



硕士毕业论文

(毕业设计)

题目：？？？

姓名：林丽

学院：信息科学与技术学院

系： 计算机科学系

专业：计算机科学与技术专业

年级：2014级

学号：23020141153152

指导教师：曲延云 职称：副教授

2017年4月10日

基于汉字组合特征的自然场景中文文本检测

**[摘要]** 自。

**[关键词]**自然场景文本检测

Chinese Scene Text Detection Based on Chinese Character

Combination Features

**[Abstract]** Scene text detection and recognition is a hot topic in the fields of computer vision. It has been widely used in intelligent navigation, image retrieval, vision assistant system, etc. Although there exist a lot of research activities on scene text detection, Chinese text detection is still remained as a challenging problem. In this paper, we proposed a text detection method based Chinese character combination features. Our contributions are as follows: 1) We proposed simple Chinese character combination features, based on which a Chinese components merging algorithm, a word analysis algorithm, and a two-layer filtering algorithm are implemented afterwards. 2) We proposed an integrated algorithm for Chinese text detection, and the experimental results demonstrate the availability of our approach. 3) We built a dataset for Chinese scene text detection and recognition to evaluate the performance of our algorithm. The recall of our approach is 81.068% and the precision is 85.978%.

**[Key Words]**Scene text detection, Chinese character combination features, context

# 绪论

## 研究背景

### 1.1.1 研究意义

# 目标检测方法

目标检测（Object detection），作为计算机视觉领域的一个重要分支，也是ILSVRC（ImageNet Large Scale Visual Recognition Challenge）比赛的任务之一。

**目标检测**指的是给定输入图像和类别（一种或多种类别），找出图像中所有属于给定类别的目标的位置和所属类别。

根据目标检测的定义，只有目标的位置和类别同时符合要求才是正确的检测。如图2.1的上图（A1-A4）为给定类别为{“Steel drum”}的目标检测标准（A1）和检测结果（A2-A4）。图A1为GroundTruth，即要检测的目标类别为“Steel drum”,且目标位置为两个绿色框所在的矩形。图A3的绿色框类别判断正确但框太大，位置错误。图A4的橙色框找到的位置是正确的，但是类别判断错误。因此，只有图A2是位置和类别都正确的。【插图2.1】



图2.1 单类目标检测与多类目标检测

目标检测按照要求类别的多少，可以分为单类目标检测（也称为“**定位”**）和多类目标检测。如图2.1，上图为给定类别为{“Steel drum”}的单类目标检测标准和检测结果，下图为给定类别为{“Steel drum”, “Person”，“Microphone”，“Folding chair”}的四类检测标准和检测结果。单类检测只需找到要求的那类的所有目标，多类检测需要找到所有要求的类别的所有目标。

目标检测按照目标的不同，又可分为行人检测，车辆检测，人脸检测，交通标志检测，文本检测等等【插图2.2】。虽然不同子领域因为目标本身属性的特殊性，有各自的不同算法。例如，基于Viola and Jones的人脸检测算法，基于HOG-SVM的行人检测算法。然而，这些检测方法同属于目标检测的领域，殊途同归，许多算法的框架和本质都是相通的。比如，HOG-SVM除了可以用在行人，用在交通标志检测中也能取到很好的效果。【引用】本文的文本检测方法正是从通用的目标检测方法中得到启发，并将其迁移至文本检测领域。因此，本章着重探讨通用的目标检测方法——从传统方法到近几年基于深度的检测方法，通过目标检测方法的发展引出本文选择SSD作为核心算法的相关背景和原因。

## 传统的目标检测方法

传统的目标检测方法一般分为两大类，基于滑动窗口和基于区域的方法。

基于滑动窗口的方法就是用不同大小的窗口在原图上进行滑动，并用分类模型判断每一个窗口是否包含目标，最后把分类模型判断为正样本的检测结果进行一定的后处理，比如非极大值抑制等。这种方法需要用不同尺度，不同长宽比的窗口对图像的所有位置进行遍历，虽然尽可能降低了目标漏检的可能性，但时间复杂度太高，产生的候选区域太多，给后续的分类器过滤噪声步骤带来很大的计算压力。

基于区域的方法首先根据低级特征（比如，光强，颜色等）把图像的像素分成不同的区域，再用分类模型对这些候选区域进行判断，过滤其中的噪声区域。该类方法得到的候选区域比滑动窗口法得到的窗口少得多，因此在时间和实用性上比基于滑动窗口的方法具有明显的优势。另一方面，候选样本的减少以及使用的低级特征的鲁棒性差也极大限制了该类算法性能的提升。

两类方法各有优劣，其共同的部分在于特征表示与模型分类，不同在于基于区域的方法需要额外的候选区域提取算法来选择样本。近年来，候选区域提取，特征表示，模型分类这三个关键技术受到了越来越多的关注，发展迅速。

* + 1. **候选区域提取**

文章【引用What makes for effective detection proposals?】对比了目标检测领域近几年有代表性的十二种候选区域提取算法，包括Bing， CPMC， Edge Boxes，Endres，Geodesic， MCG， Objectness， Rahtu， RandomizedPrim’s**，** Rantalankila**，**Rigor**，**Selective Search。【引用】其中，Selective Search和Edge Boxes相较于其他算法在综合考量性能和速度时表现较好，使用也更加广泛。Selective Search是使用图像分割技术得到初始的小区域，再根据临近区域的颜色，纹理，尺寸，填充等相似性用贪心算法进行合并。Edge Boxes是用一个框内包含的完整的轮廓个数来判断框内是否包含目标，其依据是如果一个框内包含目标，则框内应该包含足够多的完整的轮廓，而不会与框的边界有很多相交的边缘。

* + 1. **特征表示**

图像特征对图像分类，目标检测，语义分割等计算机视觉技术非常重要，好的图像特征可以极大的提高算法的性能。在没有引入深度学习之前，几乎所有好的目标检测系统都是基于HOG（Hitogram of Gradient）, SIFT（Scale-Invariant Feature Transform）这两大经典特征提取算法。【引用】HOG统计了图像的局部方向梯度的直方图，其核心思想是用光强梯度（或边缘方向）来描述物体的局部轮廓。SIFT通过在尺度空间中寻找极值点作为特征点，并用与位置，尺度，方向相关的描述子来表征目标。

* + 1. **模型分类**

常用的分类模型有很多，比如支持向量机（Support Vector Machine，简称SVM）, 随机森林（Random Forest）， AdaBoost，K近邻（K-Nearest Neighbors，简称KNN），朴素贝叶斯（Naïve Bayes）, 神经网络（Neural Network，简称NN）等等。基于结构风险最小化的支持向量机，尤其是采用了核方法的SVM，由于其鲁棒性强而被广泛应用于各种检测算法中。

## 基于R-CNN框架的目标检测方法

近年来深度学习受到了计算机视觉领域的极大关注，短短几年内在各种不同的视觉任务中取得了巨大的突破【引用】。相比于传统方法使用大量人工设计的特征，深度神经网络可以通过层次化的多层网络自动学习出具有更强鲁棒性的高级特征与语义信息。在众多深度网络结构中，CNN（Convolutional Neural Network）是发展最早、最快的结构之一。2012年，【引用】在ImageNet的图像分类比赛中大放异彩，拉开了CNN快速发展的帷幕。2014年，Ross B.Girshick【引用】使用了Region Proposal与CNN结合的方法（R-CNN）代替了传统目标检测中滑动窗口与手工设计特征的方法，提出了R-CNN框架，使得目标检测取得了巨大突破，并开启了基于深度学习目标检测的热潮。

* + 1. **R-CNN**

1. **R-CNN的主要思想**

R-CNN的主要思想是利用候选区域算法从图像中提取出一定量的候选区域，并将其分别输入到CNN网络中，进行特征提取和分类判别。

1. **R-CNN的算法流程**

图【插图】是R-CNN的算法流程图。其检测过程主要分为3步：

第一， 用Selective search提取2000个候选区域；

第二， 用7层CNN网络模型（5个卷积，2个全连接层）提取卷积特征；

第三， 用多类线性SVM对每一个区域进行分类。

1. **R-CNN的实验效果**

R-CNN算法通过把Region proposal与CNN结合起来，在Pascal VOC 2007的目标检测数据库上达到了58.5%的mAP（mean Average Precision），Pascal VOC 2012的mAP是53.3%，比当时业内最高的记录高出了30%【插入表格或图】。紧随其后的SPP, Fast R-CNN， Faster R-CNN，MultiBox正是基于R-CNN的这种框架进行的改进，这些方法被称为是R-CNN系列的目标检测法。

* + 1. **SPP**
  1. **R-CNN算法的问题**

**问题一**：需要对Selective Search提取到的每一个候选框（proposal）都用前向CNN进行特征提取，大量重叠区域的卷积重复计算，影响算法的实时性；

**问题二**：由于存在全连接层，CNN的输入要求统一大小。因此，每个proposal需要统一裁切（crop）或者缩放（warp）到相同尺度大小，如图【插图】，这在一定程度上造成了图像畸变，影响算法的性能。

* 1. **SPP解决R-CNN问题的思路**

针对问题一，SPP采用整张图像作为CNN的卷积模块的输入，得到全图的卷积特征图。再依据每个候选框在原始图像中的位置，在卷积特征图上取出对应的卷积特征。采用这种策略，整张图只需要一次卷积，避免了重叠区域的重复卷积计算。由于卷积计算是整个CNN中最耗时的部分，因此，采用共享卷积计算的方式可以大大提高计算速度。

针对问题二，SPP在卷积和全连接层之间引入了一个称为Spatial Pyramid Pooling的池化层。该池化层可以将任意维度的卷积特征转换成固定维度的特征。SPP不但可以避免图像畸变，而且满足全连接层输入大小固定的要求。

* 1. **Spatial Pyramid Pooling**

Spatial Pyramid Pooling是一种特殊的池化层，类似于图像金字塔。它的特点是不管输入图像大小如何，输出是固定维度的（一维）特征向量。

如图【插图】，Spatial Pyramid Pooling的输入是任意大小的图像（图中是卷积特征图），设为w\*h\*c，w和h分别表示每张图像的宽度和高度，c表示图像的通道数（图中c = 256）。设Spatial Pyramid Pooling的层数为k （图中k = 3），将每层的每个通道的图像分成nk\*nk个bin（图中，n1 = 1，n2 = 2，n3 = 4），在图像的每个bin中进行池化（图中采用最大池化，也可以用平均池化），并把c个通道的nk\*nk个bin的池化结果串联成一个nk\*nk\*c的一维向量（如图中的256 = 1\*1\*256， 4\*256 = 2\*2\*256， 16\*256 = 4\*4\*256分别为第1层，第2层，第3层的特征向量）。最后把k层的特征向量依次串联作为Spatial Pyramid Pooling的结果输出到全连接层。

* 1. **SPP和R-CNN的网络结构对比**

采用了Spatial Pyramid Pooling后的SPP与传统的R-CNN网络结构不同，如图【插图】。传统R-CNN中，需要对每个候选区域进行裁切或压缩后再分别输入到卷积层，全连接层得到每个候选区域的分类结果。而SPP则首先对整张图进行全卷积，再对每个候选区域提取卷积特征，并分别输入到Spatial Pyramid Pooling和全连接层，得到对应的分类结果。

* 1. **SPP的实验效果**

SPP利用了共享卷积计算和Spatial Pyramid Pooling改进R-CNN算法，不但大幅提高了算法速度，而且还提高了检测精度。【引用】中提到，SPP比R-CNN快了24~64倍的速度，同时在Pascal VOC 2007的目标检测库上精度从mAP为58.5%提高到了59.2%。

* + 1. **Fast R-CNN**
    2. **Faster R-CNN**
    3. **MultiBox**

## Single-shot的目标检测方法

* + 1. **YOLO**
    2. **R-FCN**
    3. **SSD: Single Shot MultiBox Detector**

## 不同目标检测方法的对比

## 本章总结

# 基于SSD的水平文本检测



## SSD: Single Shot MultiBox Detector

## 基于SSD的水平文本检测

* + 1. **文字检测的不同粒度**
    2. **Default box的改进**
    3. **基于投票的多尺度输入融合**

## 实验与分析

* + 1. **Default box尺度大小的影响**
    2. **Default box长宽比的影响**
    3. **数据扩增的影响**
    4. **训练尺度的影响**
    5. **多尺度输出融合的影响**
    6. **多尺度输入融合的影响**
    7. **学习率的影响**
    8. **性能可视化**
    9. **样本大小的敏感性**

## 本章总结

# 基于SSD的倾斜文本检测

## 汉字组合特征

# 实验结果分析

## 数据库

## 评价标准

## 实验环境

## 字符级，单词级，文本线级的不同库上的对比结果（表格，pr曲线，时间）

# 总结与展望

## 总结

自然场景文本检测是计算机视觉领域的研究热点之一。自然场景中的文字蕴含着丰富的信息，在很多领域有着重要的应用。本课题旨在对自然场景中的中文文本进行检测和定位。本文主要利用基于多通道的最大稳定极值区域融合技术提取候选的文本区域。然后对文本区域进行了基于组合特征的文本检测，包括汉字部件合并，词组分析，和非文本滤波三个步骤。最后，采用机器学习的思想，利用改进的方向梯度直方图（HOG）和基于卷积神经网络（CNN）的无监督特征分别训练支持向量机（SVM），对候选区域进一步滤波。本文的主要贡献如下：

1.搭建了适用于中文文本检测识别的自然场景中文文本识别的数据库；

2.提出了汉字组合特征，并基于该特征，设计了相关的汉字部件合并算法、词组分析算法和两层非文本过滤算法；

3. 提出了对候选区域进行分类时应该结合其周围相似区域的性质来判断的思想（上下文信息），并在实验中成功运用该思想提升了实验结果，

4. 提出了面向自然场景中文文本检测的完整算法，并通过实验验证了算法的有效性。

## 展望

本文的研究工作虽然取得了一定的研究成果，但是仍然存在许多需要改进的地方，未来的工作将在以下方面开展：

1. 扩建自然场景中文文本识别数据库CCSV；

2. 结合其他基于连通分量的方法或者基于区域的方法。本文采用的是连通分量分析方法，这种方法在候选区域生成时一般都存在漏检的问题，因此可以将几种连通分量方法结合起来（比如与ER结合等）；

3.改进基于组合特征的相关算法。基于组合特征设计的算法参数均由人工调整，没有很强的自适应性。可以结合无监督学习的思想，将这些算法改进成可以自动学习的模型；

4. 卷积神经网络结构的改进。本文考虑卷积神经网络的复杂性，只利用到了卷积神经网络的特征提取方法。因此未来工作可以对卷积神经网络的训练结构进行简化后使用；

5. 设计端到端的中文文本检测识别系统。将中文文本的检测与识别结合起来，设计完整的端到端中文文本检测识别系统；

6. 将中文的文本检测算法推广到英文上，使算法具备更广泛的使用价值。

# 参考文献

1. A. Coates, B. C., C. Case, et al.“Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning”[C].in ICDAR. 2011. pp.440-445.
2. Asif Shahab, Faisal Shafait, Andreas Dengel. “Reading Text in Scene Images” [C]. in ICDAR.2011. pp.1491-1496.
3. Boris Epshtein, Eyal Ofek, Yonatan Wexler. “Detecting Text in Natural Scenes with Stroke Width Transform” [C]. in CVPR.2010.pp.2963-2970.
4. Dan Cires¸an, Ueli Meier, Jonathan Masci, et al.“A Committee of Neural Networks for Traffic Sign Classification” [C]. in IJCNN.2011.pp.1918-1921.
5. J. Matas, O. Chum,M.Urban, et al.“Robust Wide Baseline Stereo fromMaximally Stable Extremal Regions” [C]. in BMVC.2002.pp.384-393.
6. Jung-Jin Lee, Jung-Jin Lee, Pyoung-HeanLee, et al. “AdaBoost for Text Detection in Natural Scene” [C]. in ICDAR.2011.pp.429-434.
7. K. C. Kim, H. R. Byun, Y. J. Song, et al. “Scene Text Extraction in Natural Scene Images using Hierarchical Feature Combining and Verification” [C]. in ICPR. vol.2. 2004. pp.679-682.
8. Kai Wang, B. Babenko, S. Belongie.“End-to-End Scene Text Recognition” [C]. in ICCV. 2011. pp.1457-1464.
9. Hyung Il Koo,Duck Hoon Kim. “Scene Text Detection via Connected Component Clustering and Nontext Filtering” [C]. in TIP. 2013. pp.2296-2305.
10. S. M. Lucas, A. Panaretos, L. Sosa, et al. “ICDAR 2003 Robust Reading Competitions” [C]. in ICDAR. 2003. pp.682-687.
11. Lukas Neumann, J. Matas. “Text Localization in Real-world Images using Efficiently Pruned Exhaustive Search” [C]. in ICDAR. 2011. pp.687-691.
12. L. Neumann, J. Matas. “Real-Time Scene Text Localization and Recognition” [C]. in CVPR. 2012. pp.3538-3545.
13. Michael R. Lyu, Jiqiang Song, Min Cai. “A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction” [C]. in TCSVT. 2004. pp.243-255.
14. Song Gao, Chunheng Wang, Baihua Xiao, et al. “Adaptive Scene Text Detection based on Transferring Adaboost” [C]. in ICDAR. 2013. pp.388-392.
15. Tao Wang, David J. Wu, Adam Coates, et al. “End-to-End Text Recognition with Convolutional Neural Networks” [C]. in ICPR. 2012. pp.3304-3308.
16. YajuanS.,YanxiangHe,QingquanLi. "ReadingText in Street Views Using Adaboost Towardsa System for Searching Target Places".[C].in IVS. 2009. pp.227-232.
17. Yi-Feng Pan, Xinwen Hou, Cheng-Lin Liu. “A Robust System to Detect and Localize Texts in Natural Scene Images” [C]. in DAS. 2008. pp.35-42.
18. 张煜东. 降维技术与方法综述.[A].2010.
19. 吕岳. 基于汉字结构特征的自由格式手写体汉字切分.[A].2000.
20. <http://baike.baidu.com/view/564370.htm?fr=aladdin>
21. <http://blog.sina.com.cn/s/blog_68b72f57010133pz.html>
22. <http://blog.sina.com.cn/s/blog_60e6e3d50101bkpn.html>
23. <http://blog.csdn.net/v_july_v/article/details/7624837>
24. <http://baike.baidu.com/view/598089.htm?fr=aladdin>
25. <http://blog.csdn.net/zouxy09/article/details/8781543>
26. <http://blog.csdn.net/onlyzkg/article/details/11538203>

# 

# 科研成果

**参与的科研活动：**

1.2013年国家级大学生创新创业训练计划项目: 基于场景文本检测识别的地图智能定位研究（项目负责人)。

**发表论文:**

1. **Li Lin**, Yanyun Qu, Weimin Liao. "Structure Context Clues for Chinese Text Detection". in International Conference on Internet Multimedia Computing and Service(ICIMCS). 2014.(EI检索,已收录)

2. YAN-YUN QU, MENG-JIE LIAO, YAN-WEN ZHOU,TIAN-ZHU FANG, **LI LIN**, and HAI- YING ZHANG. "IMAGE SUPER-RESOLUTION BASED ON DATA-DRIVEN GAUSSIAN PROCESS REGRESSION". in International Conference on Intelligence Science and Big Data Engineering(IScIDE). 2013. pp.513-520 (EI检索).

3. Weiwei Wu, Yanyun Qu, Xiaoqing Yang, and **Li Lin**. "Traffic Sign Detection based on Camera Imaging Apriority". in Pacific-Rim Conference on Multimedia(PCM). 2013. pp.720-729 (EI检索).

# 致谢语

在本篇论文完成的同时，我首先要感谢我的导师曲延云老师。在曲老师的悉心指导下，我的论文才能如此顺利的完成。曲老师的负责、认真、严谨的作风对我的影响很深，以至于在我整个论文的撰写过程中都是带着这种精益求精的态度去做。在此我由衷感谢曲老师给我的帮助和指导，我相信这些习惯、思想会陪伴我一辈子。

另外，我想感谢研究生学长廖威敏同学。他从我接触这个领域开始就一直帮助我，带着我一步步往前走。对于我的问题，学长总会悉心解答，他不但在具体思路和编程中给予我指导，而且在生活中也像一个大哥哥一样帮助我。在这里，我衷心感谢学长。

我还想感谢实验室的所有同学，李建敏、廖梦杰、宋书阳、周燕稳、吴伟伟、卢畅、杨思颖。每当我想起大家都在实验室里一起奋斗的时候就很感动。其中，尤其感谢李建敏学长和吴伟伟学长，他们在我的论文和大学生创新性实验项目中提供了很多帮助。感谢你们！

我要感谢我们的创新性实验小队的所有成员，柳冰莹、刘勇顺、罗道文、王冰、车玉鑫。没有大家，就没有这个项目，感谢大家一直坚持我们的团队。

感谢我的舍友，感谢你们一路陪伴我，支持我！

最后，感谢我的家人，你们是我坚持的动力，谢谢你们给予的支持！