

# 基于深度学习与拓展查询的商标图像检索方法<sup>\*</sup>

郭升挺 黄 晞 柯俊敏 陈宝林 吴家飞 苏浩明

(福建师范大学 医学光电科学与技术教育部重点实验室 福建省光子技术重点实验室 福州 350007)

**摘要:** 为了解决商标图像检索时,由于图像抽象性和复杂性的特点所带来检索时的“语义鸿沟”问题,提出一种基于深度学习的商标图像检索方法。由于深度学习技术在图像识别和图像分类技术中有着优秀表现,因此我们利用它来解决商标图像检索时遇到的一系列问题。本文还改进了拓展查询算法,并利用它优化检索系统。首先,我们利用商标图像的特征构建了一个多类别的商标图像数据集。其次,利用商标图像库训练一个稳定的 AlexNet 卷积神经网络模型,并构建商标图像特征库。接着,利用训练好的网络模型提取输入商标的深层语义特征并通过距离函数进行相似性度量。最后,提出一种改进的拓展查询方法,利用其对候选图像进行拓展查询,得到相似图像。经过实验结果表明,基于深度学习的商标图像检索方法比传统基于 SVM 主动学习的图像检索方法有着更好的检索效果。

**关键词:** 深度学习,基于内容的图像检索,拓展查询

## Trademark Image Retrieval Method Based on Depth Learning and Query Expansion

GUO Shengting, HUANG Xi, KE Junmin, CHEN Baolin, WU Jiafei, SU Haoming

(Key laboratory of Photon technology in Fujian Province, Key Laboratory of Optoelectronic Science and Technology for Medicine of Ministry of Education, Fujian Normal University, Fuzhou, 350007, China)

**Abstract:** In order to solve the problem of “semantic gap” in the retrieval of trademark images, this paper presents a method of trademark image retrieval based on depth learning. Because of the excellent performance of depth learning technology in image recognition and image classification, we use it to solve a series of problems encountered in trademark image retrieval. This paper also improves the query expansion algorithm and uses it to optimize the retrieval system. First of all, we construct a trademark image data set using the trademark image. Secondly, the trademark image library is used to train a stable Alex Net convolution neural network model and construct the trademark Image feature library. Then, we use the Training network model to extract the deep semantic features of the imported trademarks and measure the similarity by the distance function. Finally, we propose an improved query expansion method, which is used to expand the candidate image and get similar images. The experimental results show that the method of trademark image retrieval based on depth learning has better retrieval effect than the traditional method based on SVM active learning.

**Keywords:** depth learning, content-based image retrieval, query expansion

## 0 引言

随着国家经济的迅猛发展,企业的不断增加,商标的注册数量也逐年倍增。随着商标图像的逐渐增加,

本文于 2017-08-03 收到,2017-08-21 收到修改稿。

<sup>\*</sup> 福建省科技厅工业高校产学研合作项目(2013H6008)。

商标管理部门对商标图像管理的难度更是成倍加大。商标是商品或企业的一个重要特征,它作为一种无形的资产,对于判别企业价值有着重要意义并代表了企业的合法权益。因此,有效的判别仿冒商标显得尤为重要。商标管理部门传统的管理商标的方法主要是采用“分类号”为关键词的商标图像检索模式,然而图像之间的相似度还是主要靠人工来识别,这种方式效率低并且准确度不是最高。

传统基于内容的图像检索主要是利用图像浅层的视觉特征进行检索。例如:形状特征<sup>[1]</sup>、颜色特征、纹理特征、HU 矩特征<sup>[2]</sup>等。虽然在一些简单商标的图像检索能容易地找到相似的图像,但是对于一些商标图像的特点是抽象的、复杂的、甚至较难理解的,这些传统的图像特征在检索时就很难有较好的检索效果。因为这些方法无法克服“语义鸿沟”问题,使得计算机得到的图像特征与人从高级语义所得到的图像特征存在着差异,从而导致检索准确度不高。因此,商标检索的最关键性问题的关键在于如何找到一个合理的方法提取商标图像的特征,这也是本文研究的内容所在。

近几年来,随着 Hinton [3] 等人解决了深度学习的模型优化问题后,深度学习技术的得到了快速的发展,也给图像检索领域带来了新的希望。深度学习可以通过组合浅层特征的方式产生更为抽象的深层次的图像内容特征,利用这些深层次的特征表示图像的属性 and 内容<sup>[4]</sup>,因此,近几年来深度学习技术在人工智能领域有了巨大的突破,并在机器视觉和语音识别等方面表现优秀。本文尝试将深度学习技术应用于商标图像检索中,判断深度学习技术能否解决商标的图像的抽象性所带来的“语义鸿沟”问题。

本文研究的商标检索方法主要解决四个问题:①如何构建一个具有不同类别的商标图像数据集;②用什么方法提取商标图像深层次的特征;③如何选择相似性度量算法,让相似的商标计算得到的相似度高;④如何使用扩展查询优化检索系统,提高系统查准率。

## 1 商标图像检索系统架构

本文设计的商标图像检索系统的框架如图 1 所示,主要包括离线部分和在线部分。在离线部分中,首先将已有的商标图像数据集输入到卷积神经网络(convolutional neural network, CNN)中训练直至收敛得到稳定模型。其次,所有商标输入到得到的模型中,模型输出的向量作为商标图

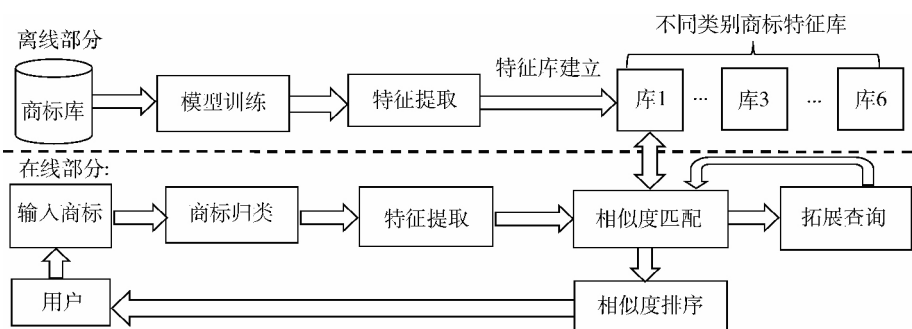


图 1 商标检索系统框架

像的特征向量从而构建不同类别的商标特征库。在线部分中,首先将待检测商标输入模型中,得到待检测商标所属的类别和商标特征向量。然后将待检测商标的特征向量与所属类别的特征库进行相似度匹配,从而得到相似度排名。最后,对得到的候选商标图像进行拓展查询后输出商标图像。

### 1.1 深度学习技术

深度学习技术是机器学习领域的一个分支,主要起源于人们对人工神经网络的研究。而在计算机视觉领域,深度学习采用的网络结构是 CNN,因为其在图像检测中具有位移不变性、缩放不变性以及扭曲不变性等特点,因此对于抽象性较大的商标图像的检测有极大优势。此外, CNN 特征检测过程是对大量数据集的学习过程,所以使用 CNN 进行图像特征检测时,避免了人为的特征设计与抽取,隐式地从训练数据中进行学习<sup>[5]</sup>。

本文使用 Alexnet<sup>[6]</sup> 卷积神经网络,它是众多 CNN 模型的一种,因其在图像识别领域优秀表现而受到人们关注。Alexnet 卷积神经网络是 8 层结构,主要由前 5 个卷积层和后 3 个全连接层组成。在每个卷积层中

都包含了激励函数(激励层),其主要作用是引入数据的非线性效果。在第1、第2和第5卷积层后面包含了一个池化层,主要是对卷积后的数据进行降采样,从而降低数据维度,而降采样的方式有 maxpooling 和 avg-pooling 两种,本文使用 maxpooling 方式进行降采样。Alexnet 网络结构模型如图2所示。

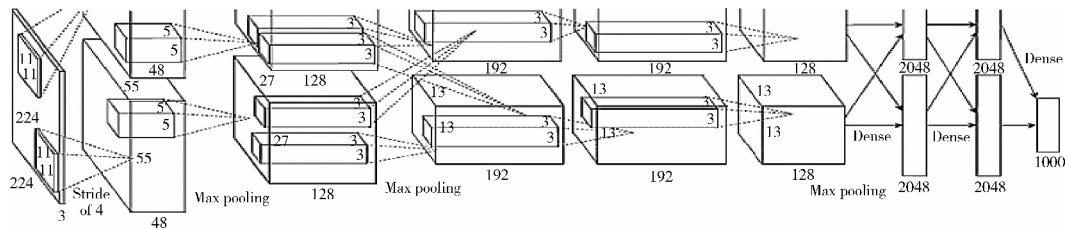


图2 AlexNet 网络结构模型图

## 1.2 基于商标图像的神经网络模型训练

本节提出并训练一种基于商标图像类别的卷积神经网络模型,并构建一个有6种类别、1万张图片的商标图像数据库。通过使用这6个类别的商标图像数据库对 AlexNet 卷积神经网络模型进行训练,并得到一个精确度最高且损失函数最低的 CNN 模型。

由于目前没有公开的商标图像数据库,因此本文使用自己收集的分类好的1万张商标图像作为商标数图像库。商标图像库主要分为六类,分别为简单型、圆型、多边形、文字型、组合型以及复杂型,如图3所示。分类的规则是按照商标的形状,轮廓以及边数特征人为进行分类,规则如下:①对于边的数量和角点的数量都少于4的商标分类为简单型;②轮廓形状为圆形或椭圆形的商标分类为圆型;③对于边数和角点数都大于4的商标分类为多边形;④当商标仅为文字或者字母组成,则分类为文字型;⑤当商标由文字和图案共同组成,且文字和图案不融合,则分类为组合型商标;⑥当商标是复杂图案或者图案和文字相互融合组成,则分类为复杂型商标。



图3 不同类别商标

笔者已经将商标图像库公开在网上,可以从 <http://download.csdn.net/download/studio-man/9940074> 里下载得到。

在构建好商标图像库后,我们使用快速卷积结构特征嵌入(convolutional architecture for fast feature embedding, Caffe)<sup>[7]</sup>深度学习框架对 AlexNet 结构的 CNN 进行训练。Caffe 框架是通过 C++ 语言开发,并支持 Python、Matlab 等接口,而且既可以在 CPU 上运行,也可以在 GPU 上运行。由于 Caffe 的高效、简洁,成为了一个在深度学习领域具有影响力的学习框架。

训练前需要定义网络模型,我们的网络模型就是 AlexNet 模型。其次,需要定义配置参数,主要参数有:学习率、权重衰减系数、迭代次数以及 CPU 和 GPU 的选择。经过大量实验比较后得到的训练参数如表1所示。

在 AlexNet 网络最后一层输出层中包含有损失函数,该损失函数用来测量训练样本的输出与真实值之间的误差大小,训练时每次迭代的误差值称为损失(Loss)。因此每次迭代得到的损失值的大小可以用来判

表1 训练参数配置表

配置参数	取值
基础学习率(base_lr)	0.0001
动量(momentum)	0.9
权重衰减(weight_decay)	0.0002
学习率变化规律(lr_policy)	"step"
学习率变化指数(gamma)	0.1
最大迭代数(max_iter)	1000
是否使用 GPU(solver_mode)	CPU

断当前训练模型的优劣程度。通常随着迭代次数的增加,损失值会不断降低直到趋于稳定。图 4 是我们用商标图像数据库训练 AlexNet 卷积神经网络模型时损失值变化曲线。从图 4 中可以看出,训练时随着迭代数不断增加,损失值不断降低,最终稳定在 0-0.25 之间。因此我们得到的 CNN 模型可以认为是稳定可靠的模型,并可以用于商标图像的分类与检索。

### 1.3 商标图像的特征提取与相似性度量

本文选用 AlexNet 卷积神经网络的第 7 层全连接层的输出作为图像的特征向量并进行特征匹配。其中商标图像的特征提取与匹配的步骤如下:

步骤 1: 建立商标特征库。将之前建立的商标库中不同类别的商标图像依次输入到训练好的 CNN 模型中,并提取网络第 7 层输出数据作为每张商标的特征向量,利用每张图片所提取的特征向量建立商标特征库。

步骤 2: 用户输入商标的特征提取与分类。当用户向系统输入商标时,系统会将商标输入到训练好的 CNN 模型中,商标图像首先会被调整到  $256 \times 256$  的大小。其次,图像经过 5 个卷积层进行卷积操作,当

图像在第二个全连接层输出时,我们将提取输出的 4096 维数据作为该商标的特征向量。最后,当商标图像到达最后一次输出层时,输出层中的 softmax 分类器将输出图像在每个类别上的概率分布,从而得到输入商标所属的类别。图 5 是输入商标图像分别在经过第 1 个卷积层、第 5 个卷积层和第 2 个全连接层时输出的图像特征可视化图。

步骤 3: 商标图像的相似性度量。在得到输入商标特征向量与所属类别后,需要计算输入商标的特征向量与所属特征库中特征向量的相似度。判断图像之间是否相似主要是通过判断两幅图像特征向量之间的距离大小,本文采用的是欧氏距离,如式(1)所示。

$$d = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \quad (1)$$

其中,  $m$  代表特征向量的维度。商标间的相似度计算方法如式(2)所示。

$$\text{相似度} = \frac{1}{d + 1} \quad (2)$$

得到库中每张商标与输入商标的相似度值,并返回相似度高的商标。

## 2 拓展查询

在基于内容的图像检索中,拓展查询被证明是提升检索性能的有效方法。图像检索系统在原始查询结束后,基于候选图像再次向系统发出一次新的“拓展”查询请求<sup>[8]</sup>。

拓展查询的方式有很多种,包括均值拓展查询(AQE)的方式<sup>[8]</sup>,其本质是指原始查询返回的前 K 张图片再加上输入商标图像本身,对它们的特征向量求和并计算均值后再做一次查询。然而对于原始查询返回的 K 张图片,有可能存在一些负样本,即在人类高级的语义特征看来与输入商标并不相似,如果将这些负样本添加到拓展查询中,将会降低拓展查询的查询率。因此本文将使用改进后的均值拓展查询方法,具体步骤如下:

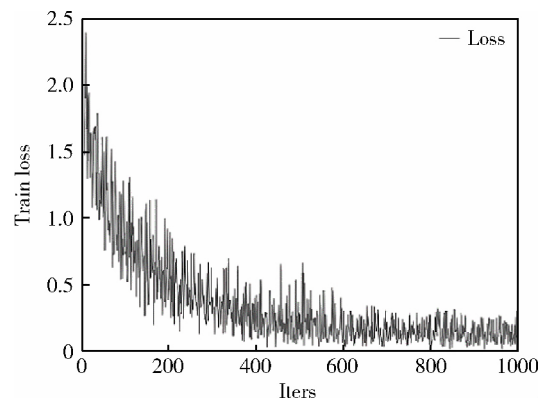


图4 迭代数与损失值关系图

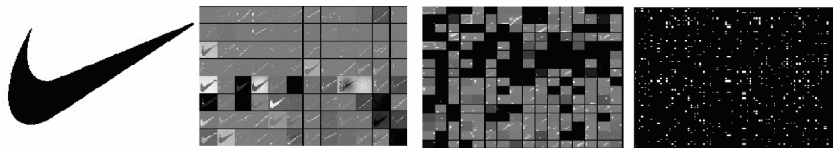


图5 输入图像在不同层的输出的图像



步骤 1: 通过上节输入商标的特征提取与相似性度量后, 得到输入商标与库内商标的相似度值和输入商标的  $m$  维特征, 令其为  $x = \{x_1, x_2, \dots, x_m\}$ 。

步骤 2: 选取相似度排名靠后的 50 张商标的特征向量, 进行求和取均值, 得到一组向量:  $y = \{y_1, y_2, \dots, y_m\}$ , 其中  $y_i$  由式 (3) 得到。

$$y_i = \frac{1}{n}(x_{1i} + x_{2i} + \dots + x_{ni}) \quad (3)$$

$x_{ji}$  代表第  $j$  组特征向量的第  $i$  维元素,  $n$  代表相似度排名靠后的  $n$  张商标, 实验中令  $n = 50$ 。

步骤 3: 进行正负样本检验。选取相似度排名前  $K$  的  $K$  张商标, 利用每张商标的特征向量分别与向量  $x$  和向量  $y$  求它们的欧氏距离, 得到  $d_x$  和  $d_y$ 。若  $d_x < d_y$ , 则当前商标判定为正样本, 若  $d_x > d_y$ , 则当前商标判定为负样本。

步骤 4: 将得到的正样本商标的特征向量进行求和取均值, 再与商标库中的商标进行相似性度量, 最终将返回的商标输出给用户。

通过实验证明改进后的均值拓展查询, 能够提升商标图像检索的查准率和相似图像的召回率, 并且能够一定程度上避免图像检索出现的“语义鸿沟”问题。因此, 在商标图像检索系统中, 对原始查询结果进行拓展查询是有必要的, 不仅能提高检索性能, 而且实现简单。

### 3 实验

#### 3.1 实验环境

软件开发环境: 基于深度学习的商标检索系统, 主要使用的是 Caffe 框架下的深度学习技术, 而 Caffe 开发的编程语言是 C++ 和 Python 两种, 并提供了 Python 接口和 MATLAB 接口, 我们使用 Python 接口。使用的编程语言主要有 C++、Python。

计算机系统环境: 由于 Caffe 最初开发者是在 Linux 系统下开发的, 因此本实验的计算机系统也是选择 Linux。其他要求: ①系统: Ubuntu 16.04; ②CPU: Intel i5 处理器; ③硬盘: 200G 以上。

#### 3.2 实验结果分析

对于判定图像检索系统检索性能的好坏主要有三个评价指标, 查全率、查准率、平均查准率, 表达方式如式 (4) 所示。

$$\text{查全率: } recall = a/b$$

$$\text{查准率: } precision = a/c$$

$$\text{平均查准率: } mAP = \frac{1}{m_p} \sum_{i \in p} \frac{1}{m_q} \sum_{j \in q} precision(k_{ij}) \quad (4)$$

其中, 参数  $a$  表示系统检索出来与输入图像相似的数目; 参数  $b$  表示图像库中与输入图像相似的图像数目; 参数  $c$  表示图像检索系统返回图像数目。在平均查准率公式中, 参数  $precision(k_{ij})$  代表第  $i$  类库的第  $j$  张图片作为输入图像所得到的查准率;  $m_p$  表示图像库中的图像类别数目;  $m_q$  表示第  $i$  类库中图像的数目<sup>[9]</sup>。

在这 3 种系统评价指标中, 平均查准率最能反映设计的检索系统性能的优劣, 因此本文使用平均查准率 (mAP) 作为实验结果的分析。

本文通过比较基于深度学习、词袋模型<sup>[10]</sup>以及 HU 矩特征模型<sup>[2]</sup>的 3 种图像检索算法, 得到 3 种算法在不同类别商标图像中 mAP 的大小, 如图 6 所示。显然深度学习检索算法的性能是最优的, 特别是在复杂类型的商标图像中, 相比于词袋模型和 HU 矩特征模型的检索, 深度学习的平均查准率分别提高了接近 5% 和 12%, 这也说明了深度学习更能克服图像之间的语义鸿沟问题。

本文采用改进后的拓展查询技术提高检索系统的查准率, 并得到在不同情况下平均查准率的大小如图 7 所示。由图 7 可看出, 在采用拓展查询后的检索性能都比原始查询的性能提高 1% - 2%, 而如果采用改进后的拓展查询, 则性能将进一步提高 2% - 5%。因此, 说明拓展查询能够在商标检索领域发挥作用, 并

且对候选图进行分类的拓展查询能够克服拓展查询盲目输入的弊端,从而进一步提高拓展查询效果。

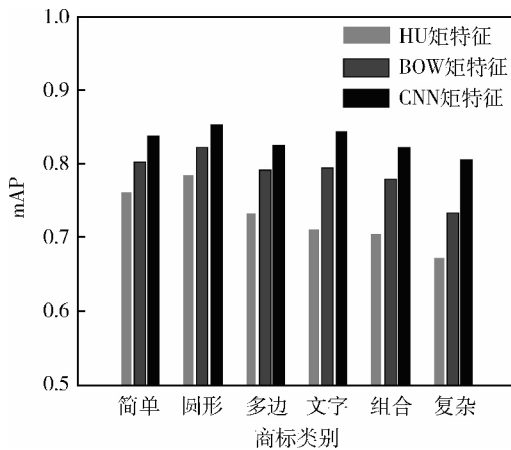


图6 不同算法性能对比图

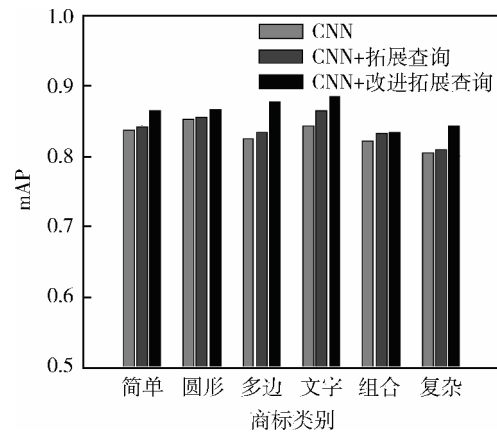


图7 拓展查询性能对比图

图8是通过比较在系统返回不同数量的图像时系统的查全率大小,图8中的查全率是所有类别下查全率的均值。从图8可以发现本文提出的CNN特征与改进的拓展查询相结合的方法在系统只返回10张图片就能得到很高的查全率,其本质是系统只需要返回少量图像就能得到库中大部分与输入商标相似的图像。另外如果使用未改进的拓展查询方法,当系统返回数量大于30张,系统查全率开始呈现下降的趋势。这也证明了上节所说的原始查询的候选图像中的负样本会降低检索系统的性能。图9是系统检索的效果图。

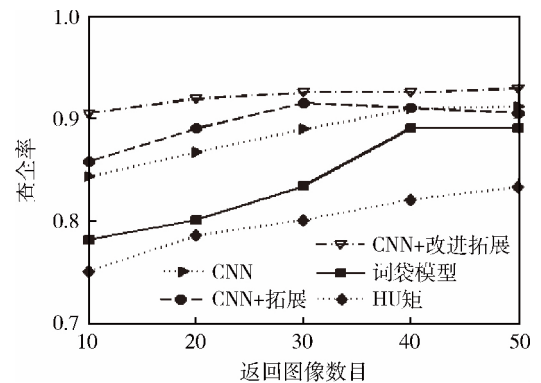


图8 返回图像与查全率关系图

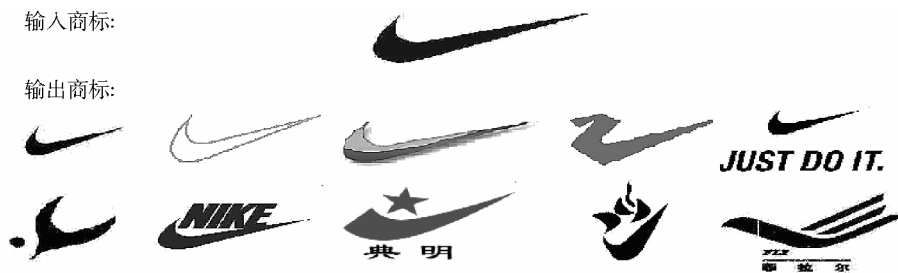


图9 系统检索效果图

#### 4 结束语

本文构建了一个6个类别,共1万张图像的商标库,并训练8层卷积神经网络,得到一个稳定的网络模型用于商标图像的检索,最后结合拓展查询技术来优化系统的检索性能。从实验结果可以看出:①深度学习技术可以从商标图像中得到更加深层次的语义特征,使得检索系统有更高的检索性能,本文所有类别商标的查准率都大于80%;②在检索系统中添加拓展查询技术可以提高系统查询率,当使用改进后的拓展查询技术能够提高2%~5%的平均查准率。

实验结果表明,深度学习技术具有解决图像“语义鸿沟”问题的潜力,在图像检索领域有着巨大的发挥空间。在如今基于内容的图像检索向着数据量大、查准度高的方向发展的情况下,利用深度学习技术来解决图像检索时的各类问题是一个具有前景的方法,值得我们更加深入的研究。

## 参 考 文 献

- [1] 邓学雄, 杨志成, 朱正海. 商标检索中形状特征描述的研究 [J]. 工程图学学报, 2011, 32(6): 21–24.
- [2] 王振海. 融合 HU 不变矩和 SIFT 特征的商标检索 [J]. 计算机工程与应用, 2012, 48(1): 187–190.
- [3] Hinton G E, Ruslan R S. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504–507.
- [4] 胡二雷, 冯瑞. 基于深度学习的图像检索系统 [J]. 计算机系统应用, 2017, 26(3): 8–19.
- [5] Russakovsky O, Deng J, Su H, et al. ImageNet Large Scale Visual Recognition Challenge [J]. International Journal of Computer Vision, 2014, 115(3): 211–252.
- [6] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C], International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012: 1097–1105.
- [7] Donahue J, Jia Y, Vinyals O, et al. DeCAF: a deep convolutional activation feature for generic visual recognition [C], International Conference on International Conference on Machine Learning. JMLR. org, 2014: 1–647.
- [8] Zhao X, Ding G. Query expansion for object retrieval with active learning using BoW and CNN feature [J]. Multimedia Tools & Applications, 2017, 76(9): 1–15.
- [9] 刘海龙, 李宝安, 吕学强, 黄跃. 基于深度卷积神经网络的图像检索算法研究 [J]. 计算机应用研究, 2017, 34(12): 73–81.
- [10] 赵永威, 李弼程, 彭天强, 高毫林等. 一种基于随机化视觉词典组和查询扩展的目标检索方法 [J]. 电子与信息学报, 2012, 34(5): 54–61.

## 作者简介

郭升挺, 男, (1993–), 硕士研究生, 主要研究方向: 计算机视觉、人工智能。

黄晞\*, 男, (1970–), 副教授, 主要研究方向: FPGA, 计算机视觉、人工智能。

柯俊敏, 女, (1994–), 硕士研究生, 主要研究方向: 计算机视觉、图像处理。

陈宝林, 男, (1992–), 硕士研究生, 主要研究方向: 计算机视觉、人工智能。

吴家飞, 男, (1993–), 硕士研究生, 主要研究方向: 计算机视觉。

苏浩明, 男, (1992–), 硕士研究生, 主要研究方向: 网络工程。

\* 为通信作者。

(上接第 26 页)

- [2] 胡二雷, 冯瑞. 基于深度学习的图像检索系统 [J]. 计算机系统应用, 2017, 26(3): 8–19.
- [3] 孙树亮, 林雪云. 基于记忆的 SVM 相关反馈算法 [J]. 计算机科学, 2011, 38(10): 256–258.
- [4] N Vasconcelos, A Lippman. Learning Over Multiple Temporal Scales in Image Databases [C] // European Conference on Computer Vision. Springer – Verlag, 2000: 33–47.
- [5] 白会肖. 基于内容图像检索中相关反馈技术研究 [D]. 石家庄: 河北科技大学, 2008.
- [6] 郭士会, 杨明, 王晓芳, 赵东方, 邓世涛. 基于 FSRM 的相关反馈图像检索算法 [J]. 计算机科学, 2012, 39(103): 540–542.
- [7] 黄晓冬, 孙亮, 刘胜蓝. 一种判别极端学习的相关反馈图像检索方法 [J]. 西安交通大学学报, 2016, 50(8): 96–102.
- [8] 陈长江, 侯进. 一种贝叶斯和支持向量机相结合的相关反馈策略 [J]. 成都信息工程学院学报, 2012, 27(1): 32–37.
- [9] 彭祖赠, 孙韞玉. 模糊数学及其应用 [M]. 武汉: 武汉大学出版社, 2003: 75–80.
- [10] 金海军. 基于语义特征的图像检索算法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2006.

## 作者简介

张水利, 女, (1974–), 副教授, 主要研究方向: 图像处理。

汪恒, 男, (1995–), 学生, 主要研究方向: 图像处理。

李蓓茹, 女, (1997–), 学生, 主要研究方向: 图像处理。

聂栋梁, 男, (1997–), 学生, 主要研究方向: 图像处理。