

基于兴趣物体检测的图像检索系统的设计与实现

大连理工大学本科毕业设计（论文）

基于兴趣物体检测的图像检索系统的设计与实现

**Design and Implementation of Image Retrieval System Based on
Interest Object Detection**

学 院（系）： 软件学院

专 业： 数字媒体技术

学 生 姓 名： 王丹丹

学 号： 201692354

指 导 教 师： 谢玲

评 阅 教 师：

完 成 日 期：

大连理工大学

Dalian University of Technology

原创性声明

本人郑重声明：本人所呈交的毕业设计（论文），是在指导老师的指导下独立进行研究所取得的成果。毕业设计（论文）中凡引用他人已经发表或未发表的成果、数据、观点等，均已明确注明出处。除文中已经注明引用的内容外，不包含任何其他个人或集体已经发表或撰写过的科研成果。对本文的研究成果做出重要贡献的个人和集体，均已在文中以明确方式标明。

本声明的法律责任由本人承担。

作者签名：

日 期：

关于使用授权的声明

本人在指导老师指导下所完成的毕业设计（论文）及相关的资料（包括图纸、试验记录、原始数据、实物照片、图片、录音带、设计手稿等），知识产权归属大连理工大学。本人完全了解大连理工大学有关保存、使用毕业设计（论文）的规定，本人授权大连理工大学可以将本毕业设计（论文）的全部或部分内容编入有关数据库进行检索，可以采用任何复制手段保存和汇编本毕业设计（论文）。如果发表相关成果，一定征得指导教师同意，且第一署名单位为大连理工大学。本人离校后使用毕业毕业设计（论文）或与该论文直接相关的学术论文或成果时，第一署名单位仍然为大连理工大学。

论文作者签名：

日 期：

指导老师签名：

日 期：

摘 要

关键词：计算机视觉；兴趣物体检测；图像检索

Design and Implementation of Image Retrieval System Based on Interest Object Detection

Abstract

Key Words: Computer Vision; Instance of Interest Detection; Image Retrieval

目 录

| | |
|----------------------------|----|
| 摘 要 | I |
| Abstract | II |
| 引 言 | 1 |
| 1 相关技术综述 | 3 |
| 1.1 计算机视觉 | 3 |
| 1.1.1 物体检测 | 3 |
| 1.1.2 语义分割与实例分割 | 3 |
| 1.1.3 全景分割 | 4 |
| 1.2 兴趣物体检测 | 4 |
| 1.3 图像检索 | 4 |
| 1.4 React 框架 | 5 |
| 1.5 Spring Boot 框架 | 5 |
| 1.6 MySQL 数据库 | 6 |
| 1.7 本章小结 | 6 |
| 2 可行性分析与需求分析 | 7 |
| 2.1 可行性分析 | 7 |
| 2.2 功能性需求分析 | 7 |
| 2.2.1 图像管理模块的需求分析 | 7 |
| 2.2.2 图像检索模块的需求分析 | 7 |
| 2.3 非功能性需求分析 | 7 |
| 2.4 本章小结 | 7 |
| 3 基于兴趣物体检测的图像检索算法的实现 | 8 |
| 3.1 兴趣物体检测 | 8 |
| 3.2 关键字检索的模糊匹配 | 8 |
| 3.3 图像检索的模糊匹配 | 8 |
| 3.4 本章小结 | 8 |
| 4 图像检索检索系统的概要设计 | 9 |
| 4.1 系统功能模块划分 | 9 |
| 4.1.1 图像管理模块 | 9 |
| 4.1.2 关键字匹配模块 | 9 |
| 4.1.3 兴趣物体匹配模块 | 9 |

| | | |
|-------------|------------------------|----|
| 4.2 | 系统用户接口模块划分 | 9 |
| 4.2.1 | 用户管理模块 | 9 |
| 4.2.2 | 图像检索模块 | 9 |
| 4.2.3 | 图像上传模块 | 9 |
| 4.3 | 本章小结 | 9 |
| 5 | 图像检索系统的详细设计与实现 | 10 |
| 5.1 | 系统实现环境 | 10 |
| 5.2 | 系统功能模块详细设计与实现 | 10 |
| 5.2.1 | 图像管理模块的详细设计与实现 | 10 |
| 5.2.2 | 关键字匹配模块的详细设计与实现 | 10 |
| 5.2.3 | 兴趣物体匹配模块的详细设计与实现 | 10 |
| 5.3 | 系统功能模块详细设计与实现 | 10 |
| 5.3.1 | 用户管理模块的详细设计与实现 | 10 |
| 5.3.2 | 图像检索模块的详细设计与实现 | 10 |
| 5.3.3 | 图像上传模块的详细设计与实现 | 10 |
| 5.4 | 系统数据库详细设计与实现 | 10 |
| 5.5 | 本章小结 | 10 |
| 6 | 图像检索系统的测试 | 11 |
| 6.1 | 系统测试环境 | 11 |
| 6.2 | 系统功能测试 | 11 |
| 6.2.1 | 用户管理模块的测试 | 11 |
| 6.2.2 | 图像检索模块的测试 | 11 |
| 6.2.3 | 图像上传模块的测试 | 11 |
| 6.3 | 系统性能测试 | 11 |
| 6.3.1 | 系统检索效果测试 | 11 |
| 6.3.2 | 系统检索速度测试 | 11 |
| 6.4 | 本章小结 | 11 |
| 结 论 | | 12 |
| 参 考 文 献 | | 13 |
| 附录 A 附录内容名称 | | 14 |
| 修改记录 | | 15 |
| 致 谢 | | 16 |

引 言

众所周知，智能来源于大脑。其中，视觉在智能中至关重要。人体工程学的研究显示，视觉是人们获取信息的主要途径，人们获取信息中约有 80% 来源于视觉。反之，视觉图像表现出来的信息是非常丰富的。随着互联网产业的蓬勃发展，每天世界上都会出现大量的信息，信息增长的速度令人瞠目。然而，如何从海量的图片数据中找到我们想要的图片，抑或是包含我们感兴趣内容的图片成为我们生活和工作中的一个难题。

传统的图像检索系统大致分为两种：一是基于上下文本的图像检索（简称 TBIR），二是基于图像内容的检索（简称 CBIR）。

基于上下文本的图像检索不对图像元素进行可视化分析，而是利用图像名称、图像尺寸、图像所在文章的标题等方面来对图像进行标注，根据这些文本分析结果推断其中的图像特征。一般以关键词的形式来查询图像，或者是根据等级目录的形式浏览查找特定类目下的图像。然而，在信息爆炸的当今社会，对每张图片进行标注的成本很高，而且我们很难十分准确地用一个图像所在文章的标题来描述图像的内容。

基于图像内容的图像检索根据图像、图像的内容语义以及上下文联系进行查找，以图像语义特征为线索从图像数据库中检索出具有相似性特征的其他图像。由于图像所具有的信息要远远大于纯粹的文本信息，所以，基于内容的图像检索在检索速度上和效率上的要求要求更高。

基于图像内容的图像检索包括相同物体的图像检索和相同类别的图像检索。在一些情境下，我们需要使用相同物体的图像检索。例如：小区监控中的图像检索，就是要检索同一个人在不同时间以及不同地点在小区的出现情况。然而，在其他情境下，也可能用到相同类别的图像检索。例如：在 3D 形状检索中是检索属于同一类的物品，如飞机等。

基于图像内容的图像检索技术一般由以下几个步骤组成：特征提取，度量学习和重排序。特征提取即为将图片进行降维，提取数据的特征信息，将一张图片降维为一个向量；度量学习一般利用度量函数，计算图片特征之间的距离，作为损失函数来训练特征提取网络，使得相似的图片提取的特征相近，不同的图片提取的特征距离较大。重排序是利用数据流间的流形关系，对度量结果进行重新排序，从而得到更好的检索结果。

在检索之前，首先需要提取到图像的特征，也就是特征提取的步骤。从 2003 年到 2012 年期间，特征提取的方法主要基于 SIFT 特征，由于 SIFT 特征在图像尺度、方向变化问题中的优异表现，十多年来基于局部描述算子的图像检索方法被广泛研究。2003 年词袋模型（BoW）^[1]进入图像检索社区的视野，并在 2004 年结合了 SIFT 方法被应用

于图像分类任务。在后来的近 10 年时间里，BoW 模型的优越性越发显著，它给图像检索任务带来了很大提升。2008 年，Jégou 等人^[2]提出了汉明嵌入，中等密码本在图像检索中开始占有一席之地，2010 年 Perronnin 等人^[3]和 Jégou 等人^[4]提出了 FV 和压缩视觉词汇表达 VLAD，小码本开始受到了欢迎。2012 年 Krizhevsky 等人^[5]提出了新的 CNN 模型，由于深度学习在 ImageNet 大赛上的优异表现，图像检索提取特征的方法逐渐从 SIFT 特征向 CNN 特征转变。2012 年以后，深度学习发展迅速，但是在图像检索领域表现平平，一直没有达到让人满意的水平，甚至低于 SIFT 时代提出的方法，但是依然有前赴后继的研究学者涌向这个新的方向，相继提出了许多有效的方法，精度也不断的逼近于 SIFT 特征。我们把基于 CNN 的方法被分为三类：使用预训练的 CNN 模型，使用微调的 CNN 模型以及使用混合模型。前两类方法使用单向传递网络来提取全局特征，混合模型方法可能需要多个网络传递，由 Razavian 等人^[6]在 2014 年最早提出。

本文了提出一种新颖的基于兴趣物体检测的图像检索算法，这是一种基于图像内容的图像检索方法，并且属于相同类别的图像检索的范畴。我们的方法首先通过兴趣物体检测方法检测出图像中的兴趣物体并对其进行编码，之后对其进行分类存入图像库中。当有关键词查询时，我们便可通过模糊匹配算法将图片数据库中具有相似关键词的图片返回给用户。当有图片查询时，我们首先使用兴趣物体检测算法检测出图像中的兴趣物体并对其进行编码，之后再计算其与图像数据库中的图像的距离，返回具有相似性的图片给用户。

另外，本文设计并实现了基于兴趣物体检测的图像检索系统，采用前后端分离的方法，基于前端 React 框架和后端 Spring Boot 框架进行开发。

本文的组织结构如下：第一章介绍了有关兴趣物体检测的背景知识知识和 React 框架及 Spring Boot 框架的背景知识。第二章对该图像检索系统的实现进行了需求分析和可行性分析。第三章详细介绍了本文提出的基于兴趣物体检测的图像检索算法。第四章介绍了该系统的概要设计。第五章介绍了该系统的详细设计与实现。第六章对该系统进行了测试。

1 相关技术综述

1.1 计算机视觉

1.1.1 物体检测

物体检测是计算机视觉中的一个经典任务。物体检测任务的目标是从图片中检测出物体的位置（一般使用矩形框来标明）并判断出其类别，流行的物体检测器将物体检测转换为矩形区域分类。说到物体检测，不得不提的是 ImageNet 大赛，ImageNet 大赛强烈地推动了物体分类和物体检测的发展。在 2012 年之前，传统的机器学习方法在 ImageNet 大赛上占据主导地位，目标检测精度的进步十分缓慢，因为传统的手工提取特征的方法来提高精度是非常困难的事。2012 年 AlexNet^[7]的卓越表现将卷积神经网络带入了大家的视野中，吸引着越来越多的专家学者对其进行研究，并将其运用到物体检测的任务中来。

目前，物体检测的方法大致可以分为两种，一种是一阶段的物体检测方法，另一种是二阶段的物体检测方法。二阶段的物体检测方法多年来一直占据着物体检测的主导地位，二阶段的物体检测方法将物体检测分为两个阶段。第一阶段是基于图片提出若干可能包含物体的区域，即 Region Proposal；第二阶段是在提出的这些区域上进行分类，得到每个区域内物体的类别。目前表现最好的二阶段物体检测方法是 Faster R-CNN^[8]，这种方法也被众多研究人员用于各种扩展任务中，二阶段物体检测方法在准确率上略胜一筹，但对计算能力的要求较高，效率相对一阶段物体检测方法较低。一阶段物体检测方法经过多年来的发展，其准确率也在向二阶段目标检测方法逼近，并且其效率更高。一阶段物体检测器没有区域裁剪模块。它们可以被视为类别特定的区域或锚点提案网络，并直接为每个正锚点分配一个类别标签。目前效果最好的一阶段目标检测方法是 ExtremeNet^[9]。

1.1.2 语义分割与实例分割

为了更好地理解图像中的内容，我们有必要对图像进行语义分割。当然，语义分割也是计算机视觉领域的一个重要任务。语义分割是为图像中的每一个像素分配一个类标签，这样我们就可以将输入的图片分为不同的语义可解释类别。与图像分类或目标检测相比，语义分割使我们对图像有更加细致的了解。这种了解在诸如自动驾驶、机器人以及图像搜索引擎等许多领域都是非常重要的。

实例分割是视觉经典四个任务（分类、检测、定位、分割）中相对最难的一个，它既具备语义分割的特点，需要做到像素层面的分类；也具备目标检测的一部分特点，需要定位出不同实例，即使它们是同一种类。

1.1.3 全景分割

在计算机视觉的早期，thing——如人、动物、工具等可计数的物体——获得了主要的注意力。研究 thing 通常被表述为对象检测或实例分割任务，目标是检测每个对象，并分别用边界框或分割掩码对其进行描述。而识别具有相似纹理或材料的无定形区域 stuff——如草地、天空、道路，通常被称为语义分割。这两个视觉识别任务的数据集和度量标准看似相关，但实际上是非常不同的。为了弥补语义分割和实例分割之间的裂缝，Alexander Kirillov^[10]等人提出了全景分割的方法，采用统一的全景度量将语义分割（为每个像素指定一个类标签）和实例分割（检测并分割每个对象实例）这两个典型的任务结合了起来，使得研究人员对这种联合任务的兴趣愈发强烈。

1.2 兴趣物体检测

图像所展现出的信息是非常丰富的，然而其中的许多细节并不是我们感兴趣的。为了检测出图像中我们感兴趣的实例，Fan Yu^[11]等人提出了一种兴趣物体检测（IOID）的方法，为图像语义描述提供了用户兴趣实例。具有更大吸引力的实例称为“兴趣实例”（IOI）。IOI 可以从更广泛的意义上被视为一种特定的感兴趣区域（ROI）。从狭义的角度看，ROI 中的“区域”可与“实例”相媲美，涵盖了 thing 和 stuff，而“兴趣”仅限于描述图像时的吸引力。显然，IOIs 是理解图像的基本元素。我们在本文中便基于兴趣物体检测来进行图像检索。

1.3 图像检索

传统的图像检索系统可以分为两种：一种是基于上下文本的图像检索，另一种是基于图像内容的检索。近年来，基于图像内容的图像检索被广泛研究。图像检索的形式也由以文字搜索图像向以图像搜索图像扩展。从 2003 年到 2012 年期间，由于 SIFT 特征不但对于旋转、尺度缩放、亮度变化保持不变性，而且对视角变化、仿射变换、噪声也保持一定程度的稳定性，并且该方法对特征点的个数和有效点的比例没有要求，图像检索中特征提取的方法主要基于 SIFT 特征。后来几年中，BoW 模型它给图像检索任务带来了很大提升。在本文中，我们采取了一种新颖的特征提取方法，即基于兴趣图物体检测的特征提取。

1.4 React 框架

React 是 Facebook 开发的一个用于构建用户界面的 JavaScript 库。React 利用声明式的语法，让开发者专注于描述用户界面“显示成什么样子”，而不是重复思考“如何去显示”，这样可以大大提高开发效率，也让代码更加容易管理。借助 Virtual DOM 技术，使得渲染效率很高。通过 React 构建组件，使得代码更加容易得到复用，能够很好地应用在大项目的开发中。React 实现了单向响应的数据流，从而减少了重复代码。

1.5 Spring Boot 框架

Spring Boot 是由 Pivotal 团队提供的全新框架，是所有基于 Spring Framework 5.0 开发的项目的起点。Spring Boot 技术的出现使得创建基于 Spring 开发的独立的生产级别的应用程序变得非常简单，大多数的 Spring Boot 应用程序只需要很少的 Spring 配置。它具有以下特征：

- 创建独立的 Spring 应用程序
- 直接嵌入 Tomcat, Jetty 或 Undertow（无需部署 WAR 文件）
- 提供“入门”依赖项，以简化构建配置
- 尽可能自动配置 Spring 和三方库
- 提供可用于生产的功能，例如指标，运行状况检查和外部化配置
- 完全没有代码生成，也不需要 XML 配置

Spring Boot 在应用中的角色是承上启下，它的底层是 Spring Framework，上层是 Spring Cloud。其中，Spring Framework 是一种 Java EE 的框架，Spring Boot 是一种快速构建的应用，Spring Cloud 是构建 Spring Boot 的分布式环境，即云应用。这种角色如图 1.1 所示：

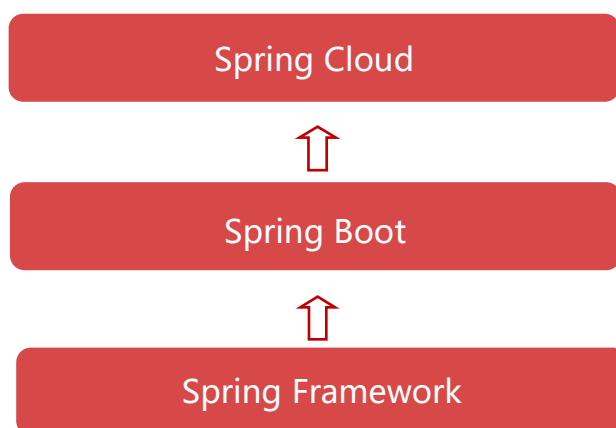


图 1.1 Spring Boot 的角色

在本项目中，我们使用 Spring Boot 的版本 2.3.0SNAPSHOT 版本，它的底层框架是 Spring Framework 5.0.x 并且依赖于 Java 8+。

1.6 MySQL 数据库

MySQL 是一种开放源代码的关系型数据库管理系统（RDBMS），使用最常用的数据库管理语言——结构化查询语言（SQL）进行数据库管理。MySQL 数据库由瑞典 MySQL AB 公司开发，目前属于 Oracle 公司。由于其体积小、速度快、成本低、开放源码、免费等优点，本文采用 MySQL 数据库来管理图像信息。

1.7 本章小结

本章中我们主要介绍了图像检索中所涉及的计算机视觉任务及方法。另外，还介绍了本文用到的目前在 Web 开发中主流的前后端开发框架。

2 可行性分析与需求分析

2.1 可行性分析

2.2 功能性需求分析

2.2.1 图像管理模块的需求分析

2.2.2 图像检索模块的需求分析

2.3 非功能性需求分析

2.4 本章小结

3 基于兴趣物体检测的图像检索算法的实现

3.1 兴趣物体检测

3.2 关键字检索的模糊匹配

3.3 图像检索的模糊匹配

3.4 本章小结

4 图像检索系统的概要设计

4.1 系统功能模块划分

4.1.1 图像管理模块

4.1.2 关键字匹配模块

4.1.3 兴趣物体匹配模块

4.2 系统用户接口模块划分

4.2.1 用户管理模块

4.2.2 图像检索模块

4.2.3 图像上传模块

4.3 本章小结

5 图像检索系统的详细设计与实现

5.1 系统实现环境

5.2 系统功能模块详细设计与实现

5.2.1 图像管理模块的详细设计与实现

5.2.2 关键字匹配模块的详细设计与实现

5.2.3 兴趣物体匹配模块的详细设计与实现

5.3 系统功能模块详细设计与实现

5.3.1 用户管理模块的详细设计与实现

5.3.2 图像检索模块的详细设计与实现

5.3.3 图像上传模块的详细设计与实现

5.4 系统数据库详细设计与实现

5.5 本章小结

6 图像检索系统的测试

6.1 系统测试环境

6.2 系统功能测试

6.2.1 用户管理模块的测试

6.2.2 图像检索模块的测试

6.2.3 图像上传模块的测试

6.3 系统性能测试

6.3.1 系统检索效果测试

6.3.2 系统检索速度测试

6.4 本章小结

结 论

参 考 文 献

- [1] J. Sivic and A. Zisserman, “Video google: A text retrieval approach to object matching in videos,” in ICCV, 2003
- [2] H. Jégou, M. Douze, and C. Schmid, “Hamming embedding and weak geometric consistency for large scale image search,” in ECCV, 2008
- [3] F. Perronnin, J. Sánchez, and T. Mensink, “Improving the fisher kernel for large-scale image classification,” in ECCV, 2010
- [4] H. Jégou, M. Douze, C. Schmid, and P. Pérez, “Aggregating local descriptors into a compact image representation,” in CVPR, 2010
- [5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in NIPS, 2012
- [6] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, “Cnn-features off-the-shelf: an astounding baseline for recognition,” in CVPR Workshops, 2014
- [7] Krizhevsky A , Sutskever I , Hinton G . ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in neural information processing systems, 2012, 25(2).
- [8] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In NIPS, 2015.
- [9] Zhou X , Zhuo J , Krhenbühl, Philipp. Bottom-up Object Detection by Grouping Extreme and Center Points[J]. 2019.
- [10] Kirillov A , He K , Girshick R , et al. Panoptic Segmentation[J]. 2018.
- [11] Yu F , Wang H , Ren T , et al. Instance of Interest Detection[C]// the 27th ACM International Conference. ACM, 2019.

附录 A 附录内容名称

修改记录

记录人（签字）：

指导教师（签字）：

致 谢