



基于 ImageNet 预训练卷积神经网络的遥感图像检索

葛 芸^{1,2} 江顺亮¹ 叶发茂¹ 许庆勇¹ 唐祎玲¹

¹ 南昌大学信息工程学院,江西 南昌,330031

² 南昌航空大学软件学院,江西 南昌,330063

摘 要:高分辨率遥感图像内容复杂,细节信息丰富,传统的浅层特征在描述这类图像上存在一定难度,容易导致检索中存在较大的语义鸿沟。本文将大规模数据集 ImageNet 上预训练的 4 种不同卷积神经网络用于遥感图像检索,首先分别提取 4 种网络中不同层次的输出值作为高层特征,再对高层特征进行高斯归一化,然后采用欧氏距离作为相似性度量进行检索。在 UC-Merced 和 WHU-RS 数据集上的一系列实验结果表明,4 种卷积神经网络的高层特征中,以 CNN-M 特征的检索性能最好;与视觉词袋和全局形态纹理描述子这两种浅层特征相比,高层特征的检索平均准确率提高了 15.7%~25.6%,平均归一化修改检索等级减少了 17%~22.1%。因此将 ImageNet 上预训练的卷积神经网络用于遥感图像检索是一种有效的方法。

关键词:遥感图像;检索;卷积神经网络;预训练

中图分类号:TP751

文献标志码:A

近年来,随着遥感技术的发展,遥感图像分辨率不断提高,数据量急速增长,如何有效地管理利用遥感图像数据,满足用户对数据的使用需求成为研究热点。其中,从海量的遥感图像中有效地检索出感兴趣的目标或场景成为关键问题之一。

基于内容的图像检索(content-based image retrieval, CBIR)在遥感图像领域得到广泛应用^[1-3]。CBIR 技术主要通过提取图像的纹理^[1]、颜色^[2]和形状^[3]等底层特征进行检索。但底层特征难以反映高层的语义信息,容易形成较大的语义鸿沟。为了缩小检索中的语义鸿沟,采用相关反馈机制^[4]、融合图像底层特征实现区域级语义特征^[5]和构建视觉词袋模型^[6-7]等都是有效方法。这些检索方法都是基于图像的浅层特征,浅层特征一方面需要人为地提取相关特征,检索性能容易受到干预。另一方面,高分辨率的遥感图像内容复杂,细节信息丰富,浅层特征在描述这类图像时存在一定难度。

卷积神经网络(convolutional neural networks, CNN)通过建立多层神经网络的训练机制来模拟人脑的学习过程,在处理复杂图像时有一定优势,特征的提取过程自动完成,减少了人为干

预;底层特征通过多层抽象机制形成高层特征,能够有效地缩小语义鸿沟。本文在遥感图像的检索中引入卷积神经网络,提取卷积神经网络中的高层特征,分别比较了不同结构的卷积神经网络提取出的不同层次的 CNN 特征的检索性能。

1 卷积神经网络

卷积神经网络是深度学习中的一种基本模型,是多层次结构的神经网络。卷积神经网络通过局部感知区域、下采样和权值共享等方式,减少了网络学习参数,提高了特征的鲁棒性和稳定性。

卷积神经网络在手写字符识别、语音识别和图像识别等领域得到广泛应用。2012 年卷积神经网络^[8]首次用于 ImageNet 大规模视觉识别挑战赛(ImageNet large scale visual recognition challenge, ILSVRC)并获得冠军,之后卷积神经网络在图像领域受到越来越多的关注。目前大多数卷积神经网络的研究是基于文献^[8]提出的网络结构进行改进,比如采用更小的局部感知区域^[9]和增加网络的深度^[10]等。

卷积神经网络需要大规模的标签数据训练大

收稿日期:2016-05-09

项目资助:国家自然科学基金(41261091);江西省教育厅科技项目(GJJ13482);江西省自然科学基金(20151BAB207062)。

第一作者:葛芸,博士生,讲师,主要从事遥感图像检索理论和方法研究。geyun@nchu.edu.cn

通讯作者:江顺亮,博士,教授。jiangshunliang@ncu.edu.cn

量的参数,当数据库图像数目较少时,卷积神经网络的优势得不到体现。针对这种情况,可以考虑在大规模数据集 ImageNet 上预训练卷积神经网络来完成相应任务。文献[11-13]研究表明,在 ImageNet 上预训练的卷积神经网络泛化能力强,可以有效地用于 SUN-397 和 PASCAL VOC 等数据集的图像分类。文献[14-15]将 ImageNet 上预训练的卷积神经网络用于遥感图像分类,取得了良好的效果。在图像检索领域,文献[16-17]将 ImageNet 上预训练的卷积神经网络用于 Oxford、Holidays 和 UKBench 等数据集中,和传统的浅层特征相比,卷积神经网络提取的特征在图像检索中的性能得到明显提高。

本文在 ImageNet 中预训练不同结构的卷积神经网络,并将预训练的网络结构和参数迁移到遥感图像数据集上,进而提取相应的高层特征进行图像检索。实验结果表明,和浅层特征相比,基于卷积神经网络高层特征的检索性能得到明显改善。

2 图像检索

2.1 特征提取

选用四种不同结构的卷积神经网络,分别提

取不同的 CNN 特征。文献[8]提出的卷积神经网络总共包含 8 层,前面 5 层为卷积层,后续 3 层为全连接层。激活函数采用目前深度网络中主流激活函数 ReLU(rectification linear unit),下采样采用最大池化(max pooling)方法。全连接层中,第 6 层和第 7 层采用 dropout 策略防止过拟合,特征维数均为 4 096 维,最后一层采用 softmax 分类器进行分类,总共有 1000 个类别。将这个网络模型记为 CNN-Alex。

文献[13]分析了多种结构的网络模型,其中性能较好的 CNN-M 和文献[9]类似,缩小了局部感知区域,CNN-M 与文献[9]的不同之处是减少了第四层卷积层的滤波器数目。文献[10]在 CNN-Alex 的基础上增加隐层的层数,网络的层次达到 16 层和 19 层,将这两种网络模型分别记为 CNN-16 和 CNN-19。

四种卷积神经网络的结构如表 1 所示。滤波器数目和局部感知区域尺寸记为“num \times size \times size”的形式。以 CNN-Alex 的结构为例,第一个卷积层 conv1 中,滤波器数目为 96,局部感知区域的尺寸为 11 \times 11。pool1~pool5 是下采样层,fc6、fc7 和 fc8 为全连接层,ReLU 和分类器没有在表中显示。

表 1 不同卷积神经网络的结构
Tab.1 Different CNN Architectures

CNN-Alex	CNN-M	CNN-16	CNN-19
conv1 96 \times 11 \times 11	conv1 96 \times 7 \times 7	conv1-1 64 \times 3 \times 3 conv1-2 64 \times 3 \times 3	conv1-1 64 \times 3 \times 3 conv1-2 64 \times 3 \times 3
pool1	pool1	pool1	pool1
conv2 256 \times 5 \times 5	conv2 256 \times 5 \times 5	conv2-1 128 \times 3 \times 3 conv2-2 128 \times 3 \times 3	conv2-1 128 \times 3 \times 3 conv2-2 128 \times 3 \times 3
pool2	pool2	pool2	pool2
conv3 384 \times 3 \times 3	conv3 512 \times 3 \times 3	conv3-1 256 \times 3 \times 3 conv3-2 256 \times 3 \times 3 conv3-3 256 \times 3 \times 3	conv3-1 256 \times 3 \times 3 conv3-2 256 \times 3 \times 3 conv3-3 256 \times 3 \times 3 conv3-4 256 \times 3 \times 3
		pool3	pool3
conv4 384 \times 3 \times 3	conv4 512 \times 3 \times 3	conv4-1 512 \times 3 \times 3 conv4-2 512 \times 3 \times 3 conv4-3 512 \times 3 \times 3	conv4-1 512 \times 3 \times 3 conv4-2 512 \times 3 \times 3 conv4-3 512 \times 3 \times 3 conv4-4 512 \times 3 \times 3
		pool4	pool4
conv5 256 \times 3 \times 3	conv5 512 \times 3 \times 3	conv5-1 512 \times 3 \times 3 conv5-2 512 \times 3 \times 3 conv5-3 512 \times 3 \times 3	conv5-1 512 \times 3 \times 3 conv5-2 512 \times 3 \times 3 conv5-3 512 \times 3 \times 3 conv5-4 512 \times 3 \times 3
		pool5	
		fc6 4096	
		fc7 4096	
		fc8 1000	

卷积神经网络中,最先学习到图像的底层特征,然后逐步形成中级特征,直到抽象出高层特

征。高层特征能有效缩小语义鸿沟,因此选用高层特征进行检索。表 1 的网络结构中 fc8 的输出

值用于图像分类,因此不考虑 fc8 的输出值。pool5 是对最后一个卷积层进行下采样,它对图像的描述比较全面。为了比较不同高层特征对检索结果的影响,使用 pool5、fc6 和 fc7 的输出值作为 CNN 特征并应用于遥感图像检索。

2.2 检索过程

通过数据集 ImageNet 预训练 § 2.1 介绍的四种卷积神经网络,将其用于遥感图像检索,检索过程如图 1 所示,检索步骤如下:

步骤 1 预训练卷积神经网络。

预训练卷积神经网络的数据集统一采用 ImageNet 数据集的子集 ILSVRC2012。ILSVRC2012 包含 1 000 种图像分类,约有 1.3 百万张训练图片、5 万验证图片和 10 万测试图片。

参数更新采用随机梯度下降算法,超参数的设置如下:冲量项 0.9,权值衰减 0.000 5,学习率初始值为 0.01,并在训练过程中动态调整。每层权重初始值服从标准差为 0.01 的零均值高斯分布。conv2、conv4、conv5 以及全连接层的偏置初始值为 1,其他层次的偏置初始值为 0。

步骤 2 将预训练的网络结构和参数分别迁移到遥感图像库 M 和查询图像 q 。

遥感图像库 M 中每幅图像输入到预训练的网络中,提取网络后面层次(pool5 或者 fc6 或者 fc7)的输出值作为 CNN 特征,形成遥感图像 CNN 特征库。pool5 层是对最后一个卷积层进行最大池化后的结果,fc6 和 fc7 的输出结果为:

$$Y_k = r(W_k \times Y_{k-1} + B_k) \quad (1)$$

式中, Y_k 代表第 k 层的输出值; Y_{k-1} 代表第 $k-1$ 层的输出值; W_k 和 B_k 分别代表第 k 层训练得到的权重和偏置,这里 k 取值为 6 和 7。激活函数 r 为 ReLU 函数。

查询图像 q 输入到预训练的网络中,提取 pool5 或者 fc6 或者 fc7 的输出值作为查询图像的高层特征,记为 q -CNN 特征。

步骤 3 分别对 CNN 特征库和 q 的特征 q -CNN 进行归一化操作。

归一化过程采用常用的高斯方法。将 CNN 特征库看作是一个矩阵,图像库中的 m 幅图像用 $\{I_1, I_2, \dots, I_i, \dots, I_m\}$ 表示,一幅图像 I_i 的某个 l 维特征向量表示为:

$$F_i = [f_{i1} \ f_{i2} \ \dots \ f_{ij} \ \dots \ f_{il}]$$

式中, f_{ij} 是 F_i 的第 j 个特征元素,将图像库中所有图像的特征向量看成一个大小为 $m \times l$ 的矩阵 F , F 的每一行是长度为 l 的特征序列。计算 F 每一行特征分量对应的平均值 μ_i 和标准差 σ_i ,通

过式(2)将特征向量进行归一化和平移,使得绝大部分的值在 $[0, 1]$ 范围。

$$f_i^* f = \frac{1 + (f_{ij} - \mu)/3\sigma_i}{2} \quad (2)$$

令 q 的特征向量 q -CNN 为:

$$F_q = [f_{q1} \ f_{q2} \ \dots \ f_{qj} \ \dots \ f_{ql}]$$

根据高斯方法对 F_q 进行归一化和平移,使得绝大部分值在 $[0, 1]$ 范围。

步骤 4 计算 q 和 M 中图像的相似度,按需求返回最相似的 n 幅图像。

根据欧氏距离计算 q 与 M 中图像的相似度,得到 q 和 M 中图像的一系列相似度,并按需求返回距离最小,即最相似的 n 幅图像。

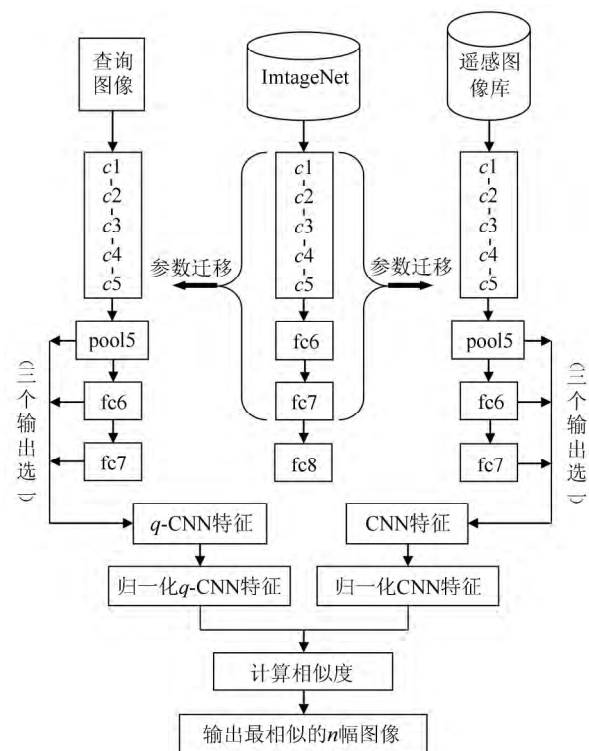


图 1 检索流程

Fig.1 Flowchart of Retrieval

3 实验结果及分析

3.1 实验数据和评估标准

预训练卷积神经网络采用 ILSVRC2012 数据集,使用 VGG 小组开发的 MatConvNet^[18] 提取预训练网络模型,并将 UC-Merced^[19] 和 WHU-RS^[20] 作为目标数据集。UC-Merced 收集了航空正射图像,总共 21 类场景,每类有 100 幅图像。WHU-RS 收集了从 Google Earth 下载的图像,总共 19 类场景,每类有 50 幅图像。图 2 第

1 行显示 UC-Merced 示例图像,第 2 行显示 WHU-RS 示例图像。

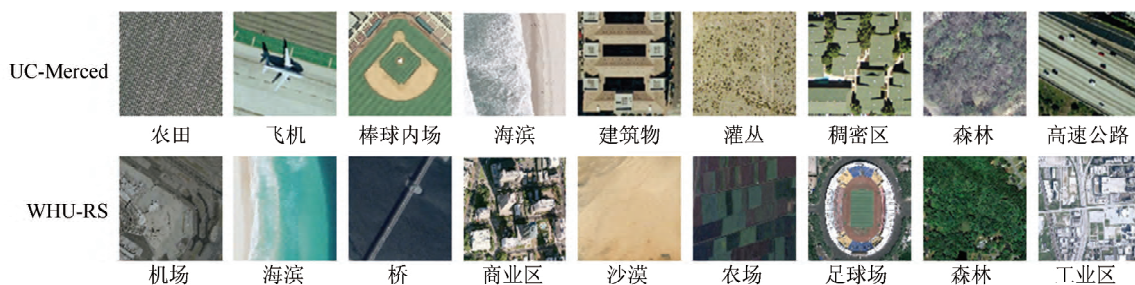


图 2 UC-Merced 和 WHU-RS 示例图像

Fig.2 Examples of the UC-Merced Dataset and the WHU-RS Dataset

为了评价检索的有效性,实验计算了图像的平均准确率(mean average precision, mAP)和平均归一化修改检索等级(average normalize modified retrieval rank, ANMRR)^[6],mAP 值越大,检索的平均准确率越高;ANMRR 取值越小,表示检索出来的相关图像越靠前。实验中还比较了采用图像检索中应用广泛的性能评价准则查准率-查全率曲线。

3.2 mAP 比较

表 2 和表 3 分别显示了 UC-Merced 和 WHU-RS 数据集中不同特征检索图像的 mAP 值。由表 2 和表 3 可以看出,每种网络的 fc6 对应的 mAP 值在 3 层中都最好。4 种网络结构中,CNN-M、CNN-16 和 CNN-19 都在 CNN-Alex 上进行改进,三者的 mAP 值也相应得到提高。

表 2 UC-Merced 不同特征的 mAP/%

Tab.2 mAPs for Different Features on the UC-Merced Dataset/%

类别	pool5	fc6	fc7
CNN-Alex	45.9	52.4	49.3
CNN-M	50.6	55.8	54.9
CNN-16	53.6	55.3	53.3
CNN-19	52.3	54.6	52.0
BoVW[6]		30.2	

表 3 WHU-RS 不同特征的 mAP/%

Tab.3 mAPs for Different Features on the WHU-RS Dataset/%

类别	pool5	fc6	fc7
CNN-Alex	55.1	62.3	62.2
CNN-M	59.2	65.6	64.6
CNN-16	58.1	64.5	63.3
CNN-19	56.6	62.5	60.8
BoVW [6]		38.9	

文献[6]使用视觉词袋(bag of visual words, BoVW)进行遥感图像检索,并取得良好的性能。

表 2 中,CNN 特征最差的 mAP 值(CNN-Alex 中的 pool5)比 BoVW 的 mAP 值提高了 15.7%,最好的 mAP 值(CNN-M 中的 fc6)比 BoVW 的 mAP 值提高了 25.6%。表 3 中,CNN 特征最差的 mAP 值(CNN-Alex 中的 pool5)比 BoVW 的 mAP 值提高了 16.2%,最好的 mAP(CNN-M 中的 fc6)比 BoVW 的 mAP 值提高了 26.7%。可见遥感图像检索中,与采用 BoVW 的方法相比,采用 CNN 特征明显改善图像的检索性能。由于 CNN 特征中 fc6 对应的结果最好,因此后面实验中,选用 fc6 的输出结果。并将 4 种网络中 fc6 的输出值分别记为 CNN-Alex 特征、CNN-M 特征、CNN-16 特征和 CNN-19 特征。

3.3 不同类别图像检索结果比较

图 3 和图 4 分别显示了 UC-Merced 和 WHU-RS 数据集中不同类别图像的检索结果。由图 3 可以看出,UC-Merced 中每类图像 CNN 特征的 mAP 值都优于 BoVW 的 mAP 值。四种 CNN 特征的 mAP 值都比较接近,其中 CNN-Alex 特征检索性能相对较差,CNN-M 特征在大多数类别中的检索性能都比较好,CNN-16 特征和 CNN-19 特征应用于遥感图像检索时,有些类别图像检索性能突出,比如海港和立交桥;而有些类别图像检索性能表现较差,比如棒球内场,从而导致 CNN-16 和 CNN-19 的平均 mAP 值低于 CNN-M。

图 4 和图 3 结果比较类似。另外,从图 4 中可以看出,在 WHU-RS 数据集中,对于绝大多数类别图像来说,用 CNN 特征检索的 mAP 值优于用 BoVW 进行检索的 mAP 值。但是对于有些图像类别,使用 BoVW 和使用 CNN 特征的检索性能接近,甚至优于 CNN 特征(比如机场和牧场),这可能与此类图像与 ImageNet 中图像相差较大,导致泛化能力降低有关。

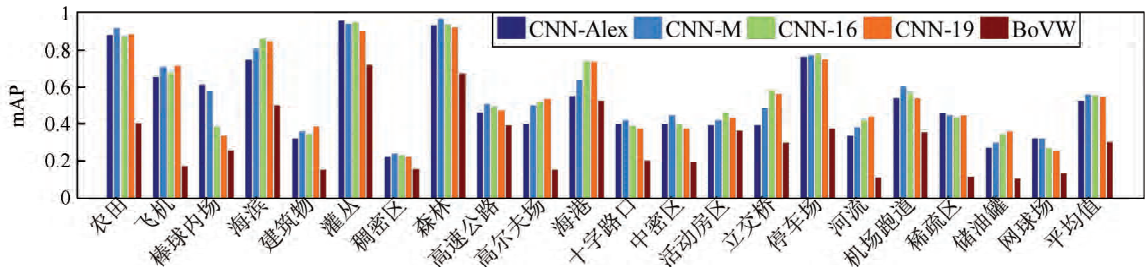


图 3 UC-Merced 每类图像不同特征的 mAP

Fig.3 Per Class mAPs for Different Features on the UC-Merced Dataset

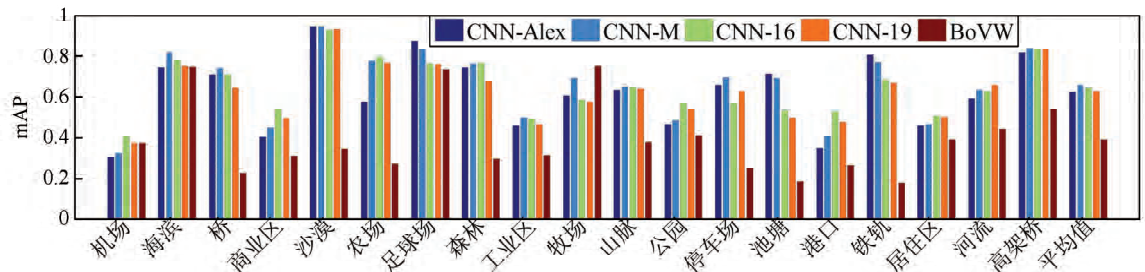


图 4 WHU-RS 每类图像不同特征的 mAP

Fig.4 Per Class mAPs for Different Features on the WHU-RS Dataset

3.4 特征维数和 ANMRR 比较

表 4 比较了 UC-Merced 和 WHU-RS 数据集中不同特征的特征维数和 ANMRR 值。4 种 CNN 特征中,以 CNN-M 特征的 ANMRR 值最小,即检索的相关图像排序更靠前。文献[1]是基于全局形态纹理描述子的遥感图像检索,它的 ANMRR 值为 0.575,只比文献[6]减少了 1.6%,表明仅在浅层特征的基础上进行改进对检索性能的提高幅度较小。

表 4 特征维数和 ANMRR 的比较

Tab.4 Feature Dimensions and ANMRRs for Different Features

类别	特征维数	ANMRR	ANMRR
		(UC-Merced)	(WHU-RS)
CNN-Alex	4 096	0.405	0.308
CNN-M	4 096	0.370	0.278
CNN-16	4 096	0.374	0.291
CNN-19	4 096	0.380	0.308
BoVW [6]	150	0.601	0.525
BoVW [7]	15 000	0.591	0.492
文献[1]	62	0.575	—

与文献[1,6]相比,虽然 CNN 特征维数较高,但是 CNN 特征的 ANMRR 值明显优于浅层特征的特征的 ANMRR 值。以 UC-Merced 为例,CNN 特征中最糟糕的结果(CNN-Alex 特征的 ANMRR 值)比 BoVW 和全局形态纹理描述子的结果分别降低了 18.6%和 17%,CNN 特征中最好

的结果(CNN-M 特征的 ANMRR 值)比 BoVW 和全局形态纹理描述子的结果分别降低了 22.1%和20.5%。这表明通过预训练的卷积神经网络提取遥感图像中的 CNN 特征检索到的相关图像排序更靠前。

3.5 查准率-查全率曲线比较

图 5 比较了 UC-Merced 和 WHU-RS 中不同特征的查准率-查全率曲线。从图 5 可以看出,CNN 特征的检索效果优于 BoVW 的检索效果。对于 BoVW 而言,当返回图像数目由 2 逐步增大到 20 时,检索性能明显降低。而 CNN 特征检索的效果则更稳定,即使返回的图像数目较多,曲线下降的速率仍然比较平缓。对于四种 CNN 特征而言,当检索返回图像数目较少时,四种 CNN 特征的结果比较接近。图 5(a)中,随着返回图像数目的增多,3 种改进的 CNN 特征的检索性能优于 CNN-Alex 特征,其中以 CNN-M 特征的检索效果最好。图 5(b)中,CNN-Alex 特征和 CNN-19 特征的检索性能很接近,CNN-M 特征的优势更明显。

一系列实验结果可以看出,从 ImageNet 数据集上预训练的卷积神经网络提取遥感图像的 CNN 特征,用于图像检索能够有效提高检索性能。CNN 特征在 mAP、ANMRR 和查准率-查全率曲线上优于传统的基于浅层特征的检索方法。

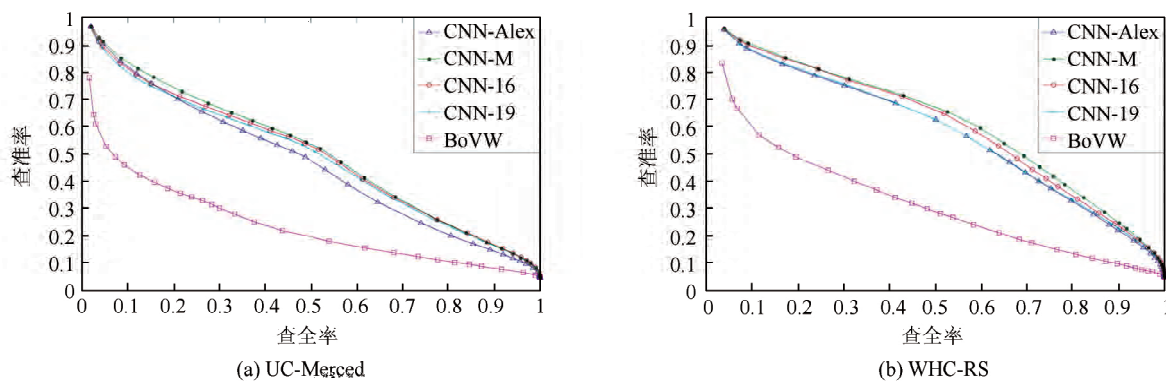


图5 UC-Merced 和 WHU-RS 数据集查准率-查全率曲线

Fig.5 Precision-Recall Curves for Different Features on the UC-Merced Dataset and WHU-RS Dataset

4 结 语

ImageNet 数据集规模大,基于其训练出来的网络具有较强的泛化能力,可以有效地迁移到遥感图像数据集中。本文将 ImageNet 数据集上预训练的四中卷积神经网络(CNN-Alex、CNN-M、CNN-16 和 CNN-19)用于遥感图像检索,提取遥感图像 pool5、fc6 和 fc7 三层的输出值作为 CNN 特征。通过 UC-Merced 和 WHU-RS 数据集上的一系列实验结果表明:pool5、fc6 和 fc7 三层中,fc6 的检索结果最好;4 种网络中,CNN-M 的检索性能更突出。与浅层特征相比,CNN 特征检索的图像准确率更高,相关图像排序更靠前;当返回图像数目增多时,CNN 特征的检索结果更稳定。以 UC-Merced 数据集为例,CNN 特征 mAP 值提高的最小幅度为 15.7%,最大幅度为 25.6%;ANMRR 值减少的最小幅度为 17%,最大幅度为 22.1%。因此,将预训练的卷积神经网络用于提取遥感图像的 CNN 特征能够有效地改进遥感图像的检索性能。下一步考虑通过数据增强等方式增加遥感图像的数量,在遥感图像库中训练卷积神经网络来实现图像的分类和检索等工作。

参 考 文 献

- [1] Aptoula E. Remote Sensing Image Retrieval with Global Morphological Texture Descriptors [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52(5): 3 023-3 034
- [2] Bretschneider T, Cavet R, Kao O. Retrieval of Remotely Sensed Imagery Using Spectral Information Content[C]. The 22nd IEEE International Conference of Geoscience and Remote Sensing Symposium, Toronto, Canada, 2002
- [3] Scott G, Klaric M, Davis C, et al. Entropy-Bal-

- anced Bitmap Tree for Shape-based Object Retrieval from Large-Scale Satellite Imagery Databases [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2011, 49(5): 1 603-1 616
- [4] Demir B, Bruzzone L. A Novel Active Learning Method in Relevance Feedback for Content-based Remote Sensing Image Retrieval[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(9): 2 323-2 334
- [5] Liu T, Zhang L, Li P, et al. Remotely Sensed Image Retrieval Based on Region-Level Semantic Mining[J]. *EURASIP Journal on Image and Video Preprocessing*, 2012, 4(1): 1-11
- [6] Yang Y, Newsam S. Geographic Image Retrieval Using Local Invariant Features[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2013, 51(2): 818-832
- [7] Yang Jin, Liu Jianbo, Dai Qin. An Improved Remote Sensing Image Retrieval Method Based on Bag of Word Framework [J]. *Geomatics and Information Science of Wuhan University*, 2014, 39(9): 1 109-1 113(杨进, 刘建波, 戴芹. 一种改进包模型的遥感图像检索方法[J]. 武汉大学学报·信息科学版, 2014, 39(9): 1 109-1 113)
- [8] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[C]. The 26th Conference on Neural Information Processing Systems, Nevada, US, 2012
- [9] Zeiler M D, Fergus R. Visualizing and Understanding Convolutional Networks[C]. The 13th European Conference on Computer Vision, Zurich, Switzerland, 2014
- [10] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C]. The 3rd International Conference on Learning Representations, San Diego, Canada, 2015
- [11] Donahue J, Jia Y, Vinyals O, et al. Decaf: A Deep

- Convolutional Activation Feature for Generic Visual Recognition[C]. The 31st International Conference on Machine Learning, Beijing, China, 2014
- [12] Oquab M, Bottou L, Laptev I, et al. Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks[C]. The 27th IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014
- [13] Chatfield K, Simonyan K, Vedaldi A, et al. Return of the Devil in the Details: Delving Deep into Convolutional Networks[C]. The 25th British Machine Vision Conference, Nottingham, England, 2014
- [14] Penatti O A B, Nogueira K, Santos J A D. Do Deep Features Generalize from Everyday Objects to Remote Sensing and Aerial Scenes Domains ? [C]. The IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, 2015
- [15] 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): 14 680-14 707
- [16] Ng J, Yang F, Davis L. Exploiting Local Features from Deep Networks for Image[C]. The IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, 2015
- [17] Babenko A, Slesarev A, Chigorin A, et al. Neural Codes for Image Retrieval[C]. The 13th European Conference on Computer Vision, Zurich, Switzerland, 2014
- [18] Vedaldi A, Lenc K. MatConvNet: Convolutional Neural Networks for MATLAB[C]. The 23rd ACM International Conference on Multimedia, Brisbane, Australia, 2015
- [19] Yang Y, Newsam S. Bag-of-Visual-Words and Spatial Extensions for Land-Use Classification[C]. The 18th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, US, 2010
- [20] Xia G S, Yang W, Delon J, et al. Structural High-Resolution Satellite Image Indexing. In *Processings of the ISPRS, TC VII Symposium Part A: 100 Years ISPRS-Advancing Remote Sensing Science* [C]. ISPRS TC VII Symposium-100 Years ISPRS 38, Vienna, Austria, 2010

Remote Sensing Image Retrieval Using Pre-trained Convolutional Neural Networks Based on ImageNet

GE Yun^{1,2} JIANG Shunliang¹ YE Famao¹ XU Qingyong¹ TANG Yiling¹

¹ Information Engineering School, Nanchang University, Nanchang 330031, China

² Software School, Nanchang Hangkong University, Nanchang 330063, China

Abstract: High resolution remote sensing images have complicated content and abundant detail information. Large semantic gaps will occur as such images are difficult to describe using traditional shallow features. This paper proposes a method using four different CNNs pre-trained on ImageNet to in remote sensing image retrieval. High-level features are extracted from different layers of the four CNNs. A Gaussian normalization method is adopted to normalize high-level features, and Euclidean distance is used as the similarity measurement. A serial of experiments carried on the UC-Merced and WHU-RS datasets show that CNN-M feature achieves the best retrieval performance with CNN features. Compared with the visual bag of words and global morphological texture descriptors, the mean average precision of CNN features was 15.7%-25.6% higher than that of shallow features. The average normalized modified retrieval rank of CNN features was 17%-22.1% lower than that of shallow features. Therefore the pre-trained convolutional neural network is effective for high-resolution remote sensing image retrieval.

Key words: remote sensing image; retrieval; convolutional neural networks; pre-trained

First author: GE Yun, PhD candidate, lecturer, specializes in remote sensing image retrieval. E-mail: geyun@nchu.edu.cn

Corresponding author: JIANG Shunliang, PhD, professor. E-mail: jiangshunliang@ncu.edu.cn

Foundation support: The National Natural Science Foundation of China, No. 41261091; the Youth Fund Project of Education Department of Jiangxi, No. GJJ13482; the National Natural Science Foundation of Jiangxi, No. 20151BAB207062.