信息系统工程重点实验室开放基金

中期总结报告

项目编号：×××××××××××

项目名称：×××××××

承研单位：××××××××××××

课题负责人：×××××××

编写日期：××××××××××

信息系统工程重点实验室

目 录

摘 要

项目基本情况：本课题以/针对×××背景/需求，力求解决×××问题，安排了×××研究/关键技术攻关。

完成情况：重点开展了×××、×××技术研究；突破/完成×××、×××技术；形成了×××等技术规范/标准/指南；研制了×××、×××研制了×××等原型系统/原理样机/原理样件，达到了×××等指标；已在×××方面（初步）应用/试用/验证（部分），达到了×××等实效，（初步）形成了×××等技术能力，其中×××指标由前期的×××提高到×××，达到了×××水平。

预期项目研制完成时，在×××方面有创新，具有×××特点，能够解决×××问题。

问题说明：存在的主要问题及技术难点，……。

**（陆）**

（以上全部800字以内）

# 一．研究总要求

## （一）研究目标

**面向低分辨率侦查视频的目标智能感知技术（陆）**

## （二）研究内容

1.低分辨率视频目标实时分类检测与跟踪技术（邹）

2.低分辨率视频目标文字检测与识别技术（徐）

研究基于深度学习框架的无人机/海警船拍摄的低分辨率视频文字目标检测和实现技术，实现不同角度、不同方向、有遮挡情况下的视频文字有效识别。

3.低分辨率目标状态估计技术（张）

## （三）关键技术

1.低分辨率视频目标实时分类检测与跟踪技术（邹）

2.低分辨率视频目标文字检测与识别技术（徐）

目标文字区域检测采用的是基于Faster RCNN模型，文字区域的OCR识别采用的是基于conv-LSTM的卷积特征序列化的 识别技术。 Faster RCNN在Fast RCNN基础上引入了RPN（Regional Proposal Networks）网络，从而替代之前的Selective Search算法，在提升性能（时间、精度）的同时，也使得模型训练更加便捷，可以由神经网络end-to-end完成。而基于卷积的长短时记忆模型conv-LSTM不仅具备空间特性（卷积），同时也引入了时间特性（LSTM），这是序列模型在图像上的应用扩展。该模块的关键技术包括：

1. 通过检测算法对文字目标区域进行获取
2. 利用Inception\_v4网络对区域文字特征进行提取
3. 基于LSTM建立字典与特征之前的映射关系，其中引入了

Attention机制，增强了对序列信息的利用

3.低分辨率目标状态估计技术（张）

## （四）主要技术指标

（陆）

## （五）进度节点安排及标志性成果

（陆）

# 二．研究工作进展情况（重点说明）

## (一)研究内容的完成情况

1.低分辨率视频目标实时分类检测与跟踪技术（邹）

2.低分辨率视频目标文字检测与识别技术（徐）

根据研究进度要求，现已完成了视频目标文字区域的多角度检测与OCR识别（但识别精度 略低于性能指标，其主要原因在于检测与识别的模型不够优，同时自然场景的复杂性以及字符种类较多且间内特征差异较小），对于检测识别结果将其保存至数据库，并可视化至Web监控界面。同时支持Web端对视频或图片，进行手动或DL辅助标注。

3.低分辨率目标状态估计技术（张）

按照研究进度要求，叙述主要研究内容完成情况，目标实现情况；

未按时完成（关键技术未突破、需求背景调整、技术发展改变）的研究内容应说明原因。

## (二)关键技术的攻关情况

重点介绍关键技术的突破现状。采用什么研究方案，通过什么技术途径，突破了哪些关键技术。对影响全局的“瓶颈”技术攻关情况进行重点说明；

对攻关不顺利（进度置后、指标未达）的关键技术应说明原因。

1.低分辨率视频目标实时分类检测与跟踪技术（邹）

2.低分辨率视频目标文字检测与识别技术（徐）

（1）文字检测

文字检测主要是提取感兴趣区域，在一定程度上也属于目标检测的范畴。目标检测的一般流程如下图所示：

图 目标检测流程图

随着深度学习的崛起，目标检测的精度不断的提升。2013年提出的R-CNN在目标检测领域取得了突破，先后出现了SPP-NET、Fast R-CNN、Faster R-CNN、R-FCN、YOLO、SSD等算法。这些创新算法把传统的计算机视觉方法和深度学习结合，效果显著。

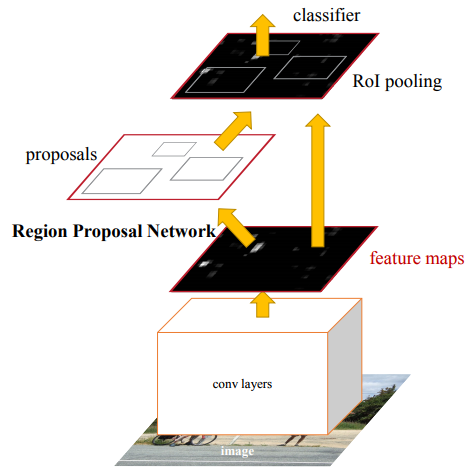
该模块采取Faster R-CNN对文字区域进行提取，模型结构如图1所示。辅助生成样本的RPN（Region Proposal Networks）网络，将算法结构分为两个部分，先由RPN网络判断候选框是否为目标，再经分类定位的多任务损失判断目标类型，整个网络流程都能共享卷积神经网络提取的的特征信息，节约计算成本，且解决Fast R-CNN算法生成正负样本候选框速度慢的问题，同时避免候选框提取过多导致算法准确率下降。

图 Faster RCNN模型结构

（2）文字OCR识别

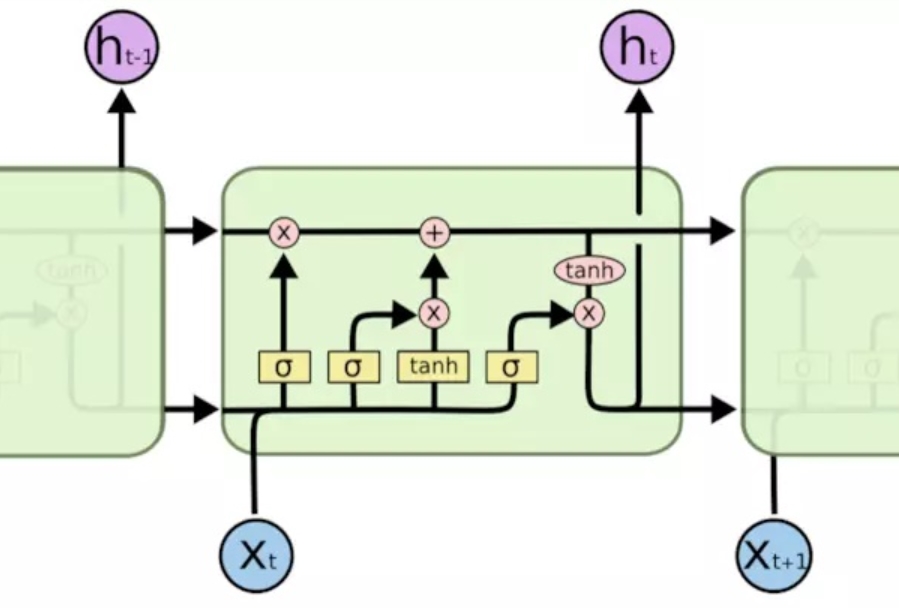
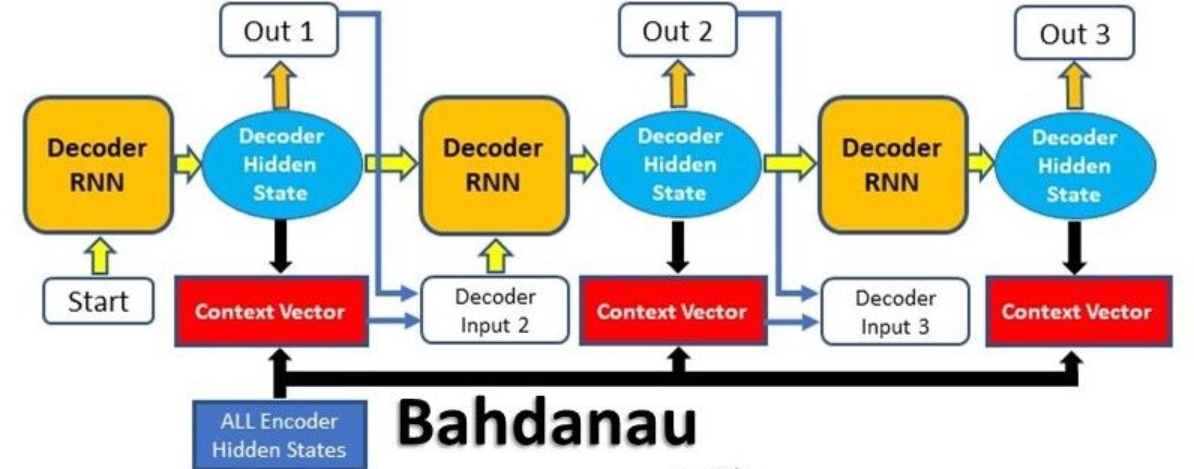
对于获取的生成的文字区域，本模块通过Inception进行特征提取，并编码嵌入LSTM长短时记忆网络，并通过Attention机制对特征信息进行回顾以防止关键信息量随着时间衰减。

图 LSTM的单元结构

如图所示，标准LSTM模型是一种特殊的RNN类型，在每一个重复的模块中有四个特殊的结构，以一种特殊的方式进行交互。在图中，每一条黑线传输整个向量，粉色的圈代表一种pointwise 操作(将定义域上的每一点的函数值分别进行运算)，诸如向量的和，而黄色的矩阵就是学习到的神经网络层。

LSTM模型的核心思想是“细胞状态”。“细胞状态”类似于传送带。直接在整个链上运行，只有一些少量的线性交互。信息在上面流传保持不变会很容易

图 Attention机制示意

如图所示是引入Attention机制的解码过程，主要在解码阶段把编码阶段的每一个隐藏层输出特征都利用上了。Attention的出现就是为了两个目的：

减小处理高维输入数据的计算负担，通过结构化的选取输入的子集，降低数据维度。

“去伪存真”，让处理系统更专注于找到输入数据中显著的与当前输出相关的有用信息，从而提高输出的质量。通过context向量，解码器可以在解码的每一步查询最相关的原文信息，从而避免模型中信息瓶颈问题。

（3）数据优化

利用爬虫技术在航运在线网站上获取9049个船名，并融合现实场景字符生成训练所用的大字典（合计5376个字符）。并将文字OCR模型在RCTW2017+LVST2019这两个官方数据集上进行预训练，使其具备更好的泛化能力

3.低分辨率目标状态估计技术（张）

## (三)技术指标实现情况

已实现的功能性能指标，主要的技术创新点等。

1.低分辨率视频目标实时分类检测与跟踪技术（邹）

2.低分辨率视频目标文字检测与识别技术（徐）

（1）功能性能指标

以NED（Normalized Edit Distance）作为差异度量化指标，100张MTWI2018图片测试的文字OCR平均精度为88.25%，略低于90%的性能指标。

由于模型采用的是采用的是四点斜框方式对文字区域进行标注，故该方法对有一定方向、角度（非水平）的文字检测与识别具有一定的泛化能力。初步效果图如3所示

图3 检测识别效果示意

（2）技术创新点

在模型构建方面，考虑到序列模型解码时输入的编码信息会随着时间步满满衰减，故该模块在conv-LSTM的基础上引入Attention机制，每个时刻对该时间步的编码信息就行回顾，以使模型对序列信息的提取更加精确，同是也提高了信息的利用率。

在训练数据方面，该模块采用大字典训练、小字典预测方式：为了更贴合船名识别，利用爬虫技术从航运在线网站上，爬取了9049个船名，并融合现实场景字符生成训练所用的大字典（合计5376个字符），融入现实场景的字符是为了增强模型的鲁棒性，虽然字符不同但在一定程度上改善船名字符低水平特征的获取。在对船名进行预测时，该模块采用的是对应数据集的小字典，这样的做法可以在一定程度上提高识别的精度。

3.低分辨率目标状态估计技术（张）

## (四)阶段成果及其应用情况

具有标志性的阶段技术成果形式、应用方向，成果转化应用情况及效益，关键技术的试验验证情况，以及对装备建设所起的积极作用等。

1.低分辨率视频目标实时分类检测与跟踪技术（邹）

2.低分辨率视频目标文字检测与识别技术（徐）

主要成果：

1. 初步完成文字区域检测与文字OCR技术
2. 完成检测识别结果MySQL数据库存储及Web窗口可视化监控
3. 完成基于文字区域检测与OCR的半自动辅助标注

应用方向：

基于视频目标文字检测与识别技术，可以应用在以字符作为身份区的

的场景，如船名识别、车牌识别等。同时数据库持久化存储，有利于后续的信息检索与查找。在数据标注方面，支持对图片或视频进行手动以及半自动辅助标注。可以为相关的监督学习提供训练样本，同时深度学习辅助的方式节省了标注数据的人工、时间成本。

关键技术实验验证情况：

从MTWI2018数据集中选取100张图片作为验证集（总共包含920个检测框，857个互不相同的字符），利用1-NED（Normalized Edit Distance）量化识别结果与实际的差异程度，字符串A与B的编辑距离lev(m, n)计算公式如下，其中m为A的长度减1，n为B的长度减1。

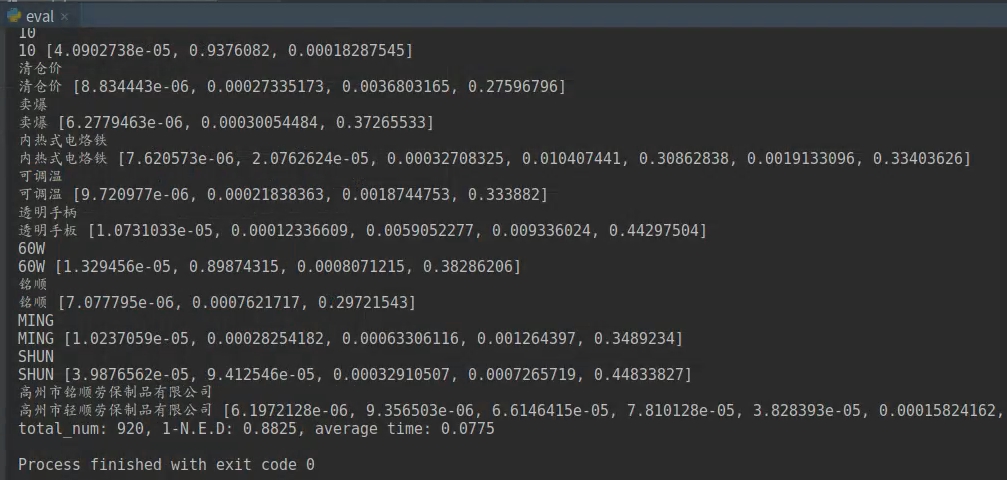


下面就以A=“abcd”，B=“axdde”为例进行简单的计算说明

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | a | x | d | d | e |
|  | 0 | 0 | 1 | 2 | 3 | 4 |
| a | 0 | 0 | 1 | 2 | 3 | 4 |
| b | 1 | 1 | 1 | 2 | 3 | 4 |
| c | 2 | 2 | 2 | 2 | 3 | 4 |
| d | 3 | 3 | 3 | 3 | 2 | 3 |

图1 两字符串的NED计算

从图中可以得出字符串A与B的编辑距离lev(m, n)=3，从而得到两字符串相似度similarity(A, B) = 1-lev(m, n)/max{m, n}= 1/4最终在该衡量指标下，对应的识别精度如图2所示。

图2 OCR识别结果

100张MTWI2018的图片，最终的在NED指标下的平均OCR精度为88.25%，略低于精度90%的性能指标。

3.低分辨率目标状态估计技术（张）

# 三．下一步工作安排与建议

## (一)下一步工作安排

说明下一步需要重点解决的问题、拟采取的措施与实施的风险，对实现原定目标、计划进行预测分析。

1.低分辨率视频目标实时分类检测与跟踪技术（邹）

2.低分辨率视频目标文字检测与识别技术（徐）

优化检测算法

Attention OCR是在检测算法的结果上进行的，所以好的检测算法可为

后续的识别提供较好的平台。最近YOLO-v5的出现掀起目标检测的又一波澜，进一步考量相关算法，看有没有替换的可能性，以提升检测性能。

优化Web界面

Web界面直接面向的是客户，良好的用户体验是产品的门面，之后会进一步根据需求进行界面的优化，借鉴优秀Web界面的形式与策略，在现有的基础上，增强信息交互以及页面模块布局的合理化。

3.低分辨率目标状态估计技术（张）

## (二)建议

1.低分辨率视频目标实时分类检测与跟踪技术（邹）

2.低分辨率视频目标文字检测与识别技术（徐）

进一步提高文字区域检测与识别精度，可以从以下方面进行考虑：

1. 数据样本的优化：数据的可靠性是方法的有效性的前提，而现实

直接相关的有效数据获取是需要较大成本的，若条件允许的话，尽可能用同分布的数据进行研究，使训练的模型对特定问题更具有针对性。

1. 端到端的训练策略：检测与识别单独训练组合，一定程度上会影

响整体性能的全局优化。所以方案许可的话可以考虑end-to-end的方式进行该部分的研究。

3.低分辨率目标状态估计技术（张）

**全文统一格式要求：**

1.页面设置：全篇A4幅面，上、下页边距3.0cm，左、右页边距2.6cm，装订线0cm，页眉1.5cm，页脚2.0cm；页码居中。

2.正文格式：

所有行间距为固定值25磅，对齐网格，首行缩进2字符。

一级标题黑体三号字；二级标题楷体三号字，加粗；三级及四级标题均为仿宋四号字，加粗；正文为仿宋四号字。

标题编号次序为：二→（二）→2→2.1→2.1.1→（1）→①……

3.页码：目录为ⅠⅡ……，正文为1 2……，居中对齐，字体为Times New Roman五号字。

4.图、表编号形成为：二级标题-序号，如图3-3、表2-2等。

5.表格：外框宽度为1 1/2磅；表头黑体小四，其他仿宋小四；表内行距为23磅。

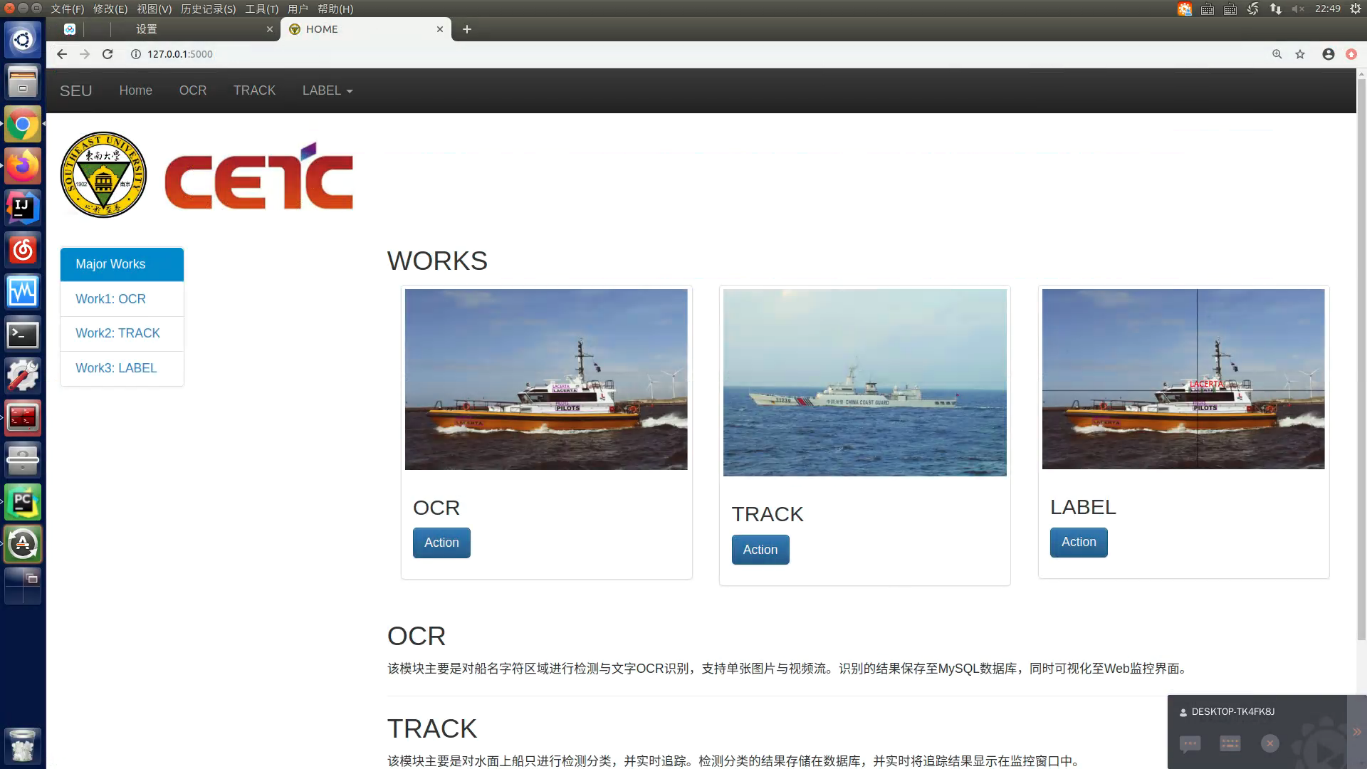
6.成文后封面、目录页和标识页单面打印，其余双面打印。

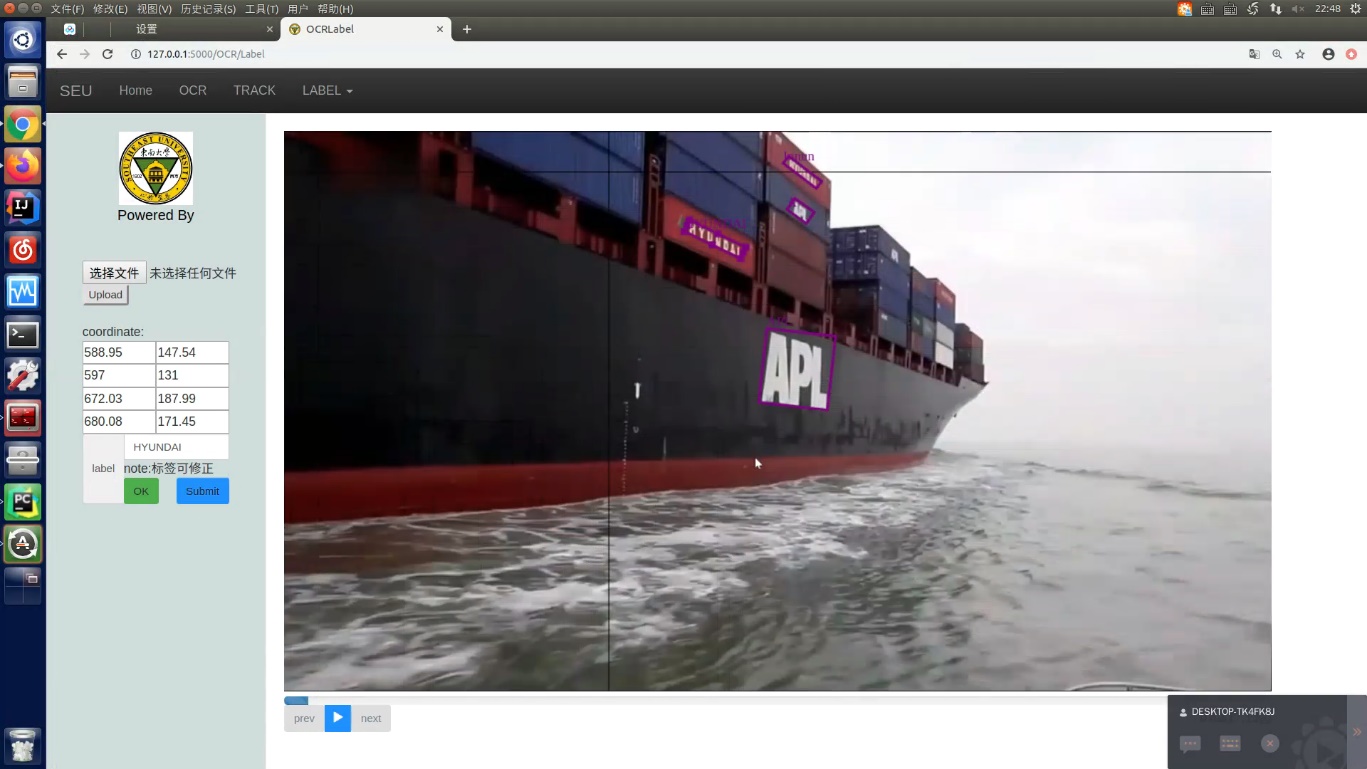
7.其他要求见标示。

8.正式成文后，删除所有红色标注和说明。

附录

部分页面展示

Home页：

OCR-LABEL页：