

基于多层次视觉语义特征融合的图像检索算法

张霞¹, 郑逢斌²

(1. 周口职业技术学院 信息工程学院, 周口 466000; 2. 河南大学 计算机科学学院, 开封 475000)

摘要: **目的** 为了解决低层特征与中层语义属性间出现的语义鸿沟, 以及在将低层特征转化为语义属性的过程中易丢失信息, 从而会降低检索精度等问题, 设计一种多层次视觉语义特征融合的图像检索算法。**方法** 首先分别提取图像的 3 种中层特征 (深度卷积神经网络 (DCNN) 特征、Fisher 向量、稀疏编码空间金字塔匹配特征 (SCSPM)); 其次, 为了对 3 种特征进行有效融合, 定义一种基于图的半监督学习模型, 将提取的 3 个中层特征进行融合, 形成一个多层次视觉语义特征, 有效结合 3 种不同中层特征的互补信息, 提高图像特征描述, 从而降低检索算法中的语义鸿沟; 最后, 引入具有视觉特性与语义统一的距离函数, 根据提取的多层次视觉语义特征来计算查询图像和训练图像的相似度量, 完成图像检索任务。**结果** 实验结果表明, 与当前检索方法对比, 文中算法具有更高的检索精度与效率。**结论** 所提算法具有良好的检索准确度, 在医疗、包装商标等领域具有一定的参考价值。

关键词: 图像检索; 深度卷积神经网络; Fisher 向量; 稀疏编码空间金字塔匹配; 多层次视觉语义特征; 半监督学习

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 1001-3563(2018)19-0223-10

DOI: 10.19554/j.cnki.1001-3563.2018.19.038

The Image Retrieval Algorithm Based on Multi-level Visual Semantic Feature Fusion

ZHANG Xia¹, ZHENG Feng-bin²

(1. School of Information Engineering, Zhoukou Vocational and Technical College, Zhoukou 466000, China;

2. School of Computer Science, Henan University, Kaifeng 475000, China)

ABSTRACT: The work aims to design an image retrieval scheme based on multi-level visual semantic feature fusion, for the purpose of solving such problems as the semantic gap between the low layer features and the middle semantic properties, and the reduced retrieval accuracy caused by the information easily lost in the process of converting low layer features into semantic properties. Firstly, three kinds of image features (deep convolutional neural network (DCNN), Fisher vector and sparse coding spatial pyramid matching (SCSPM) feature) were extracted from the middle level. Secondly, in order to effectively integrate the three kinds of features, a graph based semi supervised learning model was defined to integrate the extracted three middle features to form a multi-level visual semantic feature, so that it could improve the image feature description and thus reduce the semantic gap of the retrieval algorithm by effectively combining the complementary information of three different middle features. Finally, the distance function with visual and semantic unity was introduced and the similarity measure between the query image and the training image was calculated based on the extracted multi-layer visual semantic features to finish the image retrieval task. The experimental results showed that the proposed algorithm had higher retrieval precision and efficiency compared with current popular retrieval methods. The proposed algorithm has good retrieval accuracy, and it has certain reference value in the fields of medical treatment and packaging trademark.

KEY WORDS: image retrieval; depth convolution neural network; Fisher vector; sparse coding spatial pyramid matching; multi-level visual semantic features; semi supervised learning

收稿日期: 2017-10-23

基金项目: 国家自然科学基金 (41571417, 60973126); 河南省高等学校重点科研项目 (15A520010)

作者简介: 张霞 (1977—), 女, 硕士, 周口职业技术学院讲师, 主要研究方向为图像处理、网络信息安全、物联网技术。

随着通信、多媒体、拍照等技术的迅速发展,以图像为信息载体的应用快速膨胀,成为当今社会中不可缺少的重要部分^[1]。针对浩瀚的数字图像,迅速、精确地从图像库中寻找特定的目标图像已经成为一个重要的课题。图像检索技术在计算机视觉、信息检索、卫星遥感等领域广泛应用,具有非常重要的应用价值^[2-3]。

符祥等^[4]定义了一种兴趣点特征的图像检索方案。其通过兴趣点的局部灰度值,从而计算该点的局部 Zernike。然后,根据得到的 Zernike 矩,计算不同兴趣点的 Euclidean 距离,搜索出最佳匹配点,消除无相关特征点。最后,通过兴趣点的空间离散度对图像进行相似测量。该方法具有优异的旋转、平移不变形,检索精度得到改善。由于没有将人机交互的反馈机制在检索算法使用,导致该算法获取的相似区并不一定符合用户的兴趣区域,出现了语义鸿沟。Pallikonda 等^[5]设计了一种色彩空间累加直方图的图像检索方案。该方法根据 HSI 色彩空间,定义了图像划分技术,以计算不同块区域的颜色直方图,获取图像的 HSI 色彩空间。然后通过计算查询图像与图像数据库中的累加直方图的相似性,输出检索结果。该算法对图像颜色特征明显的图像检索效果良好。但是,颜色量化容易丢失部分颜色信息,处于量化分界处颜色易被错分到不同的量化层次。并且,彩色直方图只保留了颜色统计信息,丢失了颜色的空间分布信息,导致该算法对颜色不显著的图像检索性能不太理想。周燕等^[6]定义了 2D 压缩感知与分层特征的图像检索技术。该方法在 RGB 空间上网格离散分割,对图像分层投射,构建一个网格空间的灰度共生矩阵。通过 2D 测量模型计算纹理特征、色彩统计特征。然后,通过 AGDR 算子测量查询图像间的源信号差异与稀疏值。再结合分层特征差异、稀疏值和色彩统计特征进行相似度量。该方法能够有效检索用户需要的图像,但是计算成本较高。Yanhong 等^[7]设计了基于稀疏表示的中的词汇与特征融合的图像检索。首先,在训练数据集上生成大量词汇表,对于给定的查询图像,采用稀疏表示来选择检索词汇。在该方法中,大词汇量可以保证较高的准确度,而中等规模的词汇表则负责高召回率。此外,为了减少量化误差和提高查全率,稀疏表示方案被用于视觉词量化。同时,融合了局部特征和全局特征,提高了召回率,但是稀疏表示对特征提取不够全面,使其检索精度还待提高。

近年来,关于图像内容的检索(CBIR)方法越来越流行,CBIR 的关键技术是要准确提取图像特征。在实际过程中,单一的特征难以准确描述图像,因此,可通过融合多个特征,综合图像的颜色、形状、纹理的等特征进行图像描述,根据融合后的特征,计算查询图像与数据库图像的相似度量。对此,为解决低层

特征转化为语义概念的过程中信息丢失严重等问题,文中通过寻找不同的中层特征互补性来弥补语义鸿沟,从而设计了一种新的图像检索算法。

1 图像特征描述

1.1 DCNN 特征

语义属性是用于图像表示的中层编码方法,文中利用深层的学习语义属性特征来表示复杂背景下的的大尺度图像。为此,通过引入深度卷积神经网络(DCNN)来提取图像数据集中的语义属性特征^[7],称其为“DCNN 特征”。在文献[7]中,采用 DCNN 对测试图像分类。在文中,介绍一种在 DCNN 全连接层中生成的弱属性作为语义属性。如图 2 所示,图像上的卷积是权重向量(称为卷积核) m 和输入向量 p 之间的运算。卷积是取图像中 m 近邻权重向量的点积来生成一个新的映射 $c_m = m^T p_{j-m+1:j}$ 。

对于卷积层,有 2 个参数应当确定:内核的大小(K_x, K_y)和映射的数量。映射的数量表示在最后一个卷积层中内核生成了多少数量的不同映射。对于一个特定的映射以及在有效映射上移动的不同大小的内核,其经相同内核加权的神经元可能与不同的输入域共享相同的权重,然后使用一个最大池化层在相邻层进行下像素采样。最大池化运算则在某一矩形区域内生成了最大的活化,见图 1。

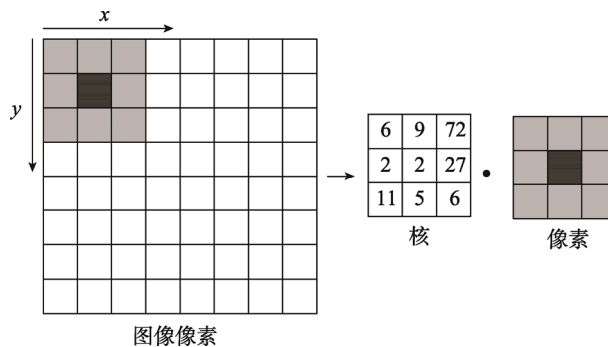


图 1 卷积运算与像素核
Fig.1 Convolution operation and pixel kernel

DCNN 的整体架构见图 2。DCNN 包含 8 个学习层,即由 5 个卷积层和 3 个全连接层构成。在 5 个卷积层中,每层的输出作为下一层的输入。在第 1 层中,96 个大小为 $11 \times 11 \times 3$ 的卷积内核被用于过滤输入图像,第 1 层的输出通过最大池化来执行。其输出的结果将作为第 2 卷积层(256 个大小为 $5 \times 5 \times 96$ 的内核)的输入。将第 2 层中所有的内核映射的输出作为第 3 层的输入,而且第 2 层的输出也同样通过最大池化来执行。需要注意的是,只有第 2 层和第 3 层间的连接是完全连接。然后第 3 层、第 4 层和第 5 层逐个连接

在一起,其间无任何的池化或标准化,它们的内核大小分别为 $3 \times 3 \times 256$, $3 \times 3 \times 384$ 和 $3 \times 3 \times 384$ 。在这 3 个神经层中,每一个都有 4096 个神经元。最后一个全连接的输出是一个 4096 维向量,其中的每个要素代表一个特定的图像相关的属性。最后,利用 4096 维向量作为语义属性来表征输出图像。

1.2 Fisher 向量特征

文中将 Fisher 向量作为第 2 个中层图像表示, 图像的 Fisher 向量表示可以看作是视觉单词包 (BOW) 的扩展^[8]。与视觉单词的出现次数 (0 阶统计) 计算有所不同, Fisher 向量是通过提取一个由生成性概率模型导出的梯度向量 (达到第 2 阶) 来表征图像。换句话说, 与 BOW 相比, Fisher 向量是利用高阶统计量来进行编码, 见图 3。另外, Fisher 向量通过一个较小词汇完成一个更高维分配, 从而能达到较低的计算成本。

图像中提取 Fisher 向量的进程见图 3。假设 $X=\{x_t, t=1 \dots T\}$ 为从图像中提取出的局部低级别描述符, 参数 $\lambda=\{w_i, \mu_i, \sigma_i, i=1 \dots N\}$ 代表在一个大的局部描述符集合上修整视觉词汇的概率, 则 X 可使用混合高斯模型 (GMM) 的梯度向量来表征^[9]:

$$\mathbf{G}_\lambda^X(I) = \frac{1}{T} \sum_{t=1}^T \nabla_\lambda \log p(x_t | \lambda) \quad (1)$$

通过式 (1) 寻找方向, 模型中参数 λ 可更好拟合数据。再将梯度向量的 Fisher 内核作为 Fisher 向量 $\mathbf{F}_i(I)$ 的点积, 表示为:

$$\Gamma_{\lambda}^X(I) = L_{\lambda} \mathbf{G}_{\lambda}^X(I) \quad (2)$$

设 GMM 模型为 $p(x|\lambda) = \sum_{i=1}^N w_i p_i(x|\lambda)$, 参数 $\lambda = \{\mu_i, \sigma_i, i=1 \dots N\}$, 其中 w_i, μ_i 和 σ 分别为高斯混合权重、均值向量和标准偏差向量 ($((\sigma_i^1)^2 \dots (\sigma_i^D)^2)$ 是协方差矩阵 Σ_i 的斜列), N 为高斯数。 w_i 的偏导数被舍弃, 因为它们仅载有很少的鉴别信息。 μ_i 和 σ_i 的梯度为:

$$\Gamma^X \mu_i^d(I) = \frac{1}{T\sqrt{w_i}} \sum_{t=1}^T \gamma_i(x_t) \left(\frac{x_t^d - \mu_t^d}{\sigma_i^d} \right) \quad (3)$$

$$\Gamma^X \sigma_i^d(I) = \frac{1}{T\sqrt{2w_i}} \sum_{t=1}^T \gamma_i(x_t) \left(\frac{x_t^d - \mu_t^d}{\sigma_i^d} - 1 \right) \quad (4)$$

式中: $\gamma_i(x_t)$ 是描述符 x_t 到高斯 i 的软分配, 因此, 图像 I 的特征是向量 $\Gamma^X \mu_i^d(I)$ 和 $\Gamma^X \sigma_i^d(I)$ 的串联。

在文中,为降低特征维数,选取密集的尺度不变特征变换(SIFT)作为局部描述符^[10],然后利用主成份分析(PCA)来进一步降低SIFT特征维数至64。

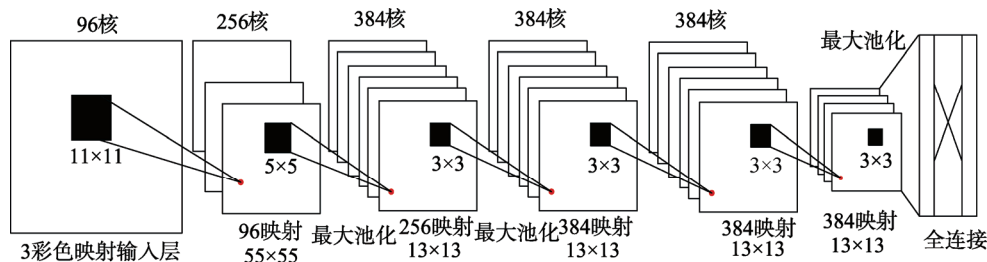


图 2 DCN 整体架构

Fig.2 Overall architecture of DCN

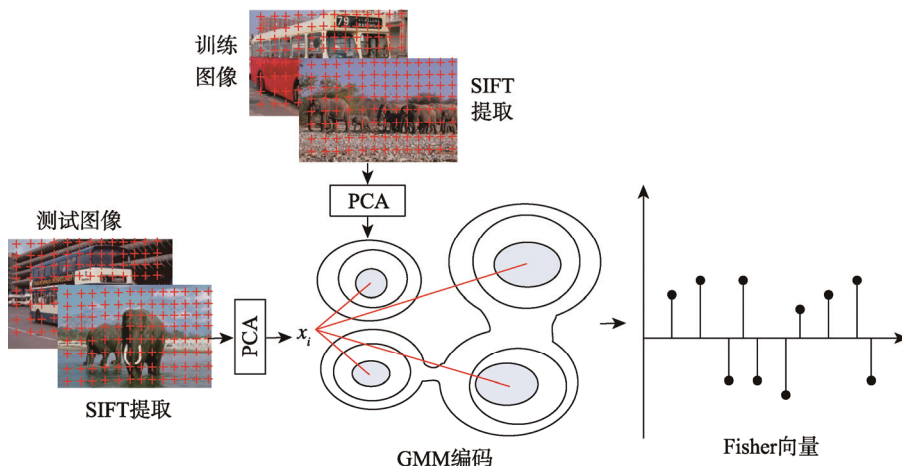


图 3 Fisher 向量提取的进程
Fig.3 Process of extracting Fisher vectors

在初步学习阶段,从训练图像 SIFT 描述符中抽取 25 万个 SIFT 描述符,然后学习 64-质心 GMM 模型,以拟合这些采样描述符的分布。利用期望最大化(EM)来学习 GMM 模型的参数。将一些空间信息整合到 Fisher 向量,把输入图像分为 4 个大小相等的小块,分别按照式(3)和(4)生成每个块的 Fisher 向量,然后从每个图像块中提取 Fisher 矢量,以组合形成了一个 $64 \times 64 \times 2 \times 4 = 32\,768$ 维的 Fisher 向量。

1.3 SCSPM 特征

稀疏编码空间金字塔匹配(SCSPM)是一种高阶统计特征^[11],它是基于属性探测器的 SIFT 特征的稀疏编码(SC)。在 SCSPM 中,K-均值矢量量化(VQ)被稀疏编码所取代,通过 SC 表示图像特征^[12]:

$$\min_{U,V} \sum_{m=1}^M \|x_m - u_m V\|^2 + \lambda |u_m| \quad (5)$$

一般地,码本 V 是一个超完备基组,换言之,它是一组用于表示数据的稀疏特征。同时, U 是将学习特征从特征空间转换为图像数据空间的关联基础。

文中从图像中提取了一组密集的 SIFT,然后通过稀疏编码来学习大小为 1024 的视觉词典。为了让该表征具有更多的不变性,在 SCSPM 中使用了多尺度空间最大池化策略。在文中 SPM 框架中,采用了三级金字塔,在第 1 层、第 2 层、第 3 层上分别有 16, 4, 1 个小块。对于每一个图像,文中从中提取了 $1024 \times (16+4+1) = 21\,504$ 维特征。

相对 SPM 算法利用的 VQ 方法,SC 的量化精度更高。由于 SC 可在量化中获到多个不同权值的码本,而不是 VQ 中单纯的码本,因此,通过 SC 的约束条件可保留更多细节信息。

2 文中图像检索算法

由于不同特征对的单图模型具有不同的响应,因此,如何对不同模型生成的结果进行融合,并对其进行优化就成了一个问题。在文中,引入了一种基于图的半监督模式,将其用于融合不同类型的图像特征,并解决多特征融合问题^[13]。在以下部分中,首先描述了单图模型,然后讨论将不同单图模型进行融合。

2.1 单图模型

在单图模型中,顶点是标签与未标签的样本,边缘则是样本对之间的相似关系。其采用标签平滑假设来评价图解模型。

将 $\chi = \{x_1 \dots x_l x_{l+1} \dots x_n\} \subset R^m$ 定义为与样本相对应的特征集,并且 W 是关联矩阵,每一个在 W 中的元素 W_{ij} 表示样本对之间的相似性;其中的一个正半径参数 δ 和一个特征之间的测距为:

$$W_{ij} = \begin{cases} \text{similarity}(x_i, x_j) & \text{if } i \neq j \\ 0 & \text{else} \end{cases} \quad (6)$$

式中: $\text{similarity}()$ 为相似函数,其计算过程为:

D 为一个对角矩阵, $D_{ii} = \sum_j W_{ij}$ 。目标函数由方程(7)的正则化来表示:

$$\arg \min \{ \beta \text{tr}(f^T L_b f) + \|f - Y\|_F^2 \} \quad (7)$$

式中: β 为一个正的正则化参数; L_b 是归一化图的 Laplace 矩阵 $L_b = D^{-1/2} (D - W) D^{-1/2}$ 。式(7)中的 Y 和 f 是 2 个大小为 n 的向量。 Y 和 f 同为样本 x_i 的对应分数, Y 为 x_i 的标签分数。当 x_i 是正数时 Y 为 1,当 x_i 是负数时 Y 为 -1,当 x_i 为未标记时 Y 为 0; f 为预测分数。式(7)的闭操作的解为:

$$f = (I + \beta L_b)^{-1} Y \quad (8)$$

式(8)可以更加有效地通过标签传播来解答,而标签传播执行一系列的迭代更新过程,直到 f 收敛为止。

2.2 多重图模型

融合过程相当于利用半监督学习来将不同的单图模型集成到一个正则化框架中。设用 G 表示不同类型图的个数, α_g 表示每个图的正权重参数,则多图图解模型的价值函数为:

$$\arg \min_{f, \alpha} \left\{ \sum_{g=1}^G \alpha_g (\beta_g \text{tr}(f^T L_{b_g} f) + \|f - Y\|_F^2) \right\} \quad (9)$$

条件 $\sum_{g=1}^G \alpha_g = 1$

式中: W_g 为图 g 的关联矩阵; D_g 为图 g 的对角矩阵。

从式(9)中看出,该函数与 α 成线性相关关系,这就意味着,当 $g = \arg \min_g f^T L_{b_g} f$ 时,则 $\alpha = 1$,只有一个图被保存。根据文献[14],可用 α 到 α^r 的变化来解决这一问题,因此,式(9)演变为:

$$\arg \min_{f, \alpha} \left\{ \sum_{g=1}^G \alpha_g^r (\beta_g \text{tr}(f^T L_{b_g} f) + \|f - Y\|_F^2) \right\} \quad (10)$$

条件 $\sum_{g=1}^G \alpha_g^r = 1$

通过式(10)可知,随着 r 的增大,模型对不同的单图更加平滑。

为了求解式(10),采用迭代更新 f 和 α 的方法来最小化方程(10)直到其收敛。过程如下所述。

首先,通过下列方程修正 f 及更新 α :

$$\alpha = \frac{\left(\frac{1}{\beta f^T L_{b_g} f + \|f - Y\|_F^2} \right)^{\frac{1}{r-1}}}{\sum_{g=1}^G \left(\frac{1}{\beta f^T L_{b_g} f + \|f - Y\|_F^2} \right)^{\frac{1}{r-1}}} \quad (11)$$

$$f = \left(I + \frac{\sum_{g=1}^G \alpha_g^r \beta L_{bg}}{\sum_{g=1}^G \alpha_g^r} \right)^{-1} Y \quad (12)$$

类似于单图模型，这里也采用归一化图的 Laplace 矩阵来找出方程的解： $L_{bg} = D_g^{-1/2} (D_g - W_g) D_g^{-1/2}$ 。

2.3 基于图的半监督学习模型及图像检索

如上所述，在构建图像表示中选择了 3 个中层特征，即 DCNN 特征、Fisher 向量和 SCSPM，以设计一种新的图像检索算法，其过程见图 4。假设已经选择了 n 个不同的图像，首先应该利用每个特征生成一个单一图模型，然后再将这些单图模型融合到多图模型中，构建了基于图的半监督学习模型。在每种类型的特征生成成为单一的图模型之前，首先，需要构建关联矩阵 W_a 。在文中， W_a 通过相似矩阵 $W_{a,ij} = \text{similarity}(x_{a,i}, x_{a,j})$ 来构建，其中 $x_{a,i}$ 和 $x_{a,j}$ 都是与图像对应的特定类型的特征。相似矩阵按照方程 (13) 来计算：

$$\text{similarity}(x_{a,i}, x_{a,j}) = \exp \left(-\frac{\|x_{a,i} - x_{a,j}\|_2^2}{2\sigma^2} \right) \quad (13)$$

其中， α 为 $\alpha_{\text{Fisher 向量(FV)}}$ ， α_{SPM} 和 α_{DCNN} ； σ^2 为所有特征距离的平均值。

然后，根据方程 (7)，生成如下 3 个单一的图模型的价值函数：

$$\rho_{\text{FV}}(f) = \beta_{\text{FV}} \text{tr}(f^T L_{\text{bFV}} f) + \|f - Y\|_F^2 \quad (14)$$

$$\rho_{\text{SPM}}(f) = \beta_{\text{SPM}} \text{tr}(f^T L_{\text{bSPM}} f) + \|f - Y\|_F^2 \quad (15)$$

$$\rho_{\text{DCNN}}(f) = \beta_{\text{DCNN}} \text{tr}(f^T L_{\text{bDCNN}} f) + \|f - Y\|_F^2 \quad (16)$$

再将 3 个单图模型整合到一个正则化框架中来生成多重图模型，其相应的价值函数为：

$$\rho(f, \alpha) = \alpha_{\text{FV}}^r \rho_{\text{FV}}(f) + \alpha_{\text{SPM}}^r \rho_{\text{SPM}}(f) + \alpha_{\text{DCNN}}^r \rho_{\text{DCNN}}(f) \quad (17)$$

条件 $\alpha_{\text{FV}} + \alpha_{\text{SPM}} + \alpha_{\text{DCNN}} = 1$

式 (17) 就是所提的基于图的半监督学习模型，它是利用 3 个特征对应的权重，联合单图模型来构建未标记样本与标记样本的权重图模型，同时，借助基于图的 Gaussian 随机场解决学习问题。每个图模型对应每个样本的邻域关系，在学习中选取置信度高的未标记样本，并根据每个特征的权重大小相互学习。基于图的半监督学习模型融合了多个特征，集合了不同特征的互补信息，解决单一信息不足的问题，显著提高了图像特征的描述能力，降低了语义鸿沟问题，为检索过程提供了有力依据。

最后，为完成查询与检索任务，引入了具有视觉特性与语义统一的距离函数 $D_j^{[15]}$ ，其函数如下：

$$D_j = \text{SIG}(j) + \beta \left\{ \frac{1}{N_R} \sum_{k \in N_R} \left[\left(1 + \alpha \sum_{i=1}^{N_j} \zeta_i \times R_j^i \right) S_{jk} \right] \right\} -$$

$$\gamma \left\{ \frac{1}{N_N} \sum_{k \in N_N} \left[\left(1 + \alpha \sum_{i=1}^{N_j} \zeta_i \times R_j^i \right) S_{jk} \right] \right\} \quad (18)$$

式中： $\text{SIG}(j)$ 为关键字与图像 j 的匹配度； N_R 与 N_N 为正反馈、负反馈的数量； α 为语义相关性在相似度的权值。例 $\alpha=0$ 时，其只能采用视觉特性相似度反馈。 S_{ij} 为图像 i 与 j 视觉的相似性。为简化计算， β, γ 均设置为 1。对于 $\text{SIG}(j)$ ，其定义如下：

$$\text{SIG}(j) = \begin{cases} \frac{M_j}{N_j} \sum_{i=1}^{M_j} R_j^i & Q \in K_j \\ 0 & Q \notin K_j \end{cases} \quad (19)$$

式中： N_j 为 j 中的关键字数量； M_j 为与 j 相符合的关键字数量。

依据式 (18) 可知，将视觉特征、语义属性与多层相关反馈整合在一起；并且调整语义相关性的权值 α 来确定相似度量中的分量，能够有效适用不同图像的相似感知。

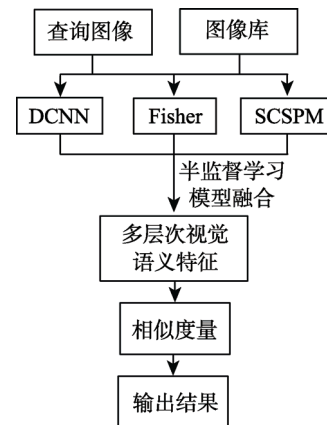


图 4 文中算法结构

Fig.4 Structure of the proposed algorithm

3 实验与讨论

为验证算法性能，利用 Corel 数据库进行测试。Corel 包含 10 种类型图像^[16]：非洲、海难、建筑、公共汽车、恐龙、大象、花卉、马、山川与食物，每类含 100 幅，共 1000 幅，见图 5。测试条件为：Intel i5 4 核 CPU，3.50 GHz，RAM 为 8 GB，64 位 WIN8 系统。测试平台为 Matlab2010。

测试之前，对 Corel 进行处理，将每类随机分成 5 份，每份 20 幅图。选取 4 份图像为训练组，1 份为测试组。为了获得最优的结果，文中的参数设置为 $r=2$ ， $\beta_{\text{DCNN}}=1$ ， $\beta_{\text{FV}}=1$ ， $\beta_{\text{SCSPM}}=1$ ， $\gamma=1$ ， $\alpha=0$ 。为突显文中算法的优异性，选取当前流行的检索算法：文献[4]、文献[5]、文献[6]和文献[7]对比，为便于描述，依次记为 A 算法、B 算法、C 算法、D 算法。其中，文献[4]的检索技术，也就是 A 算法，它是利用图像的兴



图 5 Corel 图像库
Fig.5 Corel image database

趣点来完成图像检索,利用图像兴趣点的局部灰度信息来获取对应的 Zernike 矩,通过比较各兴趣点局部 Zernike 矩的欧式距离提取最优匹配点对,去除不匹配的兴趣点,然后利用兴趣点的空间离散度来估计图像内容的相似度。该技术利用了兴趣点的抗旋转与平移特性,具有良好的鲁棒性与检索精度。B 算法是设计了一种色彩空间累加直方图的图像检索技术,根据 HSI 色彩空间,定义了图像划分技术,以计算不同块区域的颜色直方图,获取图像的 HSI 色彩空间。然后通过计算查询图像与图像数据库中的累加直方图的相似性,输出检索结果,该算法对图像颜色特征明显的图像具有较高的检索准确度。文献[6]技术,也就是 C 算法,它是利用 2D 压缩感知与分层特征来实现图像检索,该方法主要是在 RGB 空间上对图像分层投射,而且通过定义一种基于颜色网格空间的扩展灰度共生矩阵,采用二维测量模型获取图像的分层测量特征、纹理特征与分层颜色统计特征,最后,基于 AGDR 算法计算检索图像之间的原始信号差量及其稀疏值,通过这些差量、稀疏值和颜色统计特征来实现目标查询,该技术综合了图像的多种特征,能够获取良好的检索精度。文献[7],也就是 E 算法,主要是设计了基于稀疏表示的中的词汇与特征融合的图像检索,首先在训练数据集上生成大量词汇表,对于给定的查询图像,采用稀疏表示来选择检索词汇,其中的大词汇量可以保证较高的准确度,而中等规模的词汇表则负责高召回率。此外,为了减少量化误差和提高查全率,稀疏表示方案被用于视觉词量化。同时,融合了局部特征和全局特征,提高了检索精度。该技术不仅兼顾了图像的全局与局部特征,同时还引入了稀疏表示来获取查询目标的词汇,具有良好的代表性与新颖性。

3.1 性能评价

为测试算法的性能,利用图像分类、检索中最常用的评价指标:查准率(Precision)和召回率(Recall)以及精度(Accuracy)。其中,Precision 和 Recall 分别定义如下^[17]:

$$P = \frac{T_p}{T_p + F_p} \quad (33)$$

$$R = \frac{T_p}{T_p + F_N} \quad (34)$$

式中: T_p 为正确识别伪造数量; F_p 为误判为伪造数量; F_N 为漏检的伪造数量。

算法精度 A_c 计算为:

$$A_c = \frac{T_p + T_N}{T_p + F_p + F_N + T_N} \quad (35)$$

式中: T_N 为真实图像正确识别数量。

3.2 实验结果

图 6a、图 7a 是分别以恐龙、建筑为对象的查询图像。A 算法、B 算法、C 算法、D 算法以及文中算法的检索结果见图 6b—f、图 8b—f。根据图 6b—f 可知,文中算法查询的相关图像与恐龙图像相似性最高,查询返回的图像中含有了较多的恐龙图像,只有一个非相关图像-大象,见图 6f。对于 A 算法,见图 6b,其输出结果中出现了 3 幅大象图与 1 幅马儿图像。B 算法的输出结果中出现了 1 幅大象图与 1 幅马儿图,见图 6c。C 算法中出现了 2 幅大象图,见图 6d。对于 D 算法,其输出结果同样存在错误检索,见图 6e。在图 7 的结果中同样可看出,对于不同对象的查询结果,文中方案均取得了良好的检索效果,输出结果中只有 1 幅误检图像,而其他算法的检索结果均出现了多幅误检结果。根据这些实验结果可知,文中算法检索性能更优,返回的图像与目标图像相似性高。主要原因文中通过提取图像的 DCNN 特征、Fisher 向量、SCSPM 等 3 种特征,并定义一种基于图的半监督学习模型,将提取的 3 个中层特征融合,组合为一个多层次视觉语义特征,提高了图像特征描述。最后,设计了具有视觉特性与语义统一的距离函数,根据提取的多层次视觉语义特征进行查询图像与训练图像的相似度量,完成图像检索任务,从而提高了算法精度。A 算法中没有将人机交互的反馈机制在检索算法使用,导致该算法获取的相似区并不一定符合用户的兴趣区域,出现了语义鸿沟,降低了检索性能。B 算法中采用的颜色空间直方图中,颜色量化容易丢失部分颜色信息,处于量化分界处颜色易被错分到不同的量化层次,并且丢失了颜色的空间信息,导致该算法对

颜色特征不显著的图像检索性能不太理想。C 算法中得到的结果还比较理想，但是其参数敏感，过程不可控。D 算法虽然采用了词汇量与稀疏表示来精确搜索目标图像，但是稀疏表示对特征提取不够全面，忽略图像的统计特征，使其精度还待提高，仍然会出现一些不相关检索。

另外，为了验证所提算法对物互网上的图像检索

效果，文中从互联网上选择 200 幅图像作为一个数据集，其中有 20 幅图像是茄子，剩下的图像分别是洋葱、秋葵、胡萝卜、苦瓜以及南瓜。利用所提算法从此数据库中，检索出这 20 幅辣椒图像，结果见图 8。依图 8 可知，所提算法对于检索互联网上的图像仍然具有较高的检索精度，输出结果中只有 2 个错误检索目标——南瓜与洋葱。

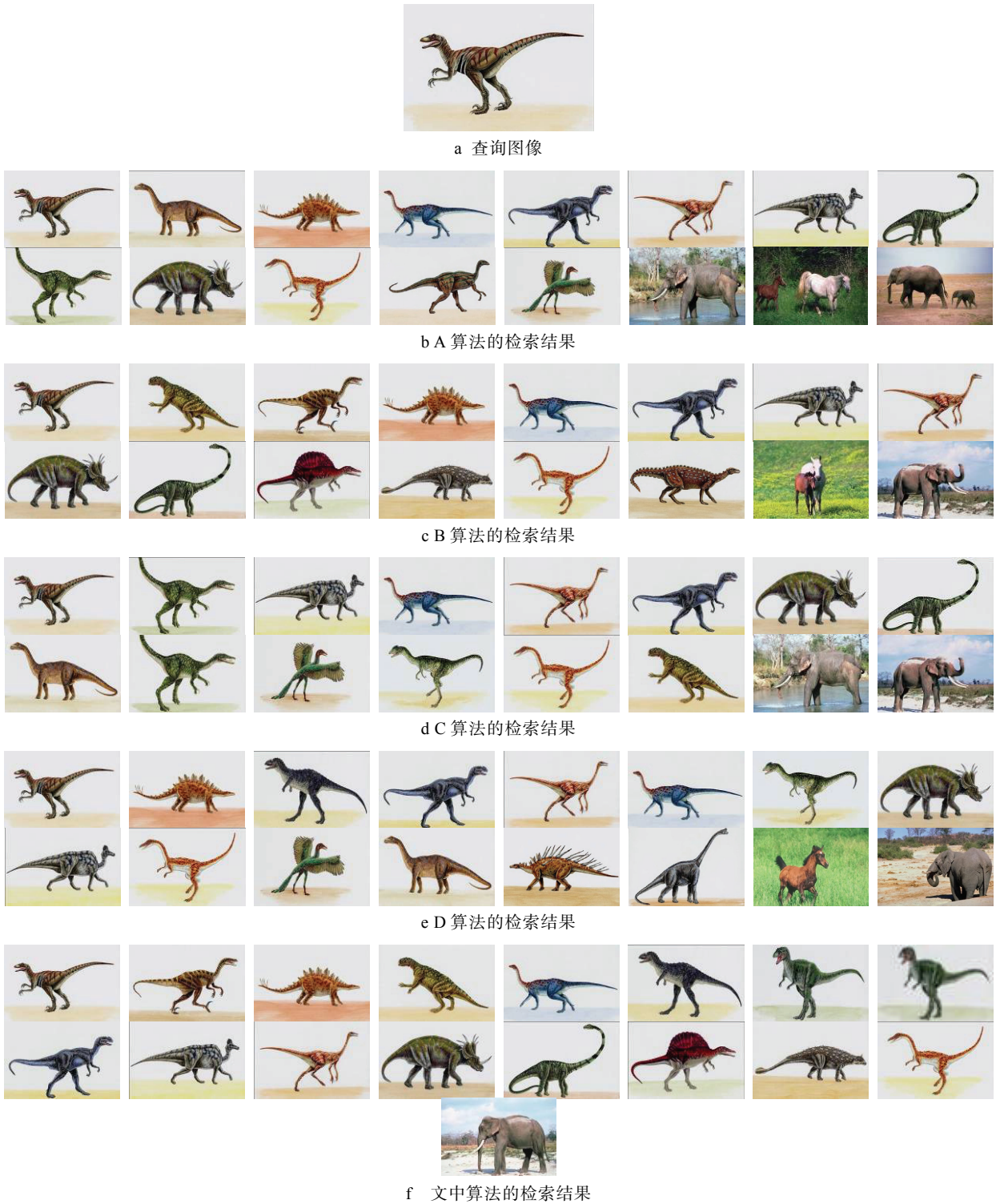


图 6 恐龙图像查询结果
Fig.6 Dinosaur image query results

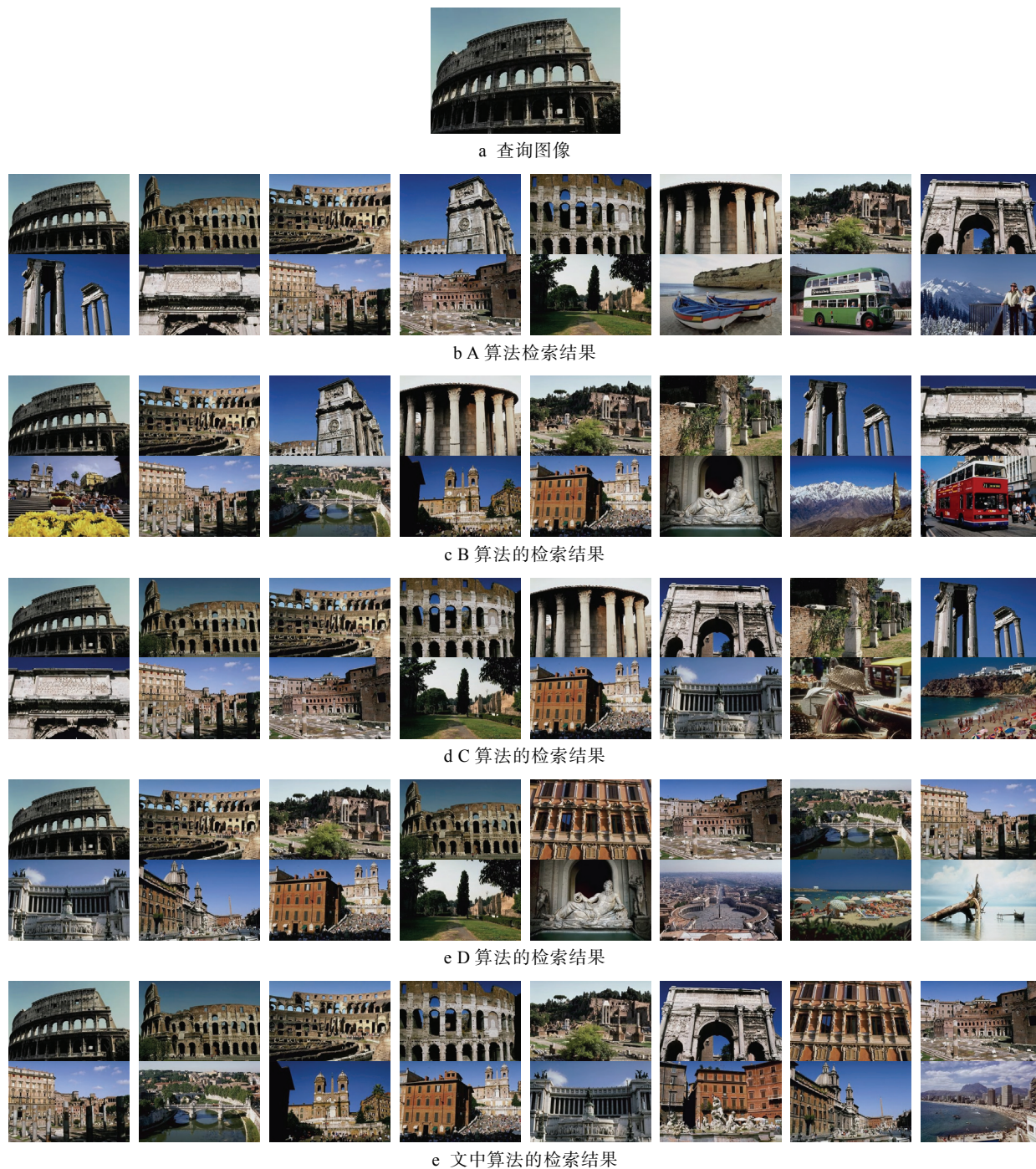


图 7 建筑图像查询结果

Fig.7 Architectural image query results

Corel 图库中不同类型图像的查准率测试见图 9, 依据图 9 中看出, 针对不同类别的图像, 不同算法的查准率有所不同, 总体来说, 文中方法的查准率优于其他 3 种对照组。不同算法的查准率-召回率曲线见图 10, 其能够综合反映算法的性能, 广泛应用于图像分类、检索中。从图 10 中看出, 文中算法的查准率-召回率曲线表现最佳, 其检索性能较好。

3.3 算法效率计算

良好的检索技术在保持较高精度的同时也要有

高的效率^[5]。为了较全面反映算法的性能, 在 Corel 库中对算法的检索效率进行统计, 结果见表 1。从表 1 得知, 文中方法检索时耗最低, 而对照组技术的效率均要低于所提算法。主要是文中采用了基于 SIFT 与稀疏编码提取特征, 降低了特征维度, 并且融合 3 种特征, 结合了视觉特性与语义统一性质, 提高了效率。A 算法中需要提取的兴趣点较多, 使其时耗较高。B 算法中色彩直方图累加过程需消耗较多时间。C 算法中在计算纹理特征、色彩统计特征时需耗费了大量



图 8 互联网上的图像检索测试
Fig.8 Image retrieval test on Internet

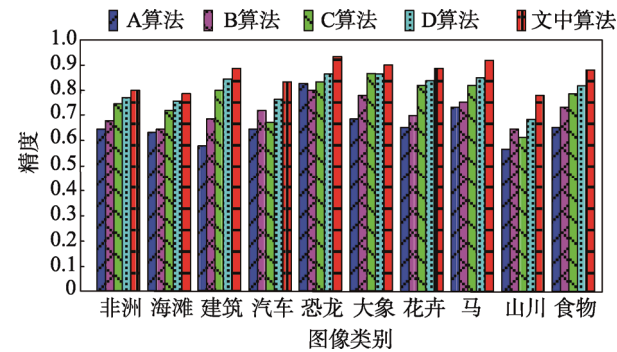


图 9 Corel 中不同类别的精度计算
Fig.9 Calculation of different classes of accuracy in Corel

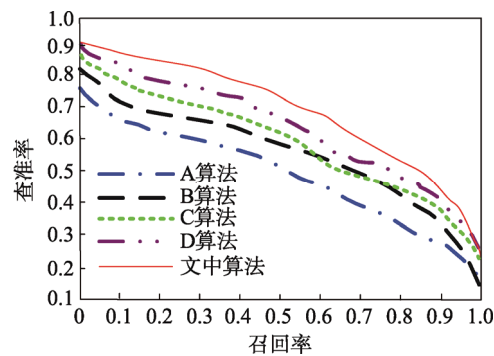


图 10 查准率-召回率曲线结果
Fig.10 P-R curve results

表 1 算法时间
Tab.1 Algorithm time

算法	时间/s	
	k=10	k=20
A 算法	4.56	11.25
B 算法	3.61	8.82
C 算法	13.69	36.55
D 算法	3.42	7.36
文中算法	2.44	5.73

的时间，降低了效率。D 算法利用了稀疏表示来完成相似度搜寻，降低了计算成本，但是该技术需要在训练数据集上生成大量词汇表，这在一定程度上增加了算法的时耗。

4 结语

为了解决低层特征转化为语义概念的过程中，易丢失信息而导致检索精度不高的问题，文中通过寻找不同的中层特征的互补性来弥补语义鸿沟，设计了一种将 3 种中层特征的多层次视觉语义特征融合的图像检索方案。提取图像 DCNN 特征、Fisher 向量、SCSPM 特征作为图像描述。为了将得到的 3 种特征有效融合，获得互补性良好的图像特征描述，设计了一种图的半监督模型，得到一个多层次视觉语义特征。为完成图像精确检索，定义了视觉特性与语义统一的距离函数，利用多层次视觉语义特征进行查询图像与训练图像的相似度量。实验证明了提出算法具有优异的检索精度与效率，可获得了良好的查准率-召回率曲线。

参考文献：

[1] XU H J, HUANG C Q, PAN P. Image Retrieval Based on Multi-concept Detector and Semantic Correlation[J]. Science China (Information Sciences), 2015(12): 104—118.

[2] 王荣芝, 王莉, 李佑婷. 基于字典统计耦合归一化多重距离的图像检索算法[J]. 包装工程, 2017, 38(5): 228—233.

WANG Rong-zhi, WANG Li, LI You-ting. Image Retrieval Algorithm Based on Dictionary Statistical Coupling Normalized Multiple Distance[J]. Packaging En-

- gineering, 2017, 38(5): 228—233.
- [3] 杨叶芬, 曾东海, 刘海. 逆向学习耦合多属性查询的图像排序/检索优化算法研究[J]. 包装工程, 2015, 36(7): 84—90.
YANG Ye-fen, ZENG Dong-hai, LIU Hai. Research on Image Sorting/Retrieval Optimization Algorithm Based on Reverse Learning Coupled with Multiple Attribute Queries[J]. Packaging Engineering, 2015, 36(7): 84—90.
- [4] 符祥, 曾接贤. 基于兴趣点匹配和空间分布的图像检索方法[J]. 中国激光, 2010, 37(3): 774—778.
FU Xiang, ZENG Jie-xian. A Novel Image Retrieval Method Based on Interest Points Matching and Distribution[J]. Chinese Journal of Lasers, 2010, 37(3): 774—778.
- [5] PALLIONDA S S, KRISHNA K S R, KRISHNA I V M. Content Based Image Retrieval Based on Different Global and Local Color Histogram Methods[J]. Journal of The Institution of Engineers, 2017, 98(1): 129—135.
- [6] 周燕, 曾凡智. 基于二维压缩感知和分层特征的图像检索算法[J]. 电子学报, 2016, 44(2): 453—460.
ZHOU Yan, ZENG Fan-zhi. An Image Retrieval Algorithm Based on Two-dimensional Compressive Sensing and Hierarchical Feature[J]. Acta Electronica Sinica, 44(2): 453—460.
- [7] WANG Y H, CEN Y G, ZHAO R Z. Separable Vocabulary and Feature Fusion for Image Retrieval Based on Sparse Representation[J]. Neurocomputing, 2016, 236(C): 14—22.
- [8] PERRONNIN F. Improving the Fisher kernel for large-scale Image Classification[J]. European Conference on Computer Vision, 2014, 35(14): 143—156.
- [9] 华媛蕾, 刘万军. 改进混合高斯模型的运动目标检测算法[J]. 计算机应用, 2014, 34(2): 580—584.
HUA Yuan-lei, LIU Wan-jun. Moving Object Detection Algorithm of Improved Gaussian Mixture Model[J]. Journal of Computer Applications, 2014, 34(2): 580—584.
- [10] 刘佳, 傅卫平, 王雯. 基于改进 SIFT 算法的图像匹配[J]. 仪器仪表学报, 2013, 34(5): 1107—1112.
LIU Jia, FU Wei-ping, WANG Wen. Image Matching Based on Improved SIFT Algorithm[J]. Chinese Journal of Scientific Instrument, 2013, 34(5): 1107—1112.
- [11] YANG J, YU K, GONG Y. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification[J]. Computer Vision and Pattern Recognition, 2016, 36(15): 1794—1801.
- [12] 赵嵩, 冯湘. 一种基于稀疏编码空间金字塔匹配的图像分类算法[J]. 应用光学, 2016, 37(5): 706—711.
ZHAO Song, FENG Xiang. Sparse Coding Based Spatial Pyramid Matching Algorithm for Image Classification[J]. Journal of Applied Optics, 2016, 37(5): 706—711.
- [13] 孔超, 张化祥, 刘丽. 基于兴趣区域特征融合的半监督图像检索算法[J]. 山东大学学报(工学版), 2014, 44(3): 22—28.
KONG Chao, ZHANG Hua-xiang, LIU Li. A Semi-supervised Image Retrieval Algorithm Based on Feature Fusion of the Region of Interest[J]. Journal of Shandong University (Engineering Edition), 2014, 44(3): 22—28.
- [14] WANG M, HUA X S, HONG R. Unified Video Annotation Via Multigraph Learning[J]. IEEE Trans Circuits System Video Technology, 2014, 19(5), 733—746.
- [15] 王宗尧, 刘金岭. 基于支持向量机的 PU 中文文本分类器构建[J]. 南京邮电大学学报(自然科学版), 2015, 35(6): 100—105.
WANG Zong-yao, LIU Jin-ling. PU Chinese Text Classifier Based on Support Vector Machine Construction[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2015, 35(6): 100—105.
- [16] 张玉兵, 宋威. 基于局部对角色差描述符的图像检索方法[J]. 计算机工程, 2016, 42(12): 248—253.
ZHANG Yu-bing, SONG Wei. Image Retrieval Method Based on Local Diagonal Color Difference Descriptor[J]. Computer Engineering, 2016, 42(12): 248—253.
- [17] 吴晓雨, 何彦, 杨磊, 等. 基于改进形状上下文特征的二值图像检索[J]. 光学精密工程, 2015, 23(1): 302—309.
WU Xiao-yu, HE Yan, YANG Lei, et al. Binary Image Retrieval Based on Improved Shape Context Algorithm[J]. Optics and Precision Engineering, 2015, 23(1): 302—309.