摘　要

近年来，移动智能设备如手机、平板电脑等，变得越来越普及，随着移动网络的发展，人们接入互联网也越来越方便和迅速。此外，越来越多的人选择分享自己生活经历相关的图像到社交网络，这些都促使网络中上传的图像数量急剧增加。然而由于用户手工输入标签的主观性，其中准确标注的图像只占很小的一部分，对于非准确标注或未标注的图像，其他用户往往不能直接从图像的视觉内容中获得图像的准确描述。而对这些图像的准确分类，是网络多媒体信息规范化和有序化的基础，也是各种网络应用如图像检索、个性化推荐以及许多潜在手机图像应用的必要技术条件。

由于上下文对图像视觉内容影响很大，大多传统的仅基于图像视觉内容的图像分类技术，无法克服上下文对于图像的分类准确性的影响。本文研究了使用地理上下文辅助图像视觉信息进行分类的方法，旨在对地理信息的约束性进行有效利用，提高图像的分类准确率。

对于不同的分类任务，由于数据规模，组织形式和信息分布的差异，上下文信息在分类中的作用大小和具体使用方法也不尽相同。在本文中，我们研究了三种方法，分别为 1）基于地理上下文的概率模型方法，2）地理局部化分类模型方法，3）视觉分类模型与地理近邻搜索相融合的方法。其中前两种方法用于餐馆菜品识别，后一种方法用于拍摄图像在常见概念上的分类。

1. 考虑到概率模型在处理复杂关系上的灵活性，我们针对餐馆菜品识别问题建立了概率模型，将地理坐标、餐馆、菜品类别和视觉特征相关联。利用这些关联，我们可以在纯视觉方法的基础上提高分类的准确率。此外，我们还可以将该模型的其他形式应用于餐馆识别和地理位置识别等目标上，相比基于单一模态信息的方法，准确率都有所提高。

2.对于餐馆菜品识别问题，考虑到在全局类别上训练的分类模型的局限性，即其与实际预测时地理近邻中的候选类别存在不匹配的问题，我们提出了地理局部化分类模型的方法，将地理上下文信息用于分类模型的训练过程，使模型从根本上对地理位置更有针对性，再根据查询图像的地理坐标将这些分类模型进行地理自适应组合。该方法避免了地理上下文不同对分类模型的影响，提高了分类准确率。同时也由于每个模型包含类别的减少，简化了分类的时间复杂度，提升了方法整体的可扩展性。

3. 作为对地理上下文辅助菜品识别的泛化，我们研究了拍摄图像在常见概念上的分类问题，这里的目标常见概念来自于 Flickr 数据集中图像的常见标签。在该工作中，我们设计了一种视觉分类和地理近邻搜索相融合的分类框架。我们考虑在两种地理尺度上对地理上下文进行利用，其中地理局部化分类模型在粗尺度上实现，而地理近邻搜索则

是在细尺度上实现，两者最终与全局类别训练的分类器相结合，形成了有效的互补，提高了分类性能。

关键词：图像分类，深度学习，地理上下文，社交网络

Research on Technique of Geographical Context Based Image Classification
Xu Ruihan (Computer Applied Technology)
Directed By Jiang Shuqiang

In recent years, mobile devices like cellphone and tablet PC have become more and more popular. With the development of mobile networks, access to Internet has also become very convenient and quick. Furthermore, more and more people choose sharing their life-related images to the social network, which make the number of uploaded images increases sharply. However, due to the subjective of manual input labels, the precisely annotated images are just a small part of the uploaded images. For the imprecisely annotated or un-annotated images, other users often cannot directly obtain any precise descriptions from their visual content. The accurate classification of these images, is the basis of standardization and ordering of internet multimedia information. And it is also the essential technical requirement for various internet applications like image retrieval and personal recommendation, as well as many potential mobile image applications.

Because of the great influence of context for image visual content, most of the traditional techniques of image classification are only based on visual content, thus it is difficult for them to overcome the contextual effect on image classification. In this thesis, we study the problem of using the geographical context to assist the visual classification, aiming at effectively leveraging the constraint of geographical information to improve the classification accuracy.

Due to the geographical scale, organization style and the information distribution of data, the effect and the usage of geo-information varies with different classification task. In this thesis, we respectively study three methods, which are 1) geographical context base probabilistic model, 2) geo-localized classification models and 3) method of fusing visual classification model and geographical search. The first two methods are used for restaurant dish recognition, and the last one is used for the classification of taken images on common concepts.

1. Considering the flexibility of probabilistic model dealing with complex correlations, we build a probabilistic model for restaurant dish recognition, which connects the geographical coordinates, restaurants, dish categories and visual features together. By using these connections, we can improve the performance of classification compared with visual only based method. Furthermore, we can extend of this model to tasks like restaurant recognition and geographical localization. The accuracy is also improved comparing with single modality based method.

2. For the same task of restaurant dish recognition, we consider the limitation of

classification model trained on global classes, i.e. there is a mismatch between classes used for training and the geographical neighboring candidate classes during prediction. We propose geo-localized models, which uses geographical context during the training stage of classification models. This can make the models instinctively specific for geo-location. During the stage of prediction, these models are adaptively combined according to the geographical coordinates of query image. This method can avoid the influence on classification models caused by different geographical context, and improve the accuracy. At the same time, due to the decrement and classes included in each model, the time complexity is reduced.

3. As a generalization of geographical context assisted dish recognition, we study the problem of image classification on common concepts, which are from the top appeared image tags in the Flickr datasets. In this work, a framework fusing visual classification and geographical neighboring search is designed. Here the geo-localized models are implemented on a coarse scale, and the geographical neighboring search is implemented on a fine scale. These two methods are finally fused with classifiers trained with global classes as complementary, and the performance of classification is improved.

Keywords: Image Classification, Deep Learning, Geographical Context, Social Network