摘要

随着智慧城市和平安城市建设的推进，监控摄像头广泛应用于城市安防的各个领域，产生了海量的多媒体数据。精准对象搜索是一项非常具有挑战性的视觉任务，其旨在找到给定对象的所有实例。与传统视觉搜索方法不同的是，在大数据场景下，表观相似对象大量存在，在特征空间很难区分开。因此，该问题具有很强的理论研究意义和实用价值。针对这一问题，本文从视觉对象的高效特征表达和度量学习这两个方向出发展开研究，缓解精准搜索问题。本文针对视觉大数据场景下对象的精准搜索问题及其应用进行研究，主要的贡献包括以下三点：

第一，在特征表达方面，本文提出基于多层次特征融合的图像表达方法。该方法在验证CNN特征与SIFT特征的互补性后，充分利用这两种特征各自在表达上的优势进行特征融合，能同时表达场景级别，对象级别以及点级别的特征信息，融合后的特征具有很强的表达力。在四个公开的对象检索数据集上的实验结果表明，使用本方法的图像表达在对象检索上能获得先进的性能，特别是在维度较低的情况下，性能尤为出色。此外，本文还在结果及得分层面进行了SIFT与CNN的融合，在大规模车辆数据集上的实验结果显示融合后检索性能明显提升，其进一步证明了SIFT与CNN的互补性。

第二，在度量学习方面，本文提出将对象之间的关系建模为多粒度关系。利用多粒度约束，本文提出两种度量学习方法，即基于多粒度二元组的度量学习和基于列表排序的度量学习方法。传统的二元组度量学习只考虑相似/不相似的二元关系，多粒度二元组的度量学习将其推广到多个粒度。基于列表排序的度量学习方法使用排列概率模型直接对给定的多粒度列表的排列进行量化评估，并利用似然损失函数来优化嵌入多粒度约束关系的特征学习。这两种度量学习方法联合多属性分类实现在一套统一的多任务深度学习框架中。为证明该框架的有效性，本文还构建了两个百万级别的车辆数据集（VD1和VD2），数据集中每张图像都标注丰富的属性信息，包括身份ID，精准车型，颜色等。据本文所知，这两个数据集是目前为止已发表的规模最大的车辆数据集。在这两个大规模数据集上的实验结果显示，本文提出的基于多粒度约束的度量学习方法获得了先进的检索性能。

第三，本文实现了一个大数据场景下的精确对象搜索系统。该系统使用深度学习框架并基于web展示，包括对象检测，对象搜索等主要功能。为提高搜索性能和效率，本文对相关模块进行了优化。在二值特征学习模块，系统使用了一种基于ReLU激活的深度二值学习方法。在搜索优化上，本文设计了一种分段距离统计方法，显著地提升了搜索效率。为验证系统性能，本文还构建了一个千万级的车辆检索数据库，库中每

张图片均已进行车辆检测，只包括一个车辆对象。实验显示，本系统在千万量级数据下，检索精度和速度均性能优良，平均搜索时间低于1秒。

总体而言，本文针对视觉大数据场景下精确对象搜索的核心问题展开了深入研究，所提方法均达到了国际先进水平，能显著提升对象检索性能。

关键词：特征表达，车辆精准搜索，深度学习

Research and Application of Precise Object Search in Vision Big Data

Ke Yan (Computer Application Technique)
Directed by Prof. Yonghong Tian

ABSTRACT

With the development of smart city and safety city, surveillance cameras have been widely used in the area of city security and protection. Precise object search is a very challenging task in computer vision, which aims at finding all instances of an object given a single query image. Conventional visual search methods have difficulty in solving this problem as a good deal of visually-similar yet unmatch objects exist in the scenario of big data. As a consequence, this problem is worth paying attention to in terms of both academic research and practical scenarios. To alleviate the problem, this thesis conducts researches in two main aspects of precise object search orienting visual big data, i.e. effective feature representation and distance metric learning. The main contributions are as follows:

Firstly, for feature representation, this thesis proposes an image representation method based on multi-level feature fusion. After validating the complementarity of CNN feature and SIFT feature, a method to fuse the two kinds of features is proposed to exploit their advantages to the full. The fused feature can be used to describe scene-level, object-level and point-level contents in images simultaneously. Extensive experiments on four object retrieval benchmarks are conducted. The experimental results show that the proposed method achieve the state-of-the-arts, especially for its remarkable performance at related low dimensions. Additionally, this thesis also fuse the SIFT feature and CNN feature at result-level and score-level. Experiments are conducted on large-scale vehicle datasets, showing that fusion at result-level and score-level can improve the retrieval performance significantly, which again validate the complementarity of CNN and SIFT.

Secondly, for distance metric learning, this thesis proposes to model the relationship of visually-similar objects as multiple grains. Following this, two metric learning approaches are proposed by exploiting multi-grain ranking constraints. One is Generalized Pairwise Ranking, which generalizes the conventional pairwise from considering only binary similar/dissimilar relations to multiple relations. The other is Multi-Grain based List Ranking, which introduces permutation probability to score a permutation of a multi-grain list, and further optimizes feature learning embedding multi-grain constraints by the likelihood loss function. The two approaches are implemented with multi-attribute classification in a unified multi-task deep learning framework. To demonstrate the effectiveness of the framework, this thesis also contribute two high-quality and well-annotated vehicle datasets, named VD1 and VD2, in which each image is annotated with diverse attributes including ID, precise vehicle model and color. To our knowledge, VD1 and VD2 are the largest high-quality annotated vehicle datasets published so far. Experimental results show that our approaches achieve promising performance.

Thirdly, this thesis implements a precise object search system towards big data scenarios. The system is based on a deep learning framework and a web architecture. Its main functions include object detection, object retrieval and others. In order to improve the performance and efficiency of the system, this thesis conducts optimization on some main modules. For binary feature learning, this paper uses a ReLU activation based deep binary feature learning method. As an optimization of searching function, this paper proposes a segmentation distance calculation method, which improve the searching efficiency significantly. To verify the overall performance of the system, we build a 10 million-level object image library, in which each image has been detected and contains only one object. Experimental results show that on this vehicle library the system performs promising performance at searching precision and speed. Moreover, the average searching time costs less than 1 second.

Overall, this thesis conducts intensive research on core problems of precise object search in vision big data. The proposed methods achieve promising results comparing with state-of-the-arts and significantly improve the performance of object search.

KEY WORDS: Feature Representation, Precise Vehicle Search, Deep Learning