

基于高层特征图组合及池化的高分辨率遥感图像检索

葛芸^{*①} 马琳^① 江顺亮^② 叶发茂^②

^①(南昌航空大学软件学院 南昌 330063)

^②(南昌大学信息工程学院 南昌 330031)

摘要: 高分辨率遥感图像内容复杂, 提取特征来准确地表达图像内容是提高检索性能的关键。卷积神经网络(CNN)迁移学习能力强, 其高层特征能够有效迁移到高分辨率遥感图像中。为了充分利用高层特征的优点, 该文提出一种基于高层特征图组合及池化的方法来融合不同CNN中的高层特征。首先将高层特征作为特殊的卷积层特征, 进而在不同输入尺寸下保留高层输出的特征图; 然后将不同高层输出的特征图组合成一个更大的特征图, 以综合不同CNN学习到的特征; 接着采用最大池化的方法对组合特征图进行压缩, 提取特征图中的显著特征; 最后, 采用主成分分析(PCA)来降低显著特征的冗余度。实验结果表明, 与现有检索方法相比, 该方法提取的特征在检索效率和准确率上都有优势。

关键词: 遥感图像检索; 迁移学习; 高层特征图; 组合; 池化

中图分类号: TP751.1

文献标识码: A

文章编号: 1009-5896(2019)10-2487-08

DOI: 10.11999/JEIT190017

The Combination and Pooling Based on High-level Feature Map for High-resolution Remote Sensing Image Retrieval

GE Yun^① MA Lin^① JIANG Shunliang^② YE Famao^②

^①(School of Software, Nanchang Hangkong University, Nanchang 330063, China)

^②(School of Information Engineering, Nanchang University, Nanchang 330031, China)

Abstract: High-resolution remote sensing images have complex visual contents, and extracting feature to represent image content accurately is the key to improving image retrieval performance. Convolutional Neural Networks (CNN) have strong transfer learning ability, and the high-level features of CNN can be efficiently transferred to high-resolution remote sensing images. In order to make full use of the advantages of high-level features, a combination and pooling method based on high-level feature maps is proposed to fuse high-level features from different CNNs. Firstly, the high-level features are adopted as special convolutional features to preserve the feature maps of the high-level outputs under different input sizes, and then the feature maps are combined into a larger feature map to integrate the features learned by different CNNs. The combined feature map is compressed by max-pooling method to extract salient features. Finally, the Principal Component Analysis (PCA) is utilized to reduce the redundancy of the salient features. The experimental results show that compared with the existing retrieval methods, the features extracted by this method have advantages in retrieval efficiency and precision.

Key words: Remote sensing image retrieval; Transfer learning; High-level feature map; Combination; Pooling

1 引言

近年来, 高分辨率遥感(High-Resolution

Remote Sensing, HRRS)图像由于识别地物能力强和信息准确等因素越来越受到重视。为了高效地利用遥感数据, 基于内容的HRRS图像检索成为了研究热点和难点之一, 其中提取有效的特征来准确表达HRRS图像的内容是提高检索性能的关键。

传统的基于内容的HRRS图像检索主要是提取图像的颜色、纹理和形状等底层特征, 但底层特征难以表达图像的高层语义信息, 为了提高特征的表达能力, 通常采用以下方法进行改进: (1)采用反

收稿日期: 2019-01-09; 改回日期: 2019-06-18; 网络出版: 2019-06-25

*通信作者: 葛芸 geyun@nchu.edu.cn

基金项目: 国家自然科学基金(41801288, 41261091, 61662044, 61663031, 61762067)

Foundation Items: The National Natural Science Foundation of China (41801288, 41261091, 61662044, 61663031, 61762067)

馈机制^[1],该方法根据对检索结果的反馈不断优化检索性能;(2)构建中层特征^[2],中层特征是在局部特征的基础上进一步构建出的高级特征,具有更好的抽象表达能力;(3)融合多种特征,多特征融合有效结合不同特征的优点,从而更加全面地表达遥感图像的内容。文献[3]通过线性加权的方式融合了颜色和纹理特征;文献[4]提出3层图的学习方法来融合局部特征和全局特征,进而提高检索性能。

上述方法对改进特征表达起到重要作用,但采用的特征都是浅层的手工设计特征。手工特征容易受到人为因素干扰,表达图像能力受限。深度卷积神经网络(Convolutional Neural Network, CNN)能够自动学习图像不同层次的特征,在图像识别领域的应用广泛^[5,6],其中,在大规模数据集(比如ImageNet)上充分训练的CNN^[7-11]具有很强的泛化能力,通过迁移学习能够有效应用于其他数据集的识别任务。迁移学习通常是将预训练的CNN或者微调后的CNN的参数迁移到其他数据集中,并提取迁移特征来表达图像的内容。迁移特征主要包括全连接层特征和卷积层特征,全连接层特征是高层输出的抽象特征,又称为高层特征,卷积层特征表达了图像的局部信息,可以看成是局部特征。

CNN能够有效地迁移到HRRS图像识别领域。Castelluccio等人^[12]针对不同的CNN,对比了迁移学习和重新训练两种方法,实验结果表明,迁移学习比重训练CNN更适用于小规模遥感数据集。在HRRS图像检索中,Alias等人^[13]和Napoletoano^[14]提取不同CNN中的全连接层特征,研究表明全连接层特征的检索性能明显优于手工特征。文献[15,16]比较了不同CNN中的全连接层特征和卷积层特征,实验结果表明,直接提取的全连接层特征的检索性能优于采用复杂编码后的卷积层特征的性能。文献[15]还提出一种低维度特征(Low Dimensional CNN, LDCNN),提高了图像的检索效率。文献[16]对全连接层特征采用了多小块均值池化的方法来提高特征表达能力。因此,CNN中的迁移特征,尤其是全连接层特征能较好地应用于HRRS图像检索。

在上述CNN迁移到HRRS图像的方法中,为了提取图像的全连接层特征,输入图像的尺寸必须调整为CNN需要的默认输入尺寸,然后将全连接层的3维输出值转换为1维的特征向量。文献[16]采用了多小块池化的方法来改进全连接层特征,但全连接层特征提取过程依然是传统方法,为了提取一幅图像多个分块的全连接层特征,需要多次输入才能完成,导致特征提取过程相对复杂。

因此CNN迁移到HRRS图像检索的研究中,缺

乏对高层特征(比如说全连接层特征)的深入研究,而高层特征相比其他层次特征而言,具有更抽象的表达能力,泛化能力更强,迁移学习效果更好。全连接层特征实际上可以看成一种特殊的卷积层特征^[17]。文献[17]的研究得出全连接层特征作为卷积层处理后有很多优点,比如全连接层的输出不再受输入图像尺寸的限制,并可以保留原始输入图像的空间信息。

受文献[17]的启发,本文将高层特征当作卷积层处理,高层的输出值不再是直接转换为特征向量,而是保留其特征图的输出形式。为了结合不同CNN中高层特征的优点,在不同的输出尺寸下,本文研究基于特征图组合及池化的方法来融合多种不同的高层特征。首先将不同CNN高层输出的特征图以第3维(即特征图数目)为基准组合成一个更大的特征图,以期在保留图像空间信息的基础上组合所有卷积核学习到的特征;然后对组合的特征图采用最大池化的方法进行压缩,提取出更适合表达HRRS图像的显著特征。最后,对池化后的特征采用主成分分析(Principal Component Analysis, PCA)的方法进行降维。研究表明,本组合及池化的方法能够有效结合不同CNN特征的优点,从而提高HRRS图像的检索性能。

2 融合高层特征的HRRS图像检索

HRRS图像检索的目的是在遥感图像数据集中检索出与查询图像类别相同的图像。本文使用示例图像查询的检索方式,即输入一幅查询图像,提取查询图像的融合特征,并与HRRS数据集中的融合特征进行相似性度量,根据计算的相似度进行排序,返回内容相似的图像。图像检索流程如图1所示,由于不同CNN的结构相差较大,为了统一表示,使用底层、中层和高层来简要说明CNN中的不同层次。

首先将预训练CNN的参数分别迁移到HRRS数据集 M 和查询图像 q (参数迁移包含直接迁移和微调CNN后再迁移两种方式)。然后,针对 M 和 q 提取 k 种CNN的高层特征,并对 k 种高层特征通过组合及池化的方式得到 M 和 q 的融合特征。为了使特征向量内部的不同分量在相似性度量时具有相同的地位,对提取的融合特征进行内部归一化;为了使不同的特征在相似性度量时具有相同的地位,对计算的相似度进行外部归一化,归一化方法采用常用的高斯归一化。最后,对相似度进行排序,并按需求输出相似的图像。该图像检索中的关键是提取融合特征来表达HRRS图像的内容。

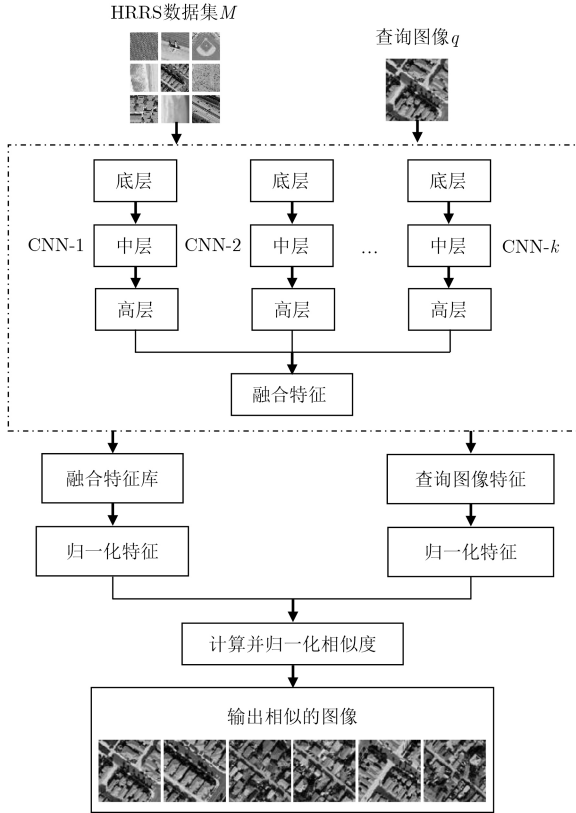


图1 图像检索流程

2.1 提取高层特征

CNN中的全连接层特征是一种典型的高层特征，全连接层可以看成一种特殊的卷积层，输出的特征图通过卷积的方式获得

$$\mathbf{x}_j = r \left(\sum_{i=1}^m \mathbf{y}_i * \mathbf{w}_{ij} + \mathbf{b}_j \right) \quad (1)$$

其中， r 为激活函数， $\mathbf{y}_i$ 为上一个层次输出的特征图， $\mathbf{w}_{ij}$ 为上层第 i 个特征图到该层第 j 个特征图的权重， m 为上层输出的特征图数目， $\mathbf{b}_j$ 是该层第 j 个特征图的偏置， $\mathbf{x}_j$ 是该层输出的第 j 个特征图。

CNN的高层特征即为高层输出的特征图

$$\mathbf{f} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_j, \dots, \mathbf{x}_C] \quad (2)$$

其中， $\mathbf{x}_j$ 为第 j 个特征图， C 为特征图的数目。

该方法提取的高层特征不受输入图像尺寸影响。在深度学习框架MatConvNet^[18]中，CNN的全连接层按照卷积层设计，表1给出了在MatConvNet中两种输入图像尺寸下不同高层特征的输出值，输入图像尺寸包括CNN的默认输入尺寸(AlexNet^[7]为 227×227 ，其他CNN^[8-11]为 224×224)和图像的原始输入尺寸，对于UC-Merced^[2]和WHU-RS7^[19]这两种常用的HRRS数据集，原始输入尺寸分别为 256×256 和 600×600 。高层特征指以AlexNet^[7]和VGG16^[9]等为代表的全连接层(fc)特征，以及以

表1 不同输入图像尺寸下高层CNN特征的输出值

输入图像尺寸	fc	G-pool5	R-pool5
默认尺寸	$1 \times 1 \times 4096$	$1 \times 1 \times 1024$	$1 \times 1 \times 2048$
$256 \times 256 \times 3$ (UC-Merced)	$2 \times 2 \times 4096$	$2 \times 2 \times 1024$	$2 \times 2 \times 2048$
$600 \times 600 \times 3$ (WHU-RS)	$13 \times 13 \times 4096$	$12 \times 12 \times 1024$	$13 \times 13 \times 2048$

GoogLeNet^[10]和ResNet^[11]为代表的最后的均值池化层特征(分别记为G-pool5和R-pool5)。从表1可以看出，当输入图像尺寸较大时，相应层次输出的特征图的尺寸也增大。

传统提取高层特征的方法是提取图像在默认输入尺寸下的高层输出值，并直接将其转换为特征向量，本文则保持这3维的输出值不变，以便对不同的高层特征进行融合。

2.2 融合高层特征

(1) 组合高层特征图

不同CNN的高层学习到的特征图也不同，本文提出组合高层特征图的方法来提高特征的表达能力。为了保留不同高层特征的所有信息，在组合不同的高层特征时，以第3维中维数最大的值为基准进行组合，较少的维数则通过补0的方式进行填充，以期在保留图像空间信息的基础上将不同的特征图组合形成一个更大的特征图。

假设组合 k 个高层特征 $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_i, \dots, \mathbf{f}_k\}$ ，最大的特征图数目为

$$C^{\max} = \max C_i \quad (3)$$

其中， C_i 为第 i 个高层特征 $\mathbf{f}_i$ 的特征图数目， $C^{\max}$ 为其中最大的特征图数目。

为了组合所有的特征图，通过补0的方式将所有特征的特征图数目都扩充到 $C^{\max}$

$$\mathbf{f}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{ij}, \dots, \mathbf{x}_{iC^{\max}}] \quad (4)$$

其中， $\mathbf{x}_{ij}$ 为第 i 个高层特征 $\mathbf{f}_i$ 中的第 j 个特征图

$$\mathbf{x}_{ij} = \begin{cases} \mathbf{x}_{ij}, & 0 \leq j \leq C_j \\ 0, & C_j \leq j \leq C^{\max} \end{cases} \quad (5)$$

将所有高层特征对应的特征图采用分块对角矩阵的形式进行合并

$$\mathbf{X}_j = \begin{pmatrix} \mathbf{x}_{1j} & & & \\ & \ddots & & \\ & & \mathbf{x}_{ij} & \\ & & & \ddots \\ & & & & \mathbf{x}_{kj} \end{pmatrix} \quad (6)$$

其中， $\mathbf{X}_j$ 是将所有 k 个特征中第 j 个特征图进行组合后得到大的特征图。

合并之后的特征为

$$\mathbf{F} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_j, \dots, \mathbf{X}_{C^{\max}}] \quad (7)$$

所有特征合并后的特征图尺寸为 $W \times H \times C^{\max}$, W 和 H 的计算公式为

$$W = \sum_{i=1}^k w_i, H = \sum_{i=1}^k h_i \quad (8)$$

其中, w_i 是第 i 个特征图的宽度, h_i 是第 i 个特征图的高度。

图2以组合两种高层特征为例进行说明。令其中一种高层特征的尺寸为 $W^a \times H^a \times C^a$, 另一种高层特征的尺寸为 $W^b \times H^b \times C^b$ 。 W^a 和 W^b 代表两种特征图的宽度, H^a 和 H^b 代表两种特征图的高度, C^a 和 C^b 则代表两种特征图的数目。特征组合时为了综合所有卷积核中的参数, 以第3维中较大的值为基准进行组合。令 $C^c = \max(C^a, C^b)$, 组合特征后的尺寸为 $(W^a + W^b) \times (H^a + H^b) \times C^c$ 。

这种组合方式将不同卷积核学习到的特征都组合在一起, 特征表达更加全面, 但也容易导致特征维数过高, 因此需要对组合特征进行压缩处理。

(2) 对组合的高层特征进行池化

CNN中卷积层后面通常跟着池化层, 池化层能够对特征图中的神经元进行有效压缩, 并能够增强特征的鲁棒性, 在CNN中起到重要作用。因为组合后的特征图可以看成是一个特殊的卷积层, 所以采用最大池化的方法来提取特征图中的显著特征, 从而达到压缩特征图的目的。

令组合特征 f 的尺寸为 $W \times H \times C$, 其中 W 和 H 分别代表组合特征图的宽度和高度, C 为特征图的数目。针对该组合特征, 采用最大池化方法

$$p_i^{\max} = \max f(\cdot, \cdot, i), i = 1, 2, \dots, C \quad (9)$$

经过池化后, 组合特征降低到 $1 \times 1 \times C$ 维, 将该组合特征直接转换为一个 C 维的特征向量, 因此融合后的特征维数和传统的直接提取高层特征的维数一样。为了进一步降低特征的冗余度, 提高特征检索效率, 可以对该组合特征采用PCA降维的方法进行优化。

3 实验结果及分析

3.1 实验数据和评估标准

实验使用MatConvNet提取基于ImageNet预训练的CNN, 再分别迁移到UC-Merced和WHU-RS中。UC-Merced为航空正射图像, 总共包含21类场景, 每类有100幅图像。WHU-RS是从Google Earth下载的19类图像, 每类有50幅图像。相似性度量标准为欧式距离, 评价标准采用了HRRS图像检索中广泛使用的平均归一化修改检索等级(Average Normalize Modified Retrieval Rank, ANMRR)和平均准确率(mean Average Precision, mAP)。ANMRR值越小, 表明检索性能越好, mAP值越大, 则检索准确率越高。

3.2 不同高层特征融合结果

为了比较不同高层特征的融合结果, 首先选用6种CNN输出的高层特征, 分别为AlexNet, VGGM, VGG16和VGG19全连接层中的fc6, GoogLeNet和ResNet(选用ResNet50)的最后池化层pool5, 为了简明表示, 上述特征依次简记为A, M, 16, 19, G和R, 组合及池化(Combination and Pooling)的融合简记为CoP; 然后在不同数据集上分别进行6种特征中任选2种、任选3种、任选4种、任选5种, 以及选择所有6种特征的融合; 最后将每类融合中检索最好的结果在图3和图4中列出, CoP(A_M)表示融合AlexNet和VGGM的高层特征, 其他特征融合的记法依此类推, CoP(ALL)是指融合6种特征。图3和图4还显示了单一特征的检索结果, 以及采用直接串联(简记为con)的方式将不同特征连接在一起的检索结果。

图3和图4显示, CoP特征的mAP值优于con特征。CoP特征的维数为4096, 而con特征的维数随着特征数目的增多而增加, con(ALL)的维数达到了19456(将每种特征的维数直接相加), 特征维数越低, 检索效率越高, 所以CoP特征的检索准确率和检索效率都优于con特征。对于CoP特征来说,

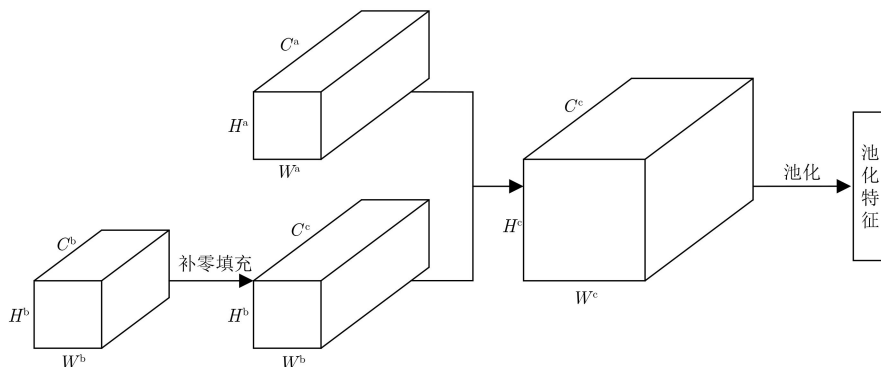


图2 融合高层特征

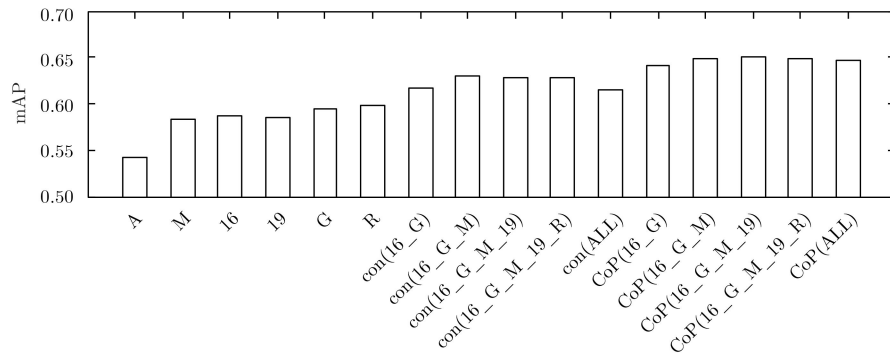


图3 UC-Merced中不同特征检索结果比较

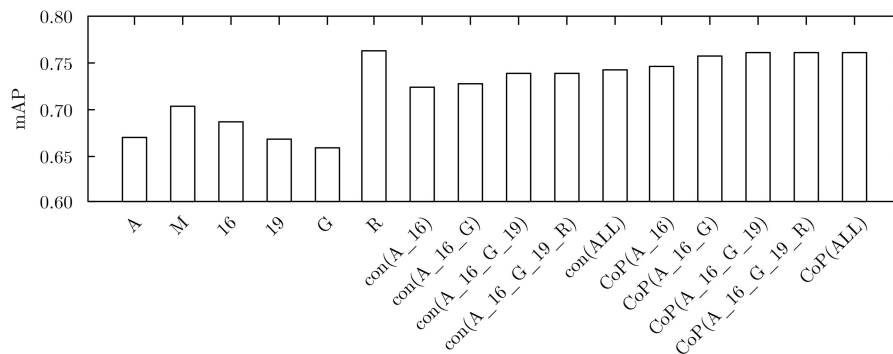


图4 WHU-RS中不同特征检索结果比较

3种、4种、5种和6种特征融合的结果非常接近，并都优于2种特征融合的结果。UC-Merced中，所有的融合特征都优于单一特征，并以4种特征融合的效果最佳。而在WHU-RS中，由于R特征的mAP值明显优于其他特征的mAP值，导致该特征与其他特征融合后的检索结果不如单一的R特征的检索结果。

图3和图4对比发现，不同数据集上的最优特征不同。为了分析其原因，计算了6种特征之间的Pearson相关系数，具体数据见表2和表3。以2种特征融合为例进行说明，UC-Merced中，检索结果最好的两种单一特征为R与G，但这两个特征的相关系数较大，导致特征融合的结果不是最优；最优的融合特征是CoP(16_G)，16和G的相关系数偏低且呈负相关，检索结果比较接近且都比较好，因此融合效果好。WHU-RS中，检索结果最好的两种单一特征为R与M，这两种特征的相关系数为0.0084，

且R的检索结果明显优于M，因此这类呈正相关且检索结果相差较大的特征融合效果并不佳；最优的融合特征是CoP(A_16)，A和16的相关系数很低，且这两种特征的检索结果比较接近，因此融合效果最好。其他数目的特征进行融合时有类似的结论，即当不同特征的相关系数较低且检索结果相近时，融合的效果较好。

对于上述6种CNN，融合3~4种高层特征，检索性能基本能达到最优，并且大多数特征融合通常选用的特征数目为2~4种。因此，选择图3和图4中对应的2种、3种和4种特征融合进行后续实验。

3.3 不同输入尺寸下的特征融合结果

表4比较了CoP特征在默认输入尺寸和原始输入尺寸下的检索结果。UC-Merced中CoP特征的检索结果在原始输入尺寸优于默认输入尺寸；WHU-RS中CoP特征的检索结果在默认输入尺寸优于原始输入尺寸。这是因为WHU-RS中原始输入尺寸较

表2 UC-Merced中特征的相关系数

特征	A	M	16	19	G
M	-0.0037				
16	0.0006	0.0028			
19	-0.0023	0.0053	0.4817		
G	0.0012	-0.0063	-0.0086	-0.0100	
R	-0.0100	0.0008	-0.0060	-0.0021	0.1175

表3 WHU-RS中特征的相关系数

特征	A	M	16	19	G
M	-0.0080				
16	-0.0009	0.0027			
19	0.0001	0.0051	0.4762		
G	-0.0024	-0.0038	-0.0110	-0.0093	
R	-0.0045	0.0084	-0.0069	-0.0022	0.1138

大,组合特征图时填充0的数目随之增多,从而影响了融合效果。

因此当输入图像的原始尺寸和默认尺寸比较接近时,选用原始尺寸下提取的CoP特征检索性能更好,而当输入图像的原始尺寸远大于默认尺寸时,则在默认尺寸下提取的CoP特征检索性能更好。

3.4 PCA降维结果

根据表4的结果,PCA降维实验中,UC-Merced中选择原始输入尺寸下的融合特征,WHU-RS中选择默认输入尺寸下的融合特征。图5比较了降维的维数从8一直到2048的结果,4096是没有降维的结果。

图5中,8维特征的检索性能明显下降,32维特征的检索结果最优。UC-Merced中,当维数降到32、64和128时,特征的检索结果优于降维前的结果,当维数为256到2048时,特征检索结果与

4096维时的结果比较接近。WHU-RS中,特征维数降到16、32和64时,检索结果优于降维前的结果,其他维数的特征检索结果与4096维的结果比较接近。因此,PCA降维能够有效降低融合特征的冗余信息。

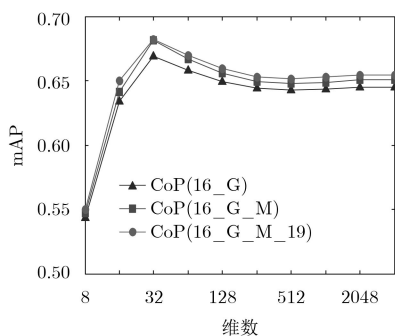
3.5 微调结果

为了进一步提高CNN的迁移学习能力,对CNN模型采用HRRS数据集RSSCN7^[20]进行微调。RSSCN7是从Google Earth下载的7类图像,每类包含了400幅图像。由于RSSCN7数据规模较小,微调CNN时,前面卷积层的参数固定不变,主要微调后面高层的参数,权重更新采用随机梯度下降的方式,冲量为0.9,权重衰减值为0.0005,一共训练30次,前面10次的学习率设置为0.001,后面20次为0.0001。

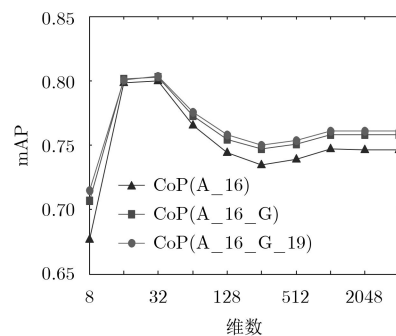
表5比较了从微调(Fine-Tune, FT)的CNN中提

表4 不同输入尺寸CoP特征检索结果比较

数据集	特征	默认尺寸		原始尺寸	
		ANMRR	mAP	ANMRR	mAP
UC-Merced	CoP(16_G)	0.2898	0.6411	0.2880	0.6446
	CoP(16_G_M)	0.2834	0.6485	0.2832	0.6504
	CoP(16_G_M_19)	0.2834	0.6496	0.2805	0.6544
WHU-RS	CoP(A_16)	0.2007	0.7466	0.2330	0.7116
	CoP(A_16_G)	0.1891	0.7582	0.2319	0.7125
	CoP(A_16_G_19)	0.1875	0.7610	0.2318	0.7124



(a) UC-Merced



(b) WHU

图5 不同CoP特征PCA降维结果

表5 UC-Merced中微调CoP特征检索结果比较

数据集	特征	默认尺寸		原始尺寸	
		ANMRR	mAP	ANMRR	mAP
UC-Merced	CoP(16_G)-FT	0.2738	0.6602	0.2777	0.6566
	CoP(16_G_M)-FT	0.2642	0.6716	0.2678	0.6683
	CoP(16_G_M_19)-FT	0.2604	0.6767	0.2561	0.6822
WHU-RS	CoP(A_16)-FT	0.1723	0.7809	0.1975	0.7501
	CoP(A_16_G)-FT	0.1582	0.7971	0.1924	0.7559
	CoP(A_16_G_19)-FT	0.1519	0.8048	0.1879	0.7615

取的CoP特征的检索结果。与表4相比,所有CoP特征微调后的检索性能都有所提高;UC-Merced中,微调CoP特征的结果优化了1.03%~2.78%,WHU-RS中,微调CoP特征的结果优化了2.84%~4.91%。RSSCN7与WHU-RS都是从GoogLe Earth下载的图像,图像比较接近,因此微调后的CoP特征在WHU-RS中的检索性能提升幅度较大。

3.6 与其他方法比较

表6将本文所提融合特征方法与现有的遥感图像检索中采用的方法进行比较,浅层特征包含了中层特征VLAD(Vector of Locally Aggregated Descriptors)和融合特征3层图;CNN特征中,包含了传统的全连接层特征VGGM-fc和VGG16-fc、采用IFK编码的卷积层特征(VGGM-conv5-IFK和VGG16-conv5-IFK)、低维度特征(LDCNN)、微调后的GoogLeNet中使用多小块均值池化特征(GoogLeNet-MultiPatch-FT)以及在原始输入尺寸下提取的CoP特征。

表 6 UC-Merced中CoP特征与其他特征检索结果比较

	特征	ANMRR	维数
浅层特征	VLAD ^[2]	0.4604	16384
	3层图 ^[4]	0.4317	—
	VGGM-fc ^[14]	0.3780	4096
	VGGM-conv5-IFK ^[15]	0.4580	102400
	VGG16-fc ^[15]	0.3940	4096
CNN特征	VGG16-conv5-IFK ^[15]	0.4070	102400
	LDCNN ^[15]	0.4390	30
	GoogLeNet-MultiPatch-FT ^[16]	0.3140	1024
	GoogLeNet-MultiPatch-FT-PCA ^[16]	0.2850	32
	CoP(16_G)	0.2880	4096
	CoP(16_G_M_19)	0.2805	4096
	CoP(16_G_M_19)-FT	0.2561	4096
	CoP(16_G_M_19)-PCA	0.2577	32

从表6可以得出,3层图的检索性能优于部分CNN特征,但VLAD的检索结果不如CNN特征。CNN特征中,CoP特征的检索性能提升明显,即便在没有进行微调的情况下,CoP的检索结果也优于GoogLeNet_MultiPatch_FT。当CoP特征降维到32时,检索结果比对GoogLeNet-MultiPatch-FT-PCA降维到32时的结果优化了2.73%。

通过一系列实验表明,CoP方法通过融合高层特征的特征图,并使用池化方法对组合特征进行压缩,可以有效结合不同CNN特征的优点,从而提高检索性能。

4 结论

本文通过组合及池化的方法来有效地融合不同CNN中的高层特征。首先将高层特征作为特殊的卷积层特征,保留高层特征输出的3维张量,并将不同的高层特征组合成一个更大的特征图,然后在该特征图的基础上使用池化方法进行压缩,最后对压缩特征使用PCA进行降维。本文融合方法比较了默认输入尺寸和原始输入尺寸下的图像检索性能,并通过微调CNN进一步提高融合特征的表达能力。实验表明,组合及池化的方法可以融合多种不同的CNN特征,当特征之间的相关系数较低且检索结果接近时,特征融合的效果较好;大多数情况下,融合特征的检索性能优于单一特征的检索性能,融合特征经过PCA降维后性能得到提升,并以32维时的提升幅度最大。因此本文融合方法利用卷积层特征的特点以及高层特征的优点,能有效地对不同的高层特征进行融合,从而提高检索性能。

参 考 文 献

- [1] DEMIR B and 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(5): 2323–2334. doi: [10.1109/TGRS.2014.2358804](https://doi.org/10.1109/TGRS.2014.2358804).
- [2] ÖZKAN S, ATEŞ T, TOLA E, *et al.* Performance analysis of state-of-the-art representation methods for geographical image retrieval and categorization[J]. *IEEE Geoscience and Remote Sensing Letters*, 2014, 11(11): 1996–2000. doi: [10.1109/LGRS.2014.2316143](https://doi.org/10.1109/LGRS.2014.2316143).
- [3] 陆丽珍, 刘仁义, 刘南. 一种融合颜色和纹理特征的遥感图像检索方法[J]. *中国图象图形学报*, 2004, 9(3): 328–333. doi: [10.3969/j.issn.1006-8961.2004.03.013](https://doi.org/10.3969/j.issn.1006-8961.2004.03.013).
LU Lizhen, LIU Renyi, and LIU Nan. Remote sensing image retrieval using color and texture fused features[J]. *Journal of Image and Graphics*, 2004, 9(3): 328–333. doi: [10.3969/j.issn.1006-8961.2004.03.013](https://doi.org/10.3969/j.issn.1006-8961.2004.03.013).
- [4] WANG Yuebin, ZHANG Liqiang, TONG Xiaohua, *et al.* A three-layered graph-based learning approach for remote sensing image retrieval[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(10): 6020–6034. doi: [10.1109/TGRS.2016.2579648](https://doi.org/10.1109/TGRS.2016.2579648).
- [5] 郭智, 宋萍, 张义, 等. 基于深度卷积神经网络的遥感图像飞机目标检测方法[J]. *电子与信息学报*, 2018, 40(11): 2684–2690. doi: [10.11999/JEIT180117](https://doi.org/10.11999/JEIT180117).
GUO Zhi, SONG Ping, ZHANG Yi, *et al.* Aircraft detection method based on deep convolutional neural network for remote sensing images[J]. *Journal of Electronics &*

- Information Technology*, 2018, 40(11): 2684–2690. doi: [10.11999/JEIT180117](https://doi.org/10.11999/JEIT180117).
- [6] YE Famao, SU Yanfei, XIAO Hui, *et al.* Remote sensing image registration using convolutional neural network features[J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, 15(2): 232–236. doi: [10.1109/LGRS.2017.2781741](https://doi.org/10.1109/LGRS.2017.2781741).
- [7] KRIZHEVSKY A, SUTSKEVER I, and HINTON G E. ImageNet classification with deep convolutional neural networks[C]. The 25th International Conference on Neural Information Processing Systems, Nevada, USA, 2012: 1097–1105. doi: [10.1145/3065386](https://doi.org/10.1145/3065386).
- [8] 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, UK, 2014. doi: [10.5244/C.28.6](https://doi.org/10.5244/C.28.6).
- [9] SIMONYAN K and ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]. The 3rd International Conference on Learning Representations, San Diego, USA, 2015.
- [10] SZEGEDY C, LIU Wei, JIA Yangqing, *et al.* Going deeper with convolutions[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 1–9. doi: [10.1109/CVPR.2015.7298594](https://doi.org/10.1109/CVPR.2015.7298594).
- [11] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [12] CASTELLUCCIO M, POGGI G, SANSONE C, *et al.* Land use classification in remote sensing images by convolutional neural networks[J]. *Acta Ecologica Sinica*, 2015, 28(2): 627–635.
- [13] ALIAS B, KARTHIKA R, and PARAMESWARAN L. Content based image retrieval of remote sensing images using deep learning with different distance measures[J]. *Journal of Advanced Research in Dynamical and Control Systems*, 2018, 10(3): 664–674.
- [14] NAPOLETANO P. Visual descriptors for content-based retrieval of remote-sensing Images[J]. *International Journal of Remote Sensing*, 2018, 39(5): 1343–1376. doi: [10.1080/01431161.2017.1399472](https://doi.org/10.1080/01431161.2017.1399472).
- [15] ZHOW Weixun, NEWSAM S, LI Congmin, *et al.* Learning low dimensional convolutional neural networks for high-resolution remote sensing image retrieval[J]. *Remote Sensing*, 2017, 9(5): 489. doi: [10.3390/rs9050489](https://doi.org/10.3390/rs9050489).
- [16] HU Fan, TONG Xinyi, XIA Guisong, *et al.* Delving into deep representations for remote sensing image retrieval[C]. The IEEE 13th International Conference on Signal Processing, Chengdu, China, 2016: 198–203. doi: [10.1109/ICSP.2016.7877823](https://doi.org/10.1109/ICSP.2016.7877823).
- [17] SHELHAMER E, LONG J, and DARRELL T. Fully convolutional networks for semantic segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(4): 640–651. doi: [10.1109/TPAMI.2016.2572683](https://doi.org/10.1109/TPAMI.2016.2572683).
- [18] VEDALDI A and LENC K. MatConvNet: Convolutional neural networks for MATLAB[C]. The 23rd ACM International Conference on Multimedia, Brisbane, Australia, 2015: 689–692. doi: [10.1145/2733373.2807412](https://doi.org/10.1145/2733373.2807412).
- [19] HU Fan, XIA Guisong, HU Jingwen, *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. doi: [10.3390/rs71114680](https://doi.org/10.3390/rs71114680).
- [20] ZOU Qin, NI Lihao, ZHANG Tong, *et al.* Deep learning based feature selection for remote sensing scene classification[J]. *IEEE Geoscience and Remote Sensing Letters*, 2015, 12(11): 2321–2325. doi: [10.1109/LGRS.2015.2475299](https://doi.org/10.1109/LGRS.2015.2475299).
- 葛 芸: 女, 1983年生, 博士, 讲师, 研究方向为遥感图像处理与机器学习。
- 马 琳: 女, 1996年生, 硕士生, 研究方向为遥感图像处理与机器学习。
- 江顺亮: 男, 1966年生, 博士, 教授, 博士生导师, 研究方向为算法设计与分析、计算机模拟与仿真、机器视觉。
- 叶发茂: 男, 1978年生, 博士, 副教授, 研究方向为遥感图像处理、计算机图形学、机器学习。