第53卷 第2期 西 安 交 通 大 学 学 报 Vol.53 No.2
2019年2月 JOURNAL OF XI'AN JIAOTONG UNIVERSITY Feb. 2019

DOI：10.7652/xjtuxb201902017

联合加权聚合深度卷积特征的图像检索方法

时璇[1]，许林松[1]，李晨[1]，王佳星[1]，李党超[2]

（1.西安交通大学软件学院，710049，西安；2.西安交通大学实践教学中心，710049，西安）

摘要：针对图像特征提取不充分影响图像检索平均精确率的问题，提出了一种基于联合加权聚合深度卷积特征的图像检索方法。该方法将图像输入到预先训练好的卷积神经网络中，提取最后一个卷积层输出作为图像的深度卷积特征；通过计算空间权重矩阵突出图像的显著性区域并抑制背景噪声区域，然后根据通道方差最大原则选取相应的特征图计算出空间权重矩阵，将原始深度卷积特征加权聚合为列向量；通过区分性地对待不同通道的特征图，计算出通道权重向量与上述列向量点乘得到最终的全局特征向量。公开数据集上的实验结果表明，本文方法能够有效地增强图像特征的表达能力，在图像检索的平均精确率上优于其他同类方法，可以有效地应用到图像检索相关领域。

关键词：图像检索；深度卷积特征；空间权重矩阵；通道权重向量；聚合

中图分类号：TP391 **文献标志码**：A **文章编号**：0253-987X(2019)02-0128-08

Joint Weighting Aggregation of Deep Convolutional Features for Image Retrieval

SHI Xuan[1], XU Linsong[1], LI Chen[1], WANG Jiaxing[1], LI Dangchao[2]

(1. School of Software Engineering, Xi'an Jiaotong University, Xi'an 710049, China;

2. School of Engineering Workshop, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: An image retrieval method based on deep convolutional features of joint weighting aggregation is proposed to solve the problem that most existing image retrieval approaches can't fully extract image features and their performance requires to be improved. Firstly, the method extracts the outputs of the last convolutional layer as deep convolutional features of an image by passing the image through a pre-trained deep convolutional neural network. Then, the spatial weight matrix is calculated to highlight significance regions of the image and to suppress the background noise of the image. The maximum-principle of channel variance is then used to select the corresponding feature map and to calculate the spatial weight matrix. The original deep convolutional features are weighted and aggregated into a feature vector. Moreover, the channel weight vector is calculated by distinguishing feature maps of different channels, and then the global feature representation of this image is obtained by multiplying the aggregated feature vector and the channel weight. Experimental results on different public available datasets for image retrieval show that the proposed approach effectively enhances the discriminative ability of image features, outperforms the state-of-the-art approaches based on pre-trained networks and can be effectively applied to related fields of image retrieval.

收稿日期：2018-08-13。 作者简介：时璇(1993—)，女，硕士生；李晨(通信作者)，女，讲师，硕士生导师。 基金项目：国家自然科学基金资助项目(61603289)。

网络出版时间：2018-12-18 网络出版地址：http://kns.cnki.net/kcms/detail/61.1069.T.20181214.0926.012.html

Keywords: image retrieval; deep convolutional features; spatial weight matrix; channel weight vector; aggregation

基于内容的图像检索技术成为了目前多媒体和计算机视觉领域的研究热点[1]，其关键步骤是得到图像的特征表示。早期研究工作是将手工设计的视觉特征(例如 SIFT 特征[2])进行嵌入和聚合[3-6]，得到图像的全局特征表示。近年来随着深度学习技术的兴起，有研究将深度卷积神经网络(CNN)[7-9]应用到图像检索技术中。此类方法将图像输入到预先训练好的卷积神经网络中，直接提取某层的输出作为图像的特征表示，此类特征与传统的手工设计的视觉特征相比具有更强的区分能力。因此，基于卷积神经网络的图像检索方法成为了目前该领域的主流方法[10]。

早期的基于卷积神经网络的图像检索方法将网络的全连接层输出作为图像的全局特征表示[11]，然而全连接层输出缺失图像的空间信息，且此类方法需要对输入图像进行缩放使其大小相同，这造成了图像内容的畸变，因此检索平均精确率不高。有研究将网络的卷积层输出，即图像的深度卷积特征作为其全局特征表示，不仅保留了图像的空间信息，且无需缩放图像，检索平均精确率得到提升[12-17]。SPoC 方法直接对深度卷积特征中每个特征图求和池化得到图像的全局特征表示[12]，是将深度卷积特征用于图像检索的开创性工作。CroW 方法基于 SPoC 方法提出空间权重和通道权重来突出图像的显著性区域[13]，检索平均精确率有了进一步提升。R-MAC 方法采用变窗口方式对图像的特征图进行滑窗提取若干局部特征，再将局部特征相加得到最终的全局特征表示[14]，该方法检索平均精确率与 CroW 方法相近。AdCoW 方法是自适应中心点与通道加权的深度卷积图像检索方法[15]，通过高斯空间权重矩阵自适应地确定显著性区域中心并分配权重，通过通道权重缓解特征爆炸现象[16]，与 SPoC 和 CroW 方法相比，检索平均精确率得到了提升。SBA 方法利用图像深度卷积特征中不同特征图的区分性生成语义探测器，通过加权聚合得到图像全局特征表示[17]。

由于上述方法仍存在背景噪声及特征提取不充分的问题，本文基于 SBA[17] 和 CroW[13] 方法提出了联合加权聚合深度卷积特征的图像检索方法，采用空间权重矩阵来突出图像的显著性区域并抑制背景区域，根据通道权重向量区分性地对待不同通道的特征图。

1　基于深度卷积特征的图像检索框架

基于深度卷积特征的图像检索包括离线和在线过程，如图 1 所示。离线过程中，先将图像数据集中的每张图像输入到预先训练好的卷积神经网络中，提取神经网络的卷积层输出得到图像的深度卷积特征，再使用合适的聚合方法将其聚合为全局特征向量并存入到全局特征向量库中；在线过程中，将查询图像输入到相同的卷积神经网络中获取图像的深度卷积特征，再使用相同的聚合方法得到查询图像的全局特征向量，通过比较查询图像的全局特征向量与全局特征向量库中的全局特征向量之间的距离，根据相似性大小进行排序，通过索引查找全局特征向量对应的图像返回检索结果。聚合方法的好坏是影响检索平均精确率的核心因素，也是本文研究的重点。

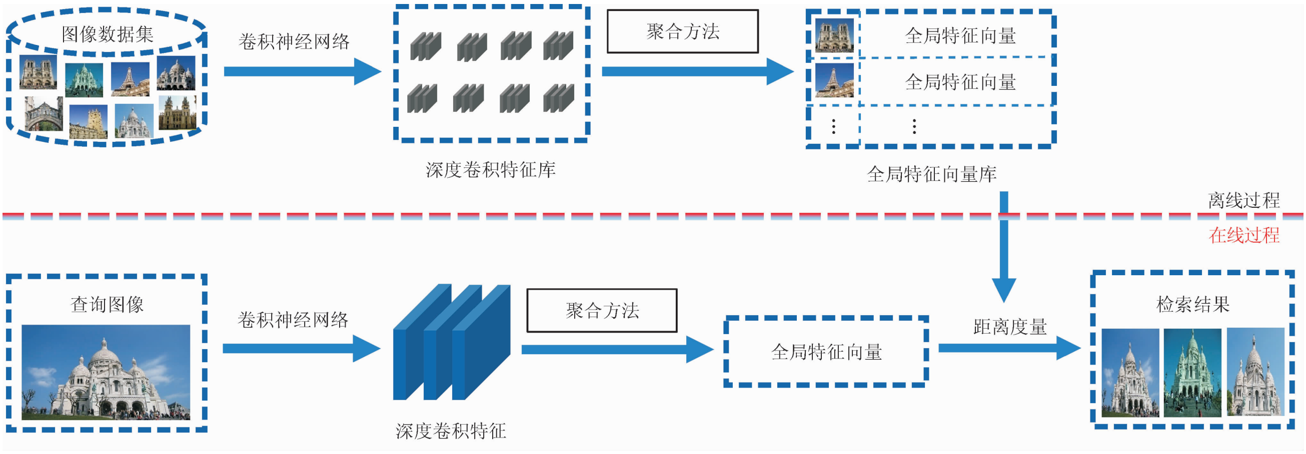


图 1　基于深度卷积特征的图像检索框架

2 联合加权聚合深度卷积特征方法

为了充分提取图像的深度卷积特征，提高图像检索的平均精确率，本文基于 SBA 方法提出了如图 2 所示的解决方法。

2.1 方法概述

本文提出了联合加权聚合深度卷积特征的方法，框架如图 2 所示，方法步骤如下：①将多个数据集输入到预先训练好的 VGG16[18] 卷积神经网络中，将 pool5 层输出作为每张图像的深度卷积特征 $\boldsymbol{X}\in R^{K\times W\times H}$，其中 $\boldsymbol{X}$ 是 $K\times W\times H$ 的三维张量，K 表示通道的个数，每个通道的特征图是 $W\times H$ 的矩阵（所有图像深度卷积特征通道个数 $K=512$，单张图像通道特征图的大小 W 和 H 是固定值）；②提出空间权重矩阵 $\boldsymbol{S}\in R^{W\times H}$，$\boldsymbol{S}$ 是 $W\times H$ 的矩阵，对 $\boldsymbol{X}$ 进行加权聚合得到图像的全局特征向量；③对数据集图像的全局特征向量按通道求方差并排序输出方差最大的前 N 个通道索引值，利用对应通道的特征图计算空间权重矩阵 $\boldsymbol{B}\in R^{W\times H}$，$\boldsymbol{B}$ 是 $W\times H$ 的矩阵；④提出通道权重向量 $\boldsymbol{P}\in R^{K}$，$\boldsymbol{P}$ 是 K 维向量；⑤对 $\boldsymbol{X}$ 进行 $\boldsymbol{B}$ 加权、聚合、$\boldsymbol{P}$ 加权和迭代拼接处理后，得到 $N\times K$ 维的全局特征向量，再经过归一化、PCA 降维和归一化处理后，将图像表示为 L 维的全局特征向量。

2.2 空间权重矩阵

本文提出空间权重矩阵 $\boldsymbol{S}$，将图像深度卷积特征 $\boldsymbol{X}$ 与 $\boldsymbol{S}$ 进行点乘、聚合后得到全局特征向量

$$\boldsymbol{\Omega}' = \sum_{i=1}^{W}\sum_{j=1}^{H}(X_{ij}S_{ij}) \tag{1}$$

式中：$\boldsymbol{S}$ 是对相应空间位置元素矩阵 $\boldsymbol{S}'$ 进行规范化得到的

$$\boldsymbol{S}' = \frac{\sum_{k=1}^{K} X_k}{K} \tag{2}$$

$$\boldsymbol{S} = \sqrt{\frac{\boldsymbol{S}'}{\sqrt{\sum_{i=1}^{W}\sum_{j=1}^{H} S'^2_{ij}}}} \tag{3}$$

空间权重矩阵 $\boldsymbol{S}$ 可以反映原始图像的语义信息，如图 3 所示，图 3a 是从 Oxford5k[19] 和 Paris6k[20] 数据集中分别随机选取的 3 张原始图像，图 3b 是原始图像所对应的空间权重矩阵 $\boldsymbol{S}$ 的热图。热图中白色区域权重大，赵靠近区域中心权重越大；黑色区域权重小，区域越黑权重越小。可以看出，空间权重矩阵在建筑物所在区域赋予较大的权重，同时在不相关区域赋予较小的权重，大致凸显了原始图像轮廓。因此，将图像的深度卷积特征与空间权重矩阵点乘后，可以突出原始图像空间位置的显著性区域，同时抑制其他区域。

SBA 方法直接将数据集图像的深度卷积特征聚合为列向量，根据通道方差最大原则选取相应的特征图。此方法认为特征图可以较好地表示图像，即激活值大的区域为图像的显著性区域，但在有些特征图中，激活值大的区域是图像的背景噪声区域。因此，本文利用空间权重矩阵 $\boldsymbol{S}$ 对图像的深度卷积特征进行加权处理，以此来突出特征图中的显著性区域并抑制背景噪声区域。

2.3 通道权重向量

SBA 方法利用方差最大的特征图计算出空间权重矩阵 $\boldsymbol{B}$，对图像深度卷积特征 $\boldsymbol{X}$ 进行加权聚合得到列向量。本文提出通道权重向量 $\boldsymbol{P}$，对上述列

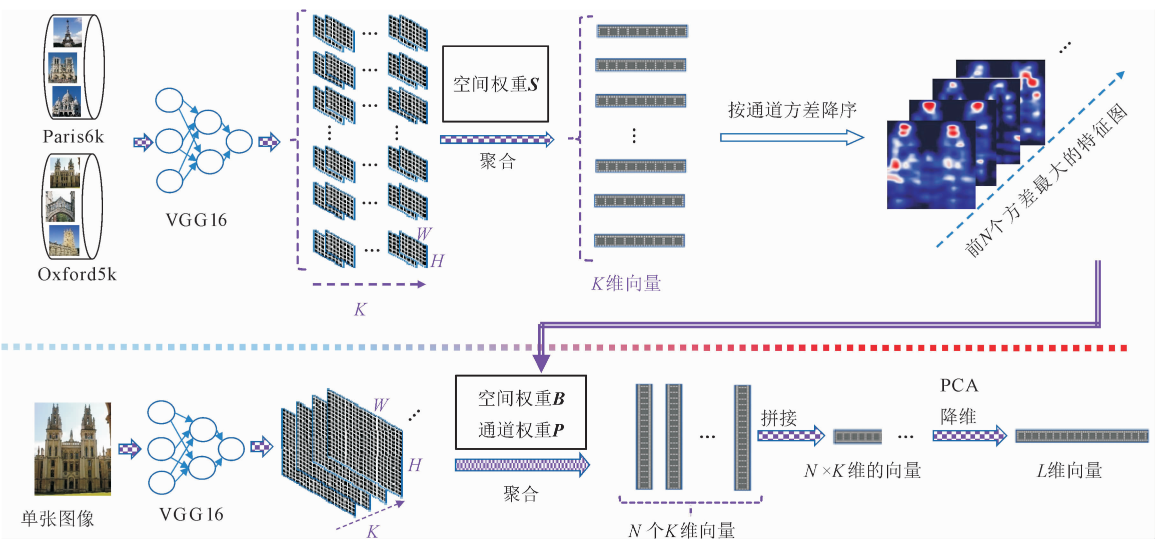


图 2 基于联合加权聚合深度卷积特征的方法框架

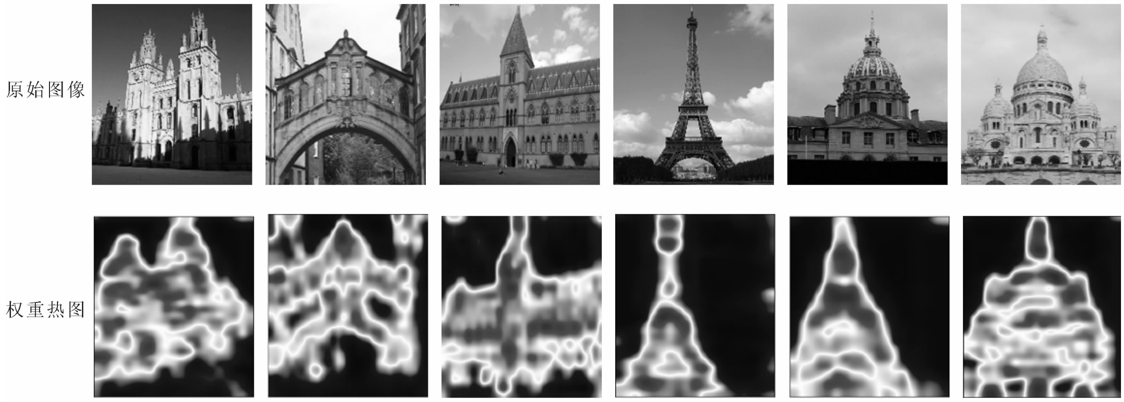


图 3　原始图像及其对应空间权重矩阵 $\boldsymbol{S}$ 的热图

向量进行加权处理得到全局特征向量

$$\boldsymbol{\Omega} = \sum_{i=1}^{W}\sum_{j=1}^{H}(X_{ij}B_{ij})\boldsymbol{P} \tag{4}$$

图像的特征图经过 $\boldsymbol{B}$ 加权后，激活值较大且激活区域较多的特征图得到了加强。然而，某些激活值较小且激活区域较少的特征图也可能提供重要信息，并对检索平均精确率产生正面影响，因此应该赋予这些特征图较大的权值作为通道权重。通道权重向量为

$$\boldsymbol{P} = \ln\left(\frac{K\varepsilon + \sum_{i=1}^{K}C_i}{\varepsilon + \boldsymbol{C}}\right) \tag{5}$$

式中：ε 是用来保持数值稳定性的常数；$\boldsymbol{C}$ 的表达式为

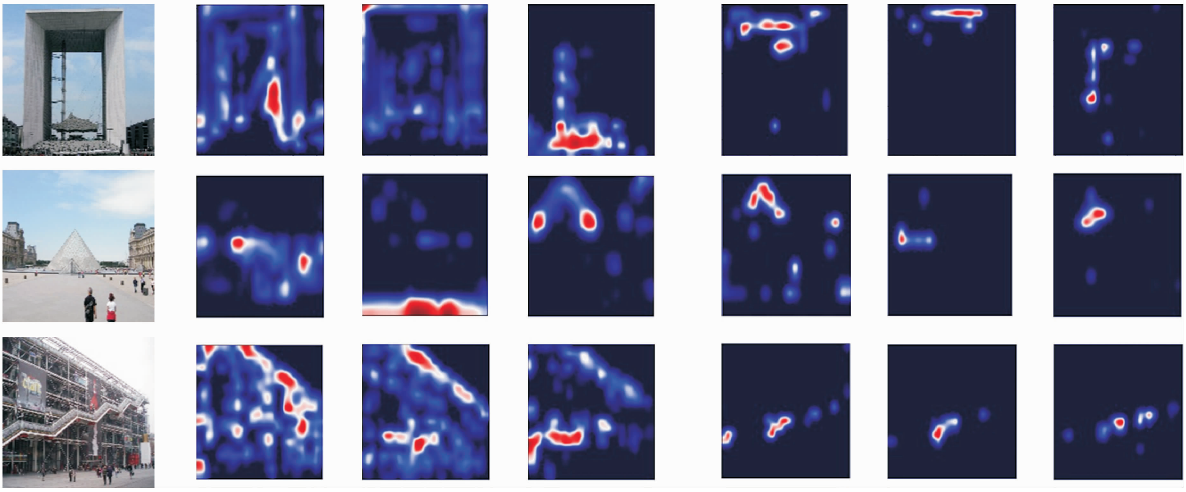
$$\boldsymbol{C} = \boldsymbol{V}\boldsymbol{Z} \tag{6}$$

其中 $\boldsymbol{V}$ 是特征图的方差值组成的向量；$\boldsymbol{Z}$ 是特征图的非零元素占比组成的向量。

图 4 为从 Oxford5k 和 Paris6k 数据集中分别随机选取的 3 张原始图像及其特征图对应的热图，图 4a 是原始图像，图 4b 是 $\boldsymbol{C}$ 中值较大的 3 张特征图所对应的热图，图 4c 是 $\boldsymbol{C}$ 中值较小的 3 张特征图所对应的热图。热图中白色区域激活值大，越靠近区域中心激活值越大；黑色区域激活值小，区域越黑激活值越小。可以看出，$\boldsymbol{C}$ 中值较小的特征图仍蕴含有用的信息，因此应赋予这些特征图较大的权重。

3　实验结果

为了验证本文所提方法的有效性，采用图像检索领域主流的 5 个数据集对该方法进行测试。Oxford5k 数据集是从 Flickr 网站上收集的 5 062 张牛津地标建筑图像，包括 55 张查询图像，并被手动标识为 11 类；Pairs6k 数据集是从 Flickr 网站上收集的 6 392 张巴黎地标建筑图像，包括 55 张查询图像，并被手动标识为 11 类；Oxford105k 和 Paris106k 数据集是分别在 Oxford5k 和 Paris6k 数据集上添加了从 Flickr 网站上收集的 10^5 张干扰图像



(a)原始图像　(b)$\boldsymbol{C}$ 中值较大的 3 张特征图对应的热图　(c)$\boldsymbol{C}$ 中值较小的 3 张特征图对应的热图

图 4　原始图像及其特征图对应的热图

扩展得到的；Holidays[21]数据集由1 491张图像组成，被手动标识为500类，包括500张查询图像。将此方法与当前主流的其他同类方法在检索平均精确率方面进行比较，本文方法采用Python语言编程实现。

3.1 参数选取

所有图像的深度卷积特征均取自基于Caffe框架VGG16网络pool5层输出，网络参数是在ImageNet数据集上预先训练好的。和其他同类方法一致，本文只对Oxford5k和Paris6k数据集上的55张查询图像进行裁剪，通过计算查询图像检索的平均精确率来度量图像检索的性能。在Oxford5k和Oxford105k数据集上测试时，使用Paris6k数据集训练PCA和白化参数；在Paris6k和Paris106k数据集上测试时，使用Oxford5k数据集训练PCA和白化参数。本文主要参数是所选语义探测器的数量N和最终表示的全局特征向量的维度D，使用本文方法和SBA方法分别在Oxford5k和Paris6k数据集上进行实验对比。表1给出了D=4 096时，选择不同N的图像检索平均精确率。

表1 D=4 096时不同N的图像检索平均精确率

N	平均精确率/%			
	Oxford5k		Paris6k	
	SBA方法	本文方法	SBA方法	本文方法
512	78.5	78.4	85.4	85.9
450	78.7	78.4	85.7	86.2
350	79.0	78.8	85.9	86.3
250	78.7	79.1	86.0	86.6
150	79.0	79.5	85.4	86.5
50	78.2	79.4	86.1	87.4
25	79.1	81.0	86.1	87.8
10	77.7	78.9	83.8	84.8

由表1可以看出，在Oxford5k和Paris6k数据集上，使用本文聚合方法将图像的全局特征向量经过PCA降维到D=4 096，当N≤25时，随着N的增加，两个数据集的图像检索平均精确率均大体呈现递增趋势；当N≥25时，随着N的增加，两个数据集的图像检索平均精确率均大体呈现递减趋势；当N=25时，本文方法在两个数据集上的图像检索平均精确率均达到最高。可以看出，本文方法即使在N较小的情况下依然能够得到较好的检索结果，考虑到内存占用和计算消耗问题，最终选择N=25。表2给出了N=25时选择不同D的图像检索的平均精确率。

由表2可以看出，在Oxford5k和Paris6k数据集上，当N=25时，随着D的增加，两个数据集的图像检索平均精确率均呈现递增趋势；当D=4 096时，本文方法在两个数据集上的图像检索平均精确率均达到最高，且本文方法在D为128～4 096这6个维度上的检索平均精确率均优于SBA方法。

表2 N=25时不同D的图像检索平均精确率

D	平均精确率/%			
	Oxford5k		Paris6k	
	SBA方法	本文方法	SBA方法	本文方法
4 096	79.1	81.0	86.1	87.8
2 048	78.2	79.6	85.4	87.5
1 024	75.3	78.2	84.2	86.9
512	72.0	75.8	82.3	85.8
256	68.7	72.6	79.6	83.9
128	64.5	67.0	76.9	82.0

综上，将本文方法的参数语义探测器数量和维度分别设置为N=25和D=4 096时，图像检索的平均精确率达到最高。

3.2 验证性实验结果

本文基于SBA方法提出空间权重矩阵$\boldsymbol{S}$加权方法，记作SBA+$\boldsymbol{S}$方法；基于SBA方法提出通道权重向量$\boldsymbol{P}$加权方法，记作SBA+$\boldsymbol{P}$方法；基于SBA方法同时使用空间权重矩阵$\boldsymbol{S}$和通道权重向量$\boldsymbol{P}$两种加权方法，记作SBA+$\boldsymbol{S}$+$\boldsymbol{P}$方法。图5是在Oxford5k数据集上，N=25条件下，不同D时图像检索的平均精确率对比结果。图6是在Paris6k数据集上，N=25条件下，不同D时图像检索的平均精确率对比结果。

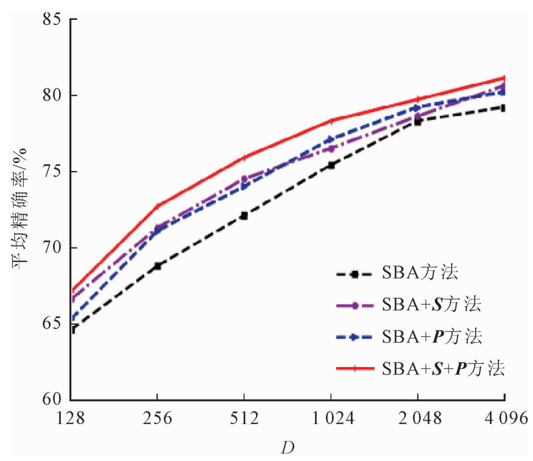


图5 Oxford5k数据集上N=25、不同D时的测试结果

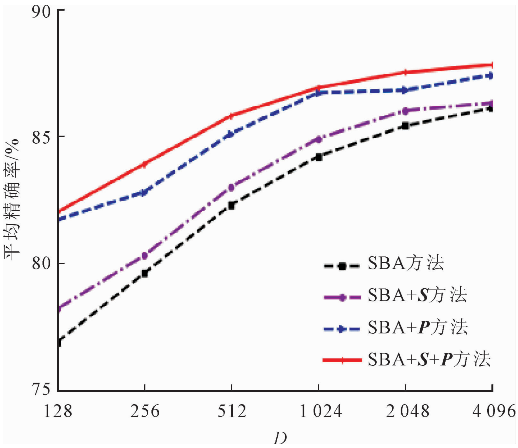


图 6　Paris6k 数据集上 $N=25$、不同 D 时的测试结果

从图 5、图 6 可以看出，本文提出的 SBA+$\boldsymbol{S}$ 和 SBA+$\boldsymbol{P}$ 方法在 Oxford5k 和 Paris6k 数据集上均能有效地提高图像检索的平均精确率，且在 D 为 128～4 096 的 6 个维度上图像检索的平均精确率均优于 SBA 方法。此外，同时使用两种加权方法即 SBA+$\boldsymbol{S}$+$\boldsymbol{P}$ 方法时，在两个数据集上检索的平均精确率均有了进一步提升，与 SBA 方法相比，在 Oxford5k 和 Paris6k 数据集上分别最多提升了 3.9% 和 5.1%。

本文提出的 SBA+$\boldsymbol{S}$ 方法赋予感兴趣区域较大权重，以突出图像的显著性区域；提出的 SBA+$\boldsymbol{P}$ 方法区分性地对待不同通道的特征图。结合两种加权方法即 SBA+$\boldsymbol{S}$+$\boldsymbol{P}$ 方法，有效地增强了图像特征的表达能力，提高了图像检索的平均精确率。

3.3　实验对比结果

在没有进行微调的情况下，本文方法与其他同类聚合方法(如 Tr. Embedding[4]、Razavian[9]、Neural code[11]、SPoC[12]、CroW[13]、R-MAC[14]、AdCoW[15]、SBA[17] 和 NetVLAD[22])在 5 个公开数据集上检索的平均精确率对比结果见表 3。

由表 3 可以看出，无论是将图像的全局特征向量表示为低维度还是高维度，本文方法在 5 个数据集上检索的平均精确率均取得了不错的结果，说明本文方法能够有效地提高图像检索的平均精确率。与 SBA 方法相比，对于扩充了 10^5 张干扰图像的 Oxford105k 和 Paris106k 数据集，本文方法在图像检索平均精确率上的提升最为显著，分别最多提升了 6.1% 和 3.9%，最少提升了 4.5% 和 2.3%。

图 7 给出了在 $N=25$ 和 $D=4\ 096$ 条件下，使用本文方法从 Oxford5k 数据集上随机选取几张查询图像得到的检索示例。图 7a 是查询图像，图 7b 是查询图像对应的排名第 1 到第 10 位的检索结果，其中实线框图像表示正确的检索结果，虚线框图像表示错误的检索结果。可以看出，利用本文方法进行图像检索时出现错误的情况较少且错误图像出现在排名较后的位置，说明利用本文方法可以有效地进行图像检索。

4　结　论

为了提高图像检索的平均精确率，本文提出了基于联合加权聚合深度卷积特征的图像检索方法——通过空间加权处理来突出图像的显著性区域并抑制背景噪声区域，赋予特征图中待检索物体区域较高的权重，同时降低不相关区域的权重；通过通道加权处理区分性地对待不同通道的特征图。在 5 个

(a)查询图像　　　　(b)查询图像对应的排名第 1 到第 10 位的检索结果

图 7　Oxford5k 数据集中查询图像对应的排名前 10 位的检索示例

公开数据集 Oxford5k、Paris6k、Oxford105k、Paris106k 和 Holidays 上的实验结果表明：与其他同类方法最优结果相比，本文方法在图像检索的平均精确率上分别最多提升了 3.0%、4.1%、5.7%、4.2% 和 16.5%。因此，采用本文所提聚合方法能够有效地进行图像检索。

表3 本文方法与其他同类聚合方法检索的平均精确率对比结果

D	方法	平均精确率/%				
		Oxford5k	Paris6k	Oxford105k	Paris106k	Holidays
4 096	SBA[17]	79.1	86.1	73.6	80.4	
	本文	81.0	87.8	78.1	82.7	89.3
2 048	SBA[17]	78.2	85.4	71.1	79.7	
	本文	79.6	87.5	76.8	82.0	88.7
1 024	SBA[17]	75.3	84.2	69.3	78.2	
	Tr. Embedding[4]	56.0		50.2		72.0
	本文	78.2	86.9	75.0	81.0	88.5
512	Tr. Embedding[4]					70.0
	CroW[13]	70.8	79.7	65.3	72.2	85.1
	Neural code[11]	55.7		52.2		78.9
	R-MAC[14]	66.9	83.0	61.6	75.7	
	Razavian[9]	46.2	67.4			74.6
	AdCoW[15]	72.8	83.0	68.1	76.3	87.4
	SBA[17]	72.0	82.3	66.2	75.8	
	NetVLAD[22]	67.6	74.9			86.1
	本文	75.8	85.8	72.3	79.7	87.6
256	Tr. Embedding[4]					65.7
	SPoC[12]	53.1		50.1		80.2
	R-MAC[14]	56.1	72.9	47.0	60.1	
	Neural Code[11]	55.7		52.4		78.9
	CroW[13]	68.4	76.5	63.7	69.1	85.1
	AdCoW[15]	70.7	80.5	66.5	74.0	87.2
	SBA[17]	68.7	79.6			
	NetVLAD[22]	63.5	73.5			84.3
	本文	72.6	83.9	68.4	77.2	86.8
128	Tr. Embedding[4]	43.3		35.3		61.7
	Neural Code[11]	55.7		52.3		78.9
	CroW[13]	64.1	74.6	59.0	67.0	82.8
	AdCoW[15]	65.8	77.9	61.4	70.5	85.7
	SBA[17]	64.5	76.9			
	NetVLAD[22]	61.4	69.5			82.6
	本文	67.0	82.0	62.1	74.7	84.8

参考文献：

[1] 徐思雨，蔡佳妮，祝继华，等. 自适应多位编码量化的哈希图像检索方法[J]. 西安交通大学学报，2017，51(8)：19-25.
XU Siyu，CAI Jiani，ZHU Jihua，et al. An adaptive hashing retrieval method of images based on multi-bit quantization [J]. Journal of Xi'an Jiaotong University，2017，51(8)：19-25.

[2] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Com-

puter Vision, 2004, 60(2): 91-110.

[3] JEGOU H, PERRONNIN F, DOUZE M, et al. Aggregating local image descriptors into compact codes [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(9): 1704-1716.

[4] JEGOU H, ZISSERMAN A. Triangulation embedding and democratic aggregation for image search [C]// Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2014: 3310-3317.

[5] JEGOU H, DOUZE M, SCHMID C, et al. Aggregating local descriptors into a compact image representation [C]// Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2010: 3304-3311.

[6] PERRONNIN F, LIU Y, SANCHEZ J, et al. Large-scale image retrieval with compressed Fisher vectors [J]. Computer Vision and Pattern Recognition, 2010, 26(2): 3384-3391.

[7] 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. Cambridge, MA, USA: MIT Press, 2012: 1097-1105.

[8] WEI Xiushen, LUO Jianhao, WU Jianxin, et al. Selective convolutional descriptor aggregation for fine-grained image retrieval [J]. IEEE Transactions on Image Processing, 2016, 26(6): 2868-2881.

[9] 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.

[10] ZHENG Liang, YANG Yi, TIAN Qi. SIFT meets CNN: a decade survey of instance retrieval [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(5): 1224-1244.

[11] BABENKO A, SLESAREV A, CHIGORIN A, et al. Neural codes for image retrieval [C]// Proceedings of the 13th European Conference on Computer Vision. Berlin, Germany: Springer, 2014: 584-599.

[12] YANDEX A B, LEMPITSKY V. Aggregating local deep features for image retrieval [C]// Proceedings of the 2016 IEEE International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2016: 1269-1277.

[13] KALANTIDIS Y, MELLINA C, OSINDERO S. Cross-dimensional weighting for aggregated deep convolutional features [C]// Proceedings of the 14th European Conference on Computer Vision. Berlin, Germany: Springer-Verlag, 2016: 685-701.

[14] TOLIAS G, SICRE R, JÉGOU H. Particular object retrieval with integral max-pooling of CNN activations [EB/OL]. (2016-02-24)[2017-08-20]. https://arxiv.org/abs/1511.05879v2.

[15] WANG Jiaxing, ZHU Jihua, PANG Shanming, et al. Adaptive co-weighting deep convolutional features for object retrieval[EB/OL]. (2018-03-20)[2018-04-20]. https://arxiv.org/abs/1803.07360.

[16] JEGOU H, DOUZE M, SCHMID C. On the burstiness of visual elements [C]// Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2009: 1169-1176.

[17] XU Jian, WANG Chunheng, QI Chengzuo, et al. Unsupervised semantic-based aggregation of deep convolutional features [J]. IEEE Transactions on Image Processing, 2018, 28(2): 601-611.

[18] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2017-07-20]. https://arxiv.org/abs/1409.1556.

[19] PHILBIN J, CHUM O, ISARD M, et al. Object retrieval with large vocabularies and fast spatial matching [C]// Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2007: 1-8.

[20] JAMES P, ONDREJ C, MICHAEL I, et al. Lost in quantization: improving particular object retrieval in large scale image databases [C]// Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2008: 1-8.

[21] JEGOU H, DOUZE M, SCHMID C. Improving bag-of-features for large scale image search [J]. International Journal of Computer Vision, 2010, 87(3): 316-336.

[22] ARANDJELOVIC R, GRONAT P, TORII A, et al. NetVLAD: CNN architecture for weakly supervised place recognition [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2016: 5297-5307.

（编辑　武红江）