

doi: 10.6046/gtzyyg.2019.01.07
引用格式: 葛芸,江顺亮,叶发茂,等. 聚合 CNN 特征的遥感图像检索[J]. 国土资源遥感,2019,31(1):49–57. (Ge Y, Jiang S L, Ye F M, et al. Aggregating CNN features for remote sensing image retrieval[J]. Remote Sensing for Land and Resources, 2019, 31(1):49–57.)

聚合 CNN 特征的遥感图像检索

葛 芸^{1,2}, 江顺亮¹, 叶发茂¹, 姜昌龙², 陈 英², 唐祎玲¹
(1. 南昌大学信息工程学院, 南昌 330031; 2. 南昌航空大学软件学院, 南昌 330063)

摘要: 针对高分辨率遥感图像检索中手工特征难以准确描述图像的问题, 提出聚合卷积神经网络 (convolutional neural network, CNN) 特征的方法来改进特征表达。首先, 将预训练的 CNN 参数迁移到遥感图像, 并针对不同尺寸的输入图像, 提取表达局部信息的 CNN 特征; 然后, 对该 CNN 特征采用池化区域尺寸不同的均值池化和视觉词袋 (bag of visual words, BoVW) 2 种聚合方法, 分别得到池化特征和 BoVW 特征; 最后, 将 2 种聚合特征用于遥感图像检索。实验结果表明: 合理的输入图像尺寸能提高聚合特征的表达能力; 当池化区域为特征图的 60%~80% 时, 绝大多数池化特征的结果优于传统均值池化方法的结果; 池化特征和 BoVW 特征的最优平均归一化修改检索等级值比手工特征分别降低了 27.31% 和 21.51%, 因此, 均值池化和 BoVW 方法都能有效提高遥感图像的检索性能。
关键词: 遥感图像; 检索; 卷积神经网络; 均值池化; 视觉词袋
中图法分类号: TP 79; TP 183 **文献标志码:** A **文章编号:** 1001-070X(2019)01-0049-09

0 引言

随着遥感技术的发展, 高分辨率遥感 (high-resolution remote sensing, HRRS) 图像数量急速增长, HRRS 图像检索技术成为了研究热点和难点之一。基于内容的遥感图像检索 (content-based remote sensing image retrieval, CBRSIR) 是目前主流的检索技术, 它包括特征提取和相似性度量 2 个部分, 其中特征提取是图像检索中的关键技术。

早期 CBRSIR 主要通过提取图像的底层特征^[1]进行检索, 但是底层特征难以表达图像的高层语义信息, 即存在严重的“语义鸿沟”问题^[2-3]。为了缩小语义鸿沟, 主要有以下 3 种方法: ①采用相关反馈机制^[2], 该方法依赖于反馈中标记的样本示例; ②融合多种特征^[4], 该方法可以有效结合不同特征的优点, 从而更加全面地描述图像信息; ③聚合特征的方法, 即在局部特征的基础上进一步构建抽象出的高一级特征, 如视觉词袋 (bag of visual words, BoVW)^[5]是在尺度不变特征转换 (scale-invariant

feature transform, SIFT) 特征的基础上通过 K 均值聚类得到的一种聚合特征, 局部结构学习 (local structure learning, LSL)^[6]是在局部特征的基础上, 结合图正则化得到的一种聚合特征。聚合特征能够减少冗余信息, 有效降低特征维度, 提高特征表达能力, 从而缩小语义鸿沟。

传统的聚合特征都是建立在手工提取特征的基础上, 但手工特征表达图像能力有限, 且容易受到人为因素干扰。目前流行的深度卷积神经网络 (convolutional neural network, CNN) 能够自动学习图像的特征, 降低了人为干扰, 在图像分类、检索和目标识别中应用广泛^[7-11], 其中在大规模数据集 (如 ImageNet) 上训练的 CNN 具有很强的泛化能力, 可以有效迁移到其他小规模数据集。CNN 迁移学习中, 全连接层的输出值首先受到关注^[7], 之后表达图像局部信息的卷积层特征越来越受到重视^[8], 卷积层特征通常采用编码^[8]和池化^[9]的方法进一步构建为聚合特征。

在遥感图像检索领域, 由于目前公开的遥感数据集规模较小, CNN 的参数得不到充分训练, 因此

收稿日期: 2017-08-21; 修订日期: 2017-12-21
基金项目: 国家自然科学基金项目“高空间分辨率遥感图像检索中卷积神经网络迁移特征改进方法的研究”(编号: 41801288)、“基于人工禁忌免疫原理的多源遥感图像自动配准研究”(编号: 41261091)、“基于多变量自然场景统计和局部均值估计的无参考立体图像质量评价”(编号: 61662044)、“基于深度神经网络和记忆机制的复杂环境目标跟踪研究”(编号: 61663031)和江西省青年科学基金项目“基于虹膜生物特征密钥的无线传感器网络用户认证和访问权限的理论与新方法研究”(编号: 20161BAB212034)共同资助。
第一作者: 葛 芸(1983-), 女, 博士研究生, 主要从事遥感图像检索和机器学习的研究。Email: geyun@ncu.edu.cn。
通信作者: 叶发茂(1978-), 男, 副教授, 主要从事遥感图像处理和人工智能方面研究。Email: yefamao@ncu.edu.cn。

相关研究主要集中于将 CNN 迁移到 HRRS 图像并进行检索^[12-14]。Napoleatano^[12]使用 CNN 中的全连接层特征进行检索; Zhou 等^[13]和 Hu 等^[14]比较了 CNN 全连接层特征和基于卷积层输出值的聚合特征,并对 CNN 进行微调; Zhou 等^[13]还提出一种低维度特征(low dimensional CNN, LDCNN), 但该特征的性能与数据集密切相关; Hu 等^[14]对卷积层特征提出了多尺度级联的方法, 对全连接层特征采用了多小块均值池化的方法, 但为了提取一幅图像的特征, 这些方法需要多个输入来重新馈送给 CNN, 导致特征提取过程相对复杂。

上述文献对 CNN 的全连接层特征和卷积特征进行了较全面的研究, 但对卷积特征采用的聚合方法均为编码方法, 缺少对卷积层特征不同聚合方法的研究。因此本文根据 HRRS 图像的特点, 研究 CNN 特征的聚合方法, 并将其用于 HRRS 图像检索。首先, 将 CNN 网络的参数迁移到 HRRS 图像, 并针对不同尺寸的输入图像, 提取表达图像局部信息的 CNN 特征; 然后, 提出池化区域不相同的均值池化和 BoVW 这 2 种方法对 CNN 特征进行聚合, 分别得到池化特征和 BoVW 特征, 并对池化区域和视觉单词数目进行了研究; 最后, 将这 2 种聚合特征用于遥感图像检索。

1 聚合 CNN 特征的图像检索

1.1 网络结构

在聚合 CNN 特征时, 选用 16 层的 VGG16 网络^[15]和 22 层的 GoogLeNet 网络^[16]。VGG16 通过扩展卷积层的数量增加了网络深度, GoogLeNet 则通过使用 inception modules 机制, 不仅增加了网络的深度, 还增加了网络的广度。因此 VGG16 和 GoogLeNet 经过前面多个层次的抽象运算, 后面的卷积层不仅仅获得更多的局部信息, 并且具有更好的泛化能力。VGG16 的 CNN 特征来自最后的卷积层(conv5-3)、激活函数层(relu5-3)和池化层(pool5)的输出值, GoogLeNet 的 CNN 特征来自倒数第二层池化层(pool4)和最后 2 个 inception 层(incep5a 和 incep5b)的输出值。

输入图像尺寸不同时, 输出值也不同, 因此不同尺寸的输入图像对检索性能有较大影响。主要考虑 2 种尺寸: ①CNN 默认的图像尺寸, 即调整后的图像尺寸, VGG16 和 GoogLeNet 的默认图像尺寸为 224 像素×224 像素(文中涉及到图像尺寸的单位均为像素, 为表达简洁, 下文省略); ②数据集集中的原图像尺寸, UC-Merced^[5]和 WHU-RS^[17]为目前

常用的 2 种 HRRS 数据集, 256×256 为 UC-Merced 中图像的原尺寸, 比较接近默认尺寸, 600×600 为 WHU-RS 中图像的原尺寸, 与默认尺寸相差较大, 因此这两种数据集中图像的不同尺寸正好可以有效比较图像尺寸对检索性能的影响。表 1 和表 2 列出了不同输入图像尺寸下相应层次的输出值。以 VGG16 中 pool5 为例, 在输入图像为 224×224×3 (3 表示对应于 R, G, B 的 3 个通道)时, pool5 的输出值为 7×7×512, 即输出值有 512 个通道, 每个通道的特征图尺寸为 7×7。

表 1 不同尺寸输入图像下 VGG16 的输出值
Tab.1 Outputs of VGG16 under different input image sizes

输入图像	conv5-3	relu5-3	pool5
224×224×3	14×14×512	14×14×512	7×7×512
256×256×3	16×16×512	16×16×512	8×8×512
600×600×3	38×38×512	38×38×512	19×19×512

表 2 不同尺寸输入图像下 GoogLeNet 的输出值
Tab.2 Outputs of GoogLeNet under different input image sizes

输入图像	pool4	incep5a	incep5b
224×224×3	7×7×832	7×7×832	7×7×1 024
256×256×3	8×8×832	8×8×832	8×8×1 024
600×600×3	18×18×832	18×18×832	18×18×1 024

1.2 特征提取

1.2.1 CNN 特征

令图像 I 某个层次 l 的输出值为

$$f^l = s^l \times s^l \times c^l, \tag{1}$$

式中: f^l 为层次 l 的 CNN 特征; $s^l \times s^l$ 为特征图的尺寸; c^l 为特征图的数目, 即通道的数目。若将 f^l 直接转化为特征向量, 则维度过高, 检索性能不佳, 因此需要将其构建为聚合特征。

1.2.2 聚合特征

HRRS 图像内容复杂, 信息丰富, 因此针对 HRRS 图像采用池化区域尺寸不同的均值池化方法, 以便找到合适的池化区域来提取区分度更好的池化特征。特征编码采用经典的 BoVW 编码方法。

1) 池化特征。目前常用的均值池化方法是令池化区域尺寸等于特征图尺寸^[9], 但针对 HRRS 图像, 由于其内容丰富, 直接令池化区域等于特征图区域, 可能会丢失一些重要信息。因此提出池化区域尺寸不相同的均值池化方法, 以获得效果更好的特征。

对于尺寸为 $s^l \times s^l$ 的图像 I 的 l 层特征图, 令池化区域为 $m^l \times m^l$, 记为 r^l ; 令步幅为 1, 则池化区域的数目为 $(s^l - m^l + 1) \times (s^l - m^l + 1)$, 将其记为 k^l , 则对于每个池化区域 i , 其池化特征为

$$p^l(i) = \frac{1}{m^l \times m^l} \sum r^l(i), \quad i = 1, 2, \dots, k^l, \tag{2}$$

式中 $m^l \times m^l \leq s^l \times s^l$, 即池化区域小于或者等于特征图区域。当池化区域尺寸比特征图小时, 可以保留更多的信息, 更适合表达内容复杂的 HRRS 图像。根据公式(2)计算的 p^l 的输出值为三维矩阵 $(s^l - m^l + 1) \times (s^l - m^l + 1) \times c^l$, 将其转换为池化特征向量, 记为 $A^p = [x_1, x_2, \dots, x_D]$, 其中 $D = (s^l - m^l + 1) \times (s^l - m^l + 1) \times c^l$, 即池化特征的维度。

因此, 本文提出的均值池化方法, 图像仅需要输入到 CNN 中一次, 通过在输出的特征图上设置较小的池化区域, 可以获取图像的很多局部信息, 从而提高图像的特征表达。

2) BoVW 特征。传统的 BoVW 特征主要基于手工提取的局部特征进行聚合, 而本文的 BoVW 特征则是基于表达图像局部信息的 CNN 特征进行聚合后的特征。

根据文献[8], 上述 CNN 特征 f^l 可理解为在特征图的每个位置 (i, j) , 能够得到一个 c^l 维的特征向量 $f_{i,j}^l$, 即 $f_{i,j}^l = f^l(i, j), i = 1, 2, \dots, s^l; j = 1, 2, \dots, s^l$ 。(3)

因此第 l 层可以看成总共输出 n^l 个 c^l 维的特征向量, 其中 $n^l = s^l \times s^l$, 将其记为 $B^l = [f_{1,1}^l, f_{1,2}^l, \dots, f_{s^l,s^l}^l]$ 。以 VGG16 的 pool5 层为例, 该层在默认图像尺寸下的 CNN 特征为 $7 \times 7 \times 512$, 即有 49 个 512 维的局部特征。

令数据集中图像总数为 N , 数据集中的所有图像按照上述方法提取相应的特征, 则提取的特征集合为 $\{B_1^l, B_2^l, \dots, B_N^l\}$; 然后, 通过 K 均值算法聚类得到 BoVW, 其中每个元素为一个视觉单词; 最后, 采用硬分配方法将每幅图像的特征向量分配到距离最近的视觉单词, 统计各个视觉单词出现的频数, 从而得到每幅图像的 BoVW 特征向量 $A^b = [y_1, y_2, \dots, y_K]$, 其中 K 为视觉单词的数目, y 为相应的视觉单词出现的频数。

1.3 检索流程

图 1 以 VGG16 为例描述了整个检索流程, 图中 $ci (i = 1, 2, 3, 4, 5 - 1, 5 - 2, 5 - 3)$ 表示卷积层。GoogLeNet 的检索流程类似, 只是提取的网络层次与 VGG16 不同。

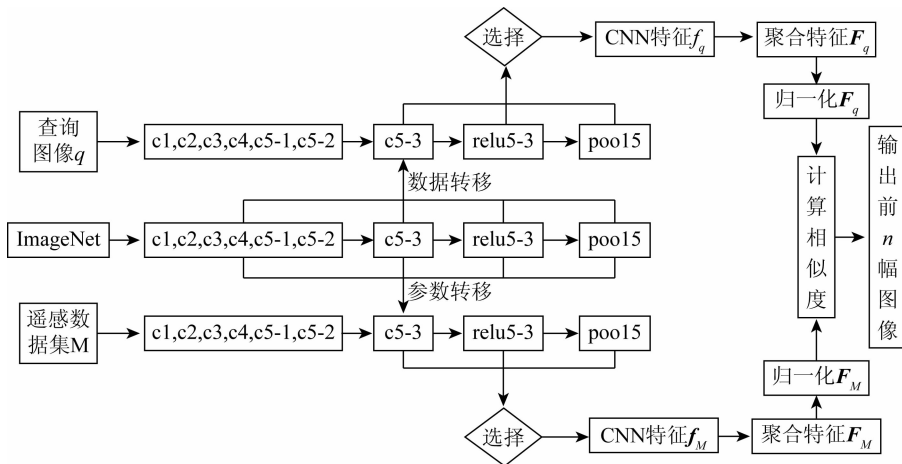


图 1 VGG16 检索流程

Fig. 1 Retrieval flow chart of VGG16

具体检索步骤如下:

1) 将预训练 CNN 的参数分别迁移到 HRRS 数据集 M 和查询图像 q 。由于聚合特征是针对卷积层特征进行的, 因此去除 VGG16 中的全连接层。将 VGG16 中卷积层的参数直接迁移到 M 和 q 。除了 conv5-3 外, 其它卷积层省略了激活函数层和池化层。

2) 提取 M 和 q 的 CNN 特征。将 M 中每幅图像和 q 分别输入到 CNN, 提取 conv5-3、relu5-3 和 pool5 层的输出值作为 M 中每幅图像和 q 的 CNN 特征。 M 中所有图像提取的 CNN 特征为 $f_M = [f_1, f_2, \dots, f_N]$, N 为数据集 M 中图像的总数量, q 的 CNN 特征记为 f_q 。

3) 提取 M 和 q 的聚合特征。 M 和 q 的 CNN 特

征分别采用池化区域不相同的均值池化和 BoVW 方法, 得到相应的池化特征和 BoVW 特征。为了简要表明, 池化特征和 BoVW 特征用统一的方式标记: q 的聚合特征记为 F_q , M 中的所有图像提取的聚合特征为 $F_M = [A_1, A_2, \dots, A_N]$ 。

4) 分别对 F_M 和 F_q 进行归一化处理。由于图像各特征向量代表的物理意义往往不同, 即使对同一个特征向量, 其各个分量的取值范围也可能存在很大差异, 因此需要对 M 和 q 的聚合特征进行归一化处理。对此, 本文采用的是常用的 L2 归一化。

5) 计算相似度, 完成图像检索。根据归一化后的聚合特征, 计算 q 和 M 中图像的相似度, 并根据相似度返回最相似的 n 幅图像。





















2 实验结果及分析

2.1 实验数据和评估标准

实验使用 MatConvNet^[18] 提取网络模型 VGG16 和 GoogLeNet。预训练 VGG16 和 GoogLeNet 的数据集采用 ImageNet 的子集 ILSVRC2012, ILSVRC2012 包含了 1 000 种图像分类, 大约有 130 万幅训练图

像, 5 万幅验证图像和 10 万幅测试图像。遥感数据集采用 UC - Merced 和 WHU - RS。UC - Merced 是从美国地质调查局收集的航空正射图像, 总共 21 类场景, 每类有 100 幅图像, 图像大小为 256×256 ; WHU - RS 是从 Google Earth 下载的 19 类场景, 每类包含 50 幅图像, 图像大小为 600×600 。表 3 显示了这 2 个数据集的示例图像。

表 3 UC - Merced 和 WHU - RS 示例图像
Tab.3 Sample images of UC - Merced and WHU - RS

UC - Merced	地物类型	农田	飞机	棒球内场	海滨	建筑物	灌木丛	稠密区	森林	高速公路	海港
	典型图像示例										
WHU - RS	地物类型	机场	海滨	桥	商业区	沙漠	农场	足球场	森林	工业区	牧场
	典型图像示例										

实验的相似度采用常用的欧式距离; 评估标准采用了近几年来 HRRS 图像中使用普遍的平均归一化修改检索等级 (average normalize modified retrieval rank, ANMRR), ANMRR 取值越小, 表明检索出来的相关图像越靠前, 即检索性能越好。同时, 实验中还比较了图像检索中重要的性能评价准则查准率—查全率曲线。

2.2 池化区域比较

采用均值池化提取聚合特征时, 池化区域的尺

寸影响网络检索性能。图 2 和图 3 分别比较了 VGG16 和 GoogLeNet 不同池化区域尺寸的检索结果。当输入图像尺寸为 224×224 时, VGG16 的 conv5 - 3 和 relu5 - 3 的特征图尺寸为 14×14 , 其他层次的特征图尺寸均为 7×7 。图中横坐标 2 ~ 7 表示池化区域尺寸从 2×2 到 7×7 。为了显示方便, 对于图 3 的 conv5 - 3 和 relu5 - 3 来说, 池化区域尺寸为横坐标值的 2 倍, 即为 4×4 到 14×14 。

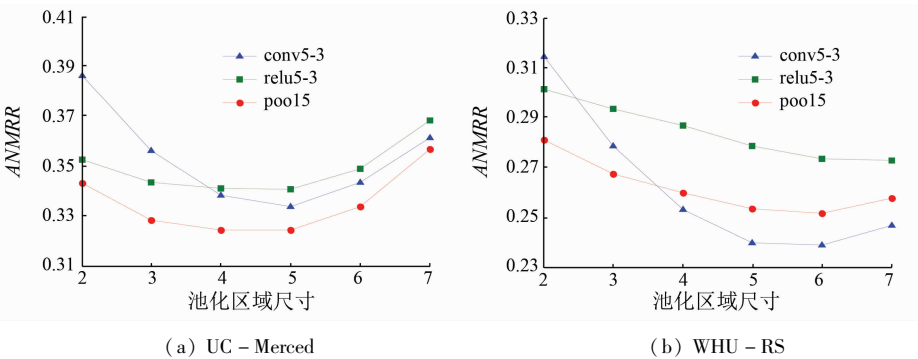


图 2 VGG16 中不同池化区域的 ANMRR

Fig.2 ANMRR with different pooling region sizes in VGG16

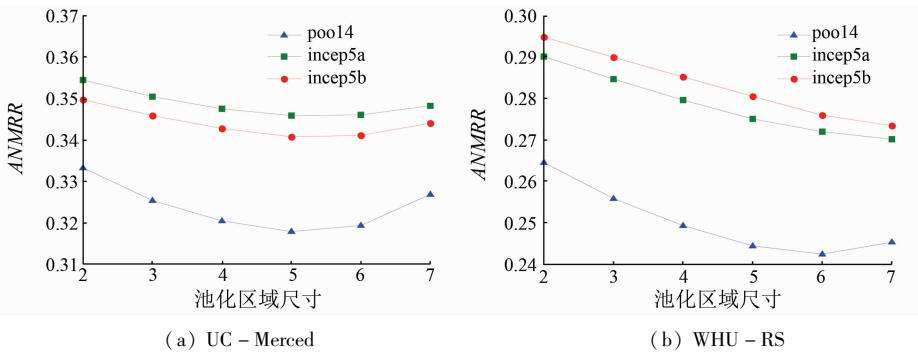


图 3 GoogLeNet 中不同池化区域的 ANMRR

Fig.3 ANMRR with different pooling region sizes in GoogLeNet

图 2(a) 显示,3 类特征的 *ANMRR* 值都呈现先降后升的趋势,其中以 conv5-3 的 *ANMRR* 值下降最快,pool5 的 *ANMRR* 值最小,即检索性能最好。图 2(b)显示,随着池化区域的增大,relu5-3 的 *ANMRR* 值呈下降趋势,而 conv5-3 和 poo5 的 *ANMRR* 值均先下降再上升。当池化区域较小时,pool5 的 *ANMRR* 值最小;随着池化区域增大,conv5-3 的 *ANMRR* 值急速下降,并小于 pool5 的值。图 3(a) 中 3 类特征的最小 *ANMRR* 值位于接近特征图的位置,其中以 pool4 的结果最好。图 3(b) 中 pool4 的最小 *ANMRR* 值位于 6×6 的位置,而其他层次的最小值位于 7×7 的位置,3 类特征中 pool4 的 *ANMRR* 值最优。

从图 2 和图 3 总体上来看,大多数特征的 *ANMRR* 值首先随着池化区域尺寸的增大而减小,到达

最低值后,再随着池化区域尺寸的增大而上升。除了 WHU-RS 上的 relu5-3,incep5a 和 incep5b 外,其他特征在池化区域尺寸小于特征图尺寸时的检索性能最好。

2.3 不同尺寸输入图像的池化特征比较

表 4 和表 5 分别显示了 UC-Merced 和 WHU-RS 中 2 种输入图像尺寸(默认尺寸和原始尺寸)下池化特征的结果。为了和传统的均值池化方法比较,对于每种特征,列出了 3 种不同池化区域尺寸的结果:前两个值是在池化区域尺寸从 2×2 增加到 $(s-1)^l\times(s-1)^l$ (特征图尺寸为 $s^l\times s^l$) 的结果中选择的 2 个最优值,第 3 个值为池化区域尺寸等于特征图尺寸的结果(即传统的均值池化方法)。表中粗体标注的值为该类特征中的最优结果,标注星号的值表示整体的最优结果。

表 4 UC-Merced 不同池化特征的 ANMRR
Tab.4 ANMRR with different pooling features on the UC-Merced

CNN	默认尺寸(224×224)						原始尺寸(256×256)					
	conv5-3		relu5-3		pool5		conv5-3		relu5-3		pool5	
	尺寸	ANMRR	尺寸	ANMRR	尺寸	ANMRR	尺寸	ANMRR	尺寸	ANMRR	尺寸	ANMRR
VGG16	9×9	0.334 1	9×9	0.340 4	*4×4	0.324 3	11×11	0.341 8	10×10	0.342 0	4×4	0.327 6
	10×10	0.333 7	10×10	0.340 8	5×5	0.324 5	12×12	0.343 0	11×11	0.342 4	5×5	0.326 2
	14×14	0.361 4	14×14	0.368 2	7×7	0.356 7	16×16	0.369 3	16×16	0.369 9	8×8	0.358 9
GoogLeNet	pool4		incep5a		incep5b		pool4		incep5a		incep5b	
	尺寸	ANMRR	尺寸	ANMRR	尺寸	ANMRR	尺寸	ANMRR	尺寸	ANMRR	尺寸	ANMRR
	*5×5	0.317 9	5×5	0.346 0	5×5	0.340 9	5×5	0.324 4	5×5	0.354 7	6×6	0.343 8
	6×6	0.319 5	6×6	0.346 1	6×6	0.341 2	6×6	0.323 5	6×6	0.354 1	7×7	0.345 1
	7×7	0.326 9	7×7	0.348 4	7×7	0.344 1	8×8	0.337 5	8×8	0.359 6	8×8	0.349 3

表 5 WHU-RS 不同池化特征的 ANMRR
Tab.5 ANMRR with different pooling features on the WHU-RS

CNN	默认尺寸(224×224)						原始尺寸(600×600)					
	conv5-3		relu5-3		pool5		conv5-3		relu5-3		pool5	
	尺寸	ANMRR	尺寸	ANMRR	尺寸	ANMRR	尺寸	ANMRR	尺寸	ANMRR	尺寸	ANMRR
VGG16	10×10	0.240 0	12×12	0.273 6	5×5	0.253 5	30×30	0.237 5	28×28	0.243 1	*14×14	0.226 2
	11×11	0.237 9	13×13	0.273 1	6×6	0.251 8	31×31	0.237 5	29×29	0.242 9	15×15	0.226 5
	14×14	0.246 8	14×14	0.272 8	7×7	0.257 6	38×38	0.251 7	38×38	0.251 6	18×18	0.238 5
GoogLeNet	pool4		incep5a		incep5b		pool4		incep5a		incep5b	
	尺寸	ANMRR	尺寸	ANMRR	尺寸	ANMRR	尺寸	ANMRR	尺寸	ANMRR	尺寸	ANMRR
	5×5	0.244 4	5×5	0.275 1	5×5	0.280 5	*14×14	0.232 1	16×16	0.262 5	16×16	0.250 9
	6×6	0.242 5	6×6	0.272 0	6×6	0.276 0	15×15	0.232 3	17×17	0.262 8	17×17	0.250 6
	7×7	0.245 3	7×7	0.270 1	7×7	0.273 4	18×18	0.241 5	18×18	0.263 4	18×18	0.251 0

表 4 中,输入图像的默认尺寸和原始尺寸比较接近,因此检索结果也很接近,整体上 256×256 的检索结果比 224×224 的结果稍差些。这种结果可能是由于与相差不大的 256×256 相比,尺寸为 224×224 的图像更适合用于 CNN 中,以便输出区别性更强的特征。表 5 中输入图像的默认尺寸和原始尺寸相差较大,因此检索结果的差异性比较明显,600×600 的结果比 224×224 的结果好,这是因为当图像尺寸从 600×600 调整到 224×

224 时,图像丢失的信息比较多,直接导致检索性能下降。

对比 2 表可知,当输入图像尺寸增大时,最优池化区域的尺寸和特征图尺寸的差距也相应增大。因此简单地令池化区域尺寸等于特征图尺寸的方法容易丢失重要的特征信息,应该根据输入图像的尺寸及网络的层次选择合理的池化区域。根据实验结果,大多数特征的最优池化区域在特征图尺寸的 60%~80% 之间。

2.4 不同尺寸输入图像的 BoVW 特征比较

表 6 和表 7 显示了 2 种输入图像尺寸下 BoVW 特征的结果。为了比较视觉单词数目 K 对检索性

能的影响,分别令 K 的取值为 100,150,1 500,2 000 和 4 000。表中粗体标注的值为该类特征中的最优结果,标星号的值表示整体的最优结果。

表 6 UC – Merced 不同 BoVW 特征的 ANMRR
Tab. 6 ANMRR with different BoVW features on the UC – Merced

CNN		默认尺寸 (224 × 224)			原始尺寸 (256 × 256)		
	K	conv5 – 3	relu5 – 3	pool5	conv5 – 3	relu5 – 3	pool5
VGG16	100	0.408 5	0.469 9	0.424 9	0.425 7	0.480 9	0.444 7
	150	*0.388 6	0.482 7	0.523 9	0.412 2	0.491 3	0.530 1
	1 500	0.410 5	0.474 8	0.482 1	0.414 5	0.477 2	0.475 2
	2 000	0.417 5	0.476 0	0.491 8	0.421 5	0.473 8	0.483 7
	4 000	0.432 1	0.478 1	0.508 6	0.436 5	0.480 0	0.497 7
	K	pool4	inception5a	inception5b	pool4	inception5a	inception5b
GoogLeNet	100	0.408 6	*0.375 9	0.402 3	0.400 5	0.394 5	0.397 5
	150	0.419 6	0.397 3	0.422 0	0.430 8	0.415 6	0.414 0
	1 500	0.536 7	0.535 4	0.480 3	0.541 4	0.532 7	0.480 1
	2 000	0.581 1	0.551 9	0.492 8	0.559 2	0.550 5	0.491 4
	4 000	0.613 8	0.616 9	0.535 3	0.616 1	0.609 8	0.540 6

表 7 WHU – RS 不同 BoVW 特征的 ANMRR
Tab. 7 ANMRR with different BoVW features on the WHU – RS

CNN		默认尺寸 (224 × 224)			原始尺寸 (600 × 600)		
	K	conv5 – 3	relu5 – 3	pool5	conv5 – 3	relu5 – 3	pool5
VGG16	100	0.280 7	0.441 7	0.352 5	0.354 7	0.424 1	0.366 3
	150	*0.249 1	0.414 2	0.370 1	0.323 9	0.462 0	0.370 7
	1 500	0.275 2	0.411 2	0.391 8	0.268 2	0.356 6	0.355 7
	2 000	0.279 6	0.431 6	0.408 6	0.270 5	0.367 6	0.352 2
	4 000	0.311 5	0.425 4	0.440 1	0.278 7	0.350 8	0.344 5
	K	pool4	inception5a	inception5b	pool4	inception5a	inception5b
GoogLeNet	100	0.287 7	0.310 5	0.267 4	0.263 1	0.262 1	0.228 2
	150	0.298 7	0.311 6	0.286 4	0.226 8	0.259 6	*0.214 9
	1 500	0.440 4	0.449 4	0.425 7	0.278 8	0.304 2	0.258 7
	2 000	0.469 5	0.488 8	0.423 7	0.289 4	0.309 0	0.262 1
	4 000	0.653 8	0.656 0	0.491 4	0.309 5	0.337 0	0.283 6

表 6 中,大多数的 BoVW 特征在 224 × 224 尺寸下的结果优于 256 × 256,VGG16 中大多数特征的最优 K 值为 100 和 150,GoogLeNet 中不同特征的最优 K 值均为 100。表 7 中,大多数的 BoVW 特征在 600 × 600 尺寸下的结果优于 224 × 224,尤其以 GoogLeNet 中的结果表现更明显。当输入图像尺寸明显增大时,用于构建视觉单词的特征数目也相应增多,相应的最优 K 值也随之增大。例如,当输入图像尺寸为 600 × 600 时,relu5 – 3 和 pool5 的最优 K 值增大到 4 000,GoogLeNet 所有层次的最优 K 值均增大到 150。

因此在 BoVW 特征中,根据图像尺寸和特征图尺寸选择一个适宜的 K 值对提高检索结果有着重要作用。当输入图像尺寸显著增大时, K 的最优取值也变大,其中以 VGG16 中 K 的最优取值变化尤为显著。

2.5 查准率—查全率曲线比较

前面实验结果中,大多数池化特征的检索结果

优于 BoVW 特征。为了进一步比较这 2 种不同的聚合特征,在每种聚合特征中分别选择最优的特征(即为表 4—7 中标记为星号的特征)比较查准率—查全率曲线。查准率是指检索返回结果中相关图像数与返回图像数的比例,反映了检索精度;查全率是指检索返回结果中相关图像数与所有相关图像总数的比值,反映了检索的全面性,与返回图像数目呈正相关。在查准率—查全率曲线中曲线比较高时,查准率和查全率都比较高,即检索性能比较好。

图 4 比较了不同特征的查准率—查全率曲线,VGG16 和 GoogLeNet 的最优池化特征记为 VGG16 – P 和 GoogLeNet – P,VGG16 和 GoogLeNet 的最优 BoVW 特征记为 VGG16 – B 和 GoogLeNet – B。UC – Merced 返回图像数目最少为 2,最多为 2 100;WHU – RS 返回图像数目最少为 2,最多为 950。在 UC – Merced 中,GoogLeNet – P 的曲线位于最顶端,因此 GoogLeNet – P 的检索性能最优,其次是 VGG16 – P。当返回图像数目较少时,GoogLeNet – B 的曲

线高于 VGG16 - B 的曲线,即 GoogLeNet - B 的检索性能优于 VGG16 - B; 当返回图像数目逐渐增多时,GoogLeNet - B 的性能迅速下降并低于 VGG16 - B。在 WHU - RS 中,VGG16 - B 的曲线位于最低端,即检索性能最差,VGG16 - P 和 GoogLeNet - P 的结果比较接近。对于 GoogLeNet - B,其检索性能

随着返回图像数目的增大逐渐变好,甚至超过 VGG16 - P 和 GoogLeNet - P; 当返回图像数目增大到一个较大值时,GoogLeNet - B 的性能又迅速下降。总体上来看,在 2 个数据集上,VGG16 - P 和 GoogLeNet - P 的检索性能优于 VGG16 - B 和 GoogLeNet - B。

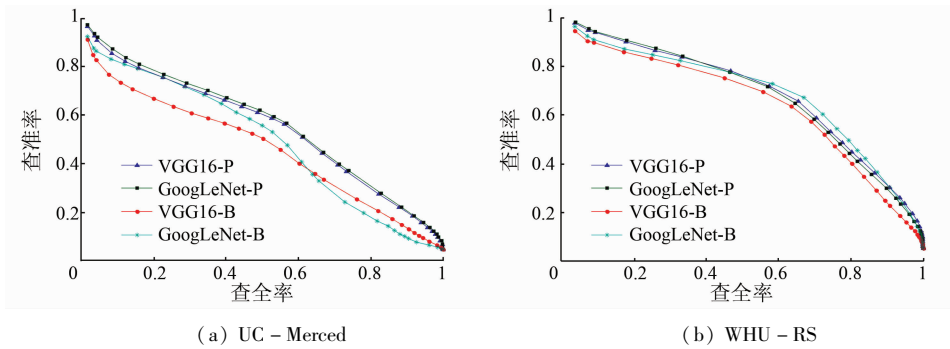


图 4 不同特征的查准率—查全率曲线

Fig.4 Precision - recall curves for different features

2.6 与其他方法比较

表 8 比较了浅层特征和 CNN 特征的 ANMRR 值和维度。浅层特征选择了 Aptoula 提出的全局形态纹理特征^[3]和基于手工特征 SIFT 构建的 BoVW^[5],以及近期提出的 LSL^[6]。CNN 特征包含了文献[12—14]提出的特征,以及本文提出的 VGG16 - P, GoogLeNet - P, VGG16 - B 和 GoogLeNet - B 特征。由于大多数其它特征使用的数据集为 UC - Merced,因此表 8 基于 UC - Merced 进行比较。

表 8 不同特征的 ANMRR 和维度

Tab.8 ANMRR and dimensions for different features

	特征	ANMRR	维度
浅层特征	Aptoula ^[3]	0.575 0	62
	BoVW ^[5]	0.591 0	15 000
	BoVW ^[5]	0.601 0	150
	LSL ^[6]	0.555 6	2 048
CNN 特征	VGGM - fc ^[12]	0.378 0	4 096
	VGGM - fc - RF ^[12]	0.316 0	4 096
	VGG16 - fc ^[13]	0.394 0	4 096
	VGGM - conv5 - IFK ^[13]	0.458 0	102
	VGG16 - conv5 - IFK ^[13]	0.407 0	102
	LDCNN ^[13]	0.439 0	30
	GoogLeNet(FT) + MultiPatch ^[14]	0.314 0	1 024
	VGG16 - P	0.324 3	8 192
	GoogLeNet - P	0.317 9	7 488
	VGG16 - B	0.388 6	150
	GoogLeNet - B	0.375 9	100

表 8 显示,CNN 特征的结果普遍优于浅层特征,与 BoVW 相比,GoogLeNet - P 和 VGG16 - P 的值分别降低了 27.31% 和 21.51%。

CNN 特征中,VGGM - fc^[12] 和 VGGM - fc -

RF^[12] 分别是 VGGM 全连接层特征及加入了反馈信息的特征; VGG16 - fc^[13] 是 VGG16 全连接层特征, VGGM - conv5 - IFK^[13] 和 VGG16 - conv5 - IFK^[13] 是对 VGGM 和 VGG16 的卷积层使用改进的费舍尔核(improved fisher kernel, IFK)编码的特征, GoogLeNet (FT) + MultiPatch^[14] 是微调后的 GoogLeNet 特征使用多个分块均值化的结果。

从表 8 中可以看出,除了 VGGM - fc - RF 和 GoogLeNet(FT) + MultiPatch 外,本文提出的 4 种 CNN 特征比其他 CNN 特征的 ANMRR 值低,而 GoogLeNet(FT) + MultiPatch 和 VGGM - fc - RF 的特征提取方法比本文方法复杂。因此选择合适的 CNN 网络以及采用合理的聚合方法能够有效提高 HRRS 图像检索性能。

3 结论

本文对 VGG16 和 GoogLeNet 中表达局部信息的 CNN 特征,采用池化区域尺寸不相同的均值池化和 BoVW 2 种方法得到不同的聚合特征,并将其用于 HRRS 图像检索。通过研究获得主要结论如下:

- 1) 针对 HRRS 图像,池化特征的检索性能比 BoVW 特征的性能好。
- 2) 池化特征中池化区域尺寸直接影响检索结果,大多数池化特征的最优池化区域尺寸为特征图尺寸的 60% ~ 80% 之间。这种尺寸既能有效地剔除 CNN 特征的冗余信息,同时也能保留一些区分度明显的特征信息。
- 3) BoVW 特征中视觉单词数目对图像检索性能

影响较大。当输入图像尺寸显著增大时,视觉单词数目的最优取值也相应增大,以 VGG16 的取值变化尤为明显。

4)不同输入图像尺寸影响聚合特征的检索性能,当默认尺寸和原尺寸相差较大时,原尺寸得到的聚合特征检索性能更好;当默认尺寸和原尺寸很接近时,默认尺寸有时更适合 CNN 网络。

5)与传统的浅层特征相比,本文提出的聚合特征的检索性能大幅度提高。GoogLeNet 的最优池化特征和 VGG16 的最优 BoVW 特征的 ANMRR 值比浅层特征 BoVW 分别降低了 27.31% 和 21.51%。与目前提出的 CNN 特征相比,本文选用的 CNN 特征更适用于聚合,采用的聚合方法简单有效。

因此本文提出的聚合特征能够有效提高 HRRS 图像的检索性能,其中池化特征提高幅度更为明显。但是池化特征的维度相对较高,今后将进一步研究如何有效降低池化特征的维度。

参考文献 (References):

- [1] 朱佳丽,李士进,万定生,等.基于特征选择和半监督学习的遥感图像检索[J].中国图象图形学报,2011,16(8):1474-1482.
Zhu J L, Li S J, Wan D S, et al. Content-based remote sensing image retrieval based on feature selection and semi-supervised learning[J]. Journal of Image and Graphics, 2011, 16(8): 1474-1482.
- [2] 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(5): 2323-2334.
- [3] Aptoula E. Remote sensing image retrieval with global morphological texture descriptors[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(5): 3023-3034.
- [4] 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.
Lu L Z, Liu R Y, Liu N. Remote sensing image retrieval using color and texture fused features[J]. Journal of Image and Graphics, 2004, 9(3): 328-333.
- [5] Yang Y, Newsam S. Geographic image retrieval using local invariant features[J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 51(2): 818-832.

- [6] Du Z X, Li X L, Lu X Q. Local structure learning in high resolution remote sensing image retrieval[J]. Neurocomputing, 2016(207): 813-822.
- [7] Babenko A, Slesarev A, Chigorin A, et al. Neural codes for image retrieval[C]//Proceedings of European Conference on Computer Vision. Springer, 2014: 584-599.
- [8] Ng J Y, Yang F, Davis L S. Exploiting local features from deep networks for image[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition workshops. IEEE, 2015: 53-61.
- [9] Babenko A, Lempitsky V. Aggregating deep convolutional features for image retrieval[C]//Proceedings of IEEE International Conference on Computer Vision. IEEE, 2015: 1269-1277.
- [10] 周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017,40(6):1229-1251.
Zhou F Y, Jin L P, Dong J. Review of convolutional neural network[J]. Chinese Journal of Computers, 2017, 40(6): 1229-1251.
- [11] 张洪群,刘雪莹,杨森,等.深度学习的半监督遥感图像检索[J].遥感学报,2017,21(3):406-414.
Zhang H Q, Liu X Y, Yang S, et al. Retrieval of remote sensing images based on semisupervised deep learning[J]. Journal of Remote Sensing, 21(3): 406-414.
- [12] Napolitano P. Visual descriptors for content-based retrieval of remote sensing images[J]. International Journal of Remote Sensing, 2018, 39(5): 1343-1376.
- [13] Zhou W X, Newsam S, Li C, et al. Learning low dimensional convolutional neural networks for high-resolution remote sensing image retrieval[J]. Remote Sensing, 2017, 9(5): 489.
- [14] Hu F, Tong X Y, Xia G S, et al. Delving into deep representations for remote sensing image retrieval[C]//Proceedings of IEEE International Conference on Signal Processing. IEEE, 2016: 198-203.
- [15] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. arXiv. <https://arxiv.org/pdf/1409.1556.pdf>.
- [16] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2015: 1-9.
- [17] Hu F, Xia G S, Hu J W, 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.
- [18] Vedaldi A, Lenc K. MatConvNet: convolutional neural networks for MATLAB [C]//Proceedings of 23rd ACM International Conference on Multimedia. ACM, 2015: 689-692.

Aggregating CNN features for remote sensing image retrieval

GE Yun^{1,2}, JIANG Shunliang¹, YE Famao¹, JIANG Changlong², CHEN Ying², TANG Yiling¹

(1. Information Engineering School, Nanchang University, Nanchang 330031, China;

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

Abstract: In the high-resolution remote sensing image retrieval, it is difficult for hand-crafted features to describe the images accurately. Thus a method based on aggregating convolutional neural network (CNN) features is

proposed to improve the feature representation. First, the parameters from CNN pre – trained on large – scale datasets are transferred for remote sensing images. Given input images with different sizes, the CNN features which represent local information are extracted. Then, average pooling with different pooling region sizes and bag of visual words (BoVW) are adopted to aggregate the CNN features. Pooling features and BoVW features are obtained accordingly. Finally, the above two aggregation features are utilized for remote sensing image retrieval. Experimental results demonstrate that the input image with reasonable size is capable of improving the feature representation. When the pooling region size is between 60% and 80% of the feature map, the vast majority of the results of pooling features are superior to those of the traditional average pooling method. The optimal average normalized modified retrieval rank values of pooling feature and BoVW feature are 27.31% and 21.51% lower than those of hand – crafted feature. Therefore, both the average pooling and BoVW can improve the remote sensing image retrieval performance efficiently.

Keywords: remote sensing image; retrieval; convolutional neural network; average pooling; bag of visual words
(责任编辑：张 仙)

=====

下期要目

程 滔	顾及时点特征的水体提取成果空间修正方法
周 阳	基于 DCNN 特征的建筑物震害损毁区域检测
胡官兵	遥感技术在滇西南植被覆盖区地质填图中的应用
尹 展	南方植被区强迫不变植被抑制技术改进与应用
陈 震	高标准农田建后利用情况遥感监测方法
董立新	三峡库区森林叶面积指数多模型遥感估算
叶发茂	卷积神经网络特征在遥感图像配准中的应用
邢学文	基于偏最小二乘法的高光谱水面油膜厚度估算
梁林林	无人机遥感影像面向对象分类的冻土热融滑塌边界提取
曲海成	基于优势集聚类和马尔科夫随机场的高光谱图像分类算法
吕金霞	基于移动窗口法雄安新区湿地景观演变及其与人为干扰间的关系
黄宝华	可燃物干燥指数在草地火险预警中的应用
毛 宁	基于 RMNE 方法的多尺度分割最优分割尺度选取
谢奇芳	基于 Faster R – CNN 的高分辨率影像目标检测技术
白泽朝	Sentinel – 1A 数据矿区地表形变监测适用性分析