



南京大學

研究生畢業論文  
(申請碩士學位)

論文題目 南京大學畢業論文 L<sup>A</sup>T<sub>E</sub>X 模板

作者姓名 管侯祺

專業名稱 計算機科學與技術

研究方向 人工智能

指導教師 申富饒 教授

2023 年 2 月 15 日

学 号：MF20330024

论文答辩日期：xxxx 年 xx 月 xx 日

指 导 教 师： (签字)

# **L<sup>A</sup>T<sub>E</sub>X Post-processing Algorithm for Video Object Detection Based on Context Information**

by

**Authoraaaa**

Supervised by

Professor

A dissertation submitted to  
the graduate school of Nanjing University  
in partial fulfilment of the requirements for the degree of

MASTER

in

Computer Science and Technology



Department of Computer Science and Technology  
Nanjing University

February 15, 2023



# 南京大学研究生毕业论文中文摘要首页用纸

毕业论文题目： \_\_\_\_\_ 标题第一行  
\_\_\_\_\_ 标题第二行用于长标题换行  
\_\_\_\_\_ 计算机科学与技术 \_\_\_\_\_ 专业 2012 级硕士生姓名： \_\_\_\_\_ 管侯祺  
指导教师（姓名、职称）： \_\_\_\_\_ 申富饶 教授

## 摘 要

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

关键词： 关键词 1    关键词 2



## 南京大学研究生毕业论文英文摘要首页用纸

THESIS: \_\_\_\_\_ englishabstracttitlea  
\_\_\_\_\_ nglishabstracttitleb  
SPECIALIZATION: \_\_\_\_\_ Computer Science and Technology  
POSTGRADUATE: \_\_\_\_\_ Authoraaaa  
MENTOR: \_\_\_\_\_ Professor

### **Abstract**

wwwwwwwwwwwwwwww

**keywords:** keyword1 keyword2





# 前言

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

作者  
20xx 年夏于南京大学



# 目 录

中文摘要 .....	i
英文摘要 .....	iii
前 言 .....	v
目 录 .....	vii
插图清单 .....	ix
附表清单 .....	xi
第一章 绪论 .....	1
1.1 研究背景 .....	1
1.2 研究现状 .....	2
1.2.1 目标检测 .....	2
1.2.2 视频目标检测后处理 .....	4
1.3 本文研究内容 .....	5
1.4 本文结构安排 .....	6
第二章 相关工作 .....	9
2.1 目标检测 .....	9
2.1.1 任务描述 .....	9
2.1.2 术语介绍 .....	9
2.1.3 性能度量 .....	10
2.2 图像目标检测 .....	11
2.2.1 R-CNN: Regions with CNN features .....	12
2.2.2 Faster R-CNN .....	12
2.2.3 YOLOv3 .....	12
2.2.4 CenterNet .....	12
2.3 视频目标检测 .....	12
2.4 后处理策略 .....	12
第三章 基于视频上下文信息的后处理框架 .....	13
3.1 问题分析 .....	13

3.2 基于视频上下文信息的后处理策略 .....	13
3.3 实验与分析 .....	13
3.4 本章小结 .....	13
<b>第四章 实时视频目标检测结果平滑及后处理框架 .....</b>	<b>15</b>
4.1 问题分析 .....	15
4.2 实时视频目标检测后处理策略 .....	15
4.3 实时视频目标检测结果的平滑处理 .....	15
4.4 实验与分析 .....	15
4.5 本章小结 .....	15
<b>第五章 基于上下文信息的后处理策略及实时后处理策略在系统中的应用 ..</b>	<b>17</b>
5.1 相关背景 .....	17
5.2 系统需求 .....	17
5.3 系统架构 .....	17
5.4 系统实现 .....	17
5.5 效果展示 .....	17
5.6 本章小结 .....	17
<b>第六章 总结与展望 .....</b>	<b>19</b>
<b>第七章 使用示例 .....</b>	<b>21</b>
7.1 示例如下 .....	21
<b>参考文献 .....</b>	<b>23</b>
<b>A 附录代码 .....</b>	<b>25</b>
A.1 main 函数 .....	25
<b>致 谢 .....</b>	<b>27</b>
<b>简历与科研成果 .....</b>	<b>29</b>
<b>版权及论文原创性说明 .....</b>	<b>31</b>
<b>《学位论文出版授权书》 .....</b>	<b>33</b>

# 插图清单

7-1	单图示例 .....	21
7-2	实验硬件设备总览 .....	22
7-3	实验测量示意图 .....	22
7-4	子图样例 .....	22



## 附表清单





# 第一章 绪论

## 1.1 研究背景

近年来，基于信息时代的海量数据以及日益提高的数据处理能力和计算能力，机器学习领域的发展尤为迅猛，特别是深度学习方面，越来越多的基于深度学习的应用出现在我们的日常生活中，降低了很多场景中的人工成本，比较典型的一些应用是计算机视觉、自然语言处理、语音信号处理、推荐系统、时间序列分析、对抗攻击安全研究等。其中，计算机视觉是较早被应用于我们日常生活的一种技术，具体来说，其任务是训练计算机和系统，使之能够从视觉输入信息中提取出用户所需要的有意义的目标信息，并在此基础上进行后续的分析并反馈给用户。计算机视觉基于大量的图像和视频等数据基础，结合以卷积神经网络 (Convolutional Neural Networks, CNN) 为例的网络结构，实现像素级的内容处理，最终实现让机器和系统模拟人类视觉系统工作机制来处理图像画面。图像数据可以采用多种形式，比如静态图像数据、动态视频序列、多个视角的摄像机图像数据以及一些通过扫描得到的多维数据，而计算机视觉技术关注的正是基于图像数据，将学科理论和模型应用于计算机视觉系统的构建。

上述内容简要介绍了计算机视觉任务目标及基础原理，而图像内容检测、目标检测、图像分类、内容跟踪、3D 场景重建等均属于比较常见的计算机视觉任务。其中，目标检测是计算机视觉领域最重要也最具有挑战性的分支之一，主要任务是检测数字图像和视频中某一类或者多类语义对象（例如汽车、动物、行人等），因此目标检测在我们日常生活中的应用也非常普遍，如自动驾驶、视频监控、场景分析、机器人视觉等。近两年也因为目标检测日益广泛的应用和快速的技术突破而受到学术界和商业界越来越多的关注。目标检测的实现方式通常分为基于神经网络的方法及非神经网络方法。对于非神经网络方法，通常需要预先定义特征然后再利用以支持向量机 (SVM) 为例的技术进行分类。而另一种基于神经网络的方法，则能够做到在不具体定义特征的情况下进行端到端的目标对象检测，这一类方法通常基于卷积神经网络 (CNN) 来实现，Region with CNN features (RCNN) 是第一个成功将深度学习应用于目标检

测上的算法，在此基础上也不断有新的目标检测模型被提出并应用到这一任务领域内，本文将要描述的后处理方法，也正是基于这类基于神经网络的目标检测方法所提出的优化机制。

目标检测的实现中，后处理是十分重要的一环，在前文提到的基于神经网络实现目标检测的方法中，以 RCNN 为例的方法主要利用选择性搜索来生成候选区域，再进行特征提取和分类，最后利用 NMS 这类方法得到修正之后的检测结果。目标检测后处理方案主要是对修正步骤的优化，前序模型检测结果包含若干个目标候选框，后处理的任务是对这些目标候选框进行筛选和过滤，得到最终的目标检测结果框。另外，目标检测领域进一步细分，可以分为基于静态图片的目标检测和基于视频流的目标检测，其中基于视频流的目标检测，由于涉及到视频时序信息及上下文信息，在后处理阶段可以设计的内容就更多，相应的后处理部分也就能提供更好的筛选过滤效果，从整体上达到提高目标检测精度的效果。

综上所述，视频目标检测后处理作为视频目标检测的技术补充，可以在计算代价低的前提下实现对视频目标检测的精度提升，从而更好的将静态图像目标检测模型用于视频目标检测的场景下，是一种高效的解决方案，具有重要的研究意义。

## 1.2 研究现状

本文主要工作是围绕视频目标检测的后处理方案展开，目标检测器的输入数据一般为静态图像或者视频流，可以粗略分为静态图像检测器和视频检测器两种。而视频目标检测有两种主流的解决方案，一种是将应用于静态图像的目标检测模型逐帧的应用于视频流，再结合后处理机制对视频数据的上下文信息加以利用，另一种则是基于视频级的数据和特征工作训练专用于视频的目标检测模型。本节内容主要介绍目标检测及视频目标检测后处理的研究现状。

### 1.2.1 目标检测

目标检测是一项重要的计算机视觉任务，主要目的是检测数字图像中特定类别的视觉对象，这一技术的应用十分广泛，包括辅助驾驶、智能避障、视频监控、图像检索等场景。在过去的几年里，深度学习技术的快速发展大大加速了目标检测的发展进程，借助深度学习网络和 GPU 的计算能力，目标检测的性

能得到了很大的提升，也奠定了许多下游计算机视觉任务的基础，例如图像字幕、对象跟踪等。在目标检测领域的研究初期，一般研究的是针对于图像数据的目标检测，后期有了针对于视频数据的目标检测，因此又把针对于图像数据的目标检测器称为静态图像目标检测器或通用检测器，把针对于视频数据的检测器称为视频目标检测器。

事实上，目标检测，也就是静态图像目标检测这一任务的提出和发展并不只是近几年的研究内容，其发展进程通常分为两个时期，以引入深度学习作为分割，分为之前和之后两个阶段。2014 年以前主要是利用传统技术手段进行目标检测，2001 年提出的 Viola-Jones 检测器开创了目标检测发展的先河，2006 年的 HOG 检测器在计算机视觉和图像处理任务中使用了特征描述符，而 2008 年的 DPM 首次引入了边界框回归。从 2014 年开始，以 RCNN 为首的检测器将目标检测带入了深度学习检测的时代，后续的检测器被分为 two-stage 和 one-stage 两个类型。通常来说，基于深度学习的对象检测器从输入图像或者视频帧中提取特征，再解决后续两个任务：查找存在于画面中的任意数量的目标；对每个目标进行分类并使用边界框预测每个目标的大小和位置信息。two-stage 类型的检测器分阶段处理这两个任务，相应的，one-stage 检测器选择将这些任务合并为一个步骤来执行，以牺牲一定的准确性为代价来获得更高的性能。如前文所述，卷积神经网络 CNN 的引入极大的促进了目标检测领域的发展，许多基于静态图像的检测模型，如 R-CNN 及其变体、YOLO 及其变体，都为静态图像目标检测建立了良好且有效的检测框架。目前较为主流的静态图像目标检测模型，其处理机制可以概括为三个阶段：建议区域生成阶段；目标分类阶段；后处理阶段。首先，在建议区域生成阶段，根据图像区域中包含目标的可能性大小，来生成一组候选区域，更早之前的建议区域生成方法是基于更低层次的图像特征，但目前的以 Faster R-CNN 为例的静态图像目标检测模型则是使用神经网络来学习并生成建议区域。随后在第二个目标分类阶段，则是给候选区域分配一个等级分数。最后在后处理阶段对多个目标检测框进行冗余过滤，得到最终的反馈结果。

静态图像目标检测在近几年的发展中虽然得到了较大性能上的提升，但将静态图像目标检测模型直接用于视频目标检测则效果欠佳，这是因为静态图像目标检测模型忽略了视频流中时间维度的信息。在视频数据中，前一帧中很容易被检测出且置信度高的目标，在下一帧中可能很难被检测到或者置信度很低，视频画面存在遮挡、剧烈运动、模糊都有可能导致这一情况，这也就使得

直接将静态图像目标检测模型应用于视频流变得更为困难。为了解决这一问题，研究者们主要探索出了两种策略。具有高计算成本的针对于视频数据的视频检测器，以及静态图像目标检测器结合快速后处理算法，这两种策略共同构成了目前视频目标检测的主流解决方案。一方面是基于视频数据的特点专门为视频目标检测设计检测模型，这一方案需要在视频数据的基础上，根据当前处理的帧及附近的帧实现特征聚合，并实现与静态目标检测相同或者更高的检测精度，而这需要大量的计算，就导致其检测速度缓慢，则无法很好的适用于低性能设备或者需要实时计算的使用场景。另一方面就是结合后处理方案，后处理方案的输入一般为目标检测模型中已检测出但未被过滤的检测结果，通过帧间关联检测对象，并使用得到的关联信息来优化检测结果，这一类方案由于计算代价低且设计逻辑更为清晰，更适用于工业生产及日常检测的场景。

### 1.2.2 视频目标检测后处理

视频目标检测与静态图像检测很大的一个不同点是，视频的关键因素是时间信息和上下文信息，因为视频中物体的位置和外观在时间上应该是一致的，换言之，在同一个视频片段内的检测结果在检测框位置和检测框置信度方面不应该发生剧烈变化。这也就使得性能良好的静态图像检测器直接运用于视频上，会因为无法利用这些时间信息和上下文信息而出现性能下滑的情况。另一方面，视频数据中存在物体遮挡、剧烈运动、模糊、罕见角度或姿势，这也加大了视频目标检测的难度，此时视频的上下文信息和时间信息就显得尤为重要。基于计算性能和研发成本的考虑，通过后处理机制弥补静态图像检测模型无法利用时间信息和上下文信息的问题，是一种在成本和收益方面都比较理想的解决方案。

后处理方案的提出是为了将静态目标检测器更好的运用于视频数据但又不在网络模型方面增加过多的负担，这类后处理方法应用于图像检测器每一帧的检测结果，结合时间信息和上下文信息，给出最终的预测输出。后处理方案不仅提高了静态检测器直接运用于视频数据的准确性，同时在计算代价和开发成本上也小于基于视频数据设计视频目标检测器所需要的代价和成本。

常见的后处理方法可以被分成两类，一类是建立帧之间的关联，结合时间信息来进行后处理；另一类则是从特征工程出发，通过视频数据和特征工作来优化检测结果。一般来说，基于特征工作的方法需要的人工设计部分比较少，也可能在准确率方面取得更好的结果，但不得不考虑的问题是，基于特征工作

的后处理方法需要视频数据集进行训练，与传统的图像数据集相比，视频数据集收集和标注所需的工作量更大，因此可用的公开数据集也少之又少。因此，基于帧间关联的后处理方法仍然十分重要，使用常见的图像数据集就可以完成训练，因而在实际工程项目中使用的也更多，是一种成效理想且实现代价较低的后处理方案。现有的基于帧间关联的方法大多依赖光流、物体跟踪或跨时间排序，但显而易见的是，基于计算复杂性的考虑，光流的计算成本更高，而且光流分析本身与目标检测这一主线任务没有太直接的关系，仅仅只是作为后处理的一种辅助手段，因此付出太过高昂的计算代价性价比很低。同理，使用目标跟踪手段也会使得计算开销变大，在获得准确率提升的同时却降低了检测效率。与上述问题相似的是，其他很多基于帧间关联的后处理方案都存在计算代价高，不能应用于在线视频检测的问题，但回溯到目标检测的常见应用场景，自动驾驶、视频监控、机器人视觉等诸多任务都是需要实时视频流检测及分析的，少有利用本地视频文件做分析的应用场景，因此基于帧间关联实现后处理的方案也仍有需要优化的地方。

### 1.3 本文研究内容

本文主要研究应用于视频目标检测的后处理方案，其目的是在静态目标检测模型的基础上，通过后处理方案的设计和实现，将静态目标检测模型过渡并应用于视频目标检测。此类方案主要利用视频检测场景中大量的上下文信息和时间信息，弥补了静态检测模型的不足，并且对于不同的静态目标检测模型具有普适性。除此之外，本文所研究的后处理方案所需要的计算成本较低，是一种高性价比的目标检测优化策略。更进一步的，本文对检测结果的优化分为非实时检测和实时检测两种，既满足了以高精度为目标的非实时检测使用场景，也满足了强调实时性的在线场景。最后，本文所设计的策略应用于实际系统工程中，通过实际效果检验其有效性和实用性。文本的主要研究内容总结如下：

- 本文利用视频流中的时空信息和上下文信息，设计了一种能应用于静态目标检测模型或视频目标检测模型的后处理策略（Context Information Based Post-Processing, CIBPP）。CIBPP 通过位置信息、语义信息、外观信息来描述一个目标检测框，并通过距离函数计算相邻视频帧之间任意两个检测框之间的距离，由此建立距离矩阵，实现跨帧的检测框连接。在建立检测框连接的同时，增加长时反馈机制，将长时间稳定出现的检测框认为

是正确的目标检测结果。建立起跨帧的检测框连接之后，每一段连接内通过类分数重置及补偿来优化该视频片段内的误检情况，与此同时在不同的两段连接之间，通过双向扩散和检测框匹配来优化漏检情况。该策略在 ImageNet VID 数据集上的实验结果证明了其有效性，并且该方法具有普适性，并不限制具体的目标检测模型选择，因此是一种成本低、效果好的目标检测优化策略。

- 基于上述非实时性视频目标检测后处理策略的设计和实现，更进一步的，本文还提出了一种适用于在线实时检测场景的后处理策略，并使用卡尔曼滤波对检测结果进行平滑（Online Post-processing and Smooth, OPPSmooth）。此类方法在处理每一帧时，仅利用当前帧及前序帧的信息，建立信息队列，在队列内对前向若干帧的检测结果进行连接，并实时重置检测框的位置、类分数等信息。并在一定程度上通过卡尔曼滤波来平衡由于像素信息引起的相邻帧之间检测结果的抖动，以达到稳定检测框的效果。和其他的在线视频目标检测后处理方案相比，本文所提方法取得了更好的效果，并且能在实时检测场景下得到更稳定的检测框输出。
- 本文将所提方法融合在了实际的系统工程中，该系统场景为井下煤矿作业场景，通过实时目标检测识别出井下的工作人员，同时标识出电缆槽的位置作为安全工作区域和危险工作区域的分界，判断当前是否有工作人员处于危险位置，若有则及时给出预警。该系统可以上传视频，进行非实时本地视频检测，也可以进行在线实时处理，即模拟井下摄像头实时输入的效果，整个系统很好的体现了本文所提的后处理方法在实际应用中的价值。

## 1.4 本文结构安排

本文主要研究的是利用视频上下文信息设计实现后处理策略，使得视频目标检测的准确度得到提升，除了聚焦于准确度的提升，本文同样关注在线检测的实时性要求，又提出了满足实时检测要求的另一种后处理方案，并最终将所提策略应用于实际的工业系统应用中。全文共有六个章节，第一章绪论部分，主要阐述本文的研究背景，从机器学习到目标检测，再到视频目标检测及后处理，一步步细化。同时还介绍了目标检测及后处理这两个领域的研究现状。最后点明本文研究内容；第二章为相关工作部分，分图像目标检测及视频目标检测两个方向介绍相关工作，对一些具有代表性的研究工作进行分析；第三章介

绍了本文提出的基于上下文信息的视频目标检测后处理策略，包括问题分析、详细设计、结果分析等内容；第四章主要介绍在线视频目标检测后处理方案，分析了实时检测这一条件的限制下，如何通过后处理实现目标检测结果优化，最后通过测试验证其有效性；第五章介绍了一个融合了前文所述的后处理策略的实际系统工程，并对整个系统的使用方式和功能进行分析，充分说了文本所提方法的实际应用价值；第六章为本文的总结部分，也包含对后续工作及研究方向的展望。





## 第二章 相关工作

本章主要介绍本文所涉及到的相关背景知识，以及国内外一些具有代表性的工作内容。首先介绍图像目标检测的相关内容，包括任务描述、发展历程、代表工作以及最新的处理模式。然后是由图像目标检测进一步引申的视频目标检测这一领域，说明视频目标检测任务的提出，以及主流的解决方案。视频目标检测的解决方案其中之一就是静态目标检测器结合后处理机制，因此对目前常见的具有代表性的一些后处理机制也进行了介绍，为后续章节中提出的后处理方案提供一定的理论基础和背景铺垫。

### 2.1 目标检测

#### 2.1.1 任务描述

目标检测是计算机视觉的一个重要分支，目标检测的任务是对图像中的物体进行识别和定位，通常包括以下步骤：首先对图像或视频进行预处理，如去噪、归一化等；接着通过特征提取算法，提取图像中的关键特征，如 SIFT、HOG 等；然后通过目标检测模型，对图像中的物体进行识别和定位，比如使用 YOLO、Faster R-CNN 等模型；最后对结果进行后处理，比如移除重叠的框，保留最可能的框。目标检测在许多领域都有广泛应用，如自动驾驶、安防监控、人脸识别等。它涉及到计算机视觉、机器学习和图像处理等多领域，是一个具有挑战性的任务。

#### 2.1.2 术语介绍

随着深度学习技术的发展，目标检测从早期的图像目标检测又拓展出了后文所述的视频目标检测这一任务。因此，为了在本文中更好的介绍目标检测及后处理技术，现列举一些常用术语及其含义：

- **锚框（Anchor Box）**：Anchor box 是一种预先定义的、固定的边界框，在目标检测中通常使用它来映射图像中的目标。例如，在 YOLO 算法中，每

个网格单元内部都有多个 Anchor box，用于将图像中的目标映射到网格单元的各个 Anchor box 中。Anchor box 这一概念在 Faster R-CNN 中首次被提出，并在后来提出的许多检测模型中得到了更广泛的应用。

- **特征图 (Feature Map)**：Feature Map 是在每一层神经网络中提取的特征的数值表示。主要的作用是将输入的图像信息转换为更加简化的特征表示，从而为目标检测任务提供有效的特征描述。例如，在图像中检测目标时，第一层卷积可能会提取出图像中的边缘信息，而第二层卷积则可能会提取出更复杂的纹理特征。每一层都是对图像特征的不断简化和抽象，从而提取出更有效的目标描述。因此，feature map 是目标检测中十分重要的一个组成部分，它对提高模型的准确性有着重要的贡献。
- **检测框交并比 (Intersection over Union, IoU)**：IoU 是衡量两个检测框的重叠情况的指标，通过计算两个边界框的重叠面积与其并集面积的比值来得到。IoU 的值在  $[0,1]$  之间，值越接近 1，说明两个边界框的重叠越大。
- **非极大值抑制 (Non-Maximum Suppression, NMS)**：NMS 是一种剔除多余边界框的方法，它通过预先选择得分最高的边界框，并删除与该边界框 IoU 超过预先定义的阈值的其他边界框，来确保最终预测结果中只包含一个目标。例如，在 YOLO 算法中，使用 NMS 可以在多个 Anchor box 中选择最优的边界框作为目标的最终预测结果。
- **两阶段检测器 (Two-Stage Detector)**：Two-stage detector 是一种常见的目标检测算法，通过两个阶段的处理来识别图像中的目标。第一阶段生成候选目标区域，第二阶段对候选区域进行评估，以确定最终的目标位置。例如，Faster R-CNN 算法就是一种 Two-stage detector。
- **一阶段检测器 (One-Stage Detector)**：One-stage detector 是一种目标检测算法，它在单个阶段内识别图像中的目标。相比于 Two-stage detector，One-stage detector 的处理速度更快，但预测精度通常略低。例如，YOLO 和 SSD 算法就是一种 One-stage detector。

### 2.1.3 性能度量

前文介绍了目标检测领域的相关术语，接下来需要介绍的是度量目标检测效果优劣的方式，即评价指标。目前有多种性能指标可以用于从不同角度评估检测器的性能优劣，比如 Precision、Recall、FPS、AP、mAP 等，其中 Precision 可以由 IoU 推导得到，mAP 则是目前最为常用的单一指标。IoU 是两

个检测框重叠的部分除以两个检测框的集合部分得出的结果。具体表示如下：

$$S = \sqrt{p(p-a)(p-b)(p-c)} \quad (2-1)$$

在目标检测框架中，通常会设定一个 IoU 阈值来判定检测结果是否正确，具体来说，当预测检测框和 Ground Truth 的 IoU 超过该阈值时，该检测结果则被分类为 True Positive (TP)，相应的当预测检测框和 Ground Truth 的 IoU 未超过该阈值时，该检测结果被分类为 False Positive (FP)，而模型未给出对应的检测结果，对应的就是 False Negative (FN) 的情况。基于上述内容，可得 Precision 和 Recall 的定义如下：

$$S = \sqrt{p(p-a)(p-b)(p-c)} \quad (2-2)$$

AP、mAP 的定义在 PASCAL VOC CHALLENGE 官方给出的定义中有过修改，按照最新的计算方法，对于每个 Recall 值，均可以计算出对应的最大精确率 Precision，然后对所有的精确率取平均即得到最后的 AP 值。AP 衡量的是学出来的模型在每个类别上的好坏，而不同的检测器对基本类别的数量和类别设定均有不同，因此就需要一个单一指标来对比不同检测器的性能，从而就引出了 mAP。mAP 衡量的是模型在所有类别上的好坏，m 是 mean 平均值，具体计算方式就是对所有的 AP 取均值。

除了上述提到的衡量检测器准确度的几个相关指标之外，目标检测算法另一个不可忽视的评估指标是计算速度。速度衡量的主要是检测器的计算性能，尤其是在实时检测的场景下，对计算速度的要求更为苛刻。Frame Per Second (FPS) 用来评估目标检测的速度，即每秒内可以处理的图片数量。FPS 越高，则处理单张图片所用时间就越少，速度越快。另外需要注意的是，对比不同检测器的速度时，需要保证硬件环境相同，硬件环境对计算速度的影响也比较大。

## 2.2 图像目标检测

2014 年以前的图像目标检测工作属于“传统目标检测”，随着 2014 年 R-CNN 的提出，目标检测正式进入了“基于深度学习的目标检测”时期。早期的目标检测代指的即为图像目标检测，因此大量的检测器也是基于这一背景下产生的。由于本文所提的后处理方法所对应的前序检测器是基于深度学习实

现的，本节主要介绍基于深度学习的目标检测相关工作。当前主流的目标检测方法可以被分为两类：两阶段检测器（Two-Stage Detector）、一阶段检测器（One-Stage Detector）。两阶段检测器有一个独立的模块用于生成区域候选框（Region Proposals）该类检测器在第一阶段产生一定数量的区域候选框，接着在第二阶段对候选框进行分类和定位。但由于整个处理逻辑被分成了两个阶段，整体架构会更为复杂，整个流程的耗时也会稍长一些。而一阶段检测器往往通过密集采样，直接对语义目标进行分类和定位，整个流程在一个阶段内完成，在处理逻辑上更为简单，计算速度也更快，但精度相比于两阶段检测器会略低。本节将分别对一些具有代表性的两阶段检测器和一阶段检测器进行介绍。

### **2.2.1 R-CNN: Regions with CNN features**

### **2.2.2 Faster R-CNN**

### **2.2.3 YOLOv3**

### **2.2.4 CenterNet**

## **2.3 视频目标检测**

## **2.4 后处理策略**

## 第三章 基于视频上下文信息的后处理框架

### 3.1 问题分析

### 3.2 基于视频上下文信息的后处理策略

### 3.3 实验与分析

### 3.4 本章小结



## 第四章 实时视频目标检测结果平滑及后处理框架

### 4.1 问题分析

### 4.2 实时视频目标检测后处理策略

### 4.3 实时视频目标检测结果的平滑处理

### 4.4 实验与分析

### 4.5 本章小结





# 第五章 基于上下文信息的后处理策略及实时后处理策略在系统中的应用

5.1 相关背景

5.2 系统需求

5.3 系统架构

5.4 系统实现

5.5 效果展示

5.6 本章小结



## 第六章 总结与展望



# 第七章 使用示例

## 7.1 示例如下

使用.bib 文件管理参考文献引用，引用示例：[1].

- 一级 item
  - 二级 item
    - ◆ 三级 item
- 一级 item

---

### 算法 7.1 算法名字

---

输入：这是输入

输出：这是输出

```
1: while flag do  
2:   这是语句  
3: end while
```

---

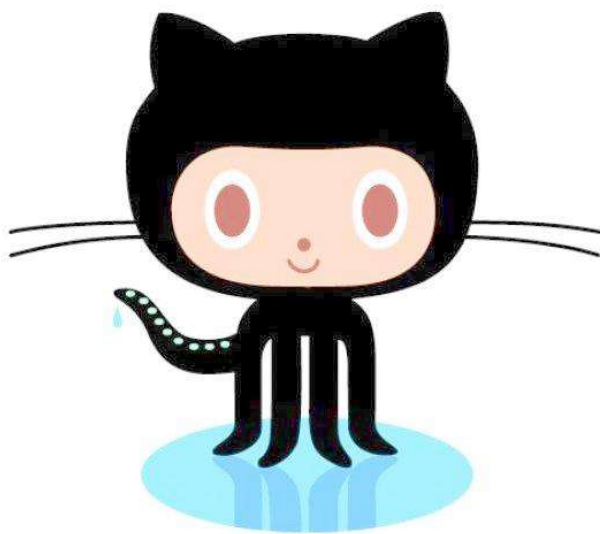


图 7-1: 单图示例

实验硬件设备如图 7-2 所示。



图 7-2: 实验硬件设备总览

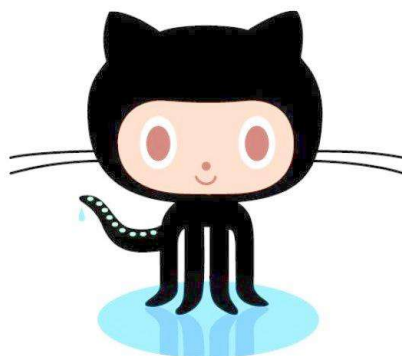


图 7-3: 实验测量示意图

图 7-4 所示子图 7-4(a) 和子图 7-4(b)。



(a) 子图



(b) 子图

图 7-4: 子图样例

## 参考文献

- [1] BONAVENTURE O, HANDLEY M, RAICIU C. An overview of Multipath TCP[J]. USENIX login;, 2012.





# 附录 A 附录代码

## A.1 main 函数

```
int main()  
{  
    return 0;  
}
```



# 致 谢

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.



# 简历与科研成果

## 基本信息

韦小宝，男，汉族，1985 年 11 月出生，江苏省扬州人。

## 教育背景

2007 年 9 月 — 2010 年 6 月	南京大学计算机科学与技术系	硕士
2003 年 9 月 — 2007 年 6 月	南京大学计算机科学与技术系	本科

## 攻读硕士学位期间完成的学术成果

1. Xiaobao Wei, Jinnan Chen, “Voting-on-Grid Clustering for Secure Localization in Wireless Sensor Networks,” in *Proc. IEEE International Conference on Communications (ICC) 2010*, May. 2010.
2. Xiaobao Wei, Shiba Mao, Jinnan Chen, “Protecting Source Location Privacy in Wireless Sensor Networks with Data Aggregation,” in *Proc. 6th International Conference on Ubiquitous Intelligence and Computing (UIC) 2009*, Oct. 2009.

## 攻读硕士学位期间参与的科研课题

1. 国家自然科学基金面上项目“问题研究”（课题年限 2010 年 1 月 — 2012 年 12 月），负责相关问题的研究。



# 版权及论文原创性说明

任何收存和保管本论文的单位和个人，未经作者本人授权，不得将本论文转借他人并复印、抄录、拍照或以任何方式传播，否则，引起有碍作者著作权益的问题，将可能承担法律责任。

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含其他个人或集体已经发表或撰写的作品成果。本文所引用的重要文献，均已在文中以明确方式标明。本声明的法律结果由本人承担。

作者签名：\_\_\_\_\_  
\_\_\_\_\_年\_\_\_\_月\_\_\_\_日





# 《学位论文出版授权书》

本人完全同意《中国优秀博硕士学位论文全文数据库出版章程》（以下简称“章程”），愿意将本人的学位论文提交“中国学术期刊（光盘版）电子杂志社”在《中国博士学位论文全文数据库》、《中国优秀硕士学位论文全文数据库》中全文发表。《中国博士学位论文全文数据库》、《中国优秀硕士学位论文全文数据库》可以以电子、网络及其他数字媒体形式公开出版，并同意编入《中国知识资源总库》，在《中国博硕士学位论文评价数据库》中使用和在互联网上传播，同意按“章程”规定享受相关权益。

作者签名：\_\_\_\_\_

\_\_\_\_\_年\_\_\_\_月\_\_\_\_日

论文题名	南京大学毕业论文 L <sup>A</sup> T <sub>E</sub> X 模板				
研究生学号	MF20330024	所在院系	计算机科学与技术系	学位年度	20xx
论文级别	<div><input type="checkbox"/> 学术学位硕士                      <input type="checkbox"/> 专业学位硕士</div> <div><input type="checkbox"/> 学术学位博士                      <input type="checkbox"/> 专业学位博士</div> <div>(请在方框内画钩)</div>				
作者 Email	sample@smail.nju.edu.cn				
导师姓名	申富饶 教授				

论文涉密情况：

☐ 不保密

☐ 保密，保密期(\_\_\_\_\_年\_\_\_\_月\_\_\_\_日至 \_\_\_\_\_年\_\_\_\_月\_\_\_\_日)

注：请将该授权书填写后装订在学位论文最后一页（南大封面）。

