清 华 大 学

综 合 论 文 训 练

题目：基于深度学习的多物体实时跟踪方法研究与实现

系 别：计算机科学与技术系

专 业：计算机科学与技术专业

姓 名：何蔚然

指导教师：姚海龙

2016年 06月 02日

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：学校有权保留学位论文的复印件，允许该论文被查阅和借阅；学校可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存该论文。

(涉密的学位论文在解密后应遵守此规定)

签 名： 导师签名： 日 期：

中文摘要

关键词：

ABSTRACT

Keywords：

目 录

第1章 引言 1

1.1 选题背景与意义 1

1.2 研究内容 1

1.3 论文结构 1

第2章 现有视频跟踪方法概述 2

2.1 小标题 2

第3章 单目标跟踪算法 3

3.1 小标题 3

第4章 多目标跟踪算法 4

4.1 小标题 4

第5章 行人数据集的应用 5

5.1 小标题 5

第6章 实验结果与分析 6

6.1 小标题 6

第7章 总结与展望 7

7.1 小标题 7

插图索引 8

表格索引 9

参考文献 11

致 谢 13

声 明 15

附录A 书面翻译 17

主要符号对照表

如果论文中使用了大量的物理量符号、标志、缩略词、专门计量单位、自定义名词和术语等，应编写成注释说明汇集表。若上述符号和缩略词使用数量不多，可以不设专门的汇集表，但必须在论文中出现时加以说明。

SOT 单物体追踪(Single Object Tracking)

MOT 多物体追踪(Multiple Objects Tracking)

CNN

FCN

RCNN

KITTI

MALF

INRIA

# 引言

## 选题背景与意义

## 研究内容

## 论文结构

# 现有视频跟踪方法概述

## 小标题

# 单目标跟踪算法

## 小标题

# 多目标跟踪算法

## 小标题

# 行人数据集的应用

## 小标题

# 实验结果与分析

## 小标题

# 总结与展望

## 小标题

插图索引

No headings found. **This is an automatic table of contents. To use it, apply heading styles (on the Home tab) to the text that goes in your table of contents, and then update this table. If you want to type your own entries, use a manual table of contents (in the same menu as the automatic one).**

表格索引

No headings found. **This is an automatic table of contents. To use it, apply heading styles (on the Home tab) to the text that goes in your table of contents, and then update this table. If you want to type your own entries, use a manual table of contents (in the same menu as the automatic one).**

参考文献

[1]

致 谢

衷心感谢姚海龙老师在本次毕业设计及综合论文训练中对我的帮助和悉心指导。姚海龙老师认真负责的研究态度以及严谨的作风都让我受益匪浅。

最后感谢班里同学在我完成综合论文训练阶段给予我的帮助。

声 明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签 名：\_\_\_\_\_\_\_\_\_\_\_\_\_ 日 期：\_\_\_\_\_\_\_\_\_\_\_\_\_\_

附录A 书面翻译

稠密盒：基于端到端物体识别的统一标志定位

A1 摘要

一个FCN（全卷积神经网络）要怎么在物体检测上工作？我们发明了稠密盒——一个统一的端到端FCN框架——可以直接预测单张图片中任意位置及大小物体的包围盒以及物体分类置信度。我们的贡献主要有两方面。一是我们证明了一个FCN，只要经过小心的设计以及优化，可以极其准确且高效地检测多个不同物体；二是我们证明在多任务学习时引入标志定位，稠密盒可以更好地提高物体检测精确度。我们展示了在包括MALF人脸检测以及KITTI车辆检测的公开数据集上的实验结果，结果表明我们的稠密盒系统在检测如人脸、车辆等有挑战性的物体方面是目前最先进的。

A2 引言

我们的日常生活充斥着大量的物体检测实例。驾驶时注意最近的车辆、找一个人以及定位一张熟悉的脸，这些都是物体检测的例子。物体检测是计算机视觉中最核心的问题之一。在CNN成功以前，物体检测常常会归类为基于移动窗的方法，也就是在图像中所有可能的位置及尺度中运用分类器抽取人工特征。最近，基于FCN的方法给物体检测领域带来了一场革命。这些FCN框架也遵循移动窗的特点，但他们在学习模型参数及抽取图像特征时端到端的方式显著提升了检测表现。

RCNN相对基于FCN的方法在物体检测准确率上有了进一步提高。理论上，RCNN包含两个部分。一是使用区域提议方法生成图像中所有潜在的候选包围盒；二是使用了CNN分类器为每个提议分辨出不同的物体。虽然RCNN在普适物体识别上成为了新的最先进系统，但它仍很难检测如人脸、远处的车等小物体，因为低分辨率及缺少背景信息的候选包围盒会显著降低分类准确率。更多地，RCNN流程上这两个不同部分不可以被合并地优化，为端到端训练RCNN留下了麻烦。

我们的工作关注一个问题：在物体检测上一个单级FCN能达到什么程度？为此我们提出了一个基于FCN的物体检测器——稠密盒——不需要提议生成并且可以在训练时被端到端地优化。尽管与很多已有移动窗FCN检测框架相似，稠密盒还是在小尺度及高污染的情况下被更仔细地设计。我们训练稠密盒并使用精细、困难的挖掘技术去压榨它的表现。为了使它表现更好，我们更通过多任务学习在系统中整合了标志定位。为了验证标志定位是有用的，我们在KITTI车辆检测上手工标注了一批关键点数据，并会在将来公开它们。

A3 稠密盒检测

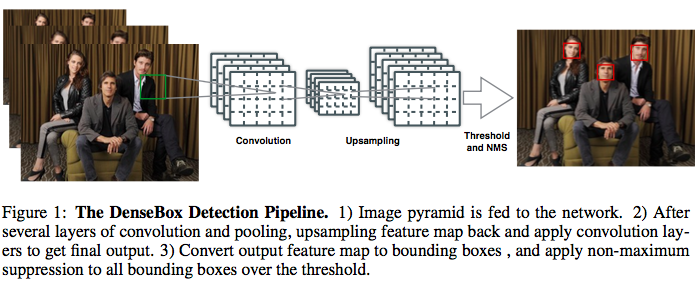


图 1

一个完整的检测系统如图 1所示。一个卷积网络同时输出多个预测包围盒以及分类置信度。稠密盒中物体检测的所有部件都被建模成一个FCN（除了非极大值抑制的部分），所以区域提议生成不是必要的。在测试中，系统把一张图片（尺寸为m×n）作为输入，并输出一张尺寸为的5通道特征图。

A3.1 生成标注值

A3.2 模型设计

A3.3 多任务训练

A3.4 改善标志定位

A3.5 对比

A4 实验

A4.1 MALF识别任务

A4.2 KITTI车辆检测任务

A5 总结