南京大学 NANJING UNIVERSITY

RINC Robotic Intelligence & Neural Computing Group

面向真实场景的跟踪算法应用研究

Tracking Algorithms in Real Scenarios

答辩人：MG20370046 许翔
导 师 ： 申富饶 教授

誠樸雄偉 勵學敦行

RINC
Robotic Intelligence & Neural Computing Group

目录

1 研究背景

2 研究内容
长时跟踪与分割算法 SRF
基于跟踪算法网络结构的智能标注算法 SiamAnno

3 实际应用
面向跟踪任务的图片标注工具系统

4 全文总结

5 研究生期间工作成果

誠樸雄偉 勵學敦行

第一部分

RINC
Robotic Intelligence & Neural Computing Group

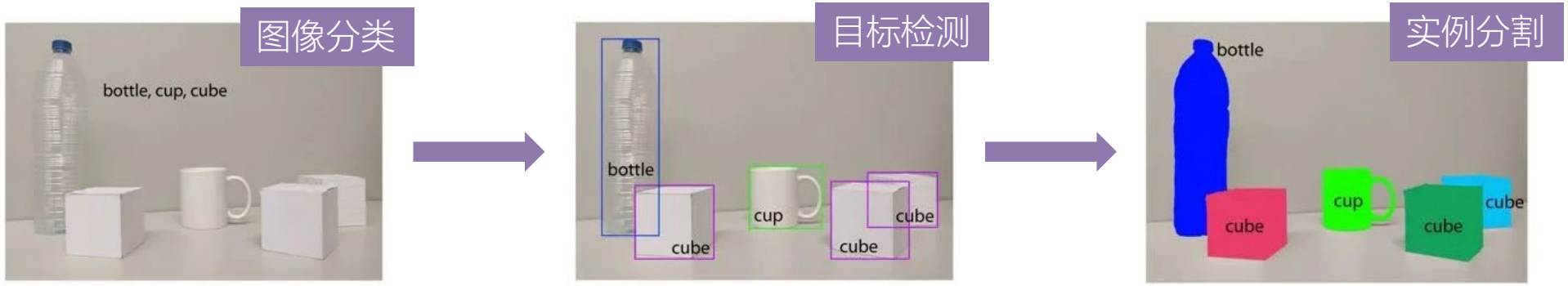
研究背景

Research Background

研究背景

计算机视觉领域的相关研究呈现两种趋势：

- 研究力度逐渐精细



- 研究主体从图像发展为视频

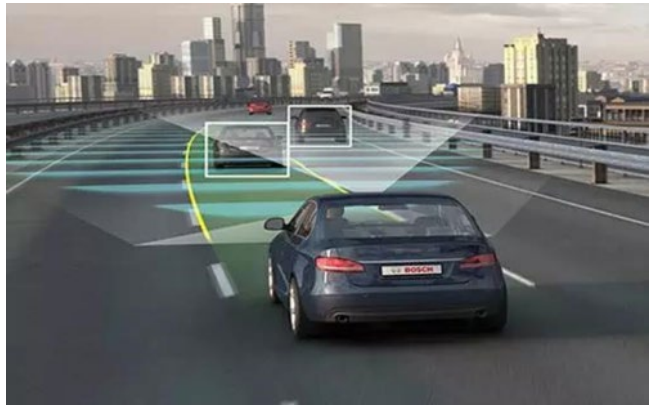
Chen K, Wang J, Yang S, et al. Optimizing video object detection via a scale-time lattice[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7814-7823.

研究意义

视频目标跟踪（Video Object Tracking, VOT）：

- 根据跟踪目标的数量，被分为单目标跟踪和多目标跟踪任务
- 根据跟踪时长的长短，被分为短时单目标跟踪和**长时单目标跟踪**
- 相关模型在多个现实任务中有巨大的应用潜力

视频监控



自动驾驶

人机交互

研究现状

长时单目标跟踪的相关研究仍有许多困难和挑战

模型

- 基于短时跟踪器的长时跟踪算法愈发臃肿，且单纯增加组件并不一定带来性能的提高
- 现有方法没有充分挖掘其中每一个组成部分的能力
- 少有具备输出像素级跟踪结果能力的跟踪器

数据

- 深度学习模型的训练依赖高质量的标注数据
- 现有跟踪数据集往往只提供包围框标注，没有像素级的标注
- 现有机器辅助标注算法没有从模型设计上考虑对新样本的学习能力
- 缺少特别考虑了单目标跟踪数据集标注需求的标注工具

本文内容

长时单目标跟踪的相关研究仍有许多困难和挑战

模型

- 基于短时跟踪器的长时跟踪算法愈发臃肿，且单纯增加组件并不一定带来性能的提高
- 现有方法没有充分挖掘其中每一个组成部分的能力
- 少有具备输出像素级跟踪结果能力的跟踪器

长时跟踪与分割算法
SRF

数据

- 深度学习模型的训练依赖高质量的标注数据
- 现有[illegible]框标注，没有像素级的标注
- 现有机器辅助算法没有从模型设计上考虑对新样本的学习能力

基于跟踪算法网络结构的智能标注算法
SiamAnno

- 缺少特别考虑了单目标跟踪数据集标注需求的标注工具

面向跟踪任务的图片标注工具系统

第二部分

RINC
Robotic Intelligence & Neural Computing Group

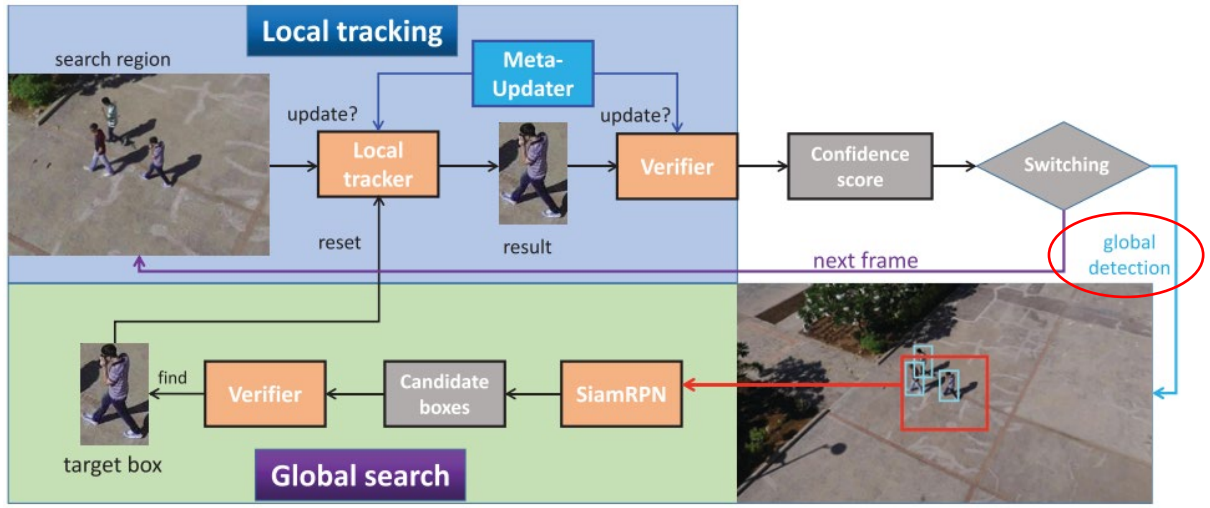
研究内容

Proposed Methods

长时跟踪与分割算法 SRF
基于跟踪算法网络结构的智能标注算法 SiamAnno

长时跟踪与分割算法：研究动机

- 后人工作往往通过对模型**“打补丁”**的方式来提高性能，使得**模型越来越臃肿**
  - 以LTMU为例，其使用了4个跟踪器，1个目标检测模型，和1个基于LSTM的元更新器



- 实验发现，**增加组件并不一定单调地提高性能**

“Many could be better than all.”

	组成	F-分数
A	SuperDiMP	0.664
B	A+SiamR-CNN	0.695
C	B+AlphaRefine	**0.705**
D	C+MetricNet	0.692
E	D+MetaUpdater	0.695

- **现有方法的能力没有被充分发挥**
- 少有可以同时完成长时目标跟踪和**视频目标分割**的方法

不引入新的网络模型，充分挖掘现有方法的潜力

设计一个简单、高效的长时跟踪与分割算法框架

长时跟踪与分割算法：整体结构

局部跟踪算法

TrDiMP

- 有足够的信心认为目标存在于画面中时使用
- 利用Transformer替代交叉相关性（cross correlation）运算

$$\min_f \left\| f * \mathrm{Dec}\left(\mathrm{Enc}(\Psi(z)), \Psi(x)\right) - y \right\|_2^2$$

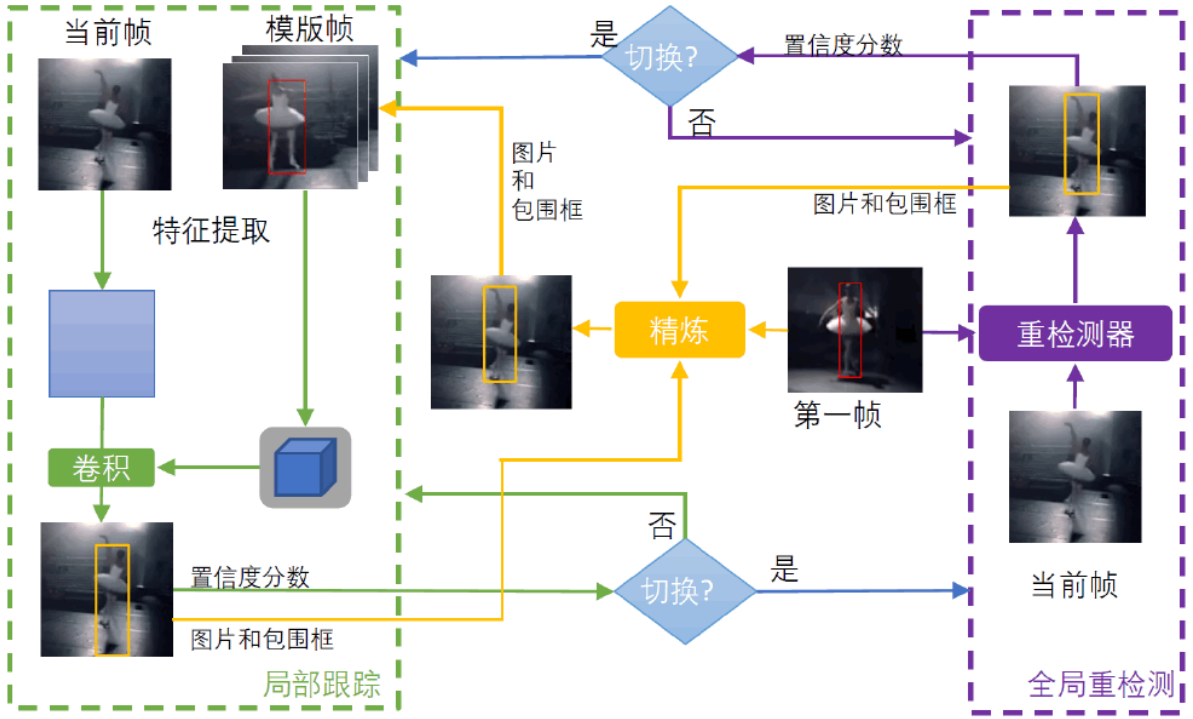


图 3-1: SRF 长时跟踪与分割算法结构图

全局重检测器

简化版SiamR-CNN

- 当目标消失时调用
- 使用RoIAlign对齐当前帧和模版帧中的目标，处理目标外观发生较大变化的情况

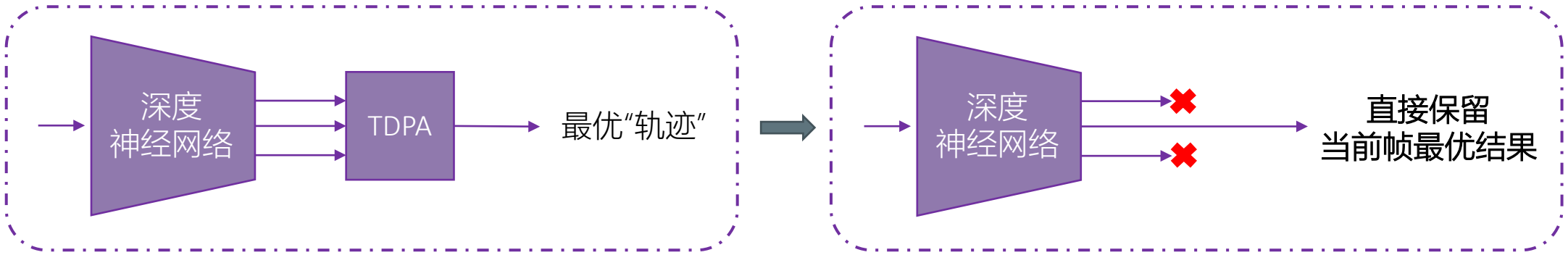
结果精炼模块

AlphaRefine

- 优化局部跟踪算法和全局重检测器的输出
- 可以输出像素级跟踪结果，支持整个框架完成视频目标分割任务

长时跟踪与分割算法：关键细节

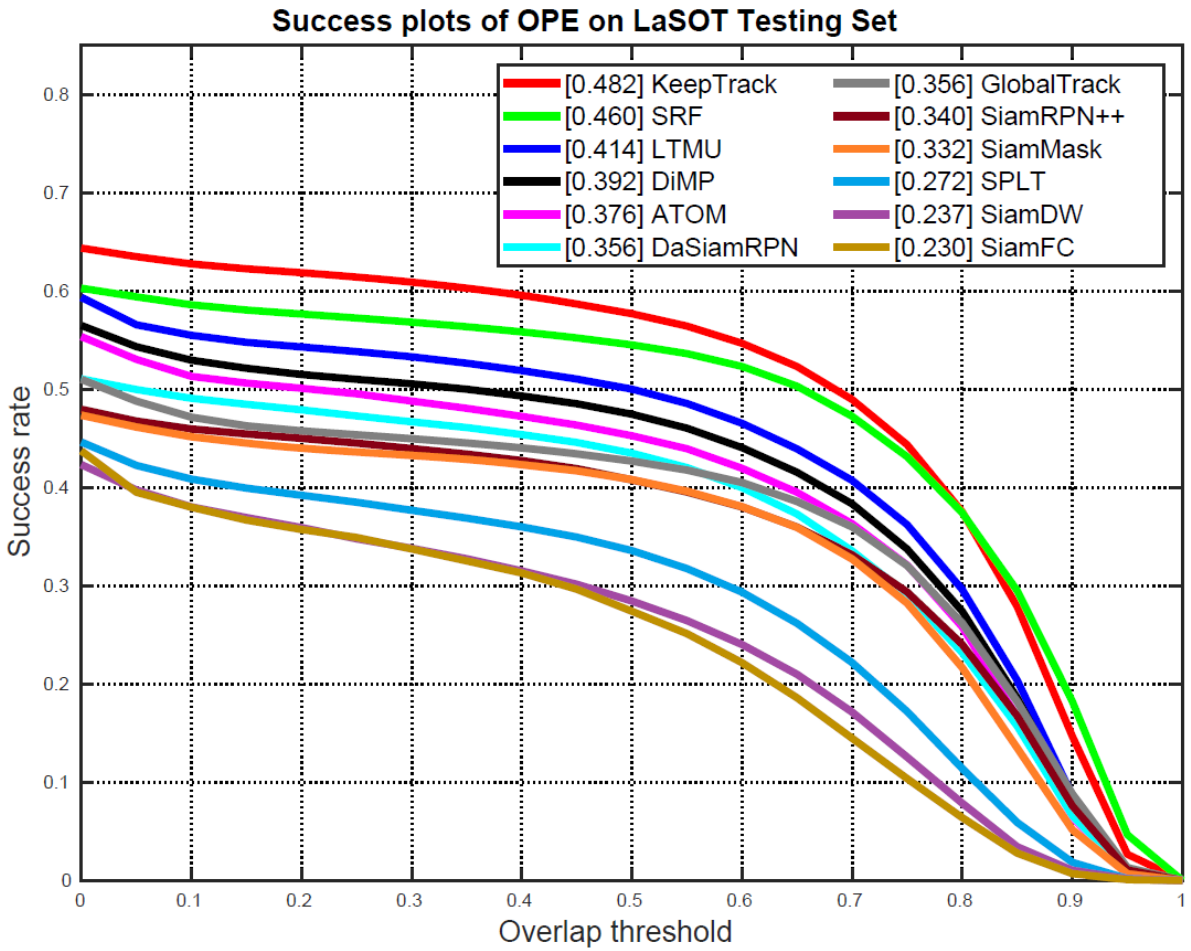
- 对SiamR-CNN的简化
  - 局部跟踪器的结果难以融合到原本的以“轨迹”为单位的TDPA最优结果匹配机制



- 时间可控的模块间切换机制
  - 局部跟踪算法速度快，但只在局部搜索；全局重检测器速度慢，但在全局范围内不遗漏目标
  - **合理地调整送入两部分的比例，可以达到调节整个框架运行速度的目的**
  - 速度控制参数SCP是一个百分位数，含义是容忍多少比例的局部跟踪器的不良结果

长时跟踪与分割算法：对比实验

- 长时目标跟踪任务：以LaSOTExtSub数据集上的结果为例



- 在追踪新物体类别的能力上，我们的方法SRF与目前最好的方法KeepTrack在**整体上表现相当。**
- 当预测包围框和真实包围框的交并比阈值高于0.8时，SRF优于Keeptrack。
- **对于一些只需要非常准确的结果的任务，SRF是一个更好的选择。**

长时跟踪与分割算法：对比实验

南京大学 NANJING UNIVERSITY

- 视频目标分割任务：以DAVIS2017数据集上的结果为例

测试时输入	训练时输入	跟踪器	$\mathcal{J}\&\mathcal{F}$	$\mathcal{J}$	$\mathcal{F}$	time
包围框	包围框	SiamR-CNN[27]	0.706	0.661	0.750	0.32
		SRF (本文方法)	0.623	0.578	0.667	0.11
		SiamMask[19]	0.558	0.543	0.585	0.02

- 速度分析：速度控制参数SCP的影响

速度控制参数（SCP）	0	0.25	0.5	0.75	1
帧每秒（FPS）	14.6	15.4	16.8	18.3	21.3
F-分数	0.708	0.707	0.706	0.703	0.697

跟踪器	LTMU [21]	Global Track [71]	Siam R-CNN [27]	Keep-Track [25]	Xuan 等人的方法 [57]	SRF 本文方法
帧每秒（FPS）	13	6	4.7	12.7	3.8	**14.6 (21.3)**
设备	2080Ti	TitanX	V100	2080Ti	2080Ti	TitanXp

- **在精度和速度之间取得了平衡**

- SCP越接近1，尽管会有更多的局部跟踪器跟踪效果不佳的帧不被送入到全局重检测器中，但仍具有较高准确率
- 据我们了解，**SRF是第一个速度可以连续调节的长时目标跟踪器**

RINC Robotic Intelligence & Neural Computing Group

长时跟踪与分割算法：消融实验

南京大学 NANJING UNIVERSITY

- 消融实验结果：以VOT2019-LT数据集上的结果为例

SiamR-CNN	TrDiMP	结果精炼模块	F-分数	跟踪查准率	跟踪查全率
✓(简化)			0.656	0.652	0.659
✓(完整)			0.663	0.658	0.669
	✓		0.653	0.673	0.633
✓(简化)		✓	0.661	0.658	0.665
✓(完整)		✓	0.669	0.664	0.675
	✓	✓	0.674	0.692	0.658
✓(简化)	✓		0.692	0.699	0.685
✓(简化)	✓	✓	**0.707**	**0.717**	**0.696**

- 简化后的SiamR-CNN精度略有损失
- SiamR-CNN的跟踪查全率高，TrDiMP的跟踪查准率高。结合使用**二者可以优势互补**

- 超参数敏感性分析：全局重检测器连续检测帧数K、置信度阈值θ对跟踪结果的影响
  - 当全局重检测器连续K帧检测到跟踪目标置信度分数大于阈值θ时，交回给局部跟踪器
  - 方法**对两个参数的敏感性不高**

K	1	2	3	4	6
F-分数	0.7082	**0.7092**	0.7088	0.7082	0.7073

θ	0.5	0.6	0.7	0.8
F-分数	0.7074	0.7088	**0.7092**	0.7056

RINC Robotic Intelligence & Neural Computing Group

长时跟踪与分割算法：可视化展示

南京大学 NANJING UNIVERSITY

绿框是我们的方法，蓝框与红框是其他既有方法

视频中存在相似物体

目标小，曾从画面中消失

目标被遮挡

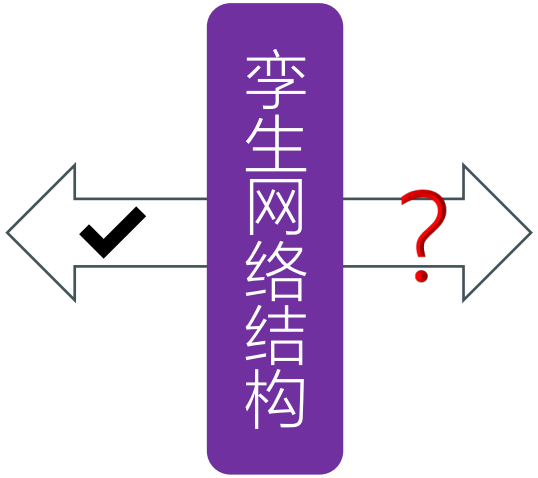
RINC Robotic Intelligence & Neural Computing Group

基于跟踪算法网络结构的智能标注算法

研究动机

目标跟踪任务

- 跟踪之前从未见过的目标物体
- 跟踪目标所处背景环境发生变化
- 基于运动连续性，跟踪目标往往出现在上一帧出现的位置附近
- 人工提供第一帧模版



智能标注任务

- 标注之前从未见过的物体类别
- 待标注物体处于新的背景中
- 人工可以较为粗糙地给出带标注物体所在位置
- 给出的位置可以看作一种“模版”

基于跟踪算法网络结构的智能标注算法：整体结构

南京大学 NANJING UNIVERSITY

❶ 基于孪生网络结构的特征提取

- 以待标注物体为中心裁剪出一个子区域作为输入

❷ 像素级相关性运算

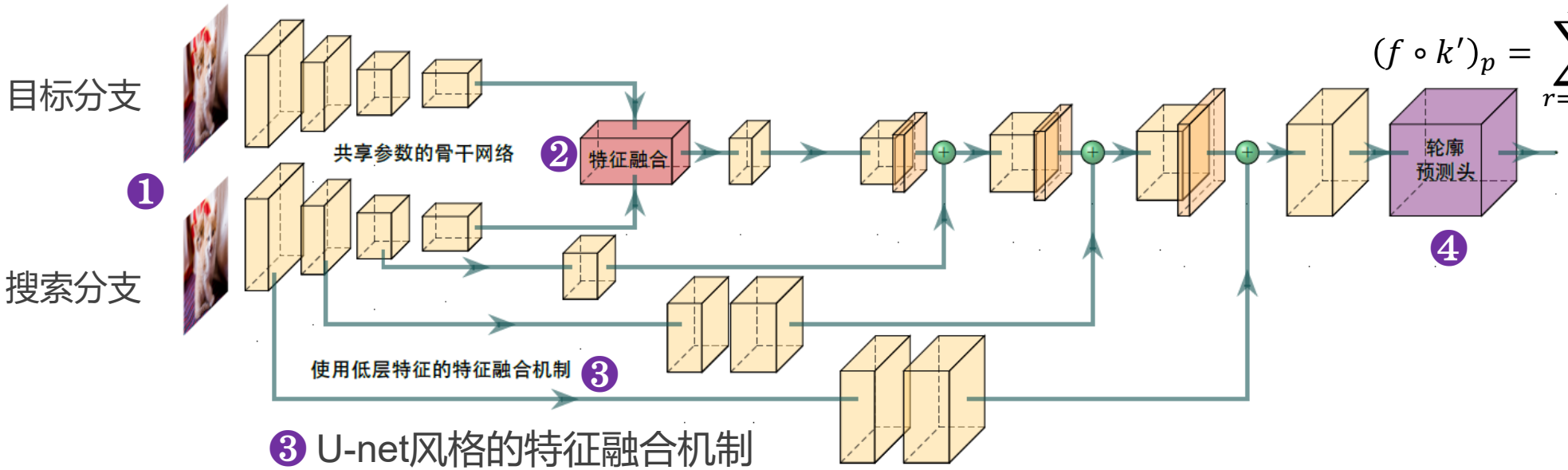
- 维持物体的空间信息

$$C = \{C_i | C_i = k_i \odot S\}$$

❹ 基于深度蛇形算法的轮廓预测头

- 从初始轮廓开始，迭代应用环形卷积（下式），不断将轮廓收缩至物体真实边界

$$(f \circ k')_p = \sum_{r=-R}^{R} f_{p+r} k'_r$$



❸ U-net风格的特征融合机制

- 高层次特征包含语义信息，低层次特征包含颜色、形状等细节信息

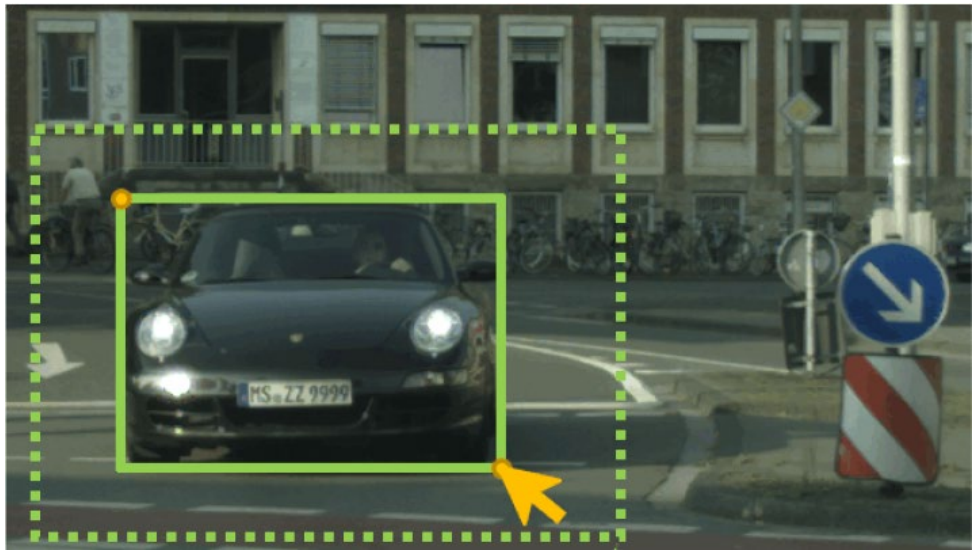
$$M_j = f_j(\text{Interpolate}(M_{j-1}) + C_j)$$

RINC Robotic Intelligence & Neural Computing Group

基于跟踪算法网络结构的智能标注算法：关键细节

- 域内（in-domain）标注任务与跨域（cross-domain）标注任务
  - 域内：标注（与模型的训练集）数据分布相同的数据集
  - 跨域：**标注新的数据集。**“新”可能是新的物体类别（domain shift）或新的环境背景（environment shift）。**跨域标注场景更符合现实需要**

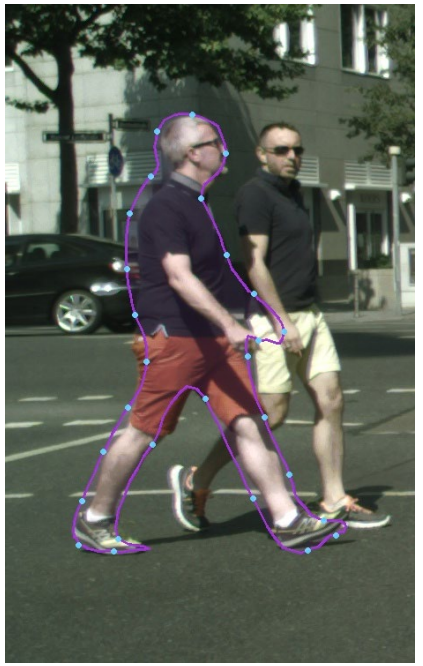
- 搜索范围参数
  - 在目标跟踪中，常定义搜索范围参数s，表示当前帧（搜索分支）的裁剪大小是模版的多少倍
  - 不同于目标跟踪，智能标注任务中的模版（目标分支的输入）是人工给定的，可以认为是绝对正确的
  - **标注任务中两个输入分支的输入来自同一张图片**
  - 跟踪任务中s常取4；在我们的算法中，$\mathbf{s} \in [\mathbf{1}, \mathbf{2}]$

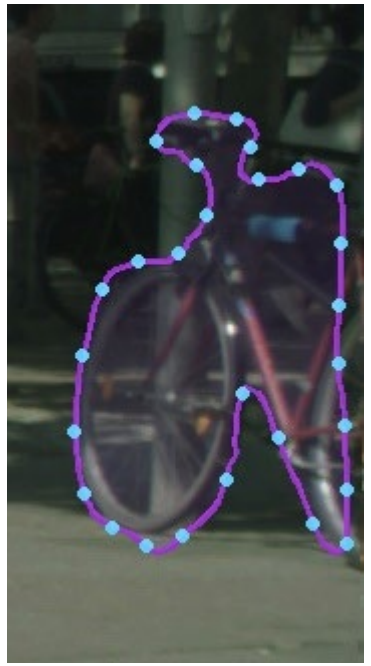


基于跟踪算法网络结构的智能标注算法：对比实验

- 域内标注任务的实验：在Cityscapes数据集上的逐类结果比较
  - 表中的平均交并比是所有类别的交并比的均值

方法	自行车	巴士	路人	火车	卡车	摩托车	乘用车	骑行者	平均交并比
Polygon-RNN	52.1	69.5	63.9	53.7	68.0	52.1	71.2	60.6	61.4
Polygon-RNN++	63.1	81.4	72.4	64.3	78.9	62.0	79.1	69.9	71.4
DACN	64.6	82.6	72.9	61.3	80.5	63.9	80.3	71.3	72.2
Polygon-GCN	64.6	85.0	72.9	61.0	79.8	63.9	81.1	71.0	72.7
PSP-DeepLab	67.2	83.8	72.6	68.8	80.5	65.9	80.5	70.0	73.7
Spline-GCN	67.4	85.4	73.7	64.4	80.2	64.9	81.9	71.7	73.7
DELSE	67.2	83.4	73.1	69.1	80.7	65.3	81.1	70.9	73.8
SiamAnno	63.9	80.6	72.1	70.3	80.1	64.0	79.4	68.2	72.3





域内标注上的效果与现有方法基本持平

基于跟踪算法网络结构的智能标注算法：对比实验

- 跨域标注任务的实验：在KITTI、ADE20k、Rooftop数据集上的结果比较（左表）
  - KITTI：街景数据集；ADE20k：通用场景数据集；Rooftop：建筑屋顶数据集
  - 模型训练仍然使用的是Cityscapes数据集
  - 表中比较的是平均交并比（mIoU）

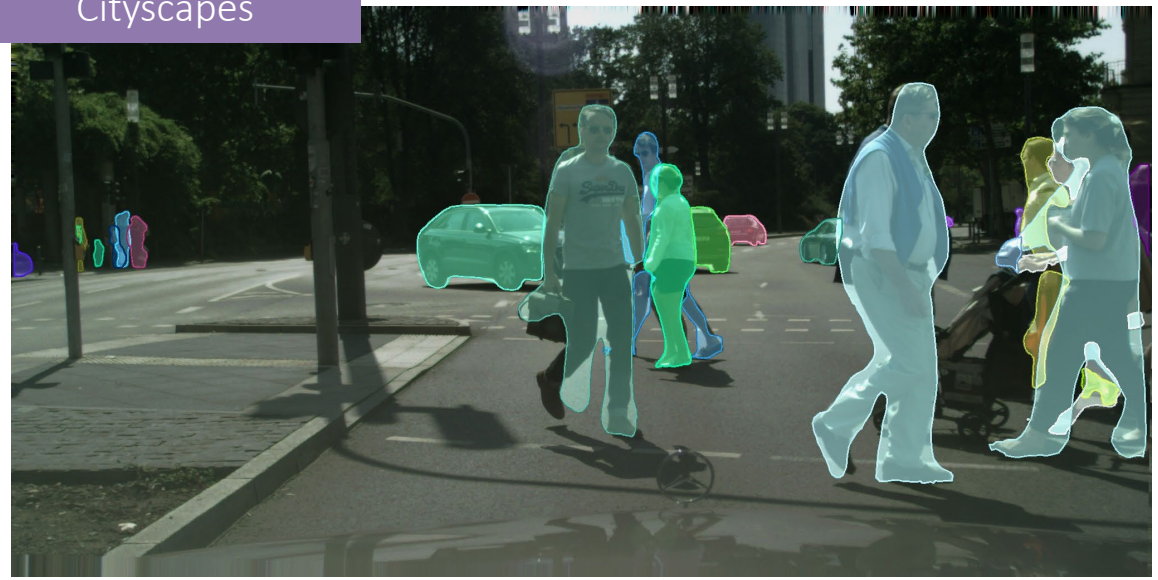
方法	KITTI	ADE20k	Rooftop
Polygon-RNN	74.22	-	-
Polygon-RNN++	83.14	71.82	65.67
PSP-Deeplab	83.35	72.70	57.91
Polygon-GCN	83.66	72.31	66.78
Spline-GCN	84.09	72.94	68.33
DACN	-	73.21	66.92
SiamAnno (本文方法)	**86.41**	**74.90**	**78.04**

在跨域任务的所有数据集上，我们的方法都取得了最好的标注准确度

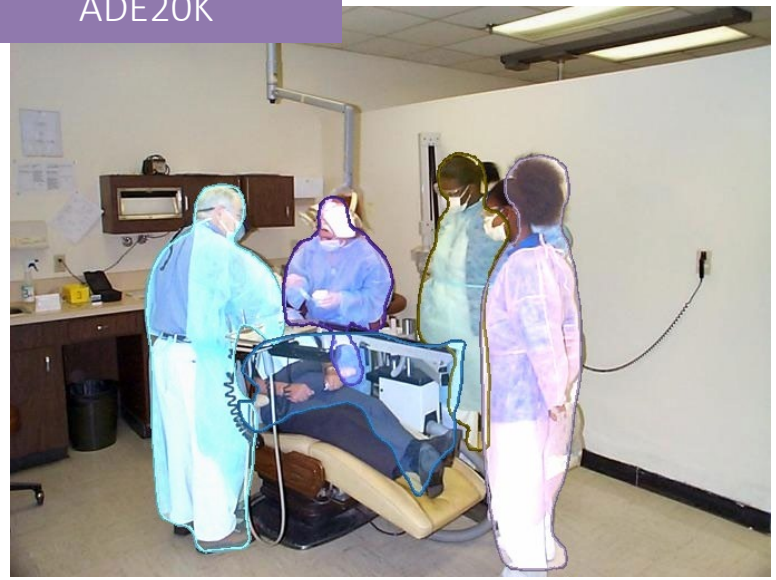
基于跟踪算法网络结构的智能标注算法：可视化展示

南京大学 NANJING UNIVERSITY

Cityscapes



ADE20K



Rooftop

KITTI

RIN Robotic Intelligence & Neural Com

第三部分

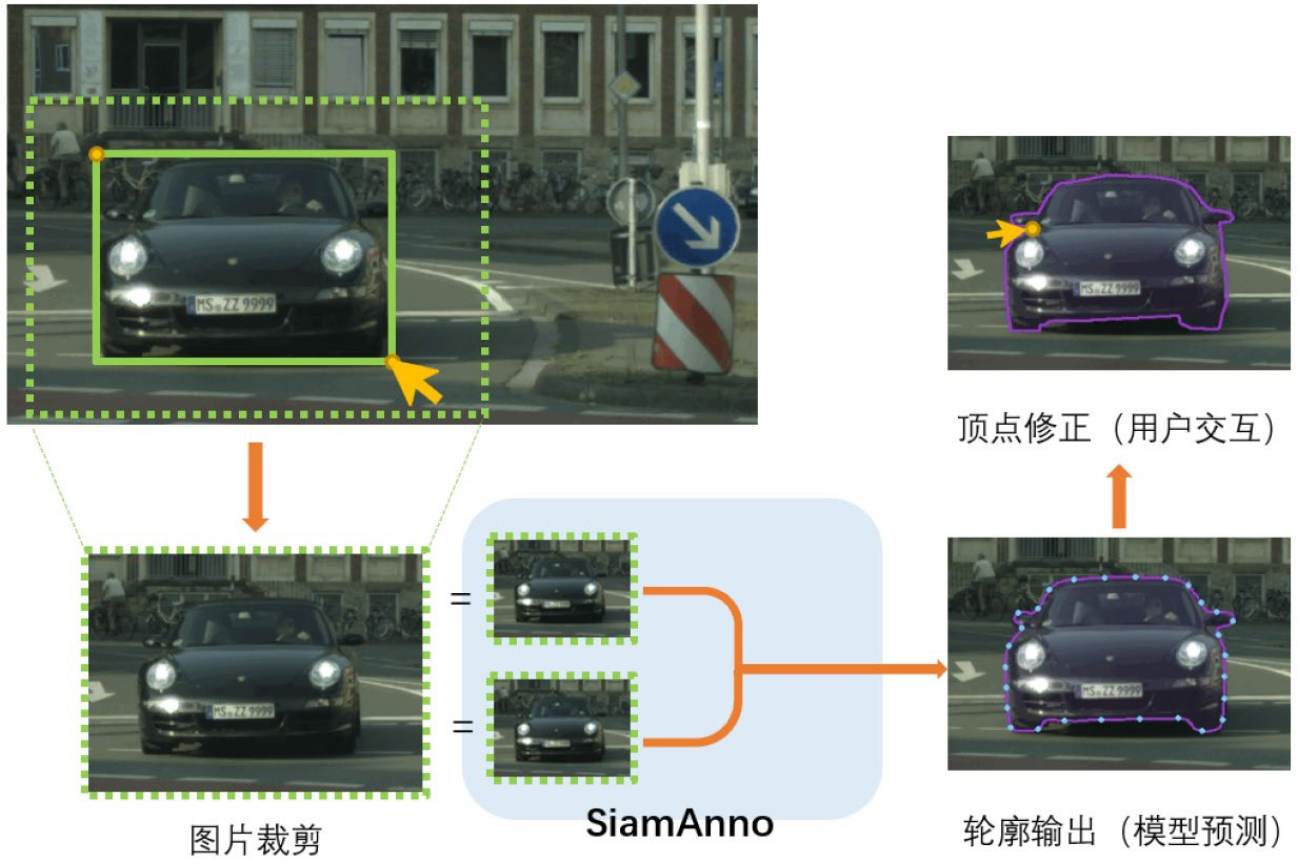
RING
Robotic Intelligence & Neural Computing Group

实际应用

Application

面向跟踪任务的图片标注工具系统

面向跟踪任务的图片标注工具系统
系统简述



一个典型的静态图片的标注过程

- 系统需求：

  针对跟踪模型的测试

  - **静态标注**：对视频的第一帧图像给出标注

  针对跟踪模型的训练

  - **动态标注**：对视频的每一帧图像给出标注

利用前文的两个方法
来支持需求的实现

面向跟踪任务的图片标注工具系统：系统功能

南京大学 NANJING UNIVERSITY

■ **智能标注**功能

用户绘制包围框，**SiamAnno算法**计算出物体边框，供用户修改

■ **连续标注**功能

SRF算法基于上一帧的标注计算物体在下一帧的标注

标注物体记录

历史动作记录

❶ 视频帧上传
❷ 标注文件下载
❸ 标注文件的上传
❹ 物体类别信息的创建与指定
❺ 包围框绘制
❻ 多边形轮廓绘制

RINC Robotic Intelligence & Neural Computing Group

面向跟踪任务的图片标注工具系统：系统效果

南京大学 NANJING UNIVERSITY

将智能标注算法应用于目标跟踪任务中

训练使用的三个数据集

LaSOT 包围框

Got10k 包围框 —SiamAnno→ 像素掩膜

Youtube-VOS 像素掩膜

在 Got10k 测试集上 SRF 结合智能标注算法完成目标跟踪任务的结果比较

用包围框训练	用像素掩膜训练	EAO	成功率
Got10k LaSOT	Youtube-VOS	84.13	93.44
LaSOT	**Got10k** Youtube-VOS	**84.47**	**94.05**

EAO：期望平均交叠（expected average overlap），每一帧预测框和实际框的交并比的均值

我们的智能标注算法对目标跟踪等下游的计算机视觉任务是有帮助的

RINC Robotic Intelligence & Neural Computing Group

第四部分

RINC
Robotic Intelligence & Neural Computing Group

全文总结

Summary

誠樸雄偉 勵學敦行

总结

南京大学 NANJING UNIVERSITY

长时跟踪与分割算法

- 不同于常见的“打补丁”的思路，对框架组件“做减法”，令整个算法简洁高效
- 可以输出像素级跟踪结果，是少有的可以同时完成长时跟踪与目标分割的算法
- 第一个速度可以连续调节的长时跟踪与目标分割算法

基于跟踪算法网络结构的智能标注算法

- 将标注人员输入的包围框自动转换成物体的多边形轮廓
- 将孪生网络结构应用于机器辅助标注任务中
- 在多个跨域标注的实验上取得了超出现有方法的轮廓预测准确度

面向跟踪任务的图片标注工具系统

- 长时跟踪与分割算法SRF为系统提供“连续标注”功能
- 智能标注算法SiamAnno为系统提供“智能标注”功能
- 实验证明算法与系统对目标跟踪等下游的计算机视觉任务是有帮助的

第五部分

RINC
Robotic Intelligence & Neural Computing Group

研究生期间工作成果

Achievement

研究生期间工作成果

南京大学 NANJING UNIVERSITY

论文

[1] **X. Xu**, J. Zhao, J. Wu and F. Shen. Switch and Refine: A Long-Term Tracking and Segmentation Framework[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(3): 1291-1304. (CCF-B，影响因子5.859)

[2]葛轶洲,**许翔**,杨锁荣等.序列数据的数据增强方法综述[J].计算机科学与探索,2021,15(07):1207-1219.

[3] **X. Xu**, R. Li, M. Yi, F. Shen and J. Zhao. Object Segmentation Annotation with SiamAnno[J]. Submitted to IEEE Transactions on Circuits and Systems for Video Technology. Under review.

项目

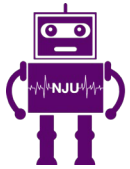
国家自然科学基金“基于深度感知增量式联想记忆神经网络的信息融合系统研究”

荣誉

2022年南京大学优秀研究生
2023届南京大学优秀毕业生

RINC Robotic Intelligence & Neural Computing Group

南京大学
NANJING UNIVERSITY



RINC
Robotic Intelligence & Neural Computing Group

谢谢！

誠樸雄偉　勵學敦行