

2023长三角[芜湖] 人工智能视觉算法大赛

赛题：基于目标检测的非机动车不戴头盔识别

团队名：积木森林

汇报人：叶顺龙、魏帅、石旭（南京理工大学）



2023长三角（芜湖）人工智能视觉算法大赛

问题分析

- 场景
- 难点

算法设计

- 选型
- 适配

总结回顾

- 提升
- 评价

我们将按照这个过程阐述我们的方案

2023长三角（芜湖）人工智能视觉算法大赛

问题分析

- 场景
- 难点

算法设计

- 选型
- 适配

总结回顾

- 提升
- 评价

问题分析

- **场景**
- **难点**

在画面中出现有**骑非机动车的人员**没有**戴头盔**时，进行报警

标注数据

2D框标注
26个类别

- (1) *bicycle*: 有人骑在车上或者有人在车上面的自行车
- (2) *motorbike*: 有人骑在车上或者有人在车上面的男式摩托车
- (3) *electric_scooter*: 有人骑在车上或者有人在车上面的电瓶车、电动车、女式摩托车
- (4) *tricycle*: 有人骑在车上或者有人在车上面的人力三轮车
- (5) *auto_tricycle*: 有人骑在车上或者有人在车上面的电动三轮小货车、电动三轮载客车
- (6) *none_person_bicycle*: 无人的自行车
- (7) *none_person_motorbike*: 无人的男式摩托车
- (8) *none_person_electric_scooter*: 无人的电瓶车、电动车、女式摩托车
- (9) *none_person_tricycle*: 无人的人力三轮车
- (10) *none_person_auto_tricycle*: 无人的电动三轮小货车、电动三轮载客车
- (11) *unknown_person_bicycle*: 未知有人的自行车
- (12) *unknown_person_motorbike*: 未知有人的男式摩托车
- (13) *unknown_person_electric_scooter*: 未知有人的电瓶车、电动车、女式摩托车
- (14) *unknown_person_tricycle*: 未知有人的的人力三轮车
- (15) *unknown_person_auto_tricycle*: 未知有人的电动三轮小货车、电动三轮载客车
- (16) *person*: 行人，路边的人、有推车动作的人等【分开标注】
- (17) *rider*: 骑车的人、坐在车上面的人【分开标注】
- (18) *bicycle_person*: 自行车上的车，一个大框框住所有人+车
- (19) *motorbike_person*: 摩托车上的车，一个大框框住所有人+车，
- (20) *electric_scooter_person*: 电瓶车、电动车、女式摩托车上的人，一个大框框住所有人+车
- (21) *tricycle_person*: 人力三轮车上的人，一个大框框住所有人+车
- (22) *auto_tricycle_person*: 电动三轮小货车、电动三轮载客车上的人，一个大框框住所有人+车
- (23) *head*: 未戴摩托车头盔的人头
- (24) *helmet*: 戴摩托车头盔、安全帽的人头
- (25) *hat*: 戴其它帽子、鸭舌帽等的人头
- (26) *bicycle_helmet*: 戴自行车头盔等的人头



2023长三角（芜湖）人工智能视觉算法大赛

问题分析

- 场景
- 难点

在画面中出现有**骑非机动车的人员**没有**戴头盔**时，进行报警

标注数据

2D框标注
26个类别

- (1) *bicycle*: 有人骑在车上或者有人在车上面的自行车
- (2) *motorbike*: 有人骑在车上或者有人在车上面的男式摩托车
- (3) *electric_scooter*: 有人骑在车上或者有人在车上面的电瓶车、电动车、女式摩托车
- (4) *tricycle*: 有人骑在车上或者有人在车上面的人力三轮车
- (5) *auto_tricycle*: 有人骑在车上或者有人在车上面的电动三轮小货车、电动三轮载客车
- (6) *none_person_bicycle*: 无人的自行车
- (7) *none_person_motorbike*: 无人的男式摩托车
- (8) *none_person_electric_scooter*: 无人的电瓶车、电动车、女式摩托车
- (9) *none_person_tricycle*: 无人的人力三轮车
- (10) *none_person_auto_tricycle*: 无人的电动三轮小货车、电动三轮载客车
- (11) *unknown_person_bicycle*: 未知有人的自行车
- (12) *unknown_person_motorbike*: 未知有人的男式摩托车
- (13) *unknown_person_electric_scooter*: 未知有人的电瓶车、电动车、女式摩托车
- (14) *unknown_person_tricycle*: 未知有人的的人力三轮车
- (15) *unknown_person_auto_tricycle*: 未知有人的电动三轮小货车、电动三轮载客车
- (16) *person*: 行人、路边的人、有推车动作的人等【分开标注】
- (17) *rider*: 骑车的人、坐在车上面的人【分开标注】
- (18) *bicycle_person*: 自行车上的行人，一个大框框住所有人+车
- (19) *motorbike_person*: 摩托车上的行人，一个大框框住所有人+车，
- (20) *electric_scooter_person*: 电瓶车、电动车、女式摩托车上的人，一个大框框住所有人+车
- (21) *tricycle_person*: 人力三轮车上的人，一个大框框住所有人+车
- (22) *auto_tricycle_person*: 电动三轮小货车、电动三轮载客车上的人，一个大框框住所有人+车
- (23) *head*: 未戴摩托车头盔的人头
- (24) *helmet*: 戴摩托车头盔、安全帽的人头
- (25) *hat*: 戴其它帽子、鸭舌帽等的人头
- (26) *bicycle_helmet*: 戴自行车头盔等的人头



2023长三角（芜湖）人工智能视觉算法大赛

问题分析

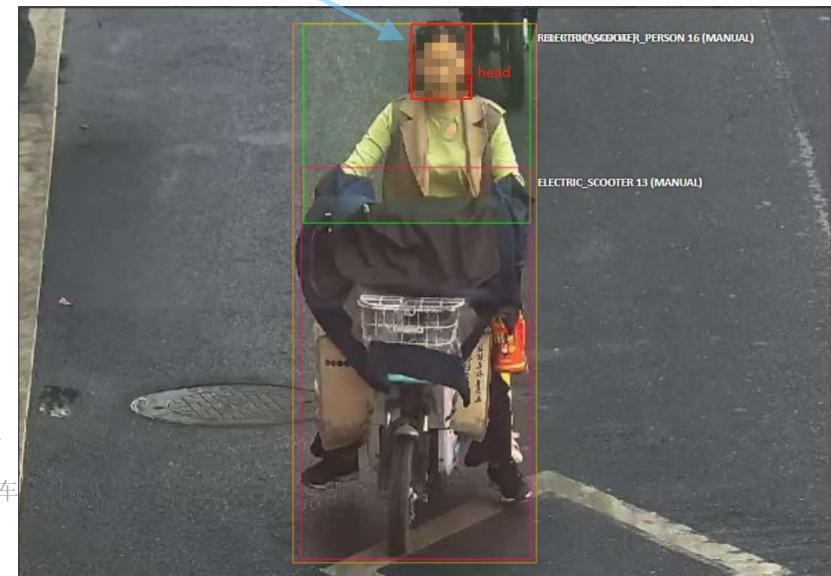
- 场景
- 难点

在画面中出现有**骑非机动车的人员**没有**戴头盔**时，进行报警

标注数据

2D框标注
26个类别

- (1) *bicycle*: 有人骑在车上或者有人在车上面的自行车
- (2) *motorbike*: 有人骑在车上或者有人在车上面的男式摩托车
- (3) *electric_scooter*: 有人骑在车上或者有人在车上面的电瓶车、电动车、女式摩托车
- (4) *tricycle*: 有人骑在车上或者有人在车上面的人力三轮车
- (5) *auto_tricycle*: 有人骑在车上或者有人在车上面的电动三轮小货车、电动三轮载客车
- (6) *none_person_bicycle*: 无人的自行车
- (7) *none_person_motorbike*: 无人的男式摩托车
- (8) *none_person_electric_scooter*: 无人的电瓶车、电动车、女式摩托车
- (9) *none_person_tricycle*: 无人的人力三轮车
- (10) *none_person_auto_tricycle*: 无人的电动三轮小货车、电动三轮载客车
- (11) *unknown_person_bicycle*: 未知有人的自行车
- (12) *unknown_person_motorbike*: 未知有人的男式摩托车
- (13) *unknown_person_electric_scooter*: 未知有人的电瓶车、电动车、女式摩托车
- (14) *unknown_person_tricycle*: 未知有人的的人力三轮车
- (15) *unknown_person_auto_tricycle*: 未知有人的电动三轮小货车、电动三轮载客车
- (16) *person*: 行人，路边的人、有推车动作的人等【分开标注】
- (17) *rider*: 骑车的人、坐在车上面的人【分开标注】
- (18) *bicycle_person*: 自行车上的车，一个大框框住所有人+车
- (19) *motorbike_person*: 摩托车上的车，一个大框框住所有人+车，
- (20) *electric_scooter_person*: 电瓶车、电动车、女式摩托车上的人，一个大框框住所有人+车
- (21) *tricycle_person*: 人力三轮车上的人，一个大框框住所有人+车
- (22) *auto_tricycle_person*: 电动三轮小货车、电动三轮载客车上的人，一个大框框住所有人+车
- (23) *head*: 未戴摩托车头盔的人头
- (24) *helmet*: 戴摩托车头盔、安全帽的人头
- (25) *hat*: 戴其它帽子、鸭舌帽等的人头
- (26) *bicycle_helmet*: 戴自行车头盔等的人头



问题分析

- **场景**
- **难点**

在画面中出现有**骑非机动车的人员**没有**戴头盔**时，进行报警

衡量指标

对于ground truth中的 motorbike_person 和 electric_scooter_person 这两个2d框标签：

- ✓ 如果这两种ground truth的2d框内，存在 **head** 的ground truth标签，那么算法也需要输出这个head 2d框的坐标
- ✗ 如果算法没有输出这个head 2d框或者坐标框不符合iou的要求，算作**漏报**
- ✗ 如果算法输出的head 2d框不匹配任何ground truth的 head 标签，算作**误报**
- ✗ 如果算法输出的head 2d框在motorbike_person和electric_scooter_person这两种2d框之外，也算做**误报**

按照以上逻辑计算的f1-score，**只有 head、hat** 标签会计入算法榜成绩

总分

算法**精度**得分 x 0.8 + 算法**性能**得分 x 0.2



问题分析

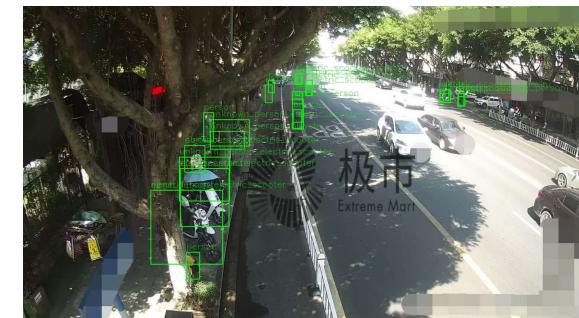
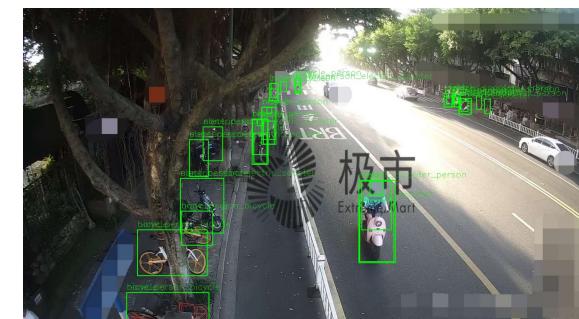
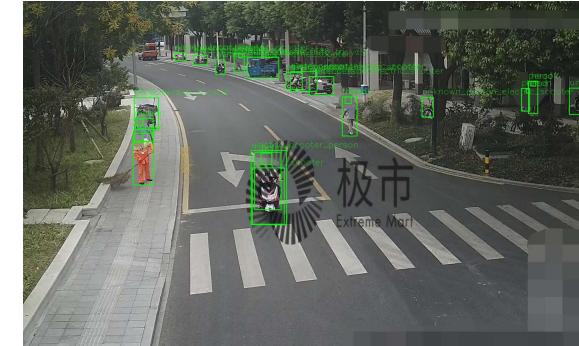
- 场景
- 难点

算法设计相关问题

- 标注数据那么多类别？用多少个类别识别？
- 最终衡量结果为head和hat两种检测框，均为小目标，怎么解决？
- 算法误报率和漏报率怎么降低？
- 在保证识别准确率的同时，性能如何兼顾？

现实场景相关问题

- 识别场景涉及警告和处罚，误报和漏报孰轻孰重？
- 切换技术选型的可行性和与其他系统的兼容性？
- 方案实现后到落地还有多远？技术上的支持？



2023长三角（芜湖）人工智能视觉算法大赛

问题分析

- 场景
- 难点

算法设计

- 选型
- 适配

总结回顾

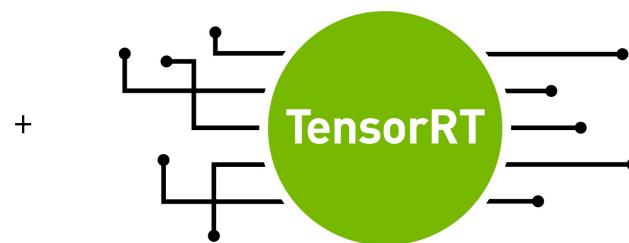
- 提升
- 评价

算法设计

- 选型
- 适配

技术选型和方案流程设计

YOLOv5



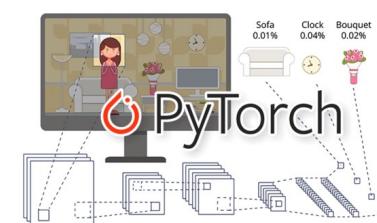
Step1. 数据预处理（清洗，筛选）



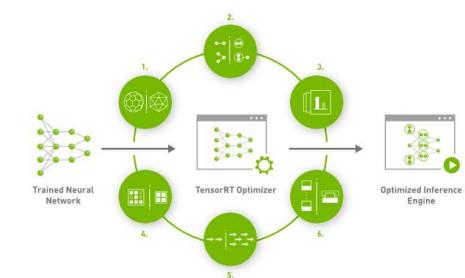
Step2. 数据集划分



Step3. 训练模型，得到pt权重



Step4. 部署模型，得到engine



算法设计

- 选型
- 适配

技术选型和方案流程设计

Step1. 数据预处理（清洗，筛选）



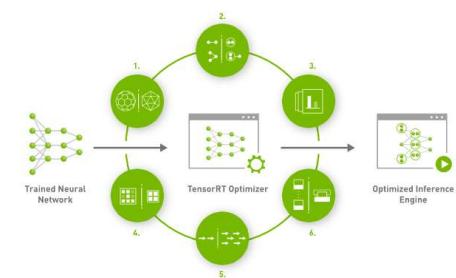
Step2. 数据集划分



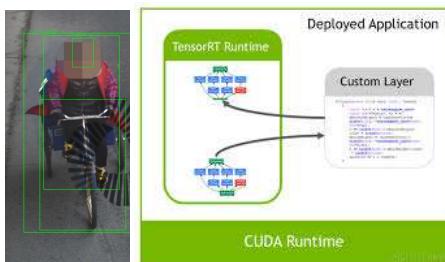
Step3. 训练模型，得到pt权重



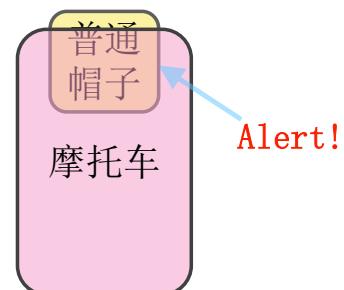
Step4. 部署模型，得到engine



Step5. 模型推理



Step6. 算法后处理



Step7. 接口结果返回



算法设计

- 选型
- 适配

- 标注数据那么多类别？用多少个类识别？
- 最终衡量结果为head和hat两种检测框，均为小目标，怎么解决？
- 算法误报率和漏报率怎么降低？
- 在保证识别准确率的同时，性能如何兼顾？

- (1) *bicycle*: 有人骑在车上或者有人在车上面的自行车
- (2) *motorbike*: 有人骑在车上或者有人在车上面的男式摩托车
- (3) *electric_scooter*: 有人骑在车上或者有人在车上面的电瓶车、电动车、女式摩托车
- (4) *tricycle*: 有人骑在车上或者有人在车上面的人力三轮车
- (5) *auto_tricycle*: 有人骑在车上或者有人在车上面的电动三轮小货车、电动三轮载客车
- (6) *none_person_bicycle*: 无人的自行车
- (7) *none_person_motorbike*: 无人的男式摩托车
- (8) *none_person_electric_scooter*: 无人的电瓶车、电动车、女式摩托车
- (9) *none_person_tricycle*: 无人的人力三轮车
- (10) *none_person_auto_tricycle*: 无人的电动三轮小货车、电动三轮载客车
- (11) *unknown_person_bicycle*: 未知有人的自行车
- (12) *unknown_person_motorbike*: 未知有人的男式摩托车
- (13) *unknown_person_electric_scooter*: 未知有人的电瓶车、电动车、女式摩托车
- (14) *unknown_person_tricycle*: 未知有人的的人力三轮车
- (15) *unknown_person_auto_tricycle*: 未知有人的电动三轮小货车、电动三轮载客车
- (16) *person*: 行人、路边的人、有推车动作的人等【分开标注】
- (17) *rider*: 骑车的人、坐在车上面的人【分开标注】
- (18) *bicycle_person*: 自行车车上的人，一个大框框住所有人+车
- (19) *motorbike_person*: 摩托车上的人，一个大框框住所有人+车，
- (20) *electric_scooter_person*: 电瓶车、电动车、女式摩托车上的人，一个大框框住所有人+车
- (21) *tricycle_person*: 人力三轮车上的人，一个大框框住所有人+车
- (22) *auto_tricycle_person*: 电动三轮小货车、电动三轮载客车上的人，一个大框框住所有人+车
- (23) *head*: 未戴摩托车头盔的人头
- (24) *helmet*: 戴摩托车头盔、安全帽的人头
- (25) *hat*: 戴其它帽子、鸭舌帽等的人头
- (26) *bicycle_helmet*: 戴自行车头盔等的人头

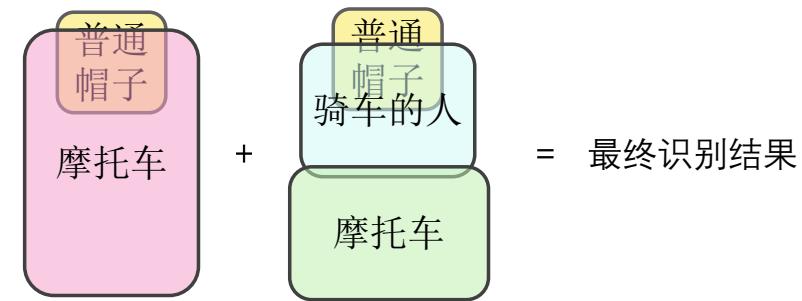
朴素识别逻辑

仅校验：所有 *head* 和 *hat*，是否包含于 *motorbike_person* 或者 *electric_scooter_person* 的检测框

更新识别逻辑

先校验：所有 *head* 和 *hat*，是否包含于 *motorbike_person* 或者 *electric_scooter_person* 的检测框

再校验：所有 *head* 和 *hat*，是否包含于 *motorbike*、*electric_scooter* 或者 *rider* 组合而成的检测框

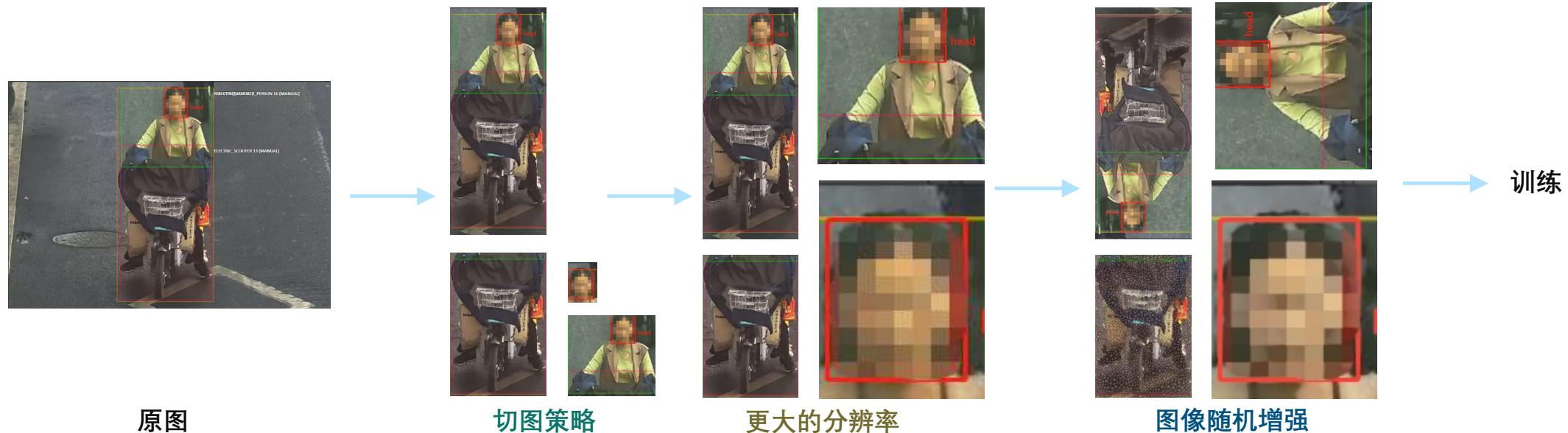


算法设计

- 选型
- 适配

- 标注数据那么多类别？用多少个类识别？
- 最终衡量结果为head和hat两种检测框，均为**小目标**，怎么解决？
- 算法误报率和漏报率怎么降低？
- 在保证识别准确率的同时，**性能**如何兼顾？

切图策略 + 更大的分辨率 + 图像随机增强 + *Slicing Aided Hyper Inference(SAHI)*



算法设计

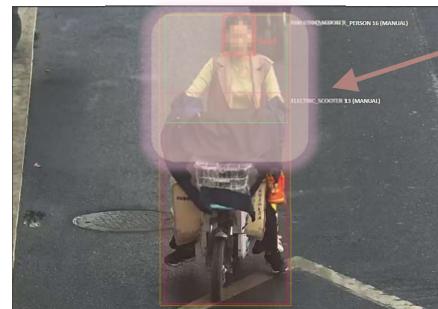
- 选型
- 适配

- 标注数据那么多类别？用多少个类识别？
- 最终衡量结果为head和hat两种检测框，均为**小目标**，怎么解决？
- 算法误报率和漏报率怎么降低？
- 在保证识别准确率的同时，~~性能如何兼顾~~

切图策略 + 更大的分辨率 + 图像随机增强 + Slicing Aided Hyper Inference(SAHI)

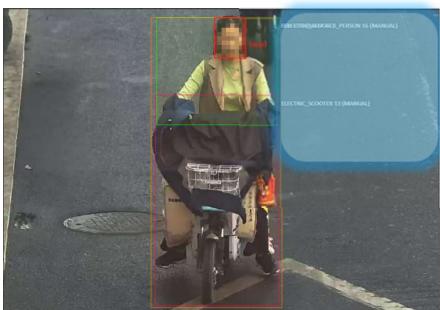


Patchify

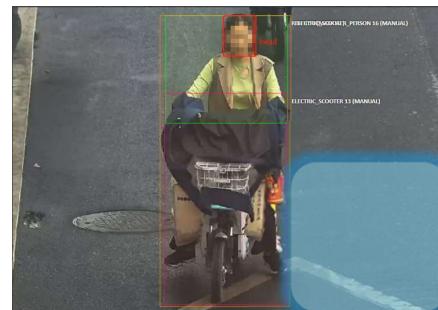


检测到目标

测试图



...



整合输出
输出

算法设计

- 选型
- 适配

- 标注数据那么多类别？用多少个类识别？
- 最终衡量结果为head和hat两种检测框，均为小目标，怎么解决？
- 算法**误报率**和**漏报率**怎么降低？
- 在保证识别准确率的同时，**性能**如何兼顾？

误报可能来源：

- 数据不平衡
- 不同环境下的变化
- 过度拟合了训练数据

误报来源对应解决方法：

- **负样本（不戴头盔的样本）增加**
- 数据随机饱和度，色相等增强
- 多尺度训练

戴头盔的
摩托车驾
驶者

戴头盔的
摩托车驾
驶者

戴头盔的
摩托车驾
驶者

不戴头盔的
摩托车
驾驶者

戴头盔的
摩托车驾
驶者

戴头盔的
摩托车驾
驶者

戴头盔的
摩托车驾
驶者

不戴头盔的
摩托车
驾驶者

不戴头盔的
摩托车
驾驶者

算法设计

- 选型
- 适配

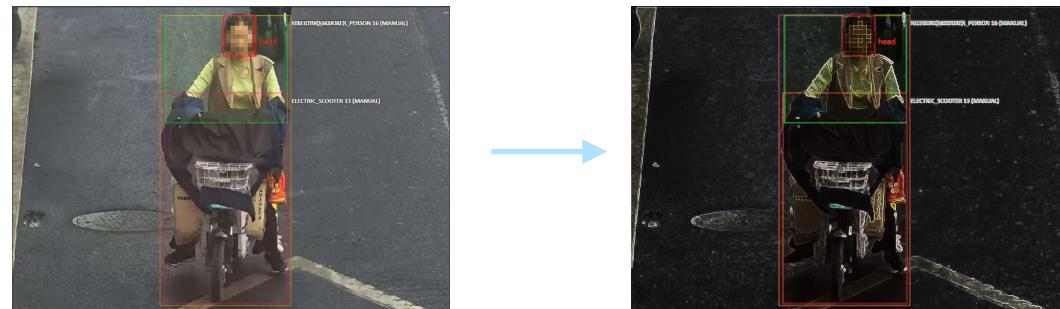
- 标注数据那么多类别？用多少个类识别？
- 最终衡量结果为head和hat两种检测框，均为小目标，怎么解决？
- 算法**误报率**和**漏报率**怎么降低？
- 在保证识别准确率的同时，**性能**如何兼顾？

误报可能来源：

- 数据不平衡
- 不同环境下的变化
- 过度拟合了训练数据

误报来源对应解决方法：

- 负样本（不戴头盔的样本）增加
- 数据**随机饱和度，色相等增强**
- 多尺度训练



算法设计

- 选型
- 适配

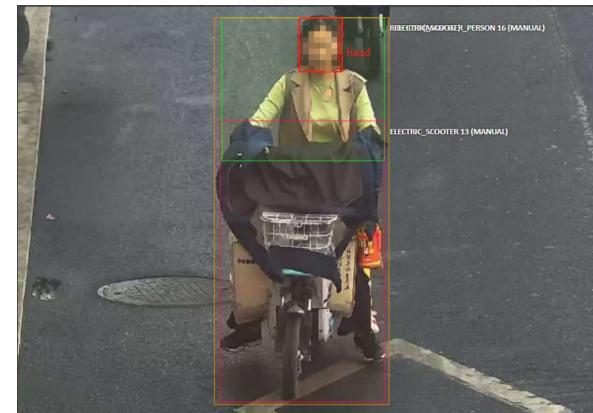
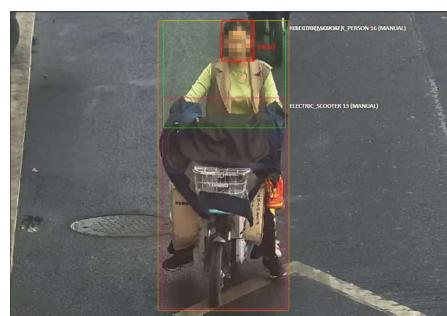
- 标注数据那么多类别？用多少个类识别？
- 最终衡量结果为head和hat两种检测框，均为小目标，怎么解决？
- 算法**误报率**和**漏报率**怎么降低？
- 在保证识别准确率的同时，**性能**如何兼顾？

误报可能来源：

- 数据不平衡
- 不同环境下的变化
- 过度拟合了训练数据

误报来源对应解决方法：

- 负样本（不戴头盔的样本）增加
- 数据随机饱和度，色相等增强
- **多尺度训练**



Multi_scale

算法设计

- 选型
- 适配

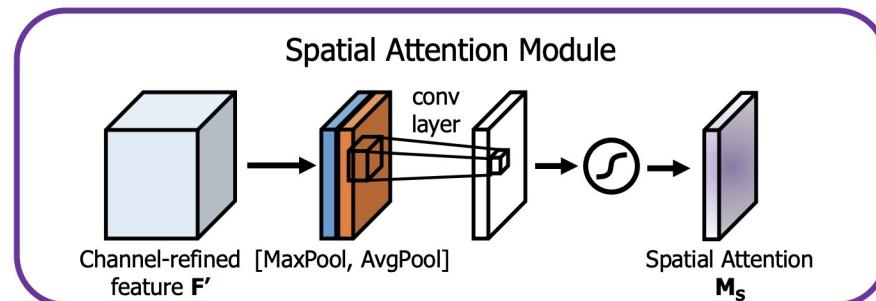
- 标注数据那么多类别？用多少个类识别？
- 最终衡量结果为head和hat两种检测框，均为小目标，怎么解决？
- 算法误报率和漏报率怎么降低？
- 在保证识别准确率的同时，性能如何兼顾？

漏报可能来源：

- 模型没有足够的训练数据来学习目标的各种变化
- 模型的架构不足以捕捉目标的复杂特征

漏报来源对应解决方法：

- **注意力机制**引入
- 使用不同输入图片大小进行微调



算法设计

- 选型
- 适配

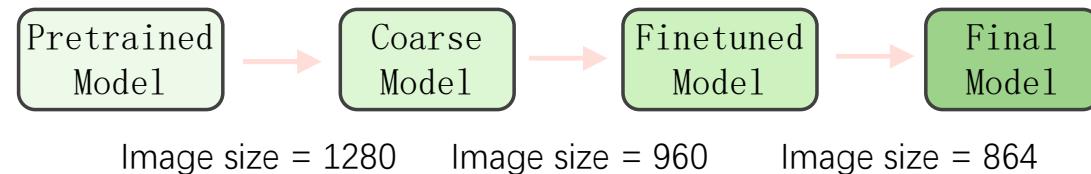
- 标注数据那么多类别？用多少个类识别？
- 最终衡量结果为head和hat两种检测框，均为小目标，怎么解决？
- 算法**误报率**和**漏报率**怎么降低？
- 在保证识别准确率的同时，**性能**如何兼顾？

漏报可能来源：

- 模型没有足够的训练数据来学习目标的各种变化
- 模型的架构不足以捕捉目标的复杂特征

漏报来源对应解决方法：

- 注意力机制引入
- 使用**不同输入图片大小进行微调**

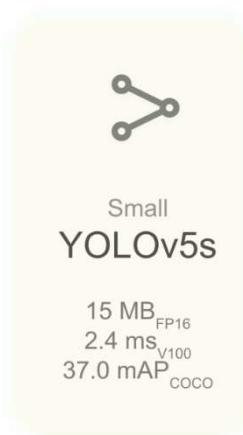


算法设计

- 选型
- 适配

- 标注数据那么多类别？用多少个类识别？
- 最终衡量结果为head和hat两种检测框，均为小目标，怎么解决？
- 算法误报率和漏报率怎么降低？
- 在保证识别准确率的同时，**性能**如何兼顾？

Yolov5s is enough



(YOLOv8 vs YOLOv5)

Model Size	YOLOv5	YOLOv8	Difference
Nano	27.6	36.7	+32.97%
Small	37.6	44.6	+18.62%
Medium	45	49.9	+10.89%
Large	49	52.3	+6.73%
Xtra Large	50.7	53.4	+5.33%

*Image Size = 640

2023长三角（芜湖）人工智能视觉算法大赛

问题分析

- 场景
- 难点

算法设计

- 选型
- 适配

总结回顾

- 提升
- 评价

总结回顾

- 提升
- 评价

方案的优点：

使用当前目标检测中我们已知的**最成熟的workflow**
对任务中几个细节问题有**较清晰的分析与算法设计**

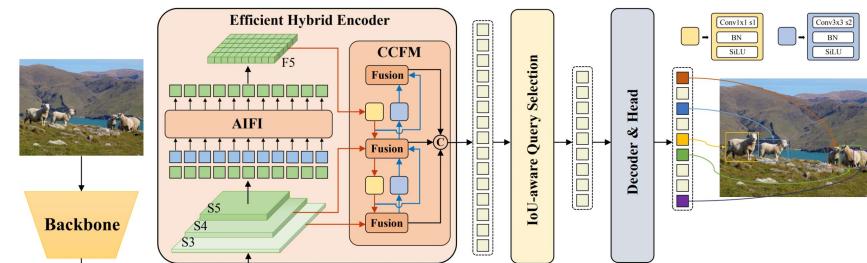
方案的不足：

未完成对其他方案的尝试，包括但不限于移动端细分模型（MobileSAM、RT-DETR）
模型参数调节部分并不充分，方案可能**并未达到实际精度上限**



MobileSAM

移动端细分模型（MobileSAM）

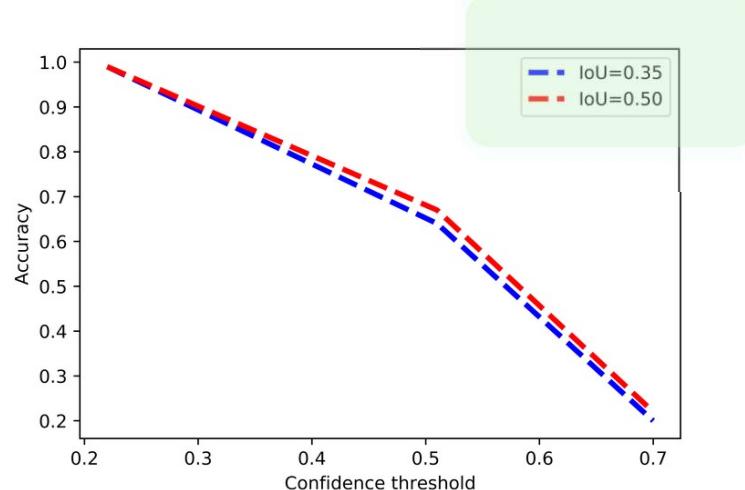


百度的RT-DETR：基于Vision Transformers的实时目标检测器

总结回顾

- 提升
- 评价

- 识别场景涉及警告和处罚，**误报**和**漏报**孰轻孰重？
- 切换技术选型的可行性和与其他系统的兼容性？
- 方案实现后到落地还有多远？技术上的支持？



降低**误报率**，是在当前衡量指标下，和精度的一场**trade-off**

完成对几个重要**阈值**参数的接口封装，包括：

Yolo系列**后处理过程自带阈值**：

NMS Conf and Bounding Box Threshold

针对**降低误报率特别设置阈值**：

Detection Person Head Threshold

Detection Person Hat Threshold

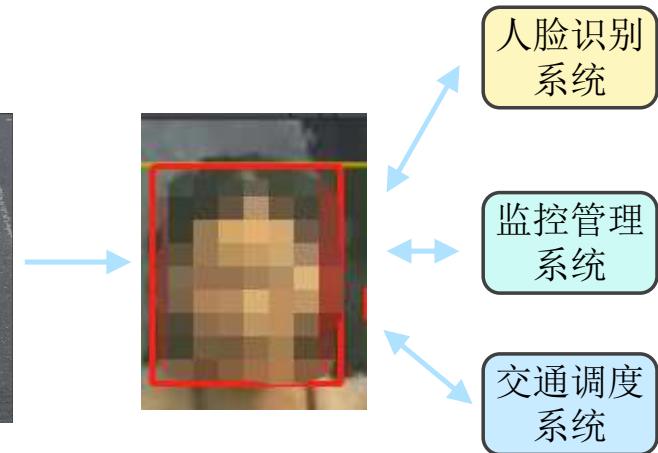
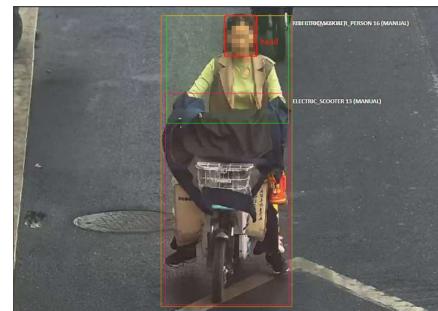
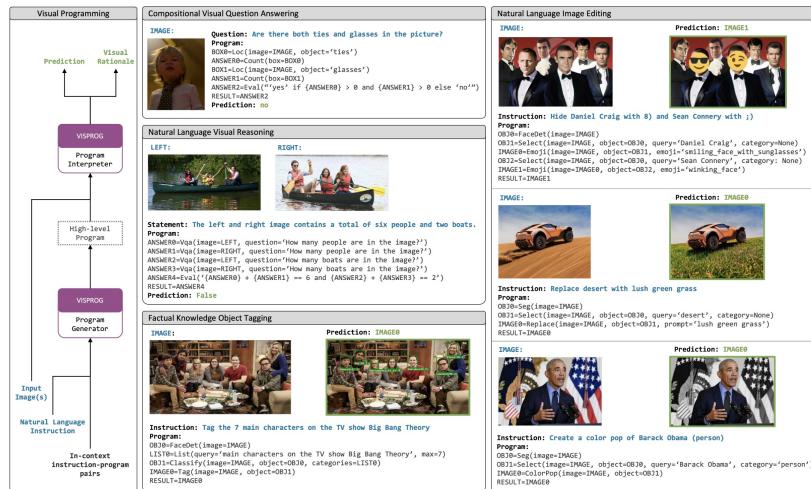
2023长三角 (芜湖) 人工智能视觉算法大赛

总结回顾

- 提升
- 评价

- 识别场景涉及警告和处罚，误报和漏报孰轻孰重？
- 切换技术选型的可行性和与其他系统的兼容性？
- 方案实现后到落地还有多远？技术上的支持？

所有Yolo系列Backbone可以轻松切换



基于神经符号结合的新方法 CVPR 2023 best paper

2023长三角（芜湖）人工智能视觉算法大赛

总结回顾

- 提升
- 评价

- 识别场景涉及警告和处罚，误报和漏报孰轻孰重？
- 切换技术选型的可行性和与其他系统的兼容性？
- 方案实现后到落地还有多远？技术上的支持？



我们虽然是学生
但**并非没有经验**
也乐于**参与项目后续的步骤**

奖项荣誉

三块**ACM-ICPC银牌/铜牌**
本硕国家奖学金获得者
华为软件精英挑战赛二等奖

开发/算法经验

阿里巴巴等企业实习工作经验
广汽传祺**车身缺陷检测**
MacBook**表面缺陷检测**
全球校园**人工智能算法精英赛**赛道第一

积木森林
THANKS

主要参考

➤ 方案设计相关

- YOLO官方文档: <https://docs.ultralytics.com/zh/>
- Montalbo F J P. A Computer-Aided Diagnosis of Brain Tumors Using a Fine-Tuned YOLO-based Model with Transfer Learning[J]. KSII Transactions on Internet & Information Systems, 2020, 14(12).
- Sun J, Ge H, Zhang Z. AS-YOLO: An improved YOLOv4 based on attention mechanism and SqueezeNet for person detection[C]//2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE, 2021, 5: 1451-1456.
- Xue M, Chen M, Peng D, et al. One spatio-temporal sharpening attention mechanism for light-weight YOLO models based on sharpening spatial attention[J]. Sensors, 2021, 21(23): 7949.

➤ 实现细节相关

- YOLO相关检测比赛冠军方案: <https://cvmart.net/community/detail/7488>

➤ 项目落地相关

- Gupta T, Kembhavi A. Visual programming: Compositional visual reasoning without training[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 14953-14962.