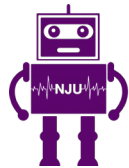




南京大學  
NANJING UNIVERSITY



RINC  
Robotic Intelligence & Neural Computing Group

# 面向实际应用场景的 图片实例标注算法研究

答辩人：李若彤 502022370022

导师：申富饶 教授

日期：2025年5月16日

誠樸雄偉 勵學敦行

## 壹 研究背景

## 貳 研究内容

- 基于交叉注意力机制的 one-shot 实例分割方法
- 基于向量夹角余弦度量辅助训练的轮廓回归算法

## 叁 实际应用

## 肆 研究生期间工作成果

## 伍 总结

# 目录

---

第一部分

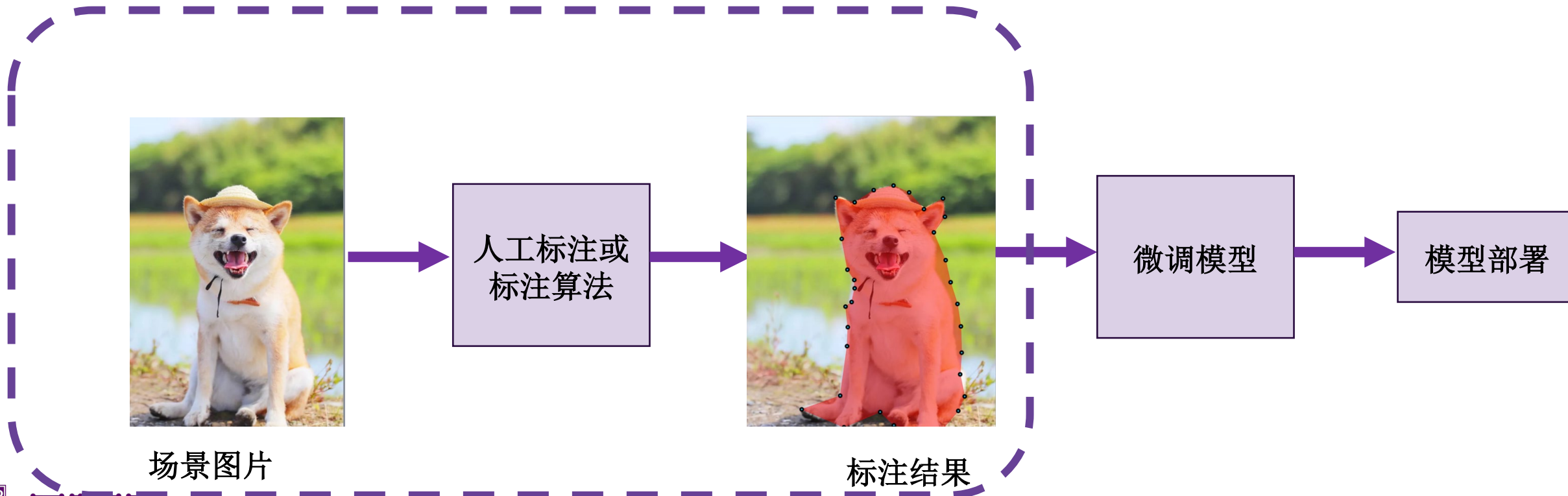
# 研究背景

Research Background

## 1.1

## 背景简介

将视觉模型应用于陌生场景时，获取场景图片的高质量标注往往是算法落地流程中遇到的首要障碍



场景图片

标注结果

## 1.2

## 研究现状

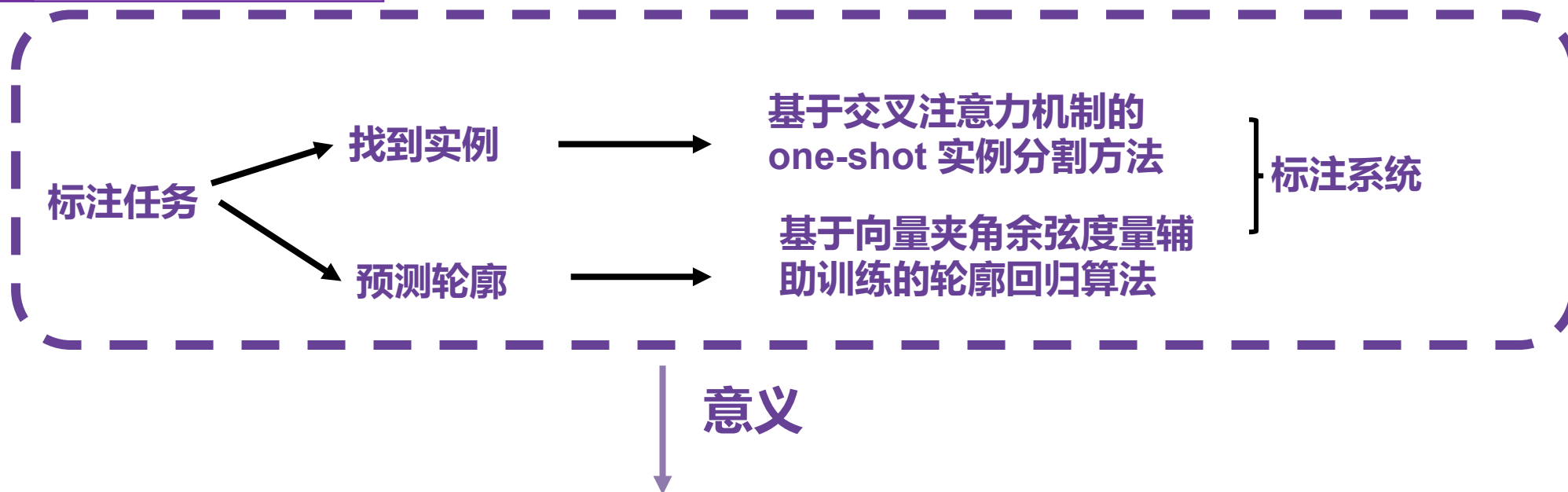
- 人工标注
- 通用实例分割方法
- 基于点击点的辅助标注方法：用户提供的少量**正负点击点**预测掩码或者轮廓
- 迭代调整多边形轮廓的辅助标注方法：调整初始**多边形轮廓**上顶点的位置预测轮廓

 挑战

- 人工标注**成本高负担重**并且专业领域知识影响**标注质量**
- 部分陌生数据直接使用实例分割方法**效果较差**
- 现有辅助标注方法**交互复杂**
- 部分方法输出掩码，标注**结果不易调整**

## 1.3

## 研究意义



- ✓ 辅助人工标注
- ✓ 快速为陌生场景中的图片数据提供准确标注
- ✓ 帮助算法快速应用于对应场景

第二部分

# 研究内容

## Research Content

基于交叉注意力机制的 one-shot

实例分割方法

基于向量夹角余弦度量辅助训练的

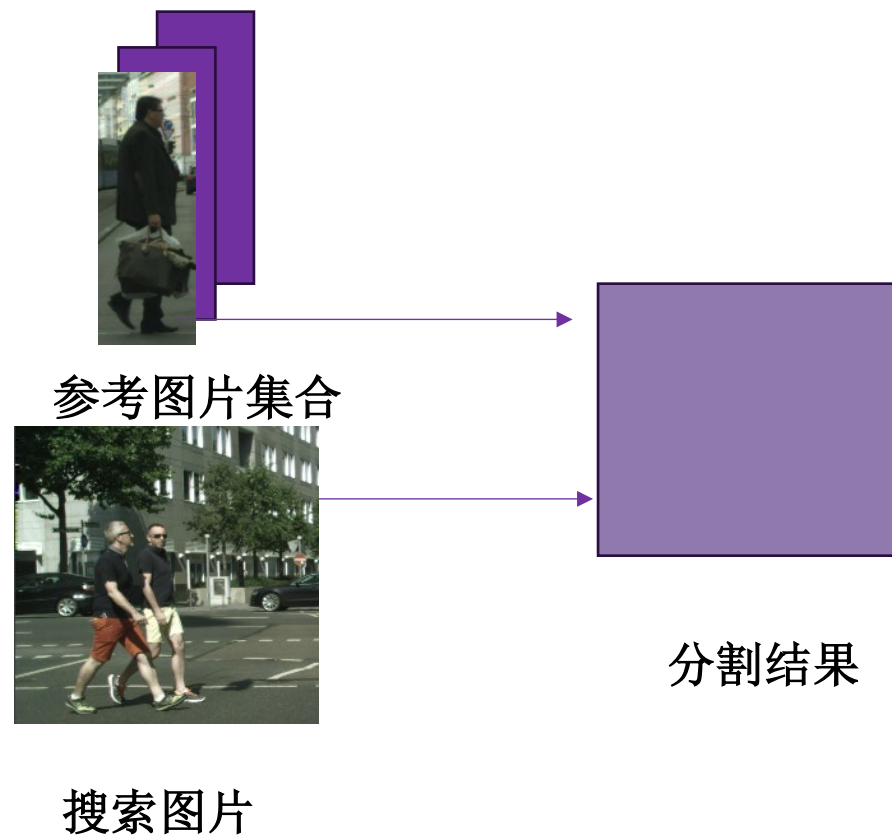
轮廓回归算法

## 2.1 基于交叉注意力机制的 one-shot 实例分割方法-研究动机

**目标：**需要给定参考图片快速在图中定位相似实例

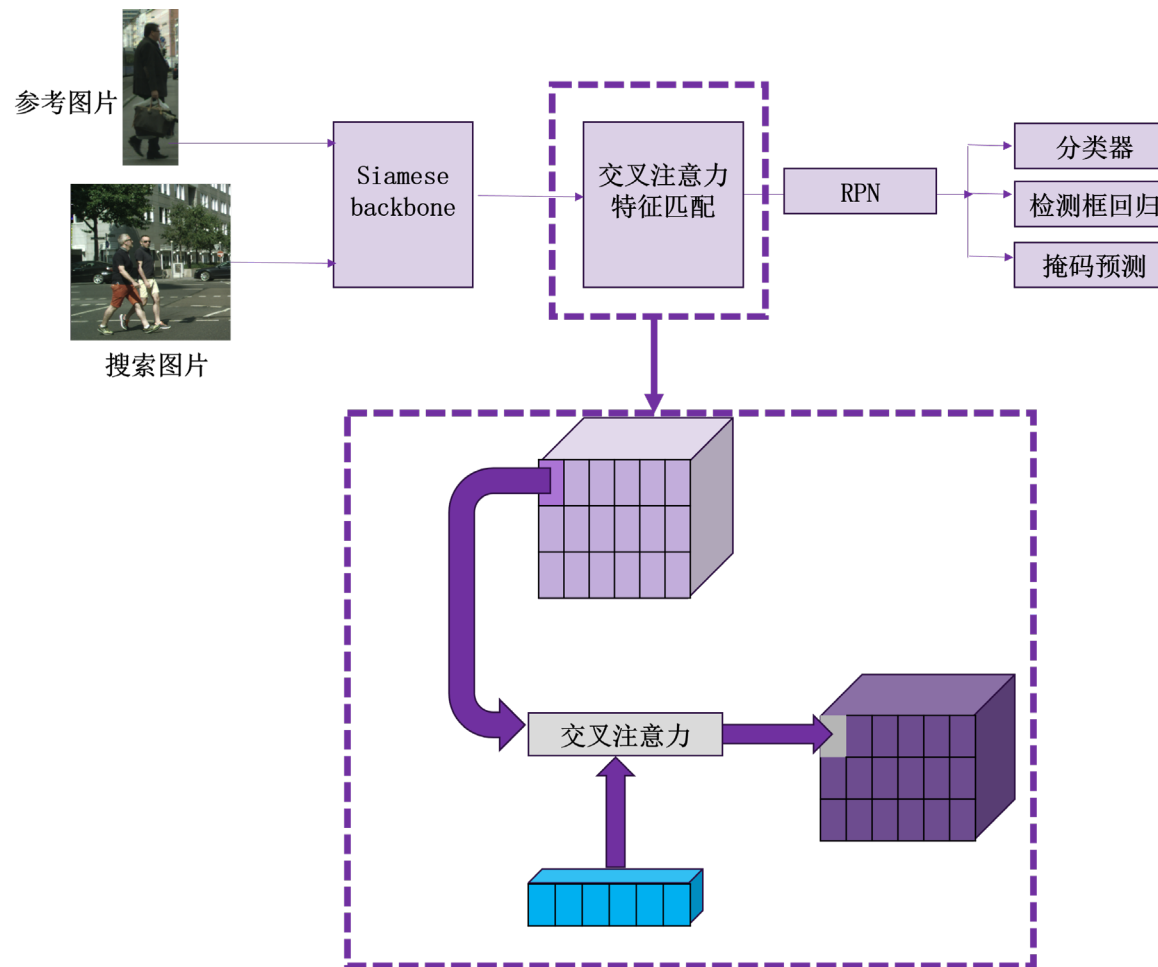
**现有方法问题：**

- 参考图片集筛选和标注质量影响模型效果
- 复杂引导机制和表示机制推理速度慢.
- 存储新类别表示内存占用大、限制模型泛化性能

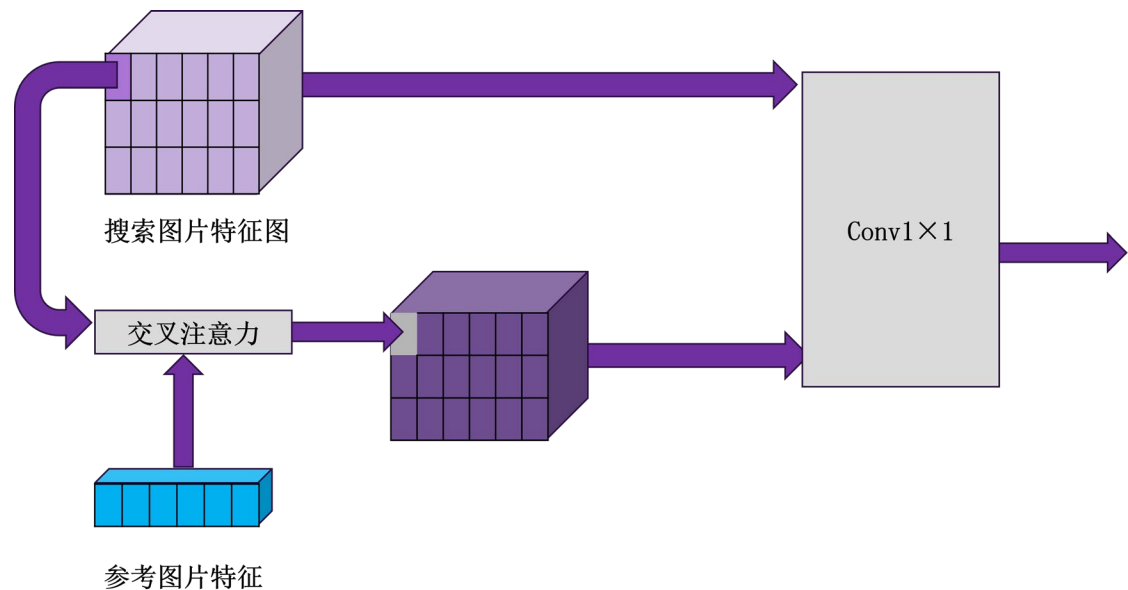


## 2.1 基于交叉注意力机制的 one-shot 实例分割方法-整体结构

- 仅通过**一张**新类别参考实例图片完成匹配，不要求提供参考示例图片标注
- **简化**参考实例和搜索图片**特征匹配**，充分匹配不同尺度的特征



## 2.1 基于交叉注意力机制的 one-shot 实例分割方法-关键部分



$$Q_{input} = GAP(feature_{ref})$$

$$K_{input} = V_{input} = feature(x, y), 0 < x < W, 0 < y < H$$

$$att = softmax\left(\frac{FC_Q(Q_{input})FC_K(K_{input})}{d}\right)$$

$$cross_{x,y} = att \cdot FC_V(V_{input})$$

- 对不同尺度的参考实例特征和搜索图片不同区域特征使用交叉注意力机制进行充分特征匹配

## 2.1 基于交叉注意力机制的 one-shot 实例分割方法：对比实验

表 3-1 Coco 数据集统计信息

划分后的数据集	$S_1$	$S_2$	$S_3$	$S_4$
图片数目	132806	72955	65508	66733
标注框数目	406602	170103	156502	177463

表 3-3 对比实验

模型	AP 检测	AP 分割
Siamese Mask R-CNN	8.6	6.7
Meta R-CNN	5.6	4.4
iMTEA	10.06	8.67
ours	9.5	6.9

表 3-2 Coco 数据集分组类别信息

$S_1$	$S_2$	$S_3$	$S_4$
Person	Bicycle	Car	Motorcycle
Airplane	Bus	Train	Truck
Boat	Traffic light	Fire Hydrant	Stop sign
Parking meter	Bench	Bird	Cat
Dog	Horse	Sheep	Cow
Elephant	Bear	Zebra	Giraffe
Backpack	Umbrella	Handbag	Tie
Suitcase	Frisbee	Skis	Snowboard
Sports ball	Kite	Baseball bat	Baseball glove
Skateboard	Surfboard	Tennis racket	Bottle
Wine glass	Cup	Fork	Knife
Spoon	Bowl	Banana	Apple
Sandwich	Orange	Broccoli	Carrot
Hot dog	Pizza	Donut	Cake
Chair	Couch	Potted plant	Bed
Dining table	Toilet	TV	Laptop
Mouse	Remote	Keyboard	Cell phone
Microwave	Oven	Toaster	Sink
Refrigerator	Book	Clock	Vase
Scissors	Teddy bear	Hair drier	Toothbrush

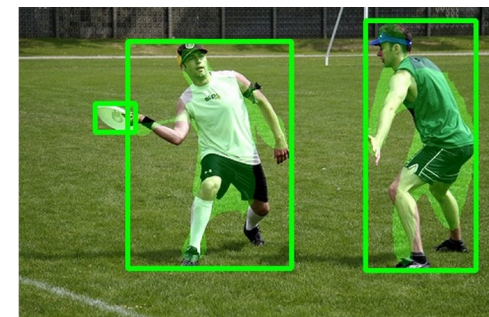
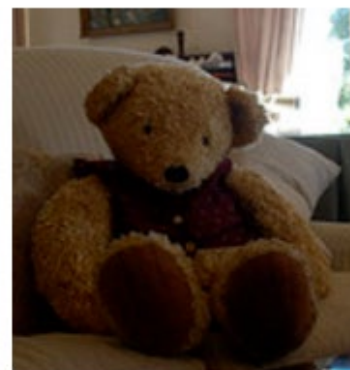
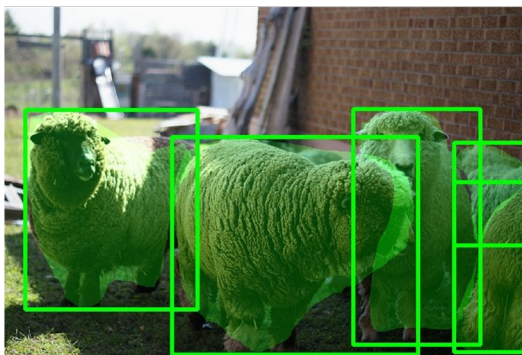
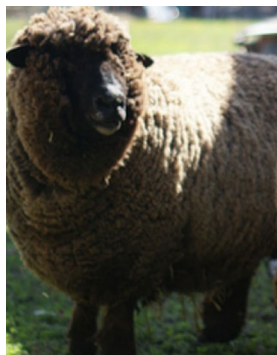
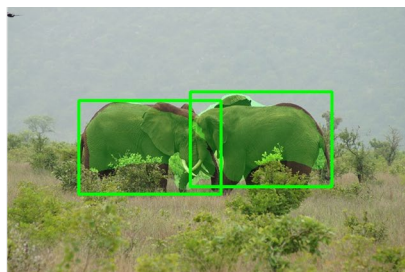
- 方法在检测和分割任务上**优于**类似的one-shot实例分割方法，在检测任务上接近在新类别上进行有策略微调的方法
- 一阶段交叉注意力机制特征匹配模块**有效**，有**更好的特征对齐效果**

表 3-4 以  $S_{1,2,3,4}$  为测试组的对比

欧式距离 cross-att	测试组 $S_1$				测试组 $S_2$				测试组 $S_3$				测试组 $S_4$			
	AP		AR		AP		AR		AP		AR		AP		AR	
$S_1$	8.0	<b>9.3</b>	10.7	<b>13.9</b>	6.0	<b>7.3</b>	5.2	<b>9.7</b>	5.1	<b>9.5</b>	9.8	<b>12.4</b>	8.3	<b>8.7</b>	8.8	<b>11.9</b>
$S_2$	6.7	<b>7.6</b>	8.9	<b>13.3</b>	6.1	<b>6.9</b>	7.6	<b>8.0</b>	3.2	<b>5.8</b>	7.1	<b>10.7</b>	6.2	<b>6.7</b>	7.0	<b>9.2</b>
$S_3$	5.9	<b>8.4</b>	10.1	<b>15.4</b>	5.8	<b>8.0</b>	9.5	<b>10.1</b>	5.1	<b>7.0</b>	12.4	<b>12.7</b>	3.9	<b>7.2</b>	9.2	<b>11.4</b>
$S_4$	7.5	7.4	9.9	<b>11.2</b>	7.1	<b>7.9</b>	8.3	<b>8.5</b>	3.7	<b>8.2</b>	11.0	<b>11.4</b>	5.0	<b>6.9</b>	9.5	<b>10.0</b>



## 2.1 基于交叉注意力机制的 one-shot 实例分割方法-结果可视化



- 在检测任务上表现良好
- 背景变化或主体被遮挡时分割效果会受到较大影响

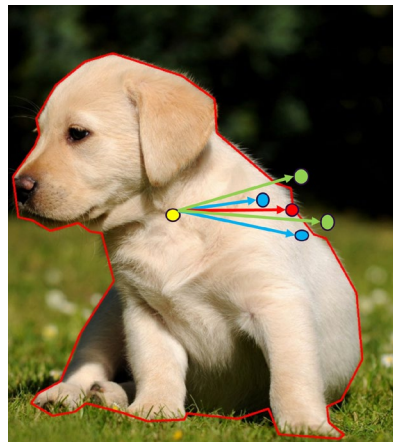
## 2.2 基于向量夹角余弦度量辅助训练的轮廓回归算法-研究动机

目标：获得便于人工交互的实例多边形轮廓

### 现有方法问题



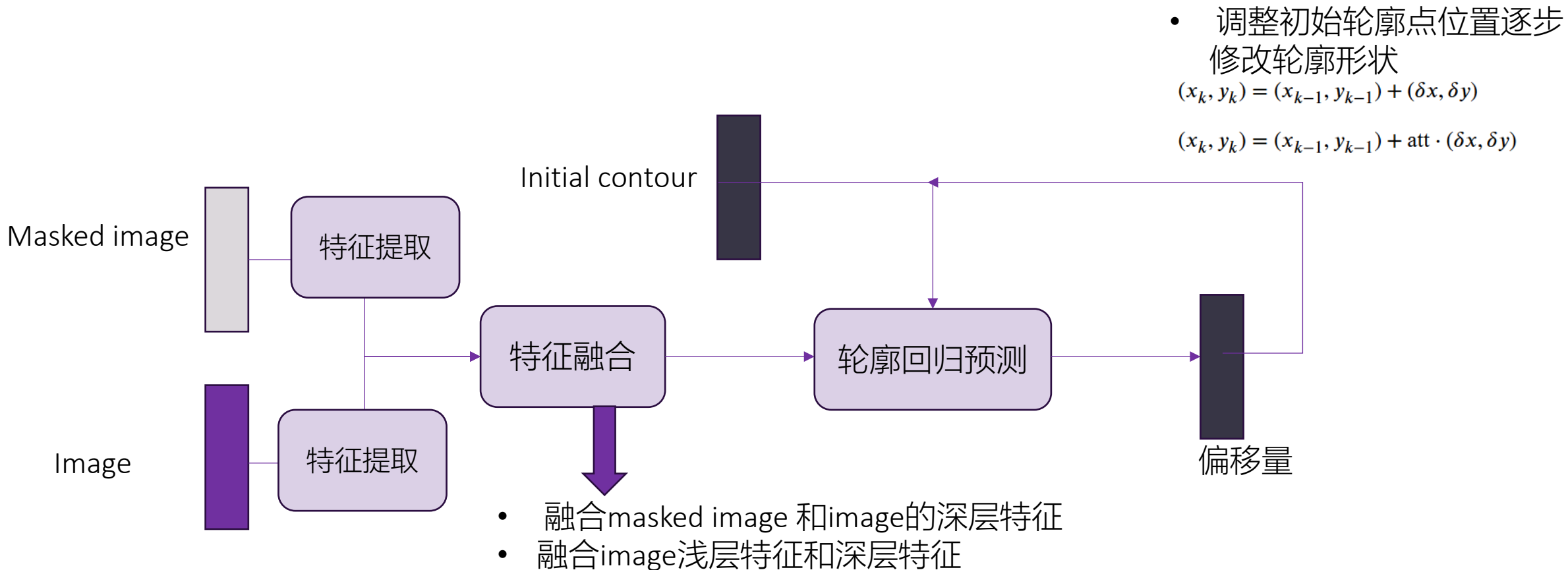
- 基于点击点的方式需要**多次点击**，并且得到掩码**不易交互**调整结果



- 迭代优化轮廓的方法**忽视点调整方向**的优化监督目标使得调整效率较低。

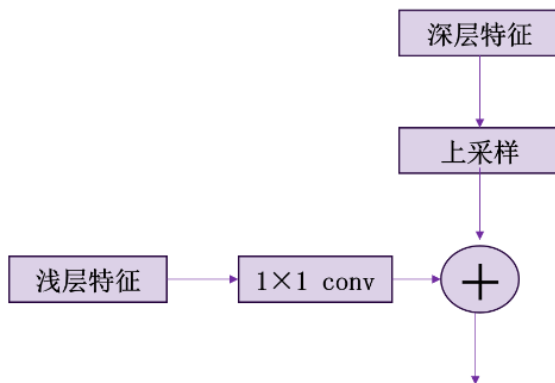
将迭代调整点的**方向监督信息**加入优化目标，辅助训练多边形轮廓回归算法，**提高**迭代调整实例轮廓的效率和有限次迭代后获得**轮廓的精度**

## 2.2 基于向量夹角余弦度量辅助训练的轮廓回归算法-整体结构



## 2.2 基于向量夹角余弦度量辅助训练的轮廓回归算法-关键细节

- 融合image浅层特征和深层特征



- 获得更鲁棒的特征

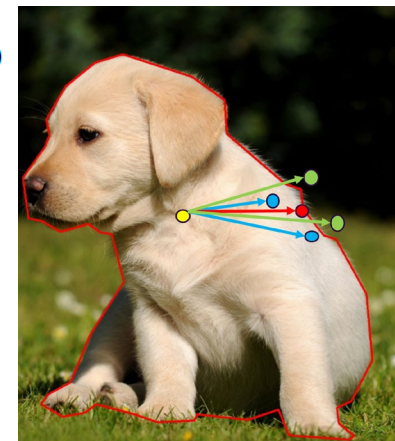
- 计算向量夹角余弦值作为迭代调整角度的方向监督信息

$$L_{\cos}(contour_{pred}, contour_{true}) = \frac{1}{K} \sum_{i=1}^K (1 - \cos(\mathbf{v}_i, \hat{\mathbf{v}}_i))$$

$$L_{vector} = \frac{1}{N} \sum_{i=1}^N Smooth_{L1}(|(x, y) - (\hat{x}, \hat{y})|)$$

$$Dice\_loss = 1 - \frac{pred \cdot GT}{\|pred\| \cdot \|GT\|}$$

$$L = L_{vector} + L_{\cos} + \alpha Dice\_loss$$



- 增加方向监督信息

## 2.2 基于向量夹角余弦度量辅助训练的轮廓回归算法-实验

表 4-1 训练集测试集来自 cityscapes 不同方法对比试验

模型	mAP	$F_{1px}$	$F_{2px}$
DACN	-	45.27	59.89
Polygon-RNN++	25.5	46.57	62.26
PSP-Deeplab	-	47.10	62.82
DELSE	-	48.59	64.45
Split-GCN	29.6	52.50	67.50
ours(ResNet34)	<b>34.7</b>	50.67	65.21
ours(ResNet50)	<b>35.8</b>	51.74	66.44

表 4-2 在 cityscapes 训练在其他数据集上测试不同方法对比实验 (mIoU)

模型	ADE20k	rooftop
polygon-RNN++	71.82	65.67
PSP-Deeplab	72.70	57.91
Polygon-GCN	72.31	66.78
Split-GCN	72.94	68.33
DACN	73.21	66.92
(train on cityscapes)		
ours(ResNet 34)	72.76	<b>75.47</b>
ours(ResNet50)	<b>74.50</b>	72.2
(train on coco)		
ours(ResNet 34)	79.05	76.78
ours(ResNet50)	79.48	85.99

- 在训练集和测试集来源相同时，在**平均精确度**上超过了其他方法
- 在训练集和测试集来源不同时，本文方法预测的轮廓能够更好的覆盖实例轮廓，**更接近真实实例轮廓**

## 2.2 基于向量夹角余弦度量辅助训练的轮廓回归算法-实验

- 实例类别有较大差距情况



Coco

	数据集	mIOU	$F_{1px}$	$F_{2px}$
train on cityscapes	Coco(ResNet34)	66.78	42.81	54.78
	Coco(ResNet50)	66.93	43.34	55.35



ADE20K

	person	bus	car	television receiver	oven	bicycle
train on coco	73.3	82.5	78.9	92.7	80.3	59.1
train on cityscapes	66.5	76.3	75.2	84.4	81.2	56.2

- 拍摄视角和实例类别有较大差距情况



	数据集	mIOU	$F_{1px}$	$F_{2px}$
train on cityscapes	Rooftop(ResNet34)	75.47	21.51	32.02
	Rooftop(ResNet50)	72.2	21.89	32.45
train on coco	Rooftop(ResNet50)	85.99	37.41	52.82

- 在训练数据和测试数据区别较大时，真实实例轮廓预测准确，模型有较好的泛化性能

## 2.2 基于向量夹角余弦度量辅助训练的轮廓回归算法-消融实验

优化目标中补充余弦度量项有效性分析，补充优化目标后轮廓预测mIoU平均**提升5%**

$$L_{\cos}(\text{contour}_{pred}, \text{contour}_{true}) = \frac{1}{K} \sum_{i=1}^K (1 - \cos(\mathbf{v}_i, \hat{\mathbf{v}}_i))$$

$$L_{vector} = \frac{1}{N} \sum_1^N \text{Smooth}_{L1}(|(x, y) - (\hat{x}, \hat{y})|)$$

$$\text{Dice}_{loss} = 1 - \frac{\text{pred} \cdot \text{GT}}{\|\text{pred}\| \cdot \|\text{GT}\|}$$

	Cityscapes	ADE20k	Rooftop
loss= $L_{vector} + \alpha \text{Dice}_{loss}$	72.33	74.9	78.04
loss= $L_{vector} + \alpha \text{Dice}_{loss} + L_{\cos}$	<b>76.69</b>	<b>79.48</b>	<b>85.99</b>

第三部分

# 实际应用

Applications

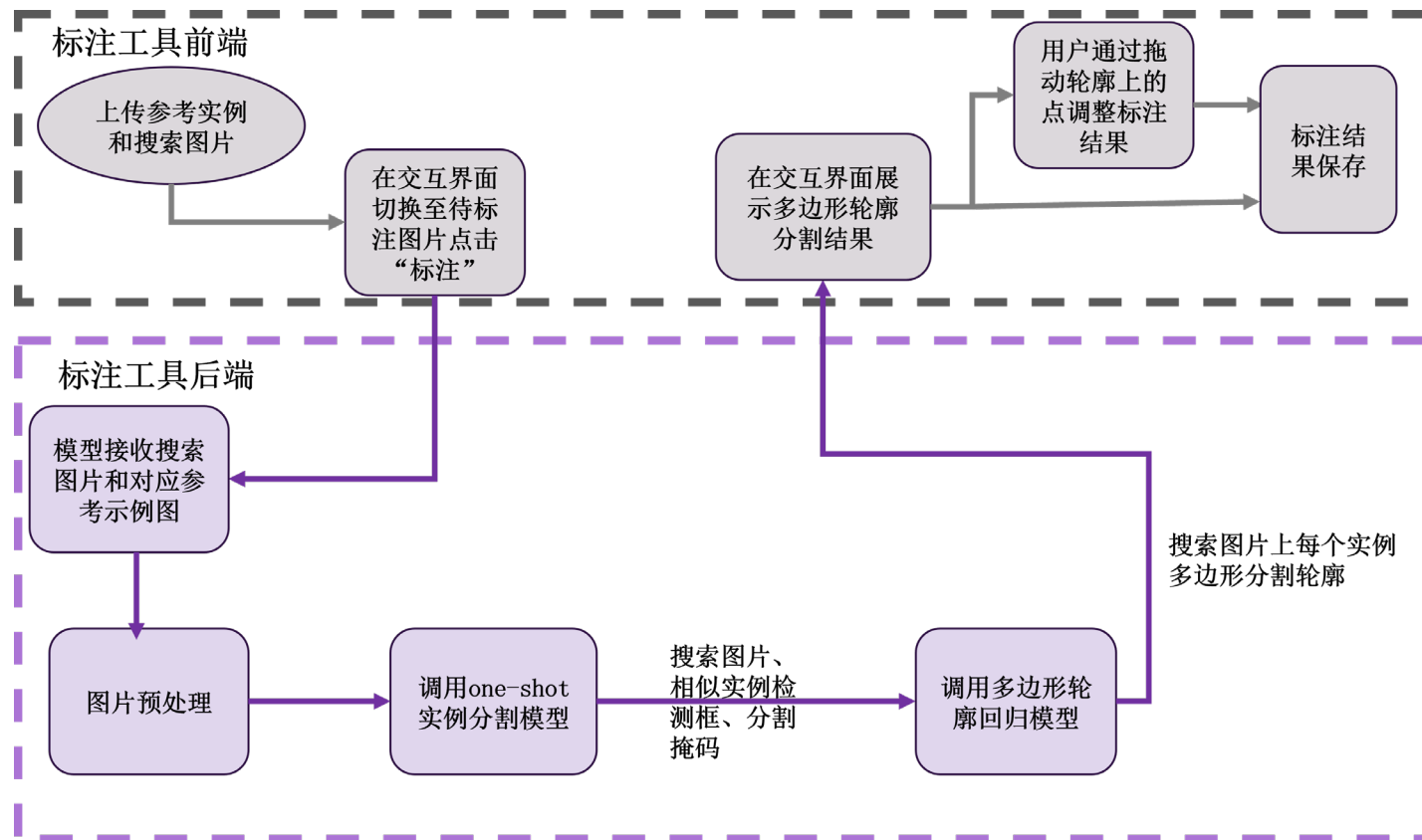
基于参考实例的图片标注系统

## 3.1

## 系统简述

## 系统需求

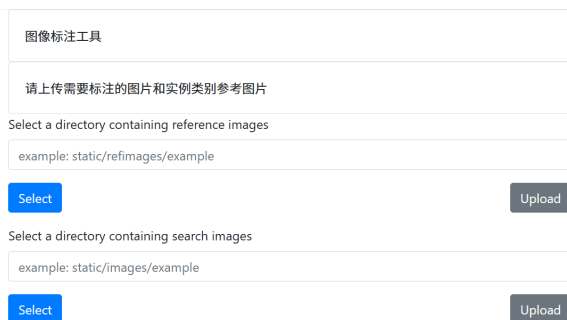
- 准确的分割结果
- 便捷的结果反馈和调整方式
- 一键保存标注结果



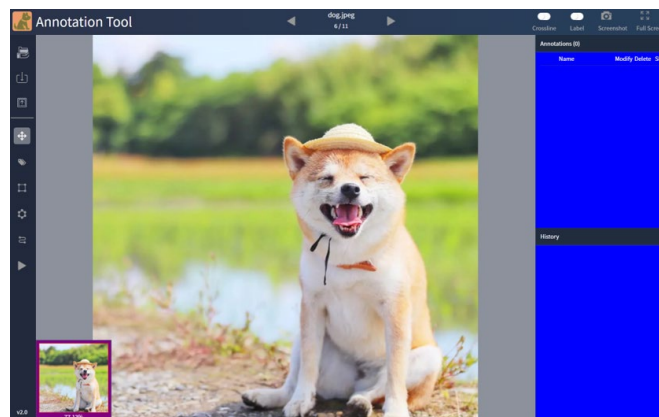
系统流程图

## 3.2

## 系统展示



1. 数据集上传界面



2. 可视化主界面



3. 获得标注结果展示界面

第四部分

# 研究生期间工作成果

Work Product

相关成果列举

## □ 论文：

1. X.Xu, **R.Li**, M.Yi, F.Shen and J. Zhao. Interactive instance annotation with siamesenetworks. arXiv preprint arXiv:2505.03184.2025.

## □ 专利：

1. 徐百乐、**李若彤**、许翔、易梦军、申富饶 “一种基于孪生神经网络的图像标注方法”(202510435652.6)

## □ 项目：

1. 科技部重大项目 “基于神经可塑性的脉冲网络高效学习机制与类脑智能系统”（参与课题年限 2022 年 9 月—2025 年 6 月），负责脉冲神经网络图像生成相关研究。

## □ 荣誉：

2023-2024学年南京大学优秀研究生

第五部分

# 总结

Summary

## 面向实际应用场景的图片实例标注算法研究

## 基于交叉注意力机制的 one-shot 实例分割方法

- 提出了一阶段不同尺度参考示例和搜索图片特征匹配模块，充分提取参考实例和搜索图片不同区域的特征，简化模型，提高算法效率

## 基于向量夹角余弦度量辅助训练的轮廓回归算法

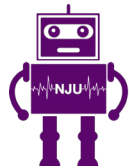
- 将实例多边形轮廓点调整方向信息加入优化目标，提高轮廓回归预测效率，使得有限次迭代预测能够获得更高精度的轮廓

## 基于参考实例的图片标注系统

- 结合两个方法开发了辅助标注系统，支持标注人员通过简单操作调整标注结果，具有实际使用价值



南京大學  
NANJING UNIVERSITY



RINC

Robotic Intelligence & Neural Computing Group

# 请各位老师批评指正

答辩人：李若彤 502022370022

导师：申富饶 教授

日期：2025年5月16日

誠樸雄偉 勵學敦行