等 45 个场景类别，每个场景类别包含 700 幅，图像，共 31 500 幅遥感图像，场景类别丰富，类内多样性和类间相似性较高，对遥感图像场景分类方法具有更高的挑战性。实验过程中，UCM 和 NWPU 数据集每类场景分别随机选取 50% 的图像作为训练数据，剩余 50% 图像作为测试数据，WHU 数据集中由于每个类别样本数量不同，每类选择 25 幅图像作为训练数据，25 幅图像作为测试数据。

针对于上述讨论，本文算法（DenseHash）分别与 BOW，VLAD 等传统特征，AlexNet，VGG19，Resnet 等主流 CNN 模型的迁移学习和深度特征编码等方法进行对比试验。UCM 和 WHU 数据集上各算法随机运行 10 次，深度学习方法训练周期设置为 20，NWPU 数据集规模较大，各算法随机运行 5 次，深度学习方法训练周期设置为 5，学习率统一为 0.000 3，验证频率设置为 10。试验硬件环境为：Xeon（R）CPU Gold－6128，GPU QUADRO P5000；软件环境为 MATLAB 2019a。实验中，数据集中所有样本均未进行数据增强。采用平均分类精度、混淆矩阵评价分类性能。分类精度 P_{overall} 的定义为：

$$P_{\text{overall}} = \frac{Z}{N} \times 100\% \quad , \tag{2}$$

式中：N 为总体样本数；Z 为所有分类正确的样本数。

3.2　实验结果与分析

3.2.1　不同映射结点数对于精确度的影响分析

全连接层结点的作用是析取密集特征，实现重要特征的低维度表示。结点数量决定特征载体的容量，如何用有限的单元实现遥感图像特征的高层语义表示尤为重要。图 3 显示了不同结点数对于分类精度的影响，结点数量增加之初，分类精度有着不同程度的提高，当结点数达到 48 时分类精度基本稳定，继续增加结点数量对于分类效果的整体提升不明显，且在结点数为 60 时还有下降趋势。因此在接下来的实验中，我们将全连接层结点数设置为 48。

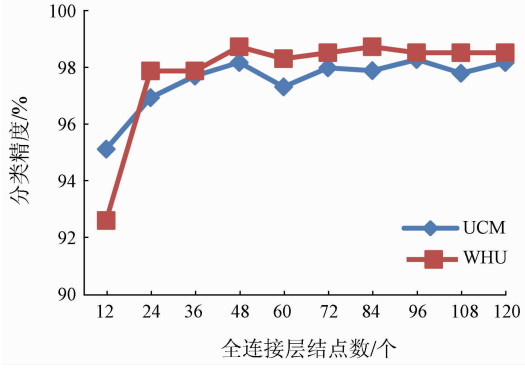


图 3　不同结点数的分类精度

Fig. 3　Classification accuracy of different nodes number

3.2.2　传统特征描述子对比

实验过程中，对比算法特征点的提取采用尺度不变特征变换（scale－invariant feature transform，SIFT）算子，BOW 的词典大小为 200，VLAD 词典大小 64 位。表 1 是传统局部特征描述算法结果，本文算法优势明显，即便与局部特征描述的改进算法 DRFV 相比，在 3 个数据集上的表现也是大幅领先，而这种优势在大数据集 NWPU 中体现的更为明显。说明密集网络的特征映射，能够有效代表遥感图像高层语义，即便是大规模数据集中也能做出很好的类别区分。

表 1　与传统特征描述子分类精度对比

Tab. 1　Comparison of traditional feature descriptors　（%）

方法	UCM	WHU	NWPU
BOW	63.80	63.96	40.11
BOW＋SPM[18]	67.72	66.29	44.78
BOW＋PCA	65.29	66.06	43.12
VLAD[19]	80.20	84.29	64.43
DRFV[20]	88.54	91.12	81.06
本文方法	98.21	99.20	96.49

3.2.3　主流 CNN 模型的迁移学习对比

应用当前主流 CNN 模型进行迁移学习，将最后的全连接层输出替换为目标数据集类别数量的结点，并在目标数据集中重新调整参数。对比结果如表 2 所示。可以看出，这些在 ImageNet 数据集中训练的 CNN 模型在不改变网络结构的情况下，经过目标数据集的参数精调后，与前面的传统描述算子相比，分类效果大幅提升，说明这些预训练模型的网络结构在不同类型的图像上具有普遍适应性。表 2 中 CNN 按照其提出的时间顺序由上至下做了排序，各数据集的整体分类精度也基本上按照时间顺序逐渐提升，说明随着深度学习网络结构的不断改进，学习能力逐渐增强，其中 DenseNet 的迁移学习效果最好。本文算法基于 DenseNet，但是通过对其网络结构的适当改进，各数据集上的分类精度再次得到了一定程度的提升。

表 2　与迁移学习算法对比

Tab. 2　Comparison of transfer learning algorithm　（%）

预训练网络	UC Merced	WHU	NWPU
AlexNet	89.80	93.16	77.81
VGG19[21]	92.00	95.03	85.84
GoogleNet[22]	92.50	93.18	88.16
ResNet－50[2]	95.50	95.56	93.68
NetVlad[23]	91.15	95.98	82.62
DenseNet[3]	96.70	96.88	94.78
本文算法	98.21	99.02	96.49

3.2.4　深度特征编码方法对比

由于版面限制，这里仅以 UCM 数据库为例对比不同编码方法对于分类准确度的影响。分别提取各 CNN 模型卷积层特征，采用 Bilinear[24]，BOW，VLAD，FV[25] 方法进行编码，结果如表 3 所示。同时分别提取 DenseNet4 个不同密集块（DenseBlock1—DenseBlock4）的特征进行编码，分析不同中间块所产生特征语义表达。

表 3 不同 CNN 模型的特征编码

Tab. 3 Feature coding of different CNN models (%)

深度特征	迁移学习	Bilinear	BOVW	VLAD	FV	编码方法平均精度
AlexNet	89.80	80.40	89.50	90.30	92.50	88.18
VGG19	92.00	90.70	92.60	93.30	93.40	92.50
Google	92.50	90.30	84.00	89.50	87.70	87.88
Resnet50	95.50	93.30	90.00	89.90	90.40	90.90
DenseNet	96.70					
DenseBlock1		54.10	78.86	84.57	87.81	76.34
DenseBlock2		71.62	85.24	87.14	92.38	84.10
DenseBlock3		90.76	91.90	92.76	95.05	92.62
DenseBlock4		95.24	93.90	92.29	93.14	93.64
本文算法						98.21

从编码方法的角度来看，相同的深度特征，FV 和 Bilinear 的编码效果相对优于其他方法，但是 Bilinear 在低层次的密集块特征表现欠佳。对于 AlexNet 和 VGG19 的卷积层特征，FV 的特征编码是最优的，超过了这两个网络的迁移学习精度。对于 Google 和 Resnet50 的卷积特征，Bilinear 编码效果较好，但是仍低于相应网络迁移学习的分类精度。对于 DenseNet 的密集块特征，密集块的依次递进，是一个不断精细化特征的过程。在 DenseBlock3 特征输出的各类编码，分类精度均已经达到 90% 以上，而 FV 编码使该层特征表现最优。从深度特征选择的角度来看，DenseNet 的 DenseBlock3 和 DenseBlock4 特征和 VGG19 的卷积层特征平均编码精度较高，能够很好地适应不同的编码方法。图 4 将不同卷积特征的最优编码，降维映射至二维空间形成散点图。相比于其他算法，本文算法的特征在同类遥感场景内聚性强，不同类别间可区分性好。

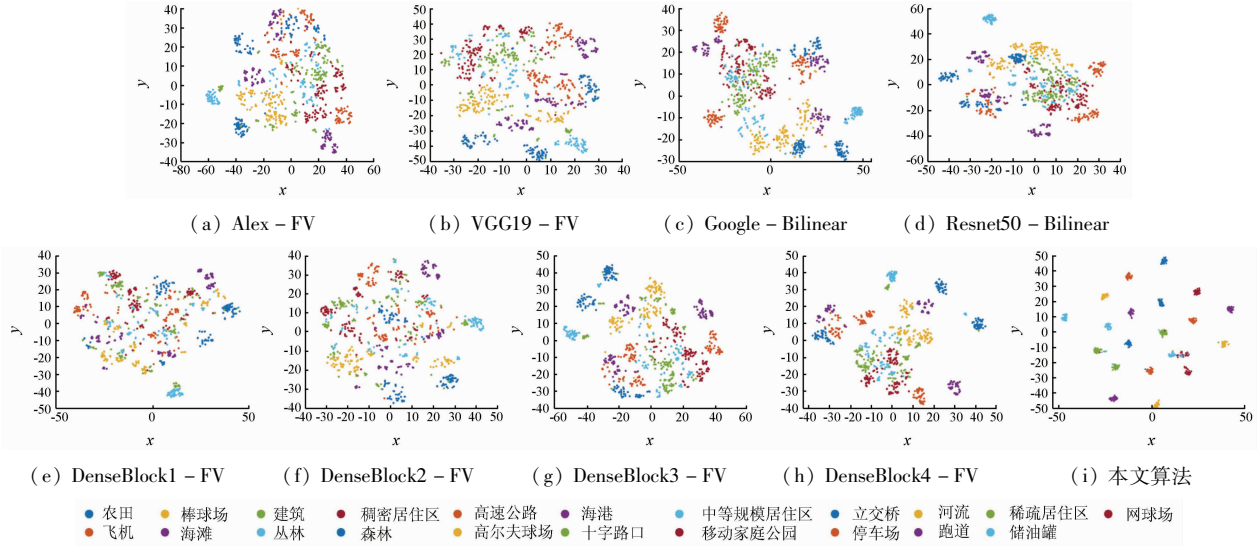


图 4 各类编码特征散点图

Fig. 4 Scatter diagram of various coding features

3.2.5 混淆分析

混淆矩阵用于定量评估各类之间的混淆程度，矩阵的行和列分别代表真实和预测场景，矩阵中任意元素 x_{ij} 代表将第 i 种场景预测为第 j 种场景类别的图像数。限于版面，这里仅以 UCM 数据集为例显示本文算法的混淆矩阵，如图 5 所示。通过对 UCM 混淆矩阵的分析，可以发现误判主要集中在稠密居住区，中等规模居住区等类别中，从遥感图像内容来看，这些类别主要是对居住区的稠密和稀疏程度分类，或者是图像中包含多个类别要素，在对不同类别要素比例划分的时候出现误判，而这些即便对于人工标注，都存在一定的辨别难度。

在 WHU 数据集中，共存在 5 个误判，分别是将农田判断为草地，河流判断为公园，工厂判断为居住区，足球场判断为池塘、机场判断为火车站。这些图像就其内容本身同样具有高度相似性和不同类别元素的交叉。

图 6 显示了 2 个数据集部分误判图像的前五类判断概率。由图可知，这些误判图像的真实标签基本上都在预测概率最高的 5 个类的范围中。这种误

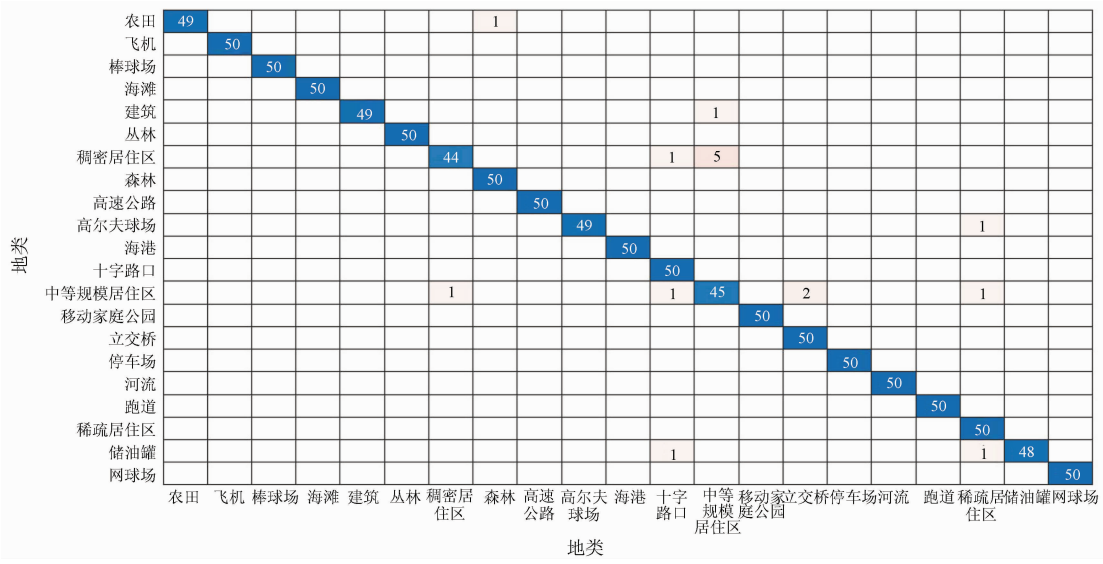
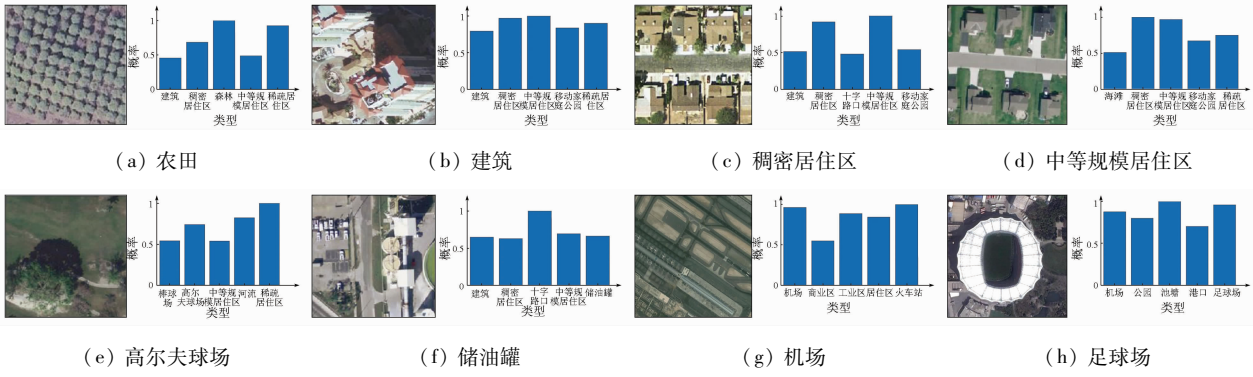


图 5　UCM 的混淆矩阵

Fig. 5　UCM confusion matrix



(a) 农田　(b) 建筑　(c) 稠密居住区　(d) 中等规模居住区

(e) 高尔夫球场　(f) 储油罐　(g) 机场　(h) 足球场

图 6　误判图像的概率分析

Fig. 6　Probability analysis of misjudgment image

判可以分为 3 种类型: 首先是语义的相似性影响,如图 6(a)将农田判断为森林,图 6(g)和(h)机场与火车站,足球场与池塘判断概率接近; 其次是图像主要类别元素的稀疏程度影响,如图 6(b)—(d)是对建筑、稠密居住区和中等规模居住区的混淆,图 6(f)因为包含路口和储油罐这 2 个元素而判断为十字路口; 最后是典型的判断错误,如图 6(e)将高尔夫球场判断为稀疏居住区,这类错误是分类器训练过程中首先应当避免的。

3.2.6　不同数据集样本的训练与测试

由于不同数据集中图像特点各不相同,遥感场景分类的一般做法都是在目标数据集上随机选择样本,训练学习网络与之适应,因此不同的数据集中需要重新进行调参,而深度学习网络的训练调参是一个需要大量时间和空间的过程。本文尝试验证 DenseHash 特征表达的通用性,即在一个数据集中训练后直接作为其他数据集的特征提取器。实验结果如表 4 所示,NWPU 数据集体量较大,训练出的密集哈希特征提取模型具有一定的通用性,迁移应用于其他 2 个数据集中,均可以获得近 90% 的分类精度,结合其紧凑的二值化表达所带来的计算时间和存储空间的降低,可以直接应用于遥感的粗分类场景。

表 4　不同数据集样本的训练与测试

Tab. 4　Training and Testing in different data sets

训练样本	测试样本	分类精度/%	时间/s
UCM	UCM	98.21	87.27
	NWPU	58.12	837.69
	WHU	77.00	26.29
WHU	UCM	71.30	55.56
	NWPU	58.23	735.37
	WHU	99.02	26.33
NWPU	UCM	88.62	54.43
	NWPU	96.49	715.13
	WHU	91.86	25.03

4　结论

本文提出了一种面向遥感场景分类的紧凑哈希特征提取网络。针对遥感图像语义丰富的特点,采

用密集网络提取高层语义特征,通过全连接层和激活函数实现低维度的特征表达,最后映射到汉明空间,生成紧凑的二值哈希码,作为支持向量机分类的特征输入。与其他方法相比,密集特征的哈希码表示一方面保留了遥感高层语义特征,另一方面二进制的特征结构,便于分类器的快速高效分类。3 个通用遥感数据集的实证分析表明了算法有效性,各数据集的分类精度得到不同程度的提升。在下一步的研究中,尝试扩大不同类别的训练样本集,构建具有广泛适用性特征提取模型,减少目标数据集的训练调参成本。

参考文献(References):

[1] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]// Proceedings of the International Conference on Neural Information Processing Systems. Lake Tahoe: NIPS, 2012: 1097 – 1105.

[2] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770 – 778.

[3] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 4700 – 4708.

[4] Wang G, Fan B, Xiang S, et al. Aggregating rich hierarchical features for scene classification in remote sensing imagery[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(9): 4104 – 4115.

[5] Razavian A S, Sullivan J, Carlsson S, et al. Visual instance retrieval with deep convolutional networks[J]. ITE Transactions on Media Technology and Applications, 2016, 4(3): 251 – 258.

[6] Zheng L, Yang Y, Tian Q. SIFT meets CNN: A decade survey of instance retrieval[J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2018, 40(5): 1224 – 1244.

[7] Hu F, Xia G S, Hu J, et al. Transferring deep convolutional neural networks for the scene classification of high – resolution remote sensing imagery[J]. Remote Sensing, 2015, 7(11): 14680 – 14707.

[8] Cheng G, Li Z, Yao X, et al. Remote sensing image scene classification using bag of convolutional features[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(10): 1735 – 1739.

[9] 王 鑫,李 可,宁 晨,等. 基于深度卷积神经网络和多核学习的遥感图像分类方法[J]. 电子与信息学报,2019,41(5):1098 – 1105.
Wang X, Li K, Ning C, et al. Remote sensing image classification method based on deep convolution neural network and multi – kernel learning[J]. Journal of Electronics & Information Technology, 2019, 41(5): 1098 – 1105.

[10] He N, Fang L, Li S, et al. Remote sensing scene classification using multilayer stacked covariance pooling[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(12): 6899 – 6910.

[11] Cheng G, Yang C, Yao X, et al. When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(5): 2811 – 2821.

[12] 刘 异,庄姊琪,闫 利,等. 联合 Fisher 核编码和卷积神经网络的影像场景分类[J]. 遥感信息,2018,33(4):8 – 15.
Liu Y, Zhuang Z Q, Yan L, et al. Combined Fisher kernel coding framework with convolutional neural network for remote sensing scene classification[J]. Remote Sensing Information, 2018, 33(4): 8 – 15.

[13] 余东行,张保明,赵 传,等. 联合卷积神经网络与集成学习的遥感影像场景分类[J]. 遥感学报,2020,24(6):717 – 727.
Yu D H, Zhang B M, Zhao C, et al. Scene classification of remote sensing image using ensemble convolutional neural network[J]. Journal of Remote Sensing, 2020, 24(6): 717 – 727.

[14] 李亚飞,董红斌. 基于卷积神经网络的遥感图像分类研究[J]. 智能系统学报,2018,13(4):550 – 556.
Li Y F, Dong H B. Classification of remote – sensing image based on convolutional neural network[J]. CAAI Transactions on Intelligent Systems, 2018, 13(4): 550 – 556.

[15] 张哲益,曹卫华,朱 蕊,等. 基于脉冲卷积神经网络稀疏表征的高分辨率遥感图像场景分类方法[J]. 控制与决策,2022,37(9):2305 – 2313.
Zhang Z Y, Cao W H, Zhu R, et al. Sparse representation with spike convolutional neural networks for scene classification of remote sensing images of high resolution[J]. Control and Decision, 2022, 37(9): 2305 – 2313.

[16] Selvaraju R R, Cogswell M, Das A, et al. Grad – cam: Visual explanations from deep networks via gradient – based localization[C]// Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 618 – 626.

[17] Lin K, Yang H F, Hsiao J H, et al. Deep learning of binary hash codes for fast image retrieval[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Boston: IEEE, 2015: 27 – 35.

[18] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories[C]// 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). New York: IEEE, 2006: 2169 – 2178.

[19] Arandjelovic R, Zisserman A. All about VLAD[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE, 2013: 1578 – 1585.

[20] 李国祥,马文斌,王继军. 稠密特征编码的遥感场景分类算法[J]. 小型微型计算机系统,2021,42(4):766 – 772.
Li G X, Ma W B, Wang J J. Remote sensing image classification based on dense feature coding[J]. Journal of Chinese Computer Systems, 2021, 42(4): 766 – 772.

[21] Simonyan K, Zisserman A. Very deep convolutional networks for large – scale image recognition[J/OL]. arxiv, 2014[2022 – 04 – 24]. https://arxiv.org/pdf/1409.1556.pdf.

[22] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 1 – 9.

[23] Arandjelovic R, Gronat P, Torii A, et al. NetVLAD: CNN architec-

ture for weakly supervised place recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas:IEEE,2016:5297 - 5307.

[24] Lin T,Roychowdhury A,Maji S. Bilinear convolutional neural networks for fine - grained visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2018,40(6):1309 - 1322.

[25] Sánchez J,Perronnin F,Mensink T,et al. Image classification with the fisher vector:Theory and practice[J]. International Journal of Computer Vision,2013,105(3):222 - 245.

Remote sensing image classification based on DenseNet feature hashing

LI Guoxiang[1,2], XIA Guo'en[2,3], BAI Liming[3], MA Wenbin[1,2]

(1. *Department of Academic Affairs Guangxi University of Finance and Economics*, *Nanning* 530003, *China*; 2. *Guangxi Engineering Research Center of Big Data Analysis of Finance and Taxation*, *Nanning* 530003, *China*; 3. *School of Business Administration*, *Guangxi University of Finance and Economics*, *Nanning* 530003, *China*)

Abstract: To achieve accurate remote sensing scene classification, this study proposed a classification algorithm based on DenseNet feature hashing. First, dimension reduction was conducted for high - level semantic features output by a DenseNet through a fully connected layer. Then, normalized feature vectors were generated as the input of the classification layer using an activation function, and an end - to - end classification network was formed. Using the trained network as a feature extractor, the features of the activation layer of test data were mapped into binary hash codes. Finally, the remote sensing scene classification was conducted using support vector machine. The new algorithm was validated on public data sets UC Merced, WHU, and NWPU - RESISC45, and its classification effect was compared with that of multiple algorithms at three levels, namely the conventional local feature descriptor, transfer learning, and depth feature coding. The experimental results are as follows. The new algorithm had significantly higher classification accuracy than conventional algorithms based on mid - and low - level semantic features. Compared with the algorithm based on transfer learning, the proposed algorithm has fine - scale DenseNet feature mapping and accumulates elements used to determine core categories of images and, thus, is more suitable for the feature distribution of remote sensing images. Compared with the depth feature coding algorithm, the new algorithm has a simple feature structure, high classification accuracy, and strong transferability and extensibility and, thus, can meet the classification requirements of different remote sensing scenarios.

Keywords: transfer learning; feature coding; DenseNet; hash code

(责任编辑: 张 仙)