

计算机应用研究 优先出版

原创性 时效性 就是科研成果的生命力
《计算机应用研究》编辑部致力于高效编排的研究
为的就是将您的成果以最快的速度
呈现于世

* 数字优先出版可将您的文章提前 10~12 个月发布于中国知网和万方数据等在线平台

基于深度卷积神经网络的图像检索算法研究

作者	刘海龙, 李宝安, 吕学强, 黄跃
机构	北京信息科技大学 网络文化与数字传播北京市重点实验室; 首都医科大学宣武医院
发表期刊	《计算机应用研究》
预排期卷	2017 年第 34 卷第 12 期
访问地址	http://www.arocmag.com/article/02-2017-12-007.html
发布日期	2017-01-23 15:03:21
引用格式	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/OL]. [2017-01-23]. http://www.arocmag.com/article/02-2017-12-007.html .
摘要	为解决卷积神经网络在提取图像特征时所造成的特征信息损失, 提高图像检索的准确率, 提出了一种基于改进卷积神经网络 LeNet-L 的图像检索算法。首先, 改进 LeNet-5 卷积神经网络结构, 增加网络结构深度;然后, 对深度卷积神经网络模型 LeNet-L 进行预训练, 得到训练好的网络模型, 进而提取出图像高层语义特征;最后, 通过距离函数比较待检图像与图像库的相似度, 得出相似图像。在 Corel 数据集上, 与原模型以及传统基于 SVM 主动学习图像检索方法相比, 该图像检索方法有较高的准确性。经实验结果表明, 改进后的卷积神经...
关键词	图像检索, 卷积神经网络, 特征提取, 深度学习
中图分类号	TP391.41
基金项目	网络文化与数字传播北京市重点实验室开放课题 (ICDD201504); 国家自然科学基金资助项目 (61271304); 北京成像技术高精尖创新中心项目 (BAICIT-2016003); 2014 年度国家社会科学基金委托课题 (14@ZH036)

基于深度卷积神经网络的图像检索算法研究^{*}

刘海龙^{1a, 1b}, 李宝安^{1b}, 吕学强^{1a}, 黄跃²

(1. 北京信息科技大学 a. 网络文化与数字传播北京市重点实验室, b. 计算机学院, 北京 100101; 2. 首都医科大学宣武医院, 北京 100053)

摘要:为解决卷积神经网络在提取图像特征时所造成的特征信息损失,提高图像检索的准确率,提出了一种基于改进卷积神经网络 LeNet-L 的图像检索算法。首先,改进 LeNet-5 卷积神经网络结构,增加网络结构深度;然后,对深度卷积神经网络模型 LeNet-L 进行预训练,得到训练好的网络模型,进而提取出图像高层语义特征;最后,通过距离函数比较待检图像与图像库的相似度,得出相似图像。在 Corel 数据集上,与原模型以及传统基于 SVM 主动学习图像检索方法相比,该图像检索方法有较高的准确性。经实验结果表明,改进后的卷积神经网络具有更好的检索效果。

关键词: 图像检索; 卷积神经网络; 特征提取; 深度学习

中图分类号: TP391.41

Image retrieval based on deep convolutional neural network

Liu Hailong^{1a, 1b}, Li Baoan^{1b}, Lyu Xueqiang^{1a}, Huang Yue²

(1. a. Beijing Key Laboratory of Internet Culture & Digital Dissemination Research, b. Computer School, Beijing Information Science & Technology University, Beijing 100101, China; 2. Xuanwu Hospital Capital Medical University, Beijing 100053, China)

Abstract: To solve the problem that the loss of image feature information and improve the accuracy of image retrieval, when the Convolutional Neural Network(CNN) is used to extract the feature information of the image, this paper proposed an image retrieval algorithm based on improved Convolutional Neural Network LeNet-L. First, improved LeNet-5 Convolution Neural Network structure, increased depth of network structure. Then, the deep Convolutional Neural Network LeNet-L was pre-trained to extract the high-level semantic features. At last, the similar images was obtained by distance function between the image being retrieved and the one in image database. In Corel dataset, compared with the original model method and the traditional image retrieval method based on SVM and active learning, the proposed method has a higher accuracy. The experimental results show that the improved Convolutional Neural Network has a better retrieval effect.

Key Words: image retrieval ; convolutional neural network; feature extraction ; deep learning

0 引言

近年来,基于内容的图像检索技术^[1](content-based image retrieval, CBIR)迅速发展,取得了大量的研究成果。CBIR 大致检索流程是通过提取诸如图像颜色、纹理、形状等图像底层特征,计算待检图像与已经构建好的图像特征库的距离,返回与待检图像最为相同或相似的图像。但底层特征与高层语义之间存在语义鸿沟,不能准确反映出用户的检索意图。支持向量机(support vector machine)对于有限样本数据具有良好的推广能力,可以有效地改善图像检索结果。部分研究者把 SVM 与主动

学习相结合通过对用户反馈信息进行图像检索,该方法的训练样本特征存在冗余、信息度低等局限性。再者,随着社会信息技术高速发展以及大数据时代的到来,图片不仅数量之多,而且图片内容也变得复杂多样,基于文本与内容的检索技术已无法满足用户的需求。

随着 Hinton 等人^[2]解决了深度学习^[3-5]模型优化问题后,深度学习得到迅速发展,至今深度学习框架已有数种,如深度神经网络、卷积神经网络(convolutional neural network,CNN)和深度信念网络等,已被广泛应用在语音识别^[6]、自然语言处理、计算机视觉、人脸识别^[7]等领域上,获得前所未有的成功。深

基金项目:网络文化与数字传播北京市重点实验室开放课题(ICDD201504);国家自然科学基金资助项目(61271304);北京成像技术高精尖创新中心项目(BAICIT-2016003);2014年度国家社会科学基金委托课题(14@ZH036)

作者简介:刘海龙(1990-),男,硕士研究生,主要研究方向为多媒体信息检索与处理(liuhailong28@163.com);李宝安(1965-),男,教授,硕士,主要研究方向为服务计算社会计算;吕学强(1970-),男,教授,博士,主要研究方向为中文与多媒体信息处理;黄跃(1987-),工程师,硕士研究生,主要研究方向为图像处理技术。

度学习发展自人工神经网络，是机器学习相关领域中一系列算法的总称，根据深度神经网络的构造结构以及训练方法的不同可分为三个大类型：生成深层结构、判别深层结构以及混合深层结构^[4]。本文所研究的卷积神经网络属于**判别深层结构**，卷积神经网络具有**位移、畸变鲁棒性、并行性**等特点，在图像识别领域得到广泛应用。

本文在研究 LeNet-5 模型的基础上，对该卷积神经网络结构在提取图像特征时所造成的**全连接层图像特征信息损失进行分析**，提出改进卷积神经网络模型 **LeNet-L**，在**与原模型以及传统基于 SVM 主动学习图像检索方法进行对比实验**，结果表明，深度卷积神经网络结构 LeNet-L 能在一定程度上减少全连接层提取图像特征时的特征信息损失，使得提取的特征信息对图像检索更有效，比传统典型图像检索方法的检索性能更好。

1 LeNet-5 及其改进网络结构

卷积神经网络的结构有多种，例如用于人脸识别的 DeepID^[8] 网络结构、用于识别数字手写体的 LeNet-5 以及 ImageNet-2010 网络结构^[9]等。文献[10]介绍深度学习在图像检索方面的一些研究，其中在经过预训练得到的卷积神经网络上提出三种图像检索方法。本文采用第一种方法：通过卷积神经网络**提取图像特征进行图像检索**。最近也有不少研究^[11-12]表明卷积神经网络全连接层提取出来的图像特征可以有效的应用在图像检索方面。

1.1 LeNet-5 网络结构

卷积神经网络 LeNet-5 的网络结构,如图 1 所示。

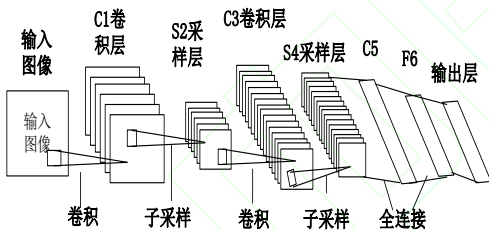


图 1 LeNet-5 网络结构图

该网络结构从输入层到输出层总共八层：除输入层与输出层之外，包括两层卷积层，两层池化层和两层全连接层。其中，

表 1 C3 特征图与 S2 特征图的连接方式

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	×				×	×	×			×	×	×	×		×	×
1	×	×				×	×	×			×	×	×	×		×
2	×	×	×				×	×	×			×		×	×	×
3		×	×	×			×	×	×	×			×		×	×
4			×	×	×			×	×	×	×		×	×		×
5				×	×	×			×	×	×	×		×	×	×

在表 1 中，0-15 列是 C3 卷积层 16 个卷积核的编号，0-5 行是 S2 池化层 6 个特征图的编号。从上表可以看出，C3 卷积层 16 个卷积核中的 0-5 号卷积核是与 S2 层中的 3 个特征图进行连接，6-14 号卷积核是与 S2 层中的 4 个特征图进行连接，最

卷积层与池化层交替进行，全连接层是放在最后池化层的后面，把池化层的二维特征图平铺成一维的特征向量。

卷积层的卷积操作如式(1)所示。

$$x_k^l = f(\sum_{i \in M_k} x_i^{l-1} * H_{ik}^l + b_k^l)$$
 (1)

l 代表模型的层，H 代表卷积核，Mk 代表 l-1 层的第 k 个特征图，b 为输出图的偏置，f 是激活函数。传统卷积神经网络的激活函数一般采用 sigmoid、tanh 等饱和非线性函数。

子采样层^[13]是进一步**减少数据量**，降低输入图像的空间分辨率，是对输入的特征图进行下采样。通过提取卷积层特征图的特征就可以简单实现下采样，从而保证模型位移缩放的鲁棒性以及畸变性。

输入层输入大小为 32*32 的图像，对输入层进行卷积操作得到 C1 层卷积特征。为保证充分提取输入图像的特征，卷积层采用多卷积核来进行提取特征，LeNet-5 采用 6 个 5*5 的卷积核进行提取特征，进而得到 6 个特征图，特征图的尺寸大小是通过公式(2)来进行计算得到，步长为 1。 n_{ij}^{k+1} 是由第 K 层通过第 i 个卷积核计算得到的第 K+1 层第 j 个特征图尺寸的大小，如输入图像大小为 32*32，采用 5*5 卷积核，通过计算可得到下一层特征图的大小尺寸为(32-5+1)*(32-5+1)=28*28。

$$n_{ij}^{k+1} = n_j^k - k_i + 1$$
 (2)

S2 层对上一层的卷积层进行下采样操作，采样操作可在保证不损失特征的情况下减少数据量。C3 与 S4 的关系和 C1 与 S2 类似，不同的是，C3 是 16 个 3*3 的卷积核，以此得到更多的**特征图，更好的描述图像内容信息**。C1 层卷积操作是直接作用在输入的一幅图像上，而 C3 层卷积操作需要作用在 S2 层上，而 S2 层是 6 个 14*14 的特征图。因此，C3 与 S2 的连接方式和 C2 与输入层的连接方式不同，C3 与 S2 的连接方式如表 1 所示：

后一个 C3 卷积核与 S2 的 6 个特征图都有连接。例如，从表 1 第十列可以看出，编号为 9 的卷积核是通过 S2 池化层编号为 0、3、4、5 四个特征图结合得到的。

1.2 改进的卷积网络结构 LeNet-L

LeNet-5 卷积神经网络最初是在识别数字手写体上,输入大小为 32*32 的数字手写体图像,输出类别为 10 类。通过计算可得出,在最后一次池化 S4 后得到 16 个 5*5 的特征图,平铺就需要 16*5*5=400 个神经元,之后 400 个神经元与 120 个神经元进行全连接。但本实验的数据集采用 Corel 图像库,该图像库图像尺寸大小是 384*256 或 256*384,这样在最后一次池化层 S4 就会得到 16 个尺寸大小为 61*93 的特征图,平铺则得到 16*61*93=90768 个神经元。考虑到 LeNet-5 网络结构 S4 层的 90768 个神经元与 C5 层的 120 个神经元进行全连接,在一定程度上会造成图像特征信息的损失,不利于图像特征的提取,从而导致图像检索效果不佳。因此本文网络结构 LeNet-L 在 LeNet-5 网络结构的 S4 层与 C5 层之间多增加一个全连接层,以减少图像特征信息的损失,提高图像检索准确率。当增加层上的神经元数目为 1000 个时,最后得到的图像检索结果与原网络结构进行对比,结果如图 2 所示:

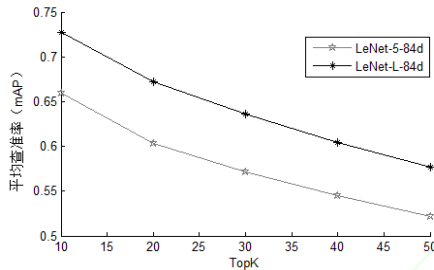


图 2 LeNet-5_84d 与 LeNet-L_84d 之间的 mAP 对比

在图 2 中, LeNet-5_84d 是 LeNet-5 卷积神经网络所提取出来的 84 维图像特征, LeNet-L_84d 是 LeNet-L 卷积神经网络所提取的 84 维图像特征,TopK 是检索图像时返回结果图像的数目。

在 LeNet-L 卷积神经网络中,激活函数采用 ReLU 函数^[14-15],如式(3)所示。该激活函数属于不饱和非线性函数,对于梯度下降训练时间而言,不饱和非线性函数要比饱和非线性函数的收敛速度快很多。

$$f(x) = \max(0, x) \quad (3)$$

本文根据实验图像特点以及实验结果的分析,提出改进 LeNet-5 卷积模型的方法。对**上述造成图像特征信息损失的神经元全连接方式进行改进**,在池化层 S4 与全连接层 C5 之间多加一层**全连接层 L-K**(K 代表增加层的神经元数量),由此,改进后的卷积网络结构,如图 3 所示:

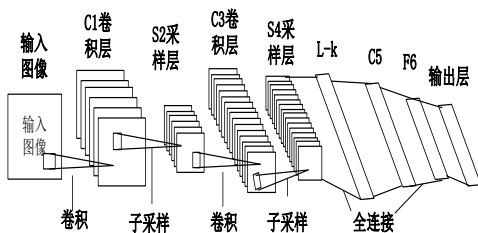


图 3 LeNet-L 卷积神经网络模型

2 实验与分析

2.1 实验数据集与实验性能评价指标

本文所采用的图像库是 Corel 数据集。Corel 图像库包含 1000 幅图像,每类包括 100 张图片,图像库所包含的类别有恐龙、建筑、鲜花等 10 个类别。

对于图像检索有查全率(recall)、查准率(precision)、F 值、平均查准率(mean Average Precision, mAP)等评价指标。

$$\text{查全率: } recall = a / b \quad (4)$$

$$\text{查准率: } precision = a / c \quad (5)$$

$$\text{平均查准率: } mAP = \frac{1}{m_p} \sum_{i \in p} \frac{1}{m_q} \sum_{j \in q} precision(k_{ij}) \quad (6)$$

其中, a: 检索出来与待检图像相似的图像数目, b: 图像库中与待检图像相似的图像总数目, c: 图像检索系统返回的图像数目,

K_{ij} 表示第 i 类的第 j 张图片作为待检图像所返回的结果。

2.2 实验数据预处理

在训练卷积神经网络阶段,通常需要对实验数据进行简单的数据预处理,以使得算法发挥出最好的效果。本文主要需要以下步骤:

a) 图像尺寸大小处理。将数据集中的图像缩放为固定大小(本文采用 256×384),在整个卷积神经网络的结构中,卷积层和池化层对图片大小是没有要求的,但是全连接层就需要图片是固定大小,因为全连接层的权值矩阵在训练后是固定大小的矩阵。

b) 归一化与零均值化。对整个图像数据消除该图像数据的均值,进行零均值化的公式如式(7)(8)所示。

$$\mu^i = \frac{1}{n} \sum_{j=1}^n x_j^i \quad (7)$$

$$x_j^i = x_j^i - \mu^i \quad (8)$$

c) 白化。白化操作类似于低通滤波器,过滤高频部分,使得特征间相关性较低,有助于降低噪声的影响。

2.3 实验参数取值

在 LeNet-L 网络结构中,有一个增加层 L-K 用以减少图像特征损失, L-K 层的神经元数目与减少图像特征损失的多少有着密切关系。本文分别对神经元数目 $K = \{1400, 1600, 1800, 2000, 2200, 2400\}$ 进行实验,

得到在 K 取不同值时的平均查准率 mAP。如图 4 所示,从图中可以看出:当 K=1800 时,无论返回检索图像的数目是多少,平均查准率都比 K 取其他值要高,Top 10 表示返回相似图像的数目为 10。

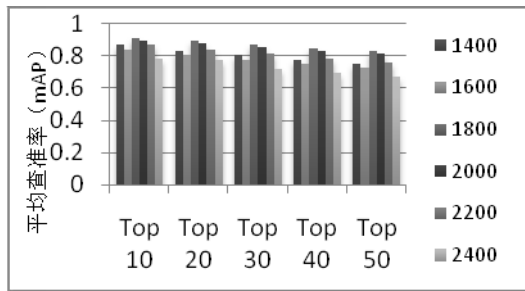


图 4 不同 K 值的平均查准率 mAP

2.4 检索示例

通过图 4 确定 K 值取 1800 可以取得较好的图像检索效果, 在 Corel 图像库中的一个检索示例, 如图 5 所示:



图 5 LeNet-L 与 LeNet-5 卷积神经网络检索结果 top10 图像对比

图 5 中 1) 的结果是通过 LeNet-L 神经网络返回的检索结果; 2) 是 LeNet-5 返回的检索结果, 其中有 4 幅图片与待检索图像不符, 从图中可以看出, LeNet-L 的检索效果要优于 LeNet-5。LeNet-5 方法返回的图像有部分结果与待检图像不相关, 检索效果不尽人意, 而改进之后的模型 LeNet-L 能减少原模型在全连接层提取图像特征的损失特征信息, 能更加有效的提取出符合图像所描述信息的特征或更加抽象的语义特征信息, 进而能够提高检索性能, 检索出令人满意的结果。

2.5 查全率

本文提出的图像检索方法是基于深度卷积神经网络来获取图像的特征表示, 获取的特征表示不仅与原始卷积神经网络模型进行对比实验, 而且还与传统基于 SVM 主动学习图像检索文献[17]所提方法(记为 Method)作对比。

在 Corel 数据集上的查全率实验结果, 如图 6 所示: 从图中可以看出 LeNet-L 的查全率始终比 LeNet-5 要高。而且当返回图像数目为 30 时, LeNet-L 比 LeNet-5 的查全率高 0.08; 当返回图像数目为 80 时, LeNet-L 比 LeNet-5 的查全率要高 0.24。本文所提出的图像检索方法与传统基于 SVM 主动学习图像检索 Method 对比, 查全率始终优于 Method 方法, 并且随着返回图像数目的增加, 效果差别在逐步明显, 也体现出深度神经网络的优势。

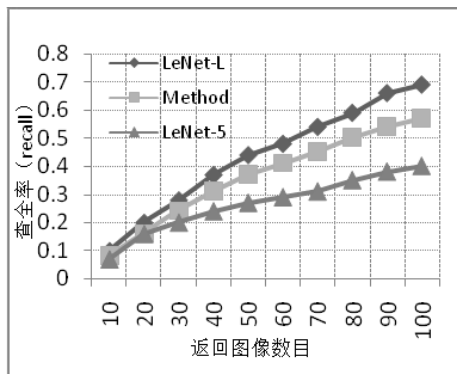


图 6 LeNet-L 与 Method、LeNet-5 查全率对比结果

2.6 检索平均查准率

在对 Corel 图像库的所有图像进行平均查准率实验中, 得到不同图像检索方法的平均查准率, 如表 2 及图 7 所示:

表 2 三种图像检索方法平均查准率对比结果 (mAP)

	top10	top20	top30	top40	top50
LeNet-L	0.9133	0.8916	0.8719	0.8499	0.8285
Method	0.9016	0.7735	0.6721	0.6332	0.5764
LeNet-5	0.7608	0.7075	0.6697	0.6411	0.6152

从表 2 可以看出本文方法的检索效果比传统基于 SVM 主动学习方法要好, 也大幅度的优于原卷积神经网络模型方法。在返回前 20 张图像时, 本文方法比 Method 方法的平均查准率高 11.8%, 当返回前 50 张图像时, 本文方法比 Method 方法要高 25.2%。

当返回图像数目在线性增长的同时, 本文方法与传统方法的平均查准率都在下降, 但是当返回图像数目不断增加的过程中, 传统 SVM 主动学习图像检索方法大概以 5%-12% 的性能损失在不断降低图像平均查准率, 图像检索性能下降迅速。而本文所提方法受返回图像数目带来的检索效果影响较小, 大概 2% 左右, 这也充分体现了卷积神经网络在学习特征所表现出来的较强的泛化能力, 如图 7 所示。

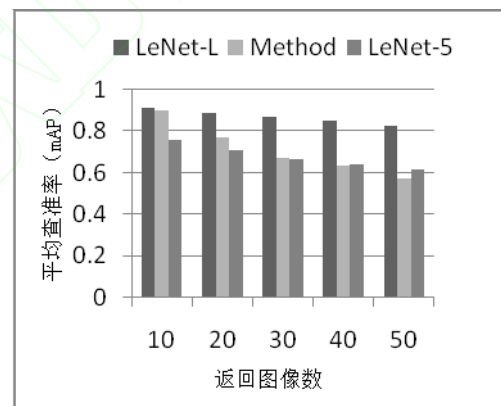


图 7 LeNet-L 与 Method、LeNet-5 平均查准率对比结果

文献[18]传统方法是通过一种基于 K-means 聚类主动反馈策略来提高反馈图片的多样性, 但是该方法容易受到 K 值和聚类中心的影响导致效果不理想。Method 方法是在文献[18]方法上提出的一种新的图像检索方法, 通过构造出信息度大、冗余度低的最优训练集来训练出更好的 SVM 分类器, 进而提高检索性能, 该方法对数据集有很高要求, 处理复杂, 有一定局限性。LeNet-5 方法的图像检索效果比较差是因为在神经网络结构卷积层与全连接层连接结构不当导致在全连接层提取图像底层特征时会损失部分特征信息, 影响图像检索性能。本文所提方法是通过深度卷积神经网络来提取图像特征, 多卷积核进行卷积操作提取更能代表图像描述的特征信息, 不需要复杂的数据预处理, 对图像位移、畸变具有鲁棒性。

卷积神经网络提取图像特征的过程是对图像库中所有图像进行不断学习得到的, 深度学习网络结构对原始图像数据进行逐层提取与抽象, 处在深度学习网络结构高层的特征具有更抽象、更高效的语义特征^[19], 能够更好的表达图像所包含信息。

当然,当返回图像检索结果数目不断增加的过程中,平均查准率也在相应的下降。

3 结束语

本文在对 LeNet-5 网络模型进行深入研究的基础上,针对卷积神经网络 LeNet-5 在提取图像特征时存在图像特征信息损失问题,提出了卷积神经网络 LeNet-5 的改进网络模型 LeNet-L。通过与 LeNet-5 神经网络结构以及传统基于 SVM 主动学习图像检索方法进行对比,最终实验结果表明了深度卷积神经网络 LeNet-L 的有效性,证实了本文方法比传统方法在图像检索方面具有更好的性能表现。

参考文献:

- [1] 杭燕, 杨育彬. 基于内容的图像检索综述[J]. 计算机应用研究, 2002, 19(9): 9-13.
- [2] Hinton G E, Ruslan R S. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504-507.
- [3] 刘建伟, 刘媛, 罗雄麟. 深度学习研究进展[J]. 计算机应用研究, 2014, 31(7): 1921-1942.
- [4] 孙志远, 鲁成祥, 史忠植, 等. 深度学习研究与进展[J]. 计算机科学, 2016, 43(02): 1-8.
- [5] 尹宝才, 王文通, 王立春. 深度学习研究综述[J]. 北京工业大学学报, 2015(01): 48-59.
- [6] 王山海, 景新幸, 杨海燕. 基于深度学习神经网络的孤立词语音识别的研究[J]. 计算机应用研究, 2015, 32(8): 2289-2291.
- [7] 杨海燕, 蒋新华, 聂作先. 基于并行卷积神经网络的人脸关键点定位方法研究[J]. 计算机应用研究, 2015, 32(8): 2517-2519.
- [8] Sun Yi, Wang Xiaogang, Tang Xiaoou. Deep Learning Face Representation from Predicting 10, 000 Classes[C]//Proc of Computer Vision and Pattern Recognition. 2014: 1891-1898.
- [9] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]// Advances in Neural Information Processing Systems. 2012: 1097-1105.
- [10] Xie Lingxi, Wang Jingdong, Zhang Bo. Fine-Grained image search[J]. IEEE Trans on Multimedia, 2015, 17(5): 636-647.
- [11] Babenko A, Slesarev A, Chigorin A, et al. Neural Codes for Image Retrieval[M]// Computer Vision. Springer International Publishing, 2014: 584-599.
- [12] Donahue J, Jia Y, Vinyals O, et al. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition[J]. Computer Science, 2013, 50(1): 815-830.
- [13] 郑启财. 基于深度学习的图像检索技术的研究[D]. 福建师范大学, 2015.
- [14] Dahl G E, Sainath T N, Hinton G E. Improving deep neural networks for LVCSR using rectified linear units and dropout[C]// 2013: 8609-8613.
- [15] Glorot X, Bordes A, Bengio Y. Deep Sparse Rectifier Networks[J]. Learning/Statistics & Optimisation, 2010.
- [16] 彭晏飞, 尚永刚, 王德建. 一种新的基于 SVM 和主动学习的图像检索方法[J]. 计算机工程与科学, 2014, 36(07): 1371-1376.
- [17] 张玉芳, 陈卓, 熊忠阳, 等. 一种基于 SVM 和主动学习的图像检索方法[J]. 计算机工程与应用, 2010, 46(24): 193-196.
- [18] 刘兵, 张鸿. 基于卷积神经网络和流形排序的图像检索算法[J]. 计算机应用, 2016, 36(2): 531-534, 540.