南京大学 NANJING UNIVERSITY

RINC Robotic Intelligence & Neural Computing Group

基于上下文信息的视频目标检测后处理研究

答辩人：管俣祺 MF20330024

导师：申富饶 教授

日期：2023年5月22日

誠樸雄偉 勵學敦行

RINC
Robotic Intelligence & Neural Computing Group

目录

誠樸雄偉 勵學敦行

RINC
Robotic Intelligence & Neural Computing Group

第一部分

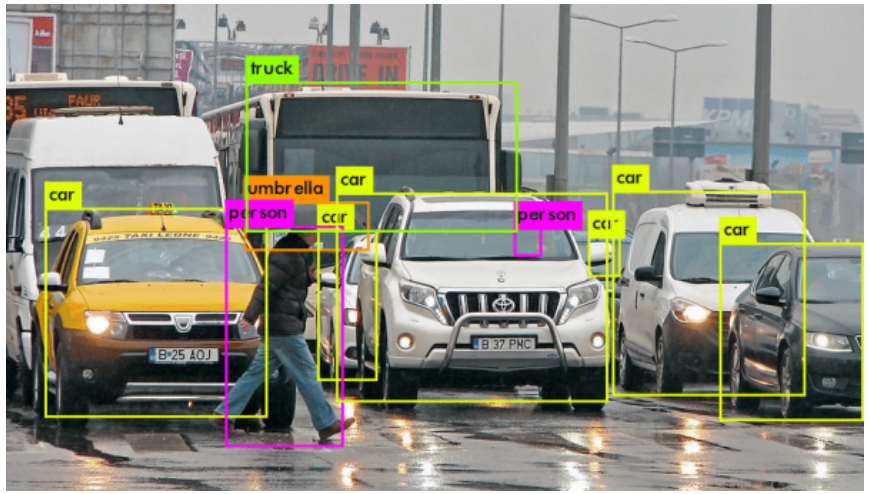
研究背景

Research Background

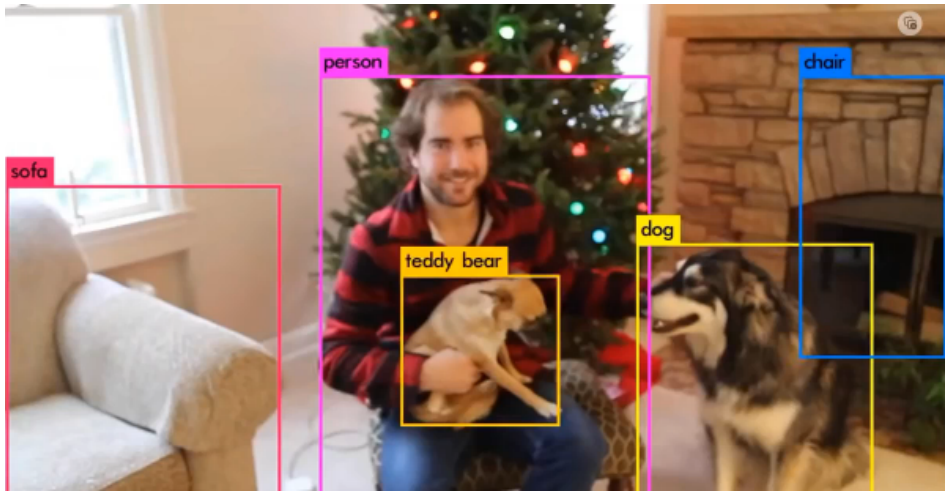
目标检测 | 后处理策略 | 研究难点与意义

誠樸雄偉 勵學敦行

1.1 图像目标检测及视频目标检测



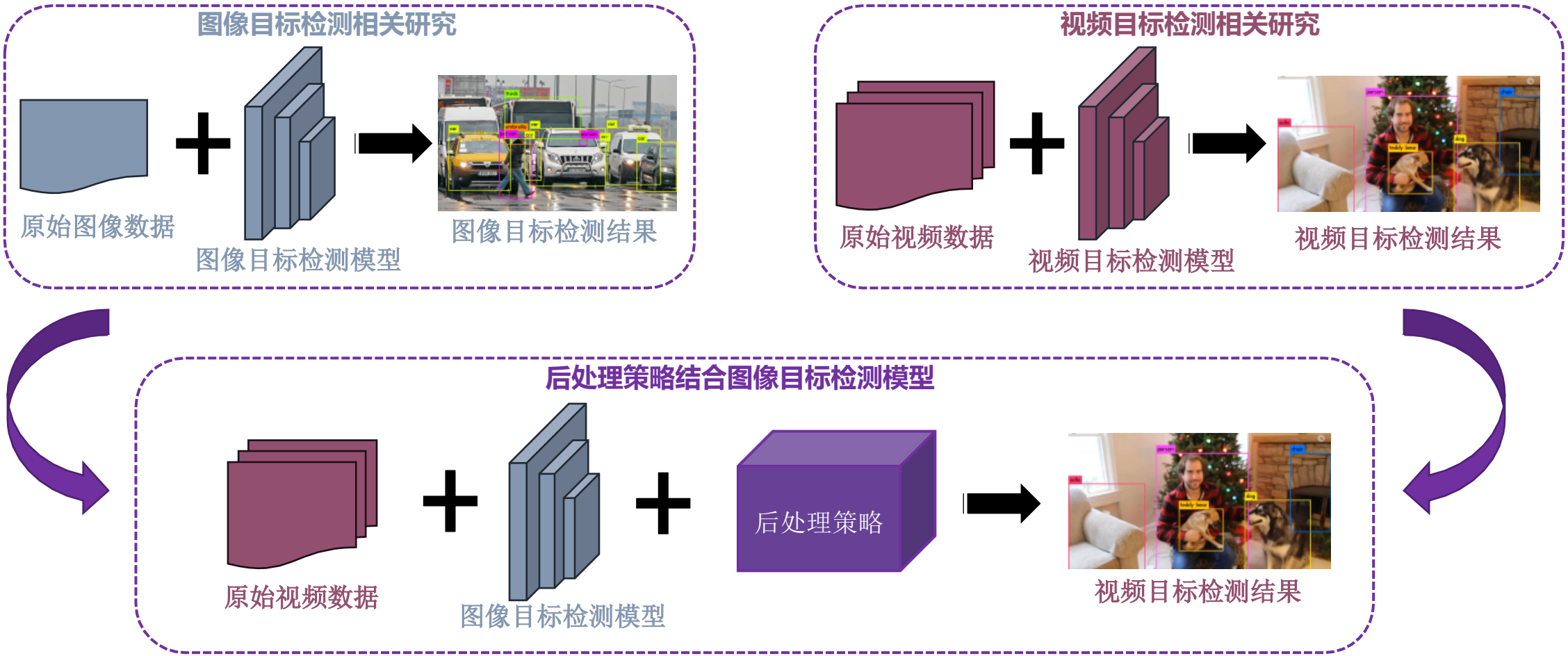
图像目标检测：面向单张图像数据



视频目标检测：面向连续视频流

共同目标：识别并定位画面中特定类别的对象

1.2 后处理策略


图像目标检测相关研究
原始图像数据
图像目标检测模型
图像目标检测结果
视频目标检测相关研究
原始视频数据
视频目标检测模型
视频目标检测结果
后处理策略结合图像目标检测模型
原始视频数据
图像目标检测模型
后处理策略
视频目标检测结果

1.3 研究难点 & 研究意义

➢ 优化目标：速度尽可能快
- 训练耗时短
- 检测耗时短

➢ 优化目标：精度尽可能高
- 召回率高
- 准确率高

现存视频目标检测模型的优劣：
✓ **检测精度高**
× 数据处理过程复杂
× 需要完成额外的计算任务，**速度慢**

直接在视频数据上使用图像检测器的优劣：
✓ **检测速度快**
× **检测精度低**
× 无法利用视频上下文信息

后处理研究意义：
发挥图像检测器的**速度优势**，弥补其无法利用视频上下文信息的缺陷，达到**与视频检测器一致的精度**

第二部分

研究内容

Research Content

- 基于视频上下文信息的非实时后处理框架
- 实时视频目标检测后处理及结果平滑框架

南京大学 NANJING UNIVERSITY

引言 两项工作的联系

后处理策略

- 非实时后处理（侧重准确性） —工作1→ CIBPP：基于视频上下文信息的非实时后处理框架
- 实时后处理（侧重实时性） —工作2→ OPP-Smooth：实时视频目标检测后处理及结果平滑框架

实时后处理 —便于用户观测结果→ 结果平滑模块（侧重结果稳定性）

2.1 基于视频上下文信息的非实时后处理框架：研究动机

研究目标：

- 发挥图像检测模型速度快的优势
- 通过后处理机制利用视频数据全局上下文信息
- 提升图像目标检测器在视频数据上的准确性

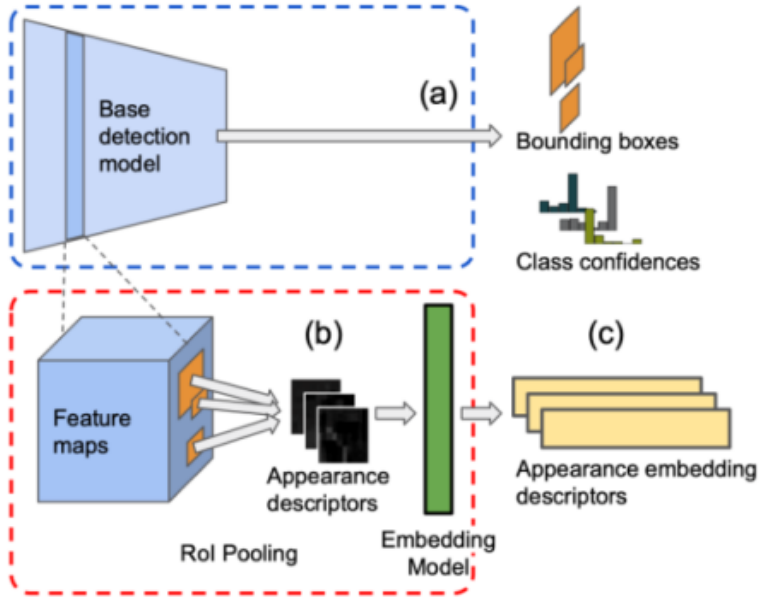
核心思路：

- 根据图像检测器在每一帧上的检测结果建立跨帧长时检测框连接
- 在每一个检测框序列内优化图像检测器的误检结果
- 在不同检测框序列间优化前序检测器的漏检结果

2.1 基于视频上下文信息的非实时后处理框架：检测框信息

对视频帧中每一个检测框，使用如下一组信息描述：

- **几何信息：** $bbox = \{x, y, w, h\}$
- **语义信息：** $cc \in \mathbb{R}^C$
- **外观信息：** $app \in \mathbb{R}^{256}$



检测框信息的形式化描述： $\{bbox, cc, app\}$

南京大学 NANJING UNIVERSITY

2.1 基于视频上下文信息的非实时后处理框架：整体结构

视频数据

基检测器

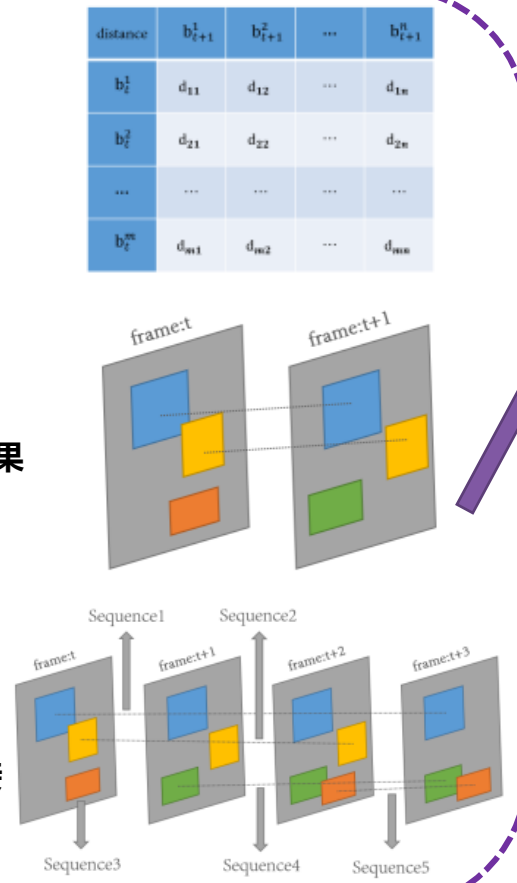
整个视频片段上形如$\{bbox, cc, app\}$的检测结果的集合

①目标检测

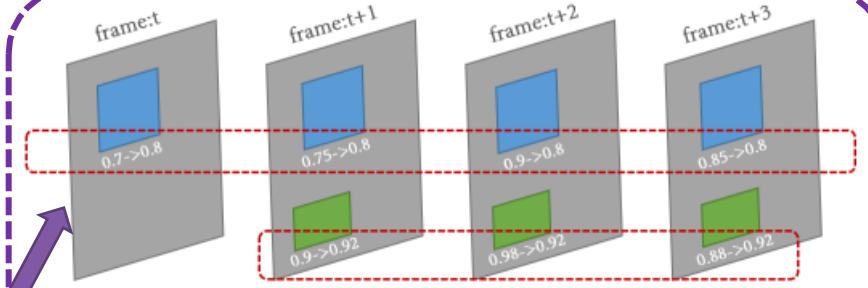
每一组相邻帧计算距离矩阵

根据距离矩阵获得相邻帧匹配结果

根据相邻帧匹配建立跨帧长时连接

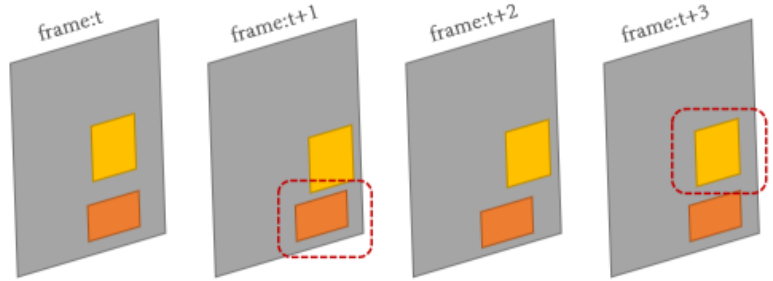


②检测框匹配及连接



在序列内实施类分数平均以优化误检

③序列内优化误检



在序列间实施双向扩散以优化漏检

④序列间优化漏检

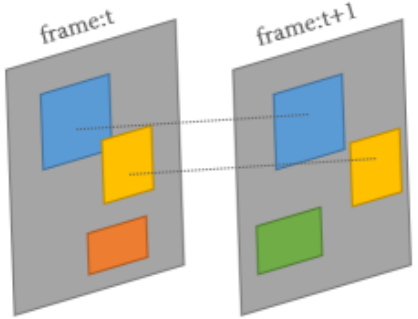
RINC Robotic Intelligence & Neural Computing Group

2.1 基于视频上下文信息的非实时后处理框架：关键模块

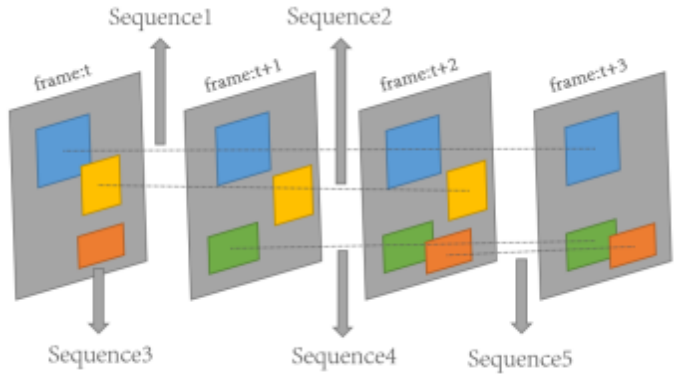
②检测框匹配及连接

distance	b_{t+1}^1	b_{t+1}^2	…	b_{t+1}^n
b_t^1	d_{11}	d_{12}	…	d_{1n}
b_t^2	d_{21}	d_{22}	…	d_{2n}
…	…	…	…	…
b_t^m	d_{m1}	d_{m2}	…	d_{mn}

每一组相邻帧计算距离矩阵



根据距离矩阵获得相邻帧匹配结果



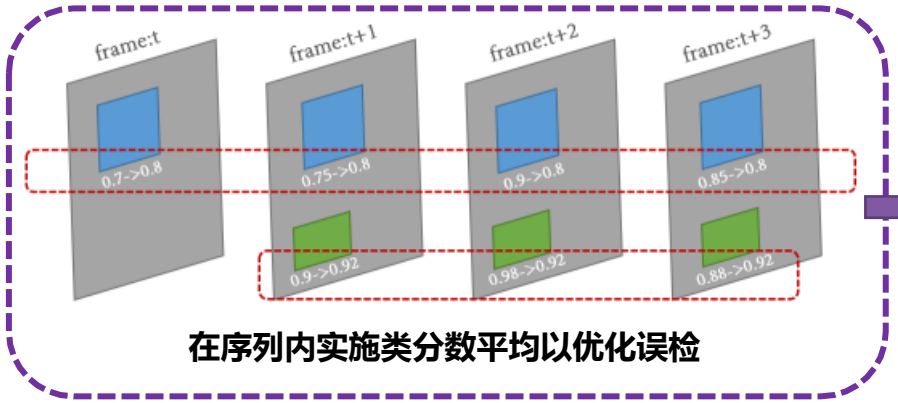
根据相邻帧匹配建立跨帧长时连接

检测框距离定义：

$$distance = \frac{1}{similarity} = \frac{1}{score_{ga} \times score_{sem}} = \frac{1}{X\left(GIoU, distance_{points}, ratio_{width}, ratio_{height}, distance_{app}\right)\left(cc_i \cdot cc_j\right)}$$

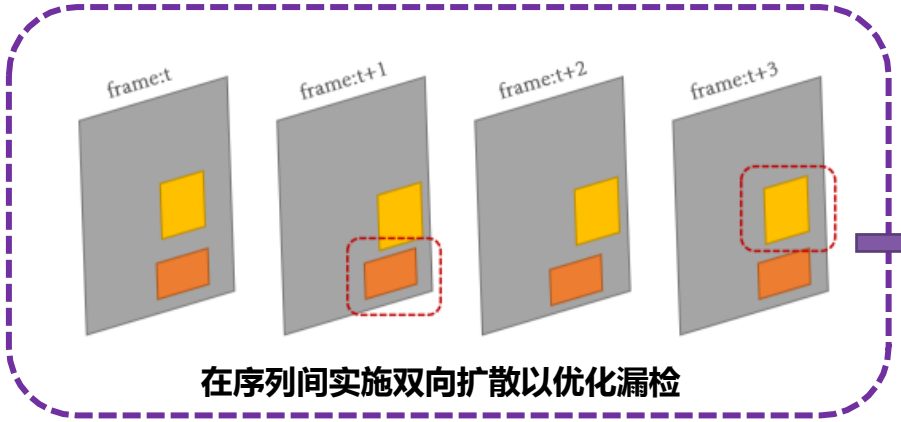
南京大学 NANJING UNIVERSITY

2.1 基于视频上下文信息的非实时后处理框架：关键模块



在序列内实施类分数平均以优化误检

③序列内优化误检



在序列间实施双向扩散以优化漏检

④序列间优化漏检

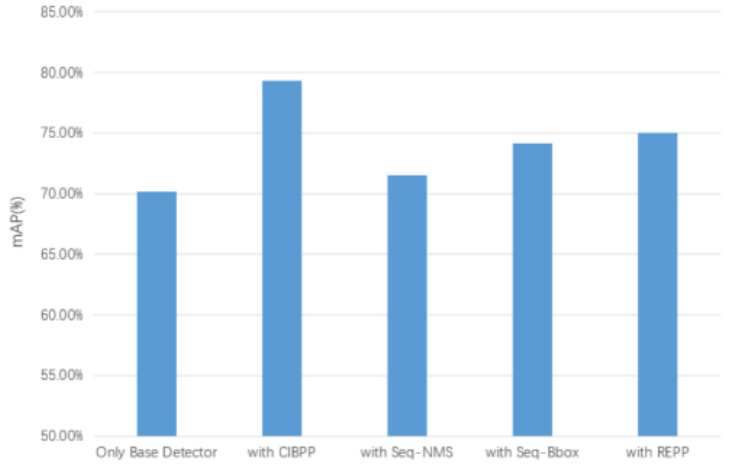
2.1 基于视频上下文信息的非实时后处理框架：实验结果

表 3-3: CIBPP 在不同基检测器上的实验结果

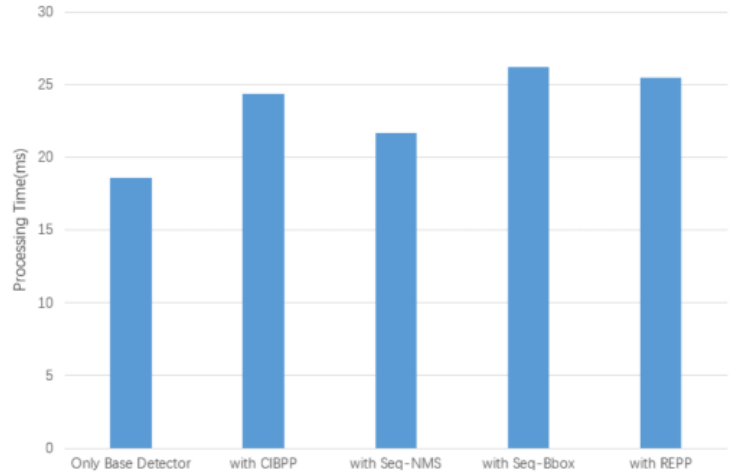
Method	Base Detector	Backbone	mAP	ProcessingTime
YOLOv3	YOLOv3	Darknet-53	70.21%	18.57
YOLOv3+CIBPP	YOLOv3	Darknet-53	79.36%	24.33
FGFA	R-FCN	ResNet-101	75.93%	91.20
FGFA+CIBPP	R-FCN	ResNet-101	82.17%	98.34
SELSA	Faster R-CNN	ResNet-101	82.01%	133.16
SELSA+CIBPP	Faster R-CNN	ResNet-101	85.69%	143.47
MEGA	Faster R-CNN	ResNeXt101	83.94%	164.58
MEGA+CIBPP	Faster R-CNN	ResNeXt101	86.21%	175.32

表 3-4: CIBPP 与其他后处理方案的结果对比

Base Detector	Post Processing Method	mAP	ProcessingTime
YOLOv3	-	70.21%	18.57
YOLOv3	CIBPP	79.36%	24.33(18.57+5.76)
YOLOv3	Seq-NMS	71.51%	21.68(18.57+3.11)
YOLOv3	Seq-Bbox	74.19%	26.19(18.57+7.62)
YOLOv3	REPP	75.06%	25.46(18.57+6.89)



(a) mAP 对比



(b) 单帧图片处理时长对比

与多个模型及后处理方案对比，我们提出的后处理方案基本都取得了更好的效果

2.2 实时视频目标检测后处理及结果平滑框架：研究动机

研究目标：

- 发挥图像检测模型速度快的优势
- 通过后处理机制利用视频数据局部上下文信息
- 在提升准确性的同时保证实时性

核心思路：

- 根据当前帧及前序帧的信息建立跨帧长时检测框连接
- 在每一个检测框序列内优化图像检测器的误检结果
- 利用卡尔曼滤波提升检测框的稳定性

2.2 实时视频目标检测后处理及结果平滑框架：卡尔曼滤波

线性高斯系统 状态方程 & 观测方程

$$x_k = Ax_{k-1} + Bu_k + q_k$$

$$s.t. q_k \sim N(0, Q)$$

$$y_k = Cx_k + r_k$$

$$s.t. r_k \sim N(0, R)$$

卡尔曼滤波 预测方程 & 更新方程

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1}$$

$$\hat{P}_k^- = A\hat{P}_{k-1}A^T + Q$$

$$K_k = \frac{\hat{P}_k^- C^T}{C\hat{P}_k^- C^T + R}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - C\hat{x}_k^-)$$

$$\hat{P}_k = (I - K_k C)\hat{P}_k^-$$

数学符号	含义
x_k	系统当前时刻，即 k 时刻的真实状态
x_{k-1}	系统前一时刻，即 $k-1$ 时刻的真实状态
$\hat{x}_k^-$	系统在 k 时刻的先验状态估计值，或称状态预测值
$\hat{x}_k$	系统在 k 时刻的后验状态估计值，或称状态最优估计值
z_k	系统在 k 时刻的状态观测值
u_k	系统控制量，若没有则设为 0
q_k	符合高斯分布的过程噪声
r_k	符合高斯分布的测量噪声
A	状态转移矩阵
B	可选的控制矩阵
C	状态观测矩阵
Q	过程噪声 q_k 的协方差矩阵
R	测量噪声 r_k 的协方差矩阵
$\hat{P}_k^-$	k 时刻的先验估计协方差
$\hat{P}_k$	k 时刻的最优估计协方差
K_k	卡尔曼增益

- 卡尔曼滤波适用于线性高斯系统的状态分析
- 该滤波器支持对过去、现在甚至未来状态的估计

南京大学 NANJING UNIVERSITY

2.2 实时视频目标检测后处理及结果平滑框架：整体结构

视频数据
逐帧处理

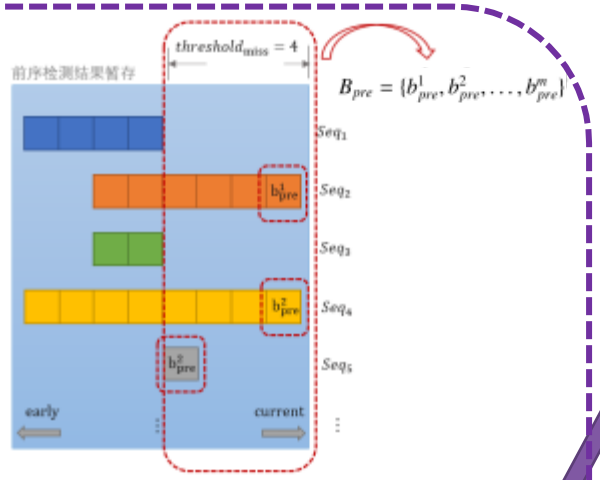
基检测器

该帧图像内
形如$\{bbox, cc, app\}$
的检测结果的集合

①目标检测

选择前序
检测结果中
参与匹配的
一组检测结果

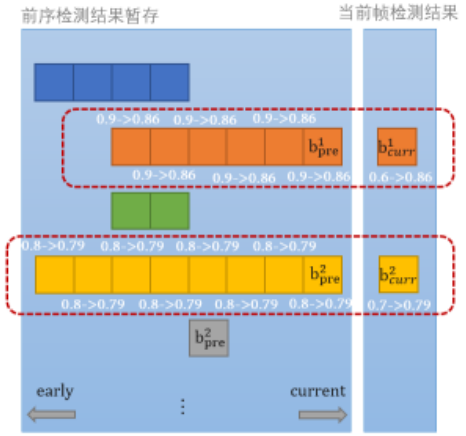
计算待匹配
检测结果与
当前帧检测结果
的距离矩阵
并获取匹配结果



distance	b^1_{curr}	b^2_{curr}	⋯	b^n_{curr}
b^1_{pre}	d_{11}	d_{12}	⋯	d_{1n}
b^2_{pre}	d_{21}	d_{22}	⋯	d_{2n}
⋯	⋯	⋯	⋯	⋯
b^m_{pre}	d_{m1}	d_{m2}	⋯	d_{mn}

②检测框匹配及连接

对于成功匹配
到现存序列的
检测结果
更新其类分数

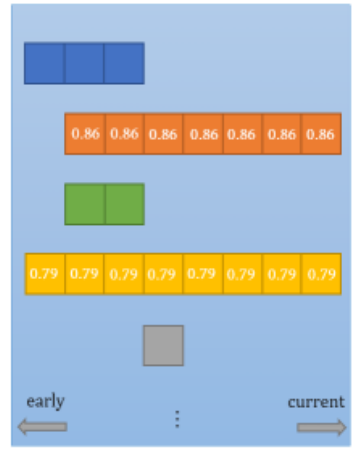


③实时类分数修正

使用卡尔曼
滤波更新当
前帧检测结
果的坐标

④实时检测结果平滑

使用当前帧
最终检测结果
刷新前序
检测结果
暂存信息

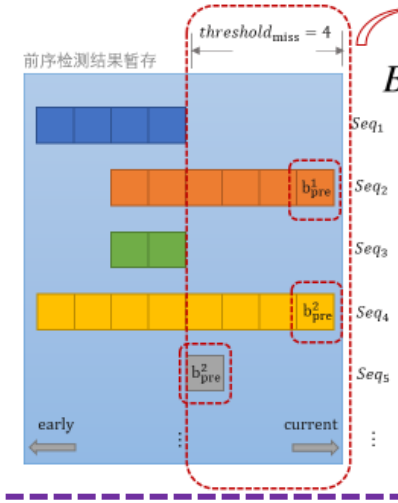


⑤刷新前序检测结果

RINC Robotic Intelligence & Neural Computing Group

2.2 实时视频目标检测后处理及结果平滑框架：关键模块

②检测框匹配及连接



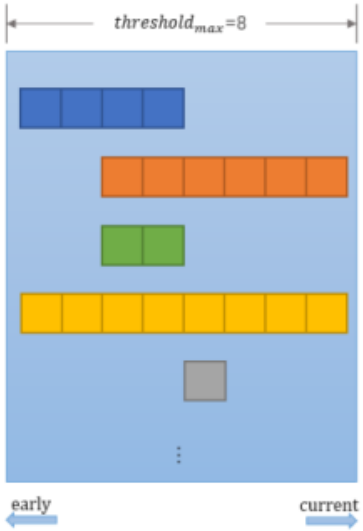
$$B_{pre} = \{b^1_{pre}, b^2_{pre}, \ldots, b^m_{pre}\}$$

选择前序检测结果中参与匹配的一组检测结果

distance	b^1_{curr}	b^2_{curr}	…	b^n_{curr}
b^1_{pre}	d_{11}	d_{12}	…	d_{1n}
b^2_{pre}	d_{21}	d_{22}	…	d_{2n}
…	…	…	…	…
b^m_{pre}	d_{m1}	d_{m2}	…	d_{mn}

计算待匹配检测结果与当前帧检测结果的距离矩阵并获取匹配结果

前序检测结果暂存机制：



	序列长度	未出现长度
Seq_1	4	4
Seq_2	6	0
Seq_3	2	4
Seq_4	8	0
Seq_5	1	3
…	…	…

检测框距离定义：

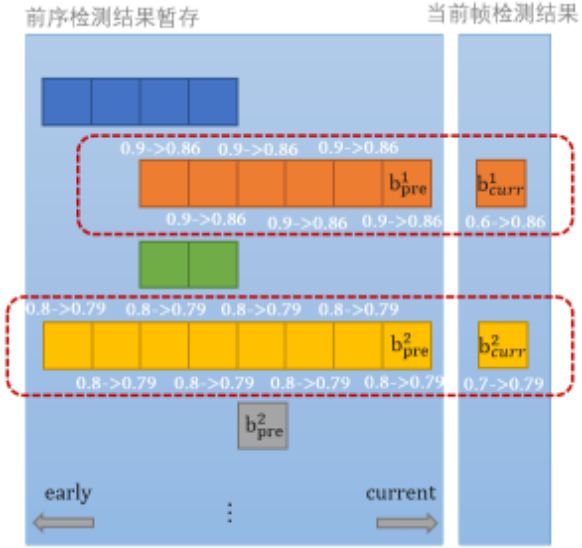
$$distance = \frac{1}{similarity} = \frac{1}{score_{ga} \times score_{sem}} = \frac{1}{X\,(GIoU, distance_{points}, ratio_{width}, ratio_{height}, distance_{app})(cc_i \cdot cc_j)}$$

南京大学 NANJING UNIVERSITY

2.2 实时视频目标检测后处理及结果平滑框架：关键模块

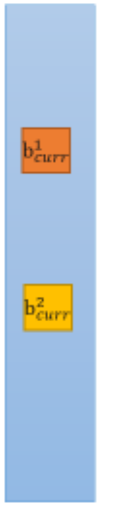
③实时类分数修正

对于成功匹配到现存序列的检测结果更新其类分数



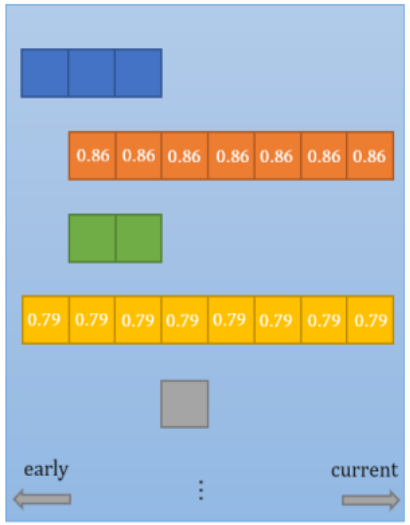
④实时检测结果平滑

使用卡尔曼滤波更新当前帧检测结果的坐标



⑤刷新前序检测结果

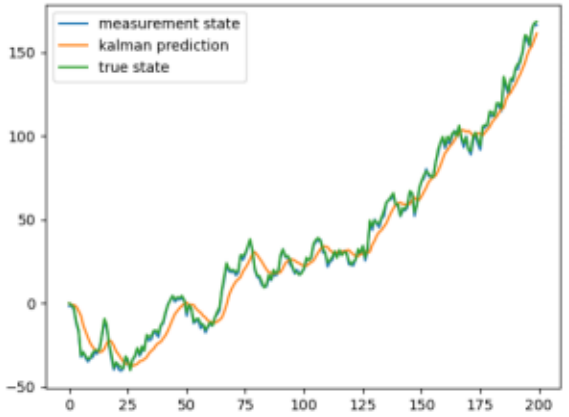
使用当前帧最终检测结果刷新前序检测结果暂存信息



类分数更新方式：

$$b.cc = \frac{1}{Seq.len + 1} b.cc + \frac{Seq.len}{Seq.len + 1} Seq.cc$$

检测结果平滑的工作原理：



RINC Robotic Intelligence & Neural Computing Group

2.2 实时视频目标检测后处理及结果平滑框架：有效性实验

表 4-3: OPP-Smooth 在不同基检测器上的实验结果

Method	Base Detector	Backbone	mAP	ProcessingTime
YOLOv3	YOLOv3	Darknet-53	70.21%	18.57
YOLOv3+ OPPSmooth	YOLOv3	Darknet-53	75.41%	20.76
FGFA	R-FCN	ResNet-101	75.93%	91.20
FGFA+ OPPSmooth	R-FCN	ResNet-101	78.8%	94.72
SELSA	Faster R-CNN	ResNet-101	82.01%	133.16
SELSA+ OPPSmooth	Faster R-CNN	ResNet-101	84.14%	138.26
MEGA	Faster R-CNN	ResNeXt101	83.94%	164.58
MEGA+ OPPSmooth	Faster R-CNN	ResNeXt101	85.32%	169.51

表 4-5: OPP-Smooth 与其他后处理方案的结果对比

Base Detector	Post Processing Method	mAP	ProcessingTime
YOLOv3	-	70.21%	18.57
YOLOv3	CIBPP	79.36%	24.33(18.57+5.76)
YOLOv3	OPP-Smooth	75.41%	20.76(18.57+2.19)
YOLOv3	Seq-Bbox	74.19%	26.19(18.57+7.62)
YOLOv3	Seq-Bbox(online)	73.11%	23.84(18.57+5.27)

与多个模型及后处理方案对比，我们提出的后处理方案基本都取得了更好的效果

2.2 实时视频目标检测后处理及结果平滑框架：对比实验

表 4-4: CIBPP 和 OPP-Smooth 在不同基检测器上的对比

Method	mAP	ProcessingTime
YOLOv3	70.21%	18.57
YOLOv3+CIBPP	79.36%	24.33
YOLOv3+ OPPSmooth	75.41%	20.76
FGFA	75.93%	91.20
FGFA+CIBPP	82.17%	98.34
FGFA+ OPPSmooth	78.8%	94.72
SELSA	82.01%	133.16
SELSA+CIBPP	85.69%	143.47
SELSA+ OPPSmooth	84.14%	138.26
MEGA	83.94%	164.58
MEGA+CIBPP	86.21%	175.32
MEGA+ OPPSmooth	85.32%	169.51

- CIBPP为第一项工作（即非实时后处理框架）
- OPP-Smooth为第二项工作（即实时后处理框架）
- OPP-Smooth对精度的提升逊色于CIBPP
- OPP-Smooth处理时长的增幅比CIBPP小

RINC
Robotic Intelligence & Neural Computing Group

第三部分

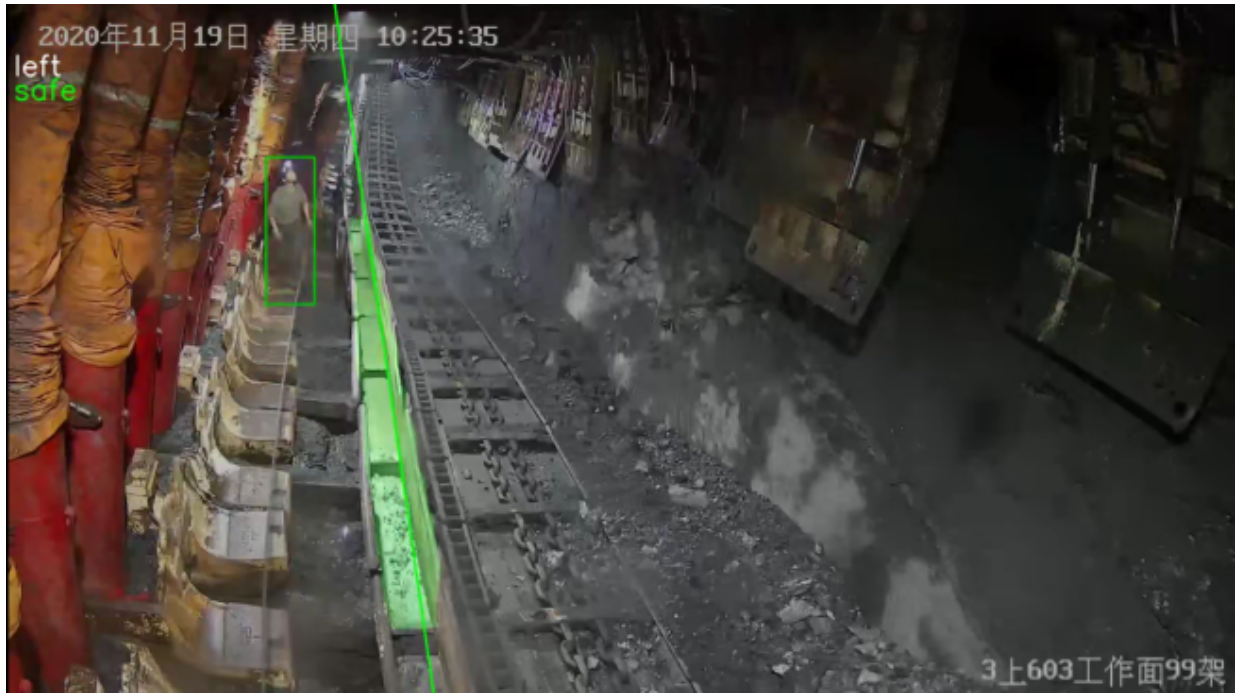
实际应用

Application

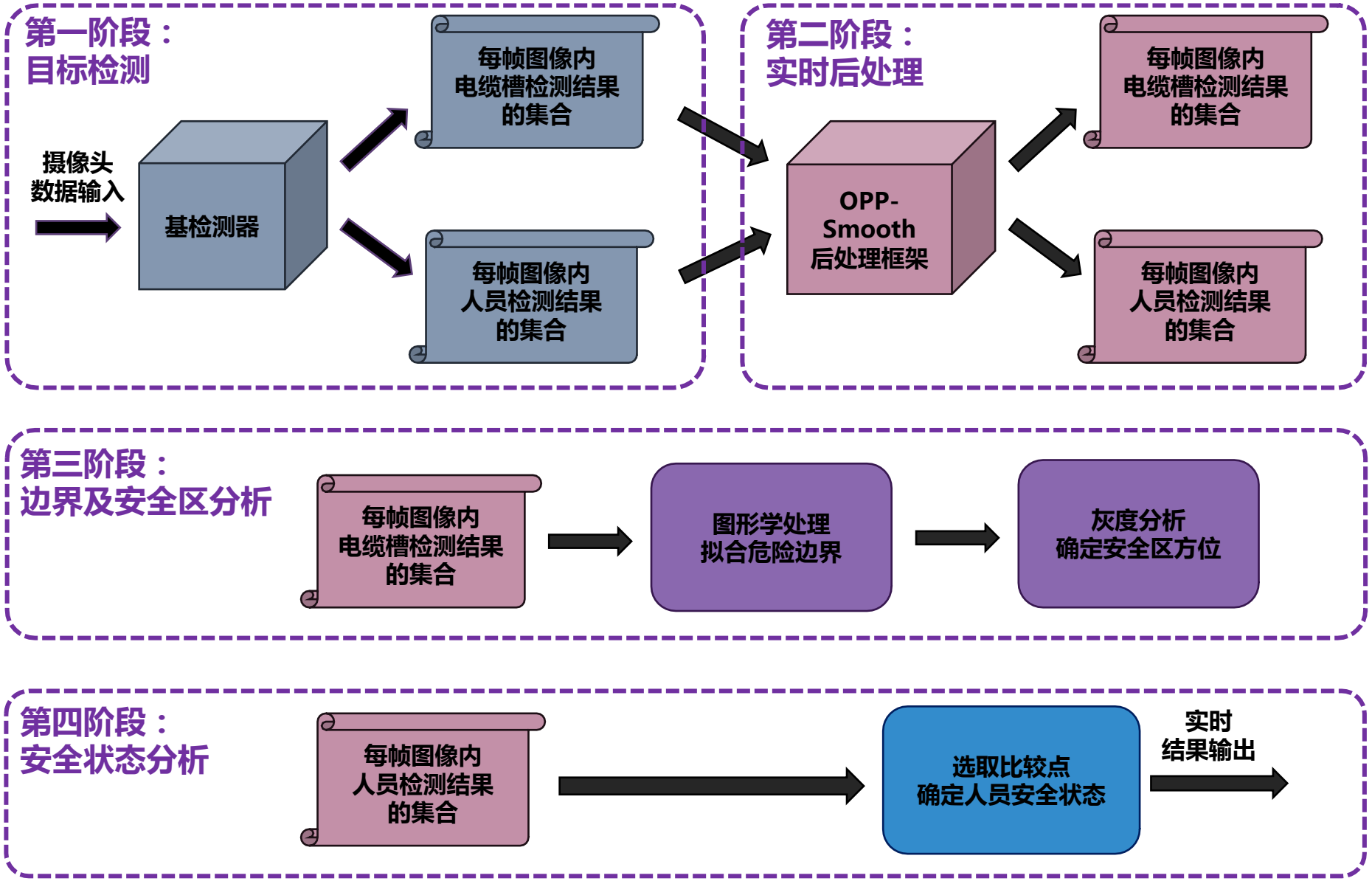
系统概述 | 系统流程 | 系统运行效果

3.1 系统概述

- 应用场景：井下煤矿作业，作为安保预警辅助工具
- 系统需求：
  - 检测工作人员位置
  - 检测电缆槽位置
  - 检测结果**后处理（本文框架的应用）**
  - 划分危险区域及安全区域
  - 判断安全状态



3.2 系统流程


第一阶段：
目标检测
摄像头
数据输入
基检测器
每帧图像内
电缆槽检测结果
的集合
每帧图像内
人员检测结果
的集合
第二阶段：
实时后处理
OPP-
Smooth
后处理框架
每帧图像内
电缆槽检测结果
的集合
每帧图像内
人员检测结果
的集合
第三阶段：
边界及安全区分析
每帧图像内
电缆槽检测结果
的集合
图形学处理
拟合危险边界
灰度分析
确定安全区方位
第四阶段：
安全状态分析
每帧图像内
人员检测结果
的集合
选取比较点
确定人员安全状态
实时
结果输出

南京大学 NANJING UNIVERSITY

3.3 系统运行效果

导入模型 等待连接摄像头

连接成功 可开始检测

开始检测 实时显示结果

RINC Robotic Intelligence & Neural Computing Group

RINC
Robotic Intelligence & Neural Computing Group

第四部分

研究生期间工作成果

Work Product

誠樸雄偉 勵學敦行

➢ 专利

申富饶，**管偲祺**，李金桥。《一种港口码头辅助障碍物过滤的快速车道线检测方法》（专利申请号 202111282391.7）

➢ 项目

国家自然科学基金面上项目 “基于深度感知增量式联想记忆神经网络的信息融合系统研究”（项目编号 61876076，课题年限 2019 年 1 月 — 2022 年 12月），本人负责目标检测相关问题的研究。

➢ 论文

管偲祺，刘雅辉，申富饶，赵健。《基于元学习和伪交互序列生成的冷启动推荐方法》（在投）

RINC
Robotic Intelligence & Neural Computing Group

第五部分

答辩总结

Summary

基于视频上下文信息的非实时后处理框架

- 形式化检测框信息
- 定义相邻帧检测框距离
- 构造跨帧长时检测框序列，利用**全局上下文**

实时视频目标检测后处理及结果平滑框架

- 引入卡尔曼滤波器
- 定义前序检测结果暂存机制
- 构造实时检测框序列，利用**局部上下文**

添加实时性约束

井下煤矿作业安全预警系统

- 开发目标为安保预警系统
- 使用**图像检测器+实时后处理**
- 结合图形学处理，结果实时显示于网页

实际应用

南京大学 NANJING UNIVERSITY

RINC Robotic Intelligence & Neural Computing Group

感谢各位老师！

答辩人：管俣祺 MF20330024

导师：申富饶 教授

日期：2023年5月22日

誠樸雄偉 勵學敦行