中国科学技术大学

工程硕士学位论文



**基于机器学习的图像检索系统研究与实现**

**作者姓名： 王 新**

**学科专业： 软件工程**

**校内导师： 郑启龙 副教授**

**企业导师： 周太川 高级工程师**

**完成时间： 二○一四年六月二日**

University of Science and Technology of China

A dissertation for master’s degree

of engineering



Research and Implementation of Image Retrieval System based on machine learning

Author’s Name： Wang Xin

speciality： software engineering

Supervisor： A.P. Qilong Zheng

Advisor： Senior Engineer Taichuan Zhou

Finished time: June 2nd, 2014

中国科学技术大学学位论文原创性声明

本人声明所呈交的学位论文,是本人在导师指导下进行研究工作所取得的成果。除已特别加以标注和致谢的地方外，论文中不包含任何他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的贡献均已在论文中作了明确的说明。

作者签名：\_\_\_\_\_\_\_\_\_\_\_ 签字日期：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

中国科学技术大学学位论文授权使用声明

作为申请学位的条件之一，学位论文著作权拥有者授权中国科学技术大学拥有学位论文的部分使用权，即：学校有权按有关规定向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅，可以将学位论文编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。本人提交的电子文档的内容和纸质论文的内容相一致。

保密的学位论文在解密后也遵守此规定。

□公开 □保密（\_\_\_\_年）

作者签名：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 签字日期：\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# 摘要

伴随着数字化进程在各行各业的深入，信息的增长不仅体现在文本类型上，更体现在图像信息中。图像在表达信息中有着直观、形象的优点，然而由于图像数据本身的特点，在海量的图像数据中搜寻有价值的图像成为了亟待解决的问题。图像检索目前主流技术有基于文本的图像检索与基于内容的图像检索两种，前者主要通过人工对图像文件标注，然后采用利用文本检索技术实现对图像的检索，但这种技术无法应对图像数据的几何级增长；后者则是对图像的特征进行提取，基于特征的相似性进行检索，但这种技术检索出的结果与人类的理解与视觉感知存在着偏差。

本文以基于内容的图像检索为背景进行研究，对图像特征（颜色、纹理、形状特征）进行深入研究，发现单一的特征在图像的描述上存在不足，并且不同图像特征在视觉判断相似时所占的权重有所不同；另外,仅使用低层的图像特征进行相似度计算的结果与视觉感知的结果存在较大差距。针对以上两点，本文采用一种基于机器学习的多特征融合检索方法。首先根据图像标注构建语义模型；其次针对语义模型中的每个节点，融合多种特征建立多层次的神经网络，使每种图像特征在不同的分类任务中的权重可以自行调整；然后利用半监督学习相关反馈算法使系统在与用户交互中将低层图像特征与高层语义之间建立关联，缩小图像检索时的语义鸿沟，使整体的检索结果更符合用户的要求；最后基于本文提出的方法，设计并实现了一个基于内容的图像检索系统，该系统能有效的提高检索性能。

通过对该本课题实现的系统的参数设置，将本文中提出的方法与传统的仅基于视觉特征的CBIR作对比实验，证明该系统充分发挥机器学习的作用，缩小语义鸿沟，使图像检索结果与人类视觉感知更为一致。

**关键字：**图像检索，机器学习，多特征

Rui等人[5]用所有正例样本在各特征分量上

取值标准差的倒数作为该分量的权值.他们的方法

是基于这样的考虑,如果在某个特征分量上,所有正

例有相似的取值,那么这个特征分量就能较好地反

映用户的需求;相反地,如果所有正例的取值很分

散,那么这个特征分量对于当前查询就不重要.

# Abstract

The paper is based on

**Key Words:**

目录

[摘要 I](#_Toc394828352)

[Abstract II](#_Toc394828353)

[第一章 绪论 1](#_Toc394828354)

[1．1 选题的依据与意义 1](#_Toc394828355)

[1．2 国内外文献资料综述 1](#_Toc394828356)

[1．3 本人的主要工作 1](#_Toc394828357)

[1．4 论文的组织 1](#_Toc394828358)

[第二章 系统相关技术 3](#_Toc394828359)

[第七章 结论 5](#_Toc394828360)

[参考文献 7](#_Toc394828361)

[附录 9](#_Toc394828362)

[致谢 11](#_Toc394828363)

# 绪论

## 1．1 选题的依据与意义

### 1.1.1 选题背景

数字媒体在各行各业中的应用使各类图像信息几何级增长，带来了巨大的利益，同时也对图像的使用带来了巨大挑战。图像可以更为直接、形象的表达信息，为了更有效的利用这些图像中所包含的信息，人们必须在大量的图像数据中快速发现对自己有用的图像，图像检索[1]问题由此而生。主流的图像检索技术包括基于文本的图像检索（Text-Based Image Retrieval）和基于内容的图像检索（Content-Based Image Retrieval）两种[1,2]。在上世纪七十年代图像检索方法主要采用的是TBIR，这主要是因为数据库技术的发展。这种检索方法根据人类对数据图像的理解，将图像所表达的语义以关键字的形式对每幅图像人工标注，然后通过传统的数据库技术以及文本检索技术对图像进行存储、索引及检索，图像检索可以完美的转换为文本检索问题。目前拥有搜索引擎的公司比如google、雅虎、百度和微软等都对这种搜索方式有很好的支持。这种检索方法的优点非常明显：检索速度快、查准率高且实现简单。但它的缺点也是显而易见的[3]，人工标注的代价过大，无法适应当前图像数据几何级的增长，虽然可以通过互联网中抓取与图像相关的语义关键字，但这种方式也可以归结为间接的人工标注且适用领域有限；人工标记存在主观性，同一幅图像在不同人眼中会有不同的意义，而且图像所表达的信息是及其丰富的，往往不是几个关键词所能够表达的；最后一点，这种检索方式无法跨越语言的界限，无法形成统一的检索方法，很难实现国际化。

为了克服基于文本的图像检索的缺点，让图像检索能够适应海量的图像数据资源，基于内容的图像检索应运而生，并成为近十几年来的研究热点。这种检索方法[4]提取图像中固有的低层视觉特征（颜色、纹理和形状特征等），并将这种特征存入数据库，在图像检索时计算输入图像特征与数据库中保存的图像特征的相似度，将具有较高相似度的图像返回给用户。但目前这种方式的图像检索系统有两点不足，其一是无法合理运用多种视觉特征[5]，在人类识别图像时，每种视觉特征在不同图像中会起到不同比重的作用[7,8,9]，目前的系统大多只是基于单一的特征进行匹配；其二是会带来语义鸿沟[6]，有些图像虽然在低层的视觉特征上非常相似，但是于人类理解上却有很大的差距，正如图2-1所示，它们的视觉特征相似度较高，但从人类视觉的理解来看，左右两幅图所描绘的是两种外形相差较大的不同动物。为解决语义鸿沟带来的问题，研究者们从两个方向入手，一是特征提取[10,11]，使提取的特征更为有效的描述图像的固有特征，并采用更合

|  |  |
| --- | --- |
| 图2-1 图像对比 | 理的方法计算相似度；二是将语义引入到检索系统中，跨越语义鸿沟，这也是本文的重要组成部分；三是将多个特征合理的融合来计算相似度。 |

机器学习[12]可以简单的理解为这样一件事，计算机程序随着经验积累自动提高性能。近年来，机器学习被成功地应用于很多领域，从检测信用卡交易欺诈的数据挖掘程序，到获取用户阅读兴趣的信息过滤系统，再到能在高速公路上自动行驶的汽车。同时，这个学科的基础理论和算法也有了重大的进展。机器学习从很多学科吸收了成果和概念，包括统计学、人工智能、哲学、信息论、生物学、认知科学、计算复杂性和控制论等。当前的社会已拥有了海量的图像数据，利用机器学习的理论与方法来模仿人类认识世界的过程必将在图像检索问题上产生巨大的推动力。

### 1.1.2 研究意义

基于内容的图像检索在各个领域中拥有广阔的应用前景[13]，目前在图像搜索引擎，在线购物，多媒体数字图书馆、军事、医学图像管理、安全生产监控、建筑及工程设计、地理信息系统等诸多领域都会涉及到基于内容的图像检索。而机器学习在多特征融合以及语义鸿沟上的重要作用也成为推动基于内容的图像检索的不可或缺的力量。

虽然利用机器学习的基于内容的图像检索技术的研究取得了长足的进步，但是如何合理运用和改进相关技术，构建实用、可靠、高速的图像检索系统仍然是一个亟待解决的问题。不难预见，在不久的将来数字化将更加深入的融入到各行各业，图像检索也将起到越来越重要的作用。因此针对此课题进行研究具有很大的理论及现实意义。

### 1.1.3 工作目标

本课题的工作目标在于探索一种合理的图像检索方法用以提高图像检索的性能，缩小视觉特征与高层语义间的鸿沟。期望能为基于内容的图像检索性能的提升上带来一定的价值，使图像检索能够在实际的使用中获得更好的效果。

## 1．2 国内外文献资料综述

图像检索目前基本可分为基于文本的图像检索与基于内容的图像检索两类。基于内容的图像检索的主要模式为首先对图像进行标注，然后采用文本检索中的技术来实现对图像的检索，基于文本的图像检索由于需要人工标注，在面对海量的图像数据时会比较乏力，因此这类图像检索技术的研究方向主要集中在图像的自动标注上，其中吴等[11]人采用机器学习的方法取得了一定的成果，本文的工作也部分受到了这方面的启发。基于内容的图像检索研究热点主要集中三个方向，第一个方向在于对图像视觉特征的提取上，图像特征的提取又可分为全局特征与局部特征的提取，其中局部特征提取是先将图像分割为若干区域，然后提取区域的局部特征，使用这种方式提取特征来进行图像检索更加符合人类理解图像的方式，在[12~14]中实现了基于显著性区域的图像检索；第二个方向是对图像语义的研究，视觉相似与语义相似之间的差距是带来CBIR性能瓶颈的主要原因，朱等[15]提出的语义网络在跨越语义鸿沟方面取得了不错的成效。第三个方向在于整个图像检索过程上，相关反馈这种模式是当前的研究热点。

自上世纪九十年代以来，各大研究机构及公司开始对基于内容的图像检索系统投入大量的人力物力，并取得了一些初步的研究成果[14],其中也不乏国内的一些研究机构。目前较为成熟的国外CBIR系统包括IBM的QBIC、麻省理工学院的Photobook系统、伊利诺大学的MARS、哥伦比亚大学的VisualSeek以及WebSeek ，国内的CBIR系统包括中科院的MIRES、浙江大学的Photo Navigator，清华大学研发的web上基于内容检索的原型系统。另外各大搜索引擎公司也对CBIR有了不同程度的支持。但这些系统多数是通过低层特征的对比来实现的，虽然具有了一定的通用性，但对没有引入高层语义特征，使检索结果与人类视觉感知效果相差较大。

图像在分类中单一的特征往往不能够在所有的分类中都获得较为理想的结果，而且对同一图像不同人理解的重点会不同，此时不同的特征会起到不同的作用，多个特征的融合问题由此而生。这方面的研究主要集中于两个方面，一个是特征的选取组合方面，在[16]中采用了多尺度Gabor小波变换作为纹理特征，在[17]则采用了医学图像中所有的特征进行融合利用，两者均取得了不错的结果；另一方面则在于如何利用多个特征，目前主流的做法是直接在基于不同特征计算的距离的基础上增加一个权值，既进行一个线性组合，例如谢[18]所做的工作。

为了跨越语义鸿沟，如何从低层的图像特征中提取高层语义已成当前的研究热点。在20世纪90年代中期文本检索领域提出的相关反馈被引入到CBIR中，在过去的近20年里各种相关反馈的算法不断涌现，将用户的反馈信息充分利用到图像索引过程中。根据相关反馈算法的检索模型可以将其分为基于概率框架的方法、基于距离向量的方法和基于机器学习的方法[19]。在基于机器学习的方向，PicSOM[20]采用了一种自组织映射图在相关反馈中进行动态聚类。Wood[21]运用了区域特征，用户要反馈图像中被选中的区域是否和目标匹配，然后用LVQ算法对这样训练样本训练聚类，将与正类为同一类的图像检出。由于在相关反馈中训练样本少的问题很突出，半监督学习算法也被用来改进相关反馈算法的性能，Wu等[22]提出了D-EM算法,它在EM迭代的内部作了一个判别分析以找到一个特征子空间使得数据分布更符合两类的假设，另外，有的工作则利用了协同学习的思想。

## 1．3 本人的主要工作

本节应对工作内容和成果进行介绍，着重分析关键问题及其解决方案的新意。

## 1．4 论文的组织

本节应分别描述每章所讨论的问题。

**相关反馈现状**

需要多次的迭代，用户体验较差

# 系统相关技术

## 2.1图像特征

特征介绍

距离计算

## 2.2神经网络

人工神经网络（Artificial Neural Networks）是一种普遍且实用的方法，它通过模拟人类的神经系统，可以在样例中学习值为离散值、实数或向量的函数。ANN具有很强的健壮性，能够容忍一定程度的错误训练样本；而且由于ANN本身的特点，它能够更容易并行实现，克服其训练时间较长的不足；ANN具有自适应、自学能力，不需要人为给出明确的特征，可以创造出设计者没有明确引入的特征；通过对神经元的选择，ANN可以学习到非线性的映射。适合神经网络学习的问题有这样几个特征，首先实例是用很多属性-值对来表示的；其次目标函数的输出可能是离散值、实数值或向量；第三，训练数据可能包含错误，需要采用健壮性较强的学习算法；最后，采用人工神经网络进行训练的时间会比较长，在进行鉴定时的计算时间较短。基于以上的特点的分析可以推知，图像的分类问题比较适合采用人工神经网络来解决，尤其是因为在图像的分类识别中还没有系统的方法能够取得非常理想效果，可以通过人工神经网络自动学习图像和其语义的复杂映射关系。

### 2.2.1 神经网络模型

神经网络本质上就是由大量神经元组成的网络，可以分为输入层、隐藏层以及输出层，如图2-1所示。具体来讲，目前主流的网络模型主要有三种，即前馈神经网络、反馈神经网络和自组织网络。

前馈网络(Feedforward Neural Networks )也称为前向网络。这类网络只在训练过程会有反馈信号，而在分类过程中数据只能向前传送，直到到达输出层，隐藏层没有向后的反馈信号，因此被称为前馈网络。像BP神经网络就属于前馈网络。

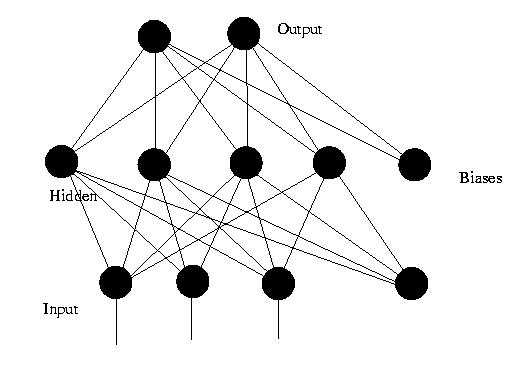


图2-1. 前馈神经网络

  反馈型神经网络(Feedback Neural Networks )是一种输出至输入具有反馈连接的神经网络，其结构比前馈网络要更为复杂。这种网络适用于时序数据，使用网络单元在时间t的输出作为其他单元在时间t+1的输入。较为典型的反馈型神经网络有Hopfield网络和Elman网络，均会在网络中使用某种形式的有向环。

       自组织神经网络 ( SOM ,Self-Organizing Neural Networks )是一种无导师学习网络模式。它通过自动的寻找样本中的内在规律和本质属性，能够自组织、自适应地改变网络结构以及相关参数。

### 2.2.2 神经元

神经元的选择是构建神经网络非常关键的一步，主流的神经单元有感知器、线性单元以及sigmod单元。

感知器可以看作是n维空间中的超平面决策面，对于超平面一侧的实例输出1，另一侧输出-1。感知器以实值向量作为输入，计算这些输入的线性组合，若计算结果大于某个阈值就输出1，否则输出-1：



其中每个为输入向量分量对应的权值。 （2.1）

感知器在训练的时候可使用感知器法则，针对每个训练样例会对权值*w*做这样的调整：



其中，。

有限次的使用感知器训练法则后，感知器会收敛到一个能正确分类所有训练样例的权向量，前提是训练样例线性可分，并且使用了充分小的*η*

线性单元可对应于感知器的第一阶段，不带有阈值：

 （2.2）

由于多个线性单元的链接仍然会产生一个线性函数，因此为了能够表征非线性的映射需要引入sigmod单元，这也正式本课题所选用的神经元，sigmod单元的定义：

 （2.3）

其中，。

sigmod单元可以讲实数域压缩到区间(0,1)，因此常常将sigmod函数成为挤压函数，sigmod函数另一个比较好的特征是他的导数可以很容易的通过它的输出来表示，即：

 （2.4）

### 2.2.3 随机梯度下降

单个神经元的误差可以表示为：

 （2.5）

其中，，*D*为训练集，为训练样例的目标输出，为神经元的实际输出。

### 2.2.4 反向传播算法

## 1.3 相关反馈

## 1.4 半监督学习

## 1.5 语义网络

# 第七章 结论

# 参考文献

11，主动学习算法中采样策略研究 吴伟宁

[12] 冯松鹤;郎丛妍;须德. 一种融合图学习与区域显著性分析的图像检索算法[J].电子学报, 2011,(10).

[13] 张菁;沈兰荪;David Dagan Feng. 基于视觉感知的图像检索的研究[J].电子学报, 2008,(03).

[14] J. Zhang;Zhou L.;L.S.Shen. Regions of interest extraction based on visual attention model and watershed segmentationIEEE International Conf on Neural Networks Signal Processing, 2008.

[16]基于多特征DS融合策略的图像检索技术研究 邵天日

[17]基于密度聚类和多特征融合的医学图像识别研究

[18]基于反馈和多特征融合的图像检索系统研究与实现 谢彰桓

[19] 基于内容图像检索中相关反馈技术的回顾\_吴洪

[20] Laaksonen J.,Koskela M.,Oja E.. PicSOM:Self-organizing maps for content-based image retrieval. In:Proceedings of

INNS-IEEE International Joint Conference on Neural Net-

works,Washington,DC,1999,2470～2479

[21] Wood M.E.J.,Campbell N.W.,Thomas B.T.. Iterative re-

finement by relevance feedback in content-based digital image

retrieval. In:Proceedings of ACM Multimedia,Bristol,UK,

1998,13～20

[22] Wu Y.,Tian Q.,Huang T.S.. Discriminant EM algorithm

with application to image retrieval. In:Preceedings of IEEE

Conference Computer Vision and Pattern Recognition,South

Carolina,2000,222～227

# 附录

# 致谢