

机房巡检机器人部分检测模块开发

摘要

机房正常运转是生产活动的基本需求，定期进行机房巡检是对财产安全、人身安全的重要保障。巡检机器人避免了人工巡检的固有弊端，为推动机房无人化、智能化发展提供了重要助力。而视觉检测作为其核心功能之一，目前发展出了大量的相关算法，但大多数算法都存在着严重的使用限制或较长的时间损耗。因此，为补充相关检测算法、提高已有算法的通用性与健壮性，文章针对指示灯、加药箱液位、机械表读数的检测算法进行了研究。

文章利用可见光或红外光信息，应用 Mask R-CNN 卷积神经网络架构，提出了：一种基于多目标识别的控制柜指示灯识别及诊断算法、一种基于实例分割的非透明加药箱液位识别及诊断算法，以及一种基于关键点检测的圆盘式机械表读数识别及诊断算法。实验结果表明，此方法有效简化了人工建模步骤，具有良好的准确性与鲁棒性，可用于复杂环境背景下的巡检任务，具有广阔的发展潜力，为其他巡检算法的实现提供了一定的参考。

关键词：机器视觉，指示灯识别，液位识别，机械表读数识别，Mask R-CNN

Development of Some Detection Modules for Inspection Robot in Machine Rooms

ABSTRACT

The proper functioning of machine rooms is the basic need of production activities. Regular inspection of machine rooms is an important guarantee for property safety and personal safety. Inspection robot avoids the inherent disadvantages of manual inspection and provides an important boost for the unattended and intelligent development of machine rooms. Detection based on computer vision, as one of its core functions, has developed numerous related algorithms. However, most of them are time consuming or have strict restrictions. Therefore, in order to supplement the related detection algorithms and improve the versatility and robustness of the existing algorithms, the article has studied the detection algorithms for indicators, liquid level of dosing boxes, and mechanical pointer meter readings.

The article, using visible or infrared image information, applying Mask R-CNN convolutional neural network architecture, proposes a control panel indicator status recognition and diagnosis algorithm based on multi-object recognition, a dosing box liquid level recognition and diagnosis algorithm based on instance segmentation, as well as a disc-type mechanical pointer meter reading recognition and diagnosis algorithm based on keypoint detection. Experiment results show that this method simplifies the manual modeling process efficiently, possesses good accuracy and robustness and can be used for inspection tasks in complex environments. It has broad development potential and provides a certain degree of reference for the implementation of other inspection algorithms.

Key words: computer vision, indicator recognition, liquid level recognition, pointer meter reading, Mask R-CNN

目 录

| | | |
|-------|--------------------|----|
| 1 | 引言..... | 1 |
| 1.1 | 机房巡检机器人的价值与意义..... | 1 |
| 1.2 | 研究现状..... | 1 |
| 1.2.1 | 关于指示灯识别..... | 1 |
| 1.2.2 | 关于液位识别..... | 2 |
| 1.2.3 | 关于机械表识别..... | 2 |
| 1.2.4 | 关于红外识别..... | 3 |
| 1.2.5 | 关于原始数据采集..... | 3 |
| 1.3 | 本文所作的工作..... | 4 |
| 2 | 基础概念..... | 6 |
| 2.1 | 部分识别任务定性..... | 6 |
| 2.2 | Mask R-CNN 原理..... | 6 |
| 2.2.1 | Backbone..... | 8 |
| 2.2.2 | RPN..... | 8 |
| 2.2.3 | ROI heads..... | 9 |
| 2.3 | 量化评估指标..... | 10 |
| 2.3.1 | 基本指标..... | 10 |
| 2.3.2 | COCO Metric..... | 13 |
| 3 | 方法理论..... | 14 |
| 3.1 | 控制柜指示灯..... | 14 |
| 3.1.1 | 识别对象..... | 14 |
| 3.1.2 | 识别方式..... | 14 |
| 3.1.3 | 诊断方式..... | 15 |
| 3.2 | 加药箱液位..... | 16 |
| 3.2.1 | 识别对象..... | 16 |
| 3.2.2 | 识别方式..... | 17 |
| 3.2.3 | 诊断方式..... | 17 |
| 3.3 | 圆盘式机械表读数..... | 20 |
| 3.3.1 | 识别对象..... | 20 |
| 3.3.2 | 识别方式..... | 21 |
| 3.3.3 | 诊断方式..... | 21 |
| 4 | 实验..... | 24 |
| 4.1 | 控制柜指示灯..... | 24 |
| 4.1.1 | 数据集制作..... | 24 |
| 4.1.2 | 超参数设置..... | 26 |
| 4.1.3 | 实验结果..... | 26 |

装
订
线

| | |
|----------------------|----|
| 4.2 加药箱液位..... | 29 |
| 4.2.1 数据集制作..... | 29 |
| 4.2.2 超参数设置..... | 30 |
| 4.2.3 实验结果..... | 30 |
| 4.3 圆盘式机械表读数..... | 32 |
| 4.3.1 数据集制作..... | 32 |
| 4.3.2 超参数设置..... | 33 |
| 4.3.3 实验结果..... | 34 |
| 5 分析与讨论..... | 37 |
| 5.1 结果分析..... | 37 |
| 5.1.1 指示灯识别结果分析..... | 37 |
| 5.1.2 液位识别结果分析..... | 38 |
| 5.1.3 机械表识别结果分析..... | 39 |
| 5.1.4 综合结果分析..... | 39 |
| 5.2 其他问题讨论..... | 39 |
| 5.3 技术效益评价..... | 41 |
| 5.3.1 经济效益..... | 41 |
| 5.3.2 社会效益..... | 41 |
| 5.3.3 环境效益..... | 41 |
| 6 结论与展望..... | 42 |
| 6.1 结论..... | 42 |
| 6.2 展望..... | 42 |
| 参考文献..... | 44 |
| 谢 辞..... | 46 |

装
订
线

1 引言

1.1 机房巡检机器人的价值与意义

为了生产活动的顺利进行以及生产安全的基本保障，机房巡检是必不可少的一环。传统的机房巡检通常为人工巡检或视频监控，不仅数据可靠性差^[1]、管理困难^[2]，更重要的是，常规巡检以结果检测为主，缺乏成因分析与过程分析，事件发现滞后^[3]；为了解决这一问题，机器人巡检的模式逐渐被开发，该模式不仅可以高效检查相关设备、优化人力成本、减少职业健康风险，而且能够收集多维建筑数据、推进微环境管理、节能减排^[2]，进行故障预测^[4]，具有深远的数据挖掘价值，目前已经成功在变电站^[5]、水电站^[6]、石油站^[7]、数据中心^[2]、公共交通^[8-9]、矿业泵房^[10]、城市基建^[11]等场景应用。

而视觉感知是巡检机器人的重要组成部分，机器人主要依靠机器视觉获取目标信息，识别目标状态，探测故障位置，推断相关设备信息，进而执行相应动作。从相关视觉算法的实现原理来看，当前主要存在两种技术路线：一是建立特征工程，即手动寻找特征的方法；二是采用机器学习，即自主学习特征的方法。就目前而言，建立特征工程方法仍是主要技术手段，自主学习特征方法相对较少。

但不可否认的是，基于特征工程的视觉识别方法大多具有通用性差、耗时长等局限性。而随着近些年深度学习的发展，神经网络在视觉识别上表现出了巨大的发展潜力与应用空间，结合神经网络与工业巡检，成为了巡检机器人的智能化发展新趋势。但是，该方面的研究目前仍处于起步阶段，主要存在着工业数据稀缺、神经网络结合程度参差不齐等问题。因此，根据具体巡检任务确定相应的神经网络迁移方式、收集工业数据建立相关数据集成为了当前迫切需要解决、也是本文主要关注的问题。

1.2 研究现状

该部分介绍了本文涉及研究内容的相关研究现状，包括指示灯识别、液位识别、机械表识别、红外识别与原始数据采集，其涵盖了手动寻找特征与自主学习特征的相关算法，并同时分析了每种算法的优势与局限。

1.2.1 关于指示灯识别

目前指示灯识别算法从人工建立特征工程到自主学习特征方法均有学者进行研究：

(1) 从完全依靠人工建立特征工程方法来看：徐波，张立群^[12]等人先根据指示灯在空间上的先验分布，确定各个指示灯的所在的相对位置及颜色。再利用灰度阈值化与形态学处理方法，结合指示灯分布的先验信息，分割出所有亮灯区域。最后将亮灯区域与其先验分布进行对比，得出指示灯状态。该方法形态学处理方案需进行大量的测试，且受光照等影响分割效果难以保证。

(2) 从特征工程结合自主学习特征的方法来看：陈卓雄，于晓丽^[13]同样基于各个颜色灯分布的先验信息，通过监督学习方法分割所有指示灯所在的整体区域，通过与相同位置拍摄的原始图像比对得到亮灯还是灭灯信息识别结果，但也因此，该方法会对拍摄角度具有一定限制；

胡灿林，肖尚华^[4]则是采用了一种人为在 LAB 颜色空间上，通过最大似然拟合不同颜色属性分布，训练多分类器的方法。该方法类似于人为指定特征空间下构建了一个多层全连接神经网络，即 RGB-LAB-基本颜色-指定颜色，但不同的是其中间穿插了形态学操作，并且参数学习不是根据反向传播（BackPropagation）同时学习所有参数，而是根据期望最大算法（Expectation Maximum）学习部分参数，其余根据人为观测经验决定。相比于完全特征自主学习，该方法一定程度上加强了监督信息，在数据较少时表现会相对较好，但当数据充足时其准确性与鲁棒性将相比较差。

(3) 从以自主学习特征为主的方法来看：Lv L, Tan Y^[5]则采用一种两阶段的方法，首先利用 AlexNet 判断当前图像中是否存在控制柜，若存在则进一步利用 SSD 对控制柜上所有指示灯所在的整体区域进行识别。该方法放宽了拍摄位置的需求，但并没有说明指示灯整体分割后进一步如何处理；

Hao J, Dai M 等人^[6]则是利用 PVAnet 对多种机柜下包含的每一类指示灯、指示灯组及其状态以训练多分类器的方式进行逐一识别，最终再结合不同机柜位置信息下预设的诊断方法，对识别的结果进行相应的逻辑判断。

1.2.2 关于液位识别

液位识别目前研究则以手动建立特征工程为主，并且大多都是针对透明容器的液位识别：

(1) 张宁，田慕琴等人^[7]采用了一种二次二值化的方法对液位对纯净背景下的玻璃管液位计液位进行识别。第一次二值化通过水平投影确定玻璃管边界，第二次在边界定位的基础上再次二值化，并通过垂直投影确定液位边界。该方法对拍摄位置、图像质量均有较高限制。

(2) 李永，东琦等人^[8]同样也是利用水平投影与垂直投影来定位液位计，不同的是他们并非对灰度图像二值化，而是在 HSV 空间内利用三维阈值二值化，但该算法依然有着类似的局限性。

1.2.3 关于机械表识别

对于机械表读取的研究当前同样也是横跨了人工建立特征工程方法到自主特征学习方法：

(1) 从人工建立特征工程方法来看：刘晓洲，龚演平^[9]在表盘灰度图像上利用圆形检测算子检测圆形表盘，利用霍夫直线检测算子检测指针所在线段、起始刻度所在线段与终止刻度所在线段，再通过线段斜率与线段中点所在表盘区域判断线段指定旋转方向下的旋转角，最后通过线性插值得到表记读数；

在 2018 年的 ARGOS Challenge 中，其冠军得主^[7]通过建立特征工程的方式确实取得了较为不错的效果。首先，他们利用 SIFT 算法进行目标图像对预设参考图像的配准并同时提取表盘区域。其次，利用霍夫变换定位指针，通过斜率表示旋转角。最后，根据预设的最大最小刻度旋转角，线性插值得到指针指示刻度。但为了得到准确、稳定的识别效果，其经过大量的调试以获得各类阈值的设定方法，并且在识别过程中也插入了大量的条件判断逻辑。因此，其实际人力投入较大，并且难以保证新环境下的应用不会出现同样繁重的二次人力投入。

(2) 从人工建立特征工程与自主特征学习结合的方法来看：Liu Y, Liu J 等人^[20]利用神经网络

络 Faster-RCNN 对机械表进行定位框回归，并针对小目标检测进行了优化，增加了锚点框的数量与不同尺度的训练数据，采用了 OHEM 算法减少了无效提议区域数量。但定位后的操作依然属于手动建立特征工程方法，且与前文方法类似——Harris 算法图像配准、霍夫变换检测表针，线性插值，同样需已知参考图像及起止刻度位置，其识别效果同样也对成像质量敏感。考虑到这一点，文章同时也提出了一种图像质量检测方法，并通过其反馈调整相机焦距、拍摄距离等，使成像质量满足要求。

1.2.4 关于红外识别

红外成像技术作为重要的工业检测手段之一，其自动识别方法同样在基于人工建立特征工程的角度与基于自主特征学习的角度衍生出了相关的研究：

(1) 从人工建立特征工程方法来看：黄嘉明，胡欣欣等人^[21]利用红外图像对电缆终端异常发热进行诊断。他们首先灰度化红外图像，通过二值化去除背景，再利用基于 canny 边缘检测算子生成的模板匹配电缆终端并对模板相应的矩形区域进行提取，然后在 R、G、B 通道内分别聚类，各通道像素值最高的聚类簇所对应点集的交集认定为疑似高温区域，最终与参考图像对应区域比对，根据标准规范进行诊断。该方法成功应用的前提在于前景背景温度差异较大，才可以有效去除背景、提取边缘，且目标图像尺度需与匹配模板尺度相一致。

(2) 从人工建立特征工程与自主特征学习结合的方法来看：李文璞，谢可等人^[22]在热像图中利用 Faster RCNN 对变电设备进行定位框回归，再利用灰度值推测设备表面温度，最终根据定位框内最高温度与热像图中最高温度进行故障诊断。但该方法识别、诊断的正确率并非十分理想，主要因为热成像相较于可见光丢失了许多目标特征，并且热像图中的温度异常点会干扰诊断。此外文章并没有说明灰度的计算方式，而并非所有方式计算的灰度均与温度存在一一对应关系。

(3) 从自主学习特征方法来看：Liu L, Yan R J 等人^[11]利用 Faster-RCNN 分别对可见光下的围护结构裂缝等外观损伤、与热成像下的墙体内部空洞进行特征识别，结果显示该方法对表面反射辐射造成的干扰具有较强的鲁棒性，且通过引入外界加热器的方式可进一步增强空洞检测的可靠性。该方法并没有直接基于温度值进行诊断，而是利用故障在热像图中的语义特征进行识别，因此降低了对于温度异常点的敏感性。

1.2.5 关于原始数据采集

对于自主学习特征方法而言，其从大量的数据中自我学习有效特征，故也有部分学者对原始数据的获取原则以及获取方法等做出了一定研究。

(1) Hao J, Dai M 等人^[16]在制作机柜、指示灯的数据集时，制定了一套数据采集原则。他们从不同的角度、高度与距离，以视频的方式拍摄所有机柜、指示灯及其状态，然后从视频中以一定帧间隔提取图像，并去除低质量图像，最后以 6: 2: 2 的比例划分数据集。

(2) 刘小波、徐波等人^[23]在制作数字表数据集时，采用了合成数据的方式增广原始数据。在实际采集的真实数据仅有 8500 张图像的情况下，他们通过多种字体与背景的组合合成了 195,000 张图像。在如此庞大的数据集下，他们最终也取得了十分优异的效果。

1.3 本文所作的工作

本文技术路线如图 1.1 所示。从机器视觉角度出发，基于多任务神经网络结构 Mask R-CNN，本文对巡检机器人的部分检测模块进行了开发，分别应用了 Mask R-CNN 的不同分支结构对控制柜指示灯、加药箱液位、圆盘式机械表读数进行了识别以及后续诊断。本文研究内容主要为：

（1）介绍相关理论基础与名词解释，包括：部分视觉识别任务类型、Mask R-CNN 网络结构、神经网络量化评估指标 COCO Metric；

（2）利用定位框分支，训练多目标识别网络对指示灯进行识别，并借助 K-means 聚类与排序的方式对识别结果进行诊断；

（3）利用掩膜分支，训练实例分割网络，分别在经过配准的自然图像与热成像下，对加药箱侧表面及加药箱内剩余液体进行分割，并以两次分割掩膜面积作比的方式对识别结果进行诊断；

（4）利用关键点分支，训练关键点检测网络对圆盘式机械表的指定特征点进行识别，并通过透视变换矫正、向量二次投影、刻度插值的方式对识别结果进行诊断；

（5）整合识别与诊断算法，总体测试检测效果，同时利用 COCO Metric 量化评估训练神经网络的各方面识别能力，针对欠缺之处提出了日后的优化方向；

（6）讨论了部分实现细节，总结本文方法优势，并对发展前景做出了一定的展望。

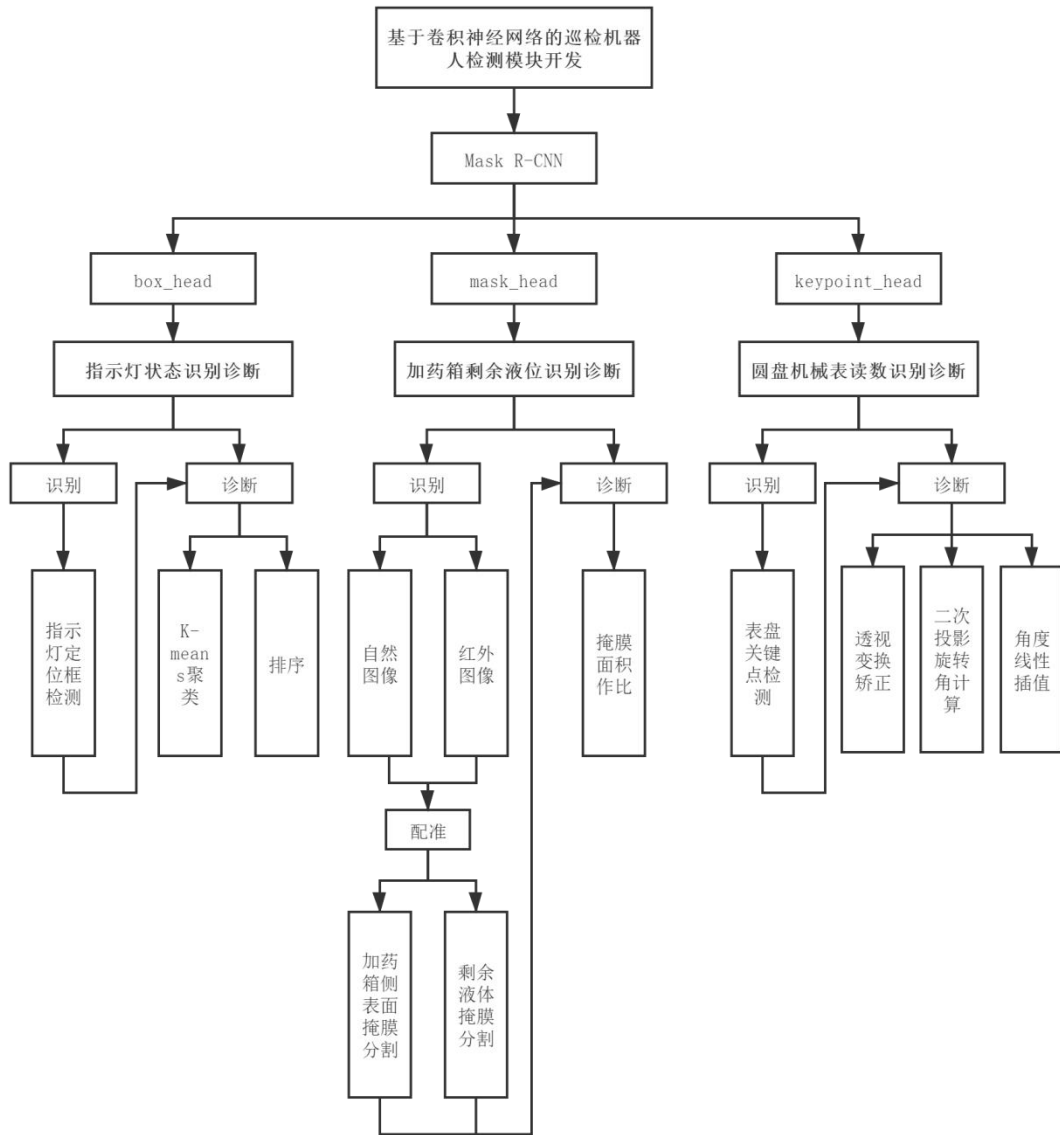


图 1.1 本文技术路线

装
订
线

2 基础概念

以下内容主要介绍本文涉及的基础概念：从基本识别任务分类出发，介绍了卷积神经网络及一种可实现多类别识别任务的神经网络架构 Mask R-CNN，最后引进了一种衡量神经网络多维性能的评估指标 COCO Metric。

2.1 部分识别任务定性

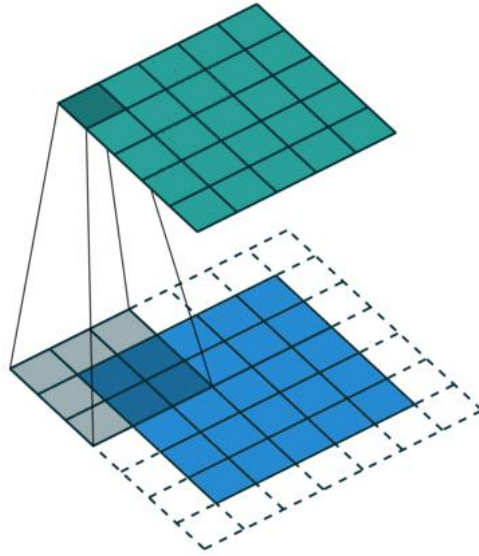
对于同一张图像而言，常常会根据不同的识别需求与后续处理方式，采取不同的技术路线、制定不同的识别任务。恰当选取识别任务有时可以起到简化算法开发、增强算法可拓展性、提高算法稳定性的作用。下面将介绍文本涉及或相关的识别任务性质：

- (1) 图像分类 (Image Classification)：只需判断给定图像的所属种类，最终输出类别标签；
- (2) 目标检测 (Object Detection)：判断给定图像中是否具有前景目标的同时，给出每个目标所在位置，类别仅有前景与背景，最终输出所有定位框坐标；
- (3) 目标识别 (Object Recognition)：判断给定图像中存在的目标类别，给出每个目标所在位置，类别可以有多种，最终输出所有定位框坐标及相应类别标签；
- (4) 语义分割 (Semantic Segmentation)：在给定图像中以像素级别分割所有属于同一类别目标的区域，在同一类别内不区分不同个体，最终输出各个类别的二进制掩膜及相应类别标签；
- (5) 实例分割 (Instance Segmentation)：在给定图像中以像素级别分割所有存在目标的区域并判断目标所属类别，在同一类别内区分不同个体，输出各个个体的二进制掩膜与相应类别标签；
- (6) 关键点检测 (Keypoint Detection)：在给定图像中定位指定关键点，若图像中存在多个含有指定关键点的个体，则分别输出各个个体的关键点坐标。

2.2 Mask R-CNN 原理

卷积神经网络 CNN (Convolutional Neural Network) 属于多层神经网络的一种，常用于图像分析，其结构受生物视觉皮层组织启发，每层单个神经元仅对下层其感受野内（即卷积核内）的刺激作出反应，底层神经网络用于逐级提取特征，上层神经网络则起到组合特征、解释特征的功用。

CNN 中的卷积核的作用与深度神经网络 DNN (Deep Neural Network) 或多层感知机 MLP (Multi-Layer Perceptron) 中相邻两层间的参数类似，同样由可学习的权重与偏置构成，因此同样也可以进行“感知”；但不同的是，卷积核每次仅局部连接上层信息，通过滑动方可遍历上层所有信息，在此过程中卷积核参数共享、维度数目与上层信息一致。一种卷积示意图如图 2.1 所示，其中下层为卷积前输入，上层为卷积后输出，卷积核相关参数：`padding=1`，`kernel_size=3`，`stride=1`。

图 2.1 卷积过程示意图^[24]

与传统图像分析算法相比，卷积神经网络需要更少的图像预处理操作，因此这意味着它可以通过学习便生成传统手工构建的滤波器。这种对特征构建先备知识、特征构建人力劳动的低依赖性成为了其相比传统算法的一大优势。

而 Mask R-CNN^[25]便是 2017 年由 He Kaiming 等人提出的一种双阶段多任务卷积神经网络结构，主体结构由 Backbone、RPN、ROI heads 三部分构成，设计初衷用于实例分割，但由于其灵活的构架亦可单独用于图像分类、目标识别、语义分割、关键点检测等其他任务。

图 2.2 为 Mask R-CNN 的一种具体网络结构示意图，其中 conv 代表卷积层、deconv 代表转置卷积层、linear 代表全连接层，括号内参数依次为输入通道数、输出通道数、卷积核大小（其余参数省略），虚线箭头表示并非直接相连。下面将逐一介绍该网络结构的各部分功能。

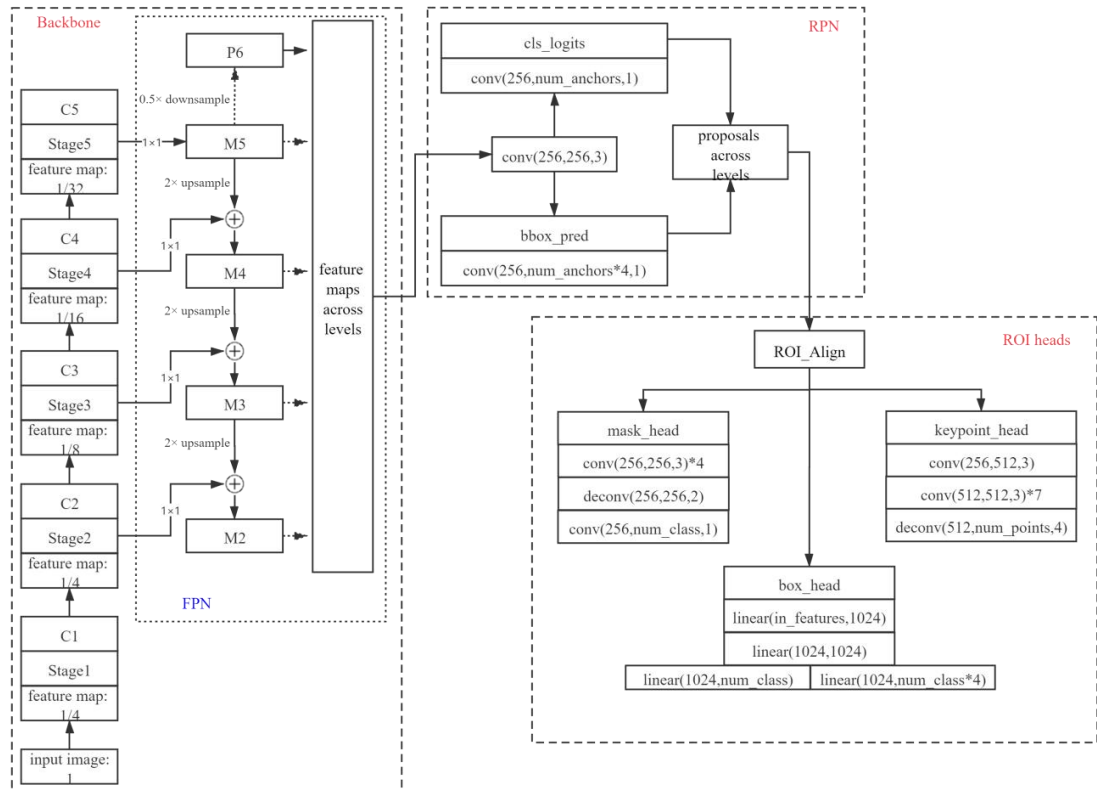


图 2.2 Mask R-CNN 结构示意图

2.2.1 Backbone

Backbone 的主要作用是特征提取器，其从原始图像中为 RPN 与 ROI heads 提供共享的图像高维度特征图，其具体结构可以是任意的卷积网络，如 ResNet、VGG 等，图 2.2 中选择的是带有 FPN 的 ResNet-50 结构。

ResNet-50 具有五个阶段——Stage1~Stage5，Stage2 开始每个阶段由一个池化块与若干个残差块组成，池化块降低分辨率、提取主要信息，同时减少计算量；残差块则通过跨层连接减轻梯度消失与退化。

FPN 则提取 ResNet 各计算过程中输出的特征图，并返回一系列新的特征图，目的在于提高不同尺度目标识别效果。高层的特征图单位感受野更大，语义信息更丰富，但分辨率较小，丢失了部分细节信息；低层特征图虽感受野较小、语义较贫瘠，但保留小尺度目标的特征。故 FPN 由上至下，通过侧向连接前一阶段的特征图，使得在保留较高语义信息的基础上获得了更高的分辨率，有效改善了小目标的检测效果。

2.2.2 RPN

RPN (Region Proposal Network) 的主要作用是生成建议区域，即对原始图像中所有可能存在目标的区域进行定位、缩小后续数据处理量，本质上为一种注意力机制。RPN 接收 Backbone 生成的特征图以及原始图像，特征图中每一点都会在原始图像的相应位置上，以定位框的方式映射

出不同大小的矩形区域，即所谓的 Anchor。RPN 便判断这些矩形区域内是否存在目标，并对存在目标的定位框进行一定的位置矫正，最终再通过置信度排序、非极大值抑制等手段去除冗余定位框，生成建议区域。

此外，Mask R-CNN 也正是基于 RPN 返回的多个建议区域，结合 ROI heads 实现了单张图像中的多目标多个体识别。

2.2.3 ROI heads

ROI heads 是 Mask R-CNN 结构中实际实现特定目标识别功能的部分。Backbone、RPN 的处理目的均在于为其减少重复、无效的计算，并向其提供富含目标信息的特征图，ROI heads 便基于这些特征图，通过相应网络结构来实现分类与回归任务。

A. ROI_Align

ROI_Align 位于每个 head 的起始部分，其主要作用是获得固定大小输出的特征图并一定程度上维持边缘像素信息不丢失，是 Mask R-CNN 网络相较先前网络结构的一个亮点。ROI_Align 直接接收 FPN 的特征图以及 RPN 在原始图像尺度下的建议区域坐标，通过一定的映射以及池化操作，最终返回不随输入变化的固定尺寸特征图。具体过程如下：

- (1) 首先，其将建议区域坐标映射到相应各尺度特征图上的坐标，并保留坐标浮点数；
- (2) 其次，在特征图上将建议区域内平均划分 $m \times m$ 个 bin， $m \times m$ 即为固定输出尺寸，具体数值与后续网络要求有关，一般 box_head 中 $m=7$ ，mask_head 与 keypoint_head 中 $m=14$ ；
- (3) 然后，在每个 bin 内均匀生成 n 个采样点，每个采样点的数值则是根据最邻近的四个像素点通过双线性插值得到，一般 $n=4$ ；
- (4) 最后，在每个 bin 内进行池化操作，即可得到 $m \times m$ 的特征图，示意图如图 2.3 所示。

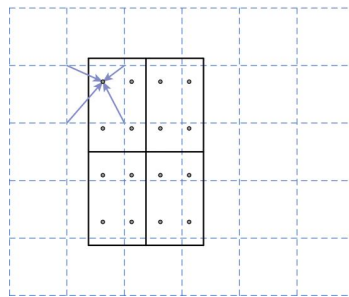


图 2.3 ROI_Align 示意图^[25]

关于双线性插值，见图 2.4 与式 (2.1~2.3)，其中 R 即等效为 bin 内采样点。

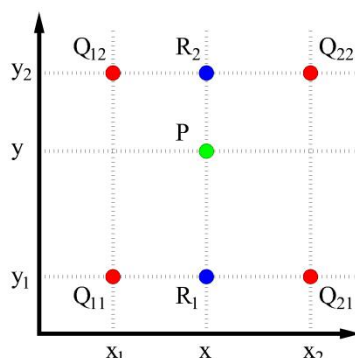


图 2.4 双线性插值示意图

$$f(R_1) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}) \quad (2.1)$$

$$f(R_2) \approx \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}) \quad (2.2)$$

$$f(P) \approx \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2) \quad (2.3)$$

B. box_head

box_head 的作用是判断目标类别，并二次修正定位框坐标。它接收 ROI_Align 返回的特征图，通过若干全连接层后，输出该特征图所属类别标签及相应定位框坐标。

C. mask_head

mask_head 的作用是在像素级别上分割目标所在区域。图 2.2 中所采用的是一种简化的 FCN (Fully Convolutional Network) 结构，即全卷积神经网络，它接收 ROI_Align 返回的特征图后，通过一系列卷积层、转置卷积层，最终输出每个像素的类别标签。值得一提的是，这里的标签并非是多分类标签，而是二元标签，这样做的目的是提高网络表现，减少类间竞争。

D. keypoint_head

keypoint_head 的作用是在像素级别上定位关键点。图 2.2 中其结构与 mask_head 类似，同样是通过卷积层、转置卷积层处理。不同的是，mask_head 中输出的每张掩膜对应每个个体目标，可以有多个像素点被标记；keypoint_head 中每张掩膜对应每个关键点，只有一个像素可以被标记。

2.3 量化评估指标

2.3.1 基本指标

A. IOU

对于定位框回归、掩膜分割问题而言，IOU (Intersection over Union) 可以作为检测值与真实值之间距离衡量的一个指标。其计算如图 2.5，即相交部分面积与相并部分面积比值，其值常作

装订线

为判断检测结果是否为阳性的阈值。

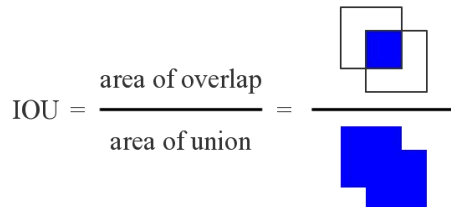


图 2.5 IOU 计算示意图

B. OKS

对于关键点检测问题而言，OKS（Object Keypoint Similarity）起到与 IOU 类似的作用，用于衡量检测值与真实值之间距离。其计算遵循式（2.4~2.5）：

$$OKS = \frac{\sum_i ks_i \delta(v_i > 0)}{\sum_i \delta(v_i > 0)} \quad (2.4)$$

$$ks_i = e^{-\frac{d_i^2}{2s^2k^2}} \quad (2.5)$$

其中 d 为检测关键点到真实关键点之间的欧式距离；

s 用于描述关键点从属对象尺度，可以用定位框面积表示；

k 为一常数，用于描述人为标注不同关键点时的落点范围偏差，一般需大量冗余注释的图像，再根据式（2.6）计算每类关键点在各自尺度 s 下的标准差 σ ，最终取 $k=2\sigma$ ，其中 d' 为各散点到计算中心点的欧式距离，脚标 j 表示第 j 类关键点；

$$\sigma_j^2 = E\left[\frac{d_j'^2}{s^2}\right] \quad (2.6)$$

v 表示可见性， $v=0$ 表示未标注； $v=1$ 表示标注但不可见； $v=2$ 表示标注且可见；

$\delta(\cdot)$ 在内部条件成立时等于 1，否则等于 0；

脚标 i 表示同一个对象的第 i 个关键点。

C. 混淆矩阵

可假定 IOU 或 OKS 值域内的某一值为临界阈值，当 IOU 或 OKS 大于该阈值时认为检测结果为阳性，否则为阴性，两种情况下分别对应真实值的阳性与阴性，便产生四种结果：TP（True Positive），FP（False Positive），FN（False Negative），TN（True Negative），其共同构成混淆矩阵（confusion matrix），如图 2.6。

| | | | |
|------|----|-----|----|
| 混淆矩阵 | | 预测值 | |
| | | 正例 | 反例 |
| 真实值 | 正例 | TP | FP |
| | 反例 | FN | TN |

图 2.6 混淆矩阵

D. Precision 与 Recall

对某一数据集进行检测，可得到混淆矩阵。准确率 Precision 为所有正确识别的阳性样本与所有识别为阳性样本的比值，表示模型识别正确的比重，可根据式（2.7）计算；召回率 Recall 为正确识别阳性样本与所有真实阳性样本比值，表示模型查全的比重，可根据式（2.8）计算。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} = \frac{\text{TP}}{\text{all detections}} \quad (2.7)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{\text{TP}}{\text{all ground truths}} \quad (2.8)$$

E. AP 与 AR

a. AP（Average Precision）与 mAP（Mean Average Precision）

针对某一类别，通过由高至低调整识别置信度阈值，可以得到一系列关于该类别的 Precision 与 Recall 值。二者一般相互制约，一者较高时，另一者往往较低，并且容易出现局部波动，因此为了比较不同模型的 Precision-Recall 曲线，便提出了另一种度量方式——AP。AP 具有多种计算方法，但中心思想均在于对 Precision-Recall 曲线积分。

这里介绍一种插值计算 AP 的方法，其计算公式为式（2.9~2.10），其中 p 为 Precision， r 为 Recall，脚标表示根据 Recall 升序排列的散点序号。

$$\text{AP} = \sum_{r=0}^1 (r_{n+1} - r_n) p_{\text{interp}}(r_{n+1}) \quad (2.9)$$

$$p_{\text{interp}}(r_{n+1}) = \max(p(r \geq r_{n+1})) \quad (2.10)$$

mAP 即为 AP 在所有类别上的平均值。

a. AR（Average Recall）与 mAR（Mean Average Recall）

同样，对于某一类别，调整临界阈值 IOU(OKS)，可得到一系列 Recall 值，AR 根据 Recall-IOU (OKS) 曲线也有不同的积分计算方法，这里给出其中两种^[26]，分别为式（2.11）与式（2.12），其中 r 为 Recall， o 为 IOU，脚标表示散点序号。

$$\text{AR} = 2 \int_{0.5}^1 r(o) do \quad (2.11)$$

$$\text{AR} = \frac{2}{n} \sum_{i=1}^n \max(o(r_i) - 0.5, 0) \quad (2.12)$$

mAR 即为 AR 在所有类别上的平均值。

2.3.2 COCO Metric

COCO 是微软团队发布的一个大规模目标检测、分割、注释的数据集，而 COCO Meric 便是用于评估不同模型在该数据集下表现的衡量指标。该指标由多个子指标构成，分别衡量指定模型不同维度的识别能力优劣，如识别大中小目标能力、识别值与真实值重合程度等，因此不仅可用于比较模型好坏，亦可为模型进一步优化提供指导，本文便采用其为网络量化评估指标。

该评价指标大体上可划分为 mAP 与 mAR 两个方向，每个方向下通过更改不同属性最终构成所有子指标，其中属性包括：

A. IOU/OKS

即前文所提及的判定识别目标阳性/阴性的临界阈值，一般为 0.5，阈值越高说明对识别结果的要求越严格、越需要更高的定位精度；反之，若对定位精度无过高要求，可放低该阈值，但一般不低于 0.5。

在该指标中共出现 0.5 到 0.95 十个阈值，“0.50:0.95”表示相应指标为根据这十个阈值计算的平均值。

B. Area

Area 用于控制采样目标大小。该指标中以像素为单位，默认规定标注目标定位框或掩膜的面积小于 32^2 ，其值为“small”；面积大于 32^2 小于 96^2 ，其值为“medium”；面积大于 96^2 ，其值为“large”；若无论大小，其值为“all”。

C. MaxDets

MaxDets 表示从每张图像中按置信度降序最多取多少个识别结果用于指标计算，指标默认值有 1、10、20、100。

此外，有两点仍需说明：一，该指标中的 AR 并非为 Recall-IOU（OKS）曲线积分而来，而是在指定临界阈值下，根据 MaxDets 从数据集中所采的全部样本计算 Recall 而来；二，若某项指标值为-1，表示无满足该指标描述样本，该指标未曾执行计算。

3 方法理论

机房巡检包含多项巡检任务，即使是相同的巡检任务也存在不同的技术路线，该部分内容主要介绍本文如何实现相关算法开发、如何具体迁移神经网络用于相关识别，以及如何根据识别结果进行相应诊断。

3.1 控制柜指示灯

3.1.1 识别对象

如图 3.1，机房机电设备一般配有相应控制柜，常见的控制柜有电气控制柜、变频控制柜、电源控制柜、水泵控制柜、空调控制柜等等。控制柜上存在多种指示灯用于描述相应设备的当前状态。指示灯具有不同的颜色以及明灭状态，不同光照条件下的显示效果也略有差异。机房巡检中通过指示灯的显示状态来判断相应设备是否符合正常工作预期，以起到设备检修、预防生产灾害、校验 BA 系统等作用。

该部分工作的识别对象即为控制柜上的所有指示灯。考虑到机器人检测的控制柜具有不同的指示灯分布，且出于定位精度问题，对于同一控制柜，机器人每次拍摄图像角度、方向等可能略有差异，因此设定算法需识别指示灯的位置、颜色及相应明暗状态。



图 3.1 机房控制柜

3.1.2 识别方式

对于指示灯识别，将任务定性为多目标识别问题，即训练神经网络定位并识别不同颜色、明暗的指示灯，不同颜色指示灯及其相应明暗状态应分属于不同目标类别，如有 n 种颜色指示灯，则类别标签总数应为 $2n+1$ ，其中单独的“+1”表示背景标签。

基于 Mask R-CNN 网络架构，具体指示灯识别网络结构 Backbone 采用 Resnet50+FPN，ROI heads 仅保留 box_head 分支，采用迁移学习的方式仅训练该分支最后两层参数（图 2.2）。

3.1.3 诊断方式

经神经网络识别后，将按置信度降序返回一系列检测到的指示灯标签及相应定位框坐标，但此时的指示灯信息各自孤立。为整体识别指示灯面板状态，仍需对其进行编码，其原理如图 3.2。

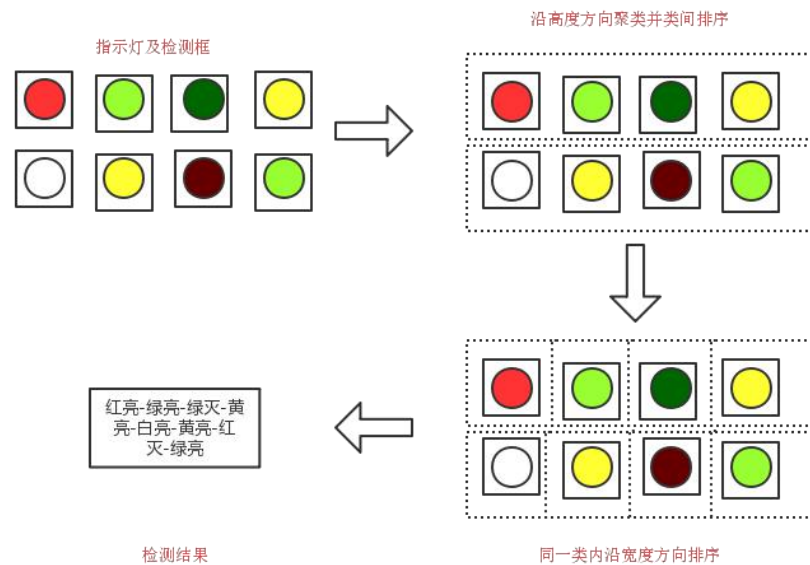


图 3.2 指示灯识别结果诊断原理

(1) 首先，人为给定聚类中心数，根据定位框垂直方向坐标，对所有标签进行 K-means 聚类；

(2) 其次，在垂直方向上对所有标签进行排序，此时聚类信息保持不变，可与第一步操作顺序互换；

(3) 最后，在同一类内，根据水平方向坐标对标签进行排序。

此时即可按照一定规则，如由上至下、从左至右的顺序，对所有指示灯信息进行编码，进而整体识别指示灯面板状态。

识别诊断总流程如图 3.3 所示。

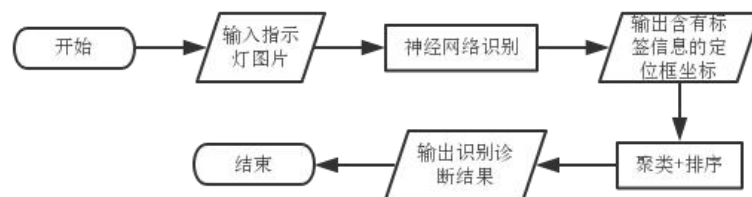


图 3.3 指示灯识别诊断总流程

关于 K-means 聚类，则是一种基于距离最短原则，对相似样本进行分类的方法。其流程，如图 3.4，可大致划分为：聚类中心初始化、距离计算、聚类、收敛判断、聚类中心再生成。本文初始化采用的方法是在最大最小坐标值之间等距离取点，距离采用是欧式距离，收敛条件为所有点到其聚类中心距离均小于 1 个像素距离。

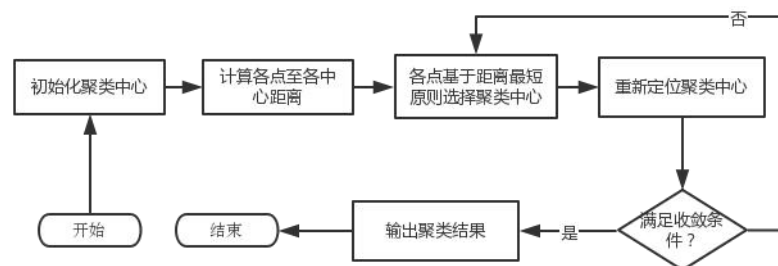


图 3.4 K-means 聚类原理

3.2 加药箱液位

3.2.1 识别对象

如图 3.5，当机房存在水处理设备时，会配备存放药液的加药箱。加药箱一般由 PE 塑料制成，表面耐腐蚀且非透明。随着水处理过程的进行，药液会不断减少，当药液减少到一定程度时需启动加药泵补充药液。机房巡检中需检测加药箱内液位是否处于正常范围，进而控制加药泵启停或推断水处理设备是否正常工作。

为实现液位识别，该部分工作的识别对象分为两部分：

一是自然图像中加药箱侧表面的识别，考虑到加药箱一般摆放在同一位置、机器人拍摄图像中包含多个加药箱，因此需设定算法在像素级别上逐一分割图像中每个加药箱的侧表面；

二是热成像中剩余液体的识别，考虑到在自然图像中已将不同加药箱区分开来，因此仅需设定算法在像素级别上分割所有剩余液体所在区域。



图 3.5 加药箱自然图像与热成像

3.2.2 识别方式

因加药箱表面非透明，所以液位的识别需借助双光摄像机，结合可见光信息与红外信息对剩余液位进行识别。其原理在于空气与药液的热物性不同^[27]，当存在外界热源干扰，如阳光照射时，热量会被加药箱内部介质吸收，一般药液的比热容较大，升温较慢，其覆盖的侧表面温度较低；而空气比热容较小，升温较快，其对应的侧表面温度较高。因此，加药箱的侧表面就会表现出一定的温度梯度，即可利用热成像技术对其内部的剩余液体进行观测。

液位识别分为两阶段，第一阶段进行加药箱定位，第二阶段对定位加药箱内的剩余液体进行检测。第一阶段是在自然图像下的识别，任务定性为实例分割，即训练网络分辨所有加药箱，并分割每个加药箱的侧表面；第二阶段在第一阶段生成的单个加药箱掩膜抑制下，对热成像进行识别，任务定性为语义分割，即训练网络分割所有液体区域。但当目标仅有一个时，语义分割与实例分割目的、效果相同，因此从简化工作的角度来看，实际上可以继续沿用第一阶段的实例分割网络结构。

基于 Mask R-CNN 网络架构，可见光与热成像下识别网络结构 Backbone 均采用 Resnet50+FPN，ROI heads 保留 box_head 与 mask_head 分支（图 2.2）。训练同样采用迁移学习的方式，经过测试，对于 box_head，二者均仅训练最后两层；而对于 mask_head，可见光需训练整个分支参数，热成像可仅训练最后两层参数。

3.2.3 诊断方式

A. 图像配准

双光摄像机拍摄的自然图像与热成像具有不同的成像范围以及分辨率，为使可见光下加药箱的位置信息有效映射到热成像下，并去除热成像中的温度标尺，需对二者进行配准，使二者像素点一一对应，成像范围基本一致。

具体配准参数，可根据双光摄像机的固有参数确定。对于本文而言，红外摄像仪型号为“FLIR E6 Wifi”，其拍摄的自然图像分辨率为 640×480 、热成像分辨率为 320×240 ，其余技术参数见表 3.1。

表 3.1 FLIR E6 红外热像仪参数

| 参数 | 性能 |
|-------|--------------|
| 红外分辨率 | 160×120 |
| 热灵敏度 | <0.06 °C |
| 空间分辨率 | 5.2 mrad |
| 测温范围 | -20~250 °C |
| 精度 | ±2 °C或读数的 2% |

本文参考罗为^[28]的方式，配准时先对自然图像进行裁剪，垂直方向取第 100 行到第 554 行像素，水平方向取第 90 列到第 459 列像素；对热成像则先缩放到分辨率为 480×370，再进行裁剪，垂直方向取所有像素，水平方向取倒数第 25 列前所有像素。配准效果如图 3.6 所示。

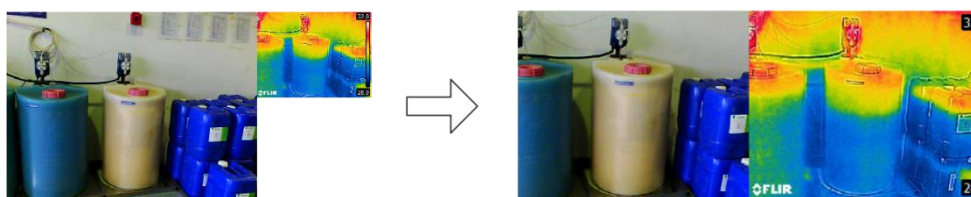


图 3.6 通过裁剪与缩放对自然图像与热成像配准

除上述方法以外，若无法获得双光摄像机相关参数，尚存在一种更为通用的配准方式，即借助仿射变换。该方法可适用于各类双光摄像机，且无需反复调试裁剪区域与缩放区域大小，仅需人为在成组的自然图像与热成像间标注匹配的三组关联点，即可计算仿射变换矩阵，配准自然图像与热成像。

仿射变换是在二维平面上的一种变换，是线性映射与平移的组合，可借助变换矩阵，即式 (3.1~3.3)，来描述。

$$(x, y, w') = (u, v, w) \cdot T = (u, v, w) \cdot \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad (3.1)$$

$$(X, Y) = \left(\frac{x}{w'}, \frac{y}{w'} \right) \quad (3.2)$$

$$(U, V) = \left(\frac{u}{w}, \frac{v}{w} \right) \quad (3.3)$$

其中， T 为仿射变换矩阵， (U, V) 为变换前点坐标， (X, Y) 为变换后点坐标， $w=w'=a_{33}=1$ ， $a_{31}=a_{32}=0$ ， T 中其余 6 个 a_{ij} 则是通过标注的三组关联点，联立六个独立方程解出（一组关联点可

以生成两个独立方程）。

通过仿射变换进行配准的效果如图 3.7 所示，其中左边为热像仪拍摄的一组自然图像与热成像，右边为自然图像配准可见光图像后重叠显示的效果。配准后温标等水印已被去除，热成像中存在部分白色边缘，为热像仪勾勒的物体边缘，但勾勒位置并不准确。

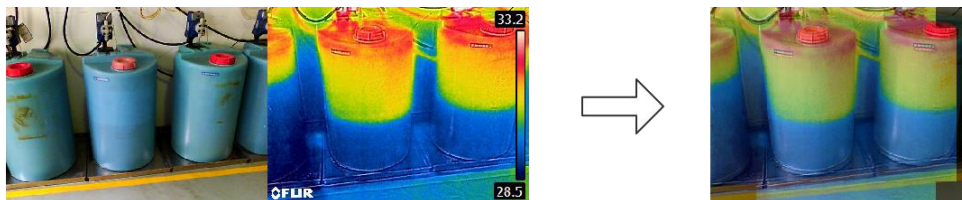


图 3.7 通过仿射变换对自然图像与热成像配准

B. 面积作比

经图像配准后，首先将自然图像送入加药箱分割网络，网络将返回图像中每个加药箱侧表面的二值掩膜位置；其次，逐一地利用这些掩膜图像对热成像进行抑制，并送入液体分割网络中，网络则将返回剩余液体的二值掩膜位置；记掩膜图像面积为二值图像中所有像素点像素和，如图 3.8，A 表示剩余液体掩膜面积，A+B 表示加药箱侧表面掩膜面积。最后，根据式（3.4）将剩余液体掩膜图像面积与相应加药箱掩膜面积作比，即可诊断当前剩余液体百分比。

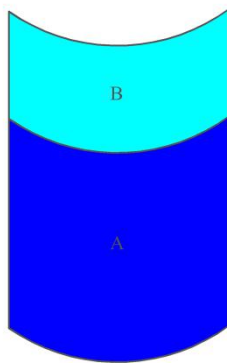


图 3.8 掩膜面积

$$\text{Remain} = \frac{A}{A+B} \times 100\% \quad (3.4)$$

识别诊断总流程如图 3.9 所示。

装订线

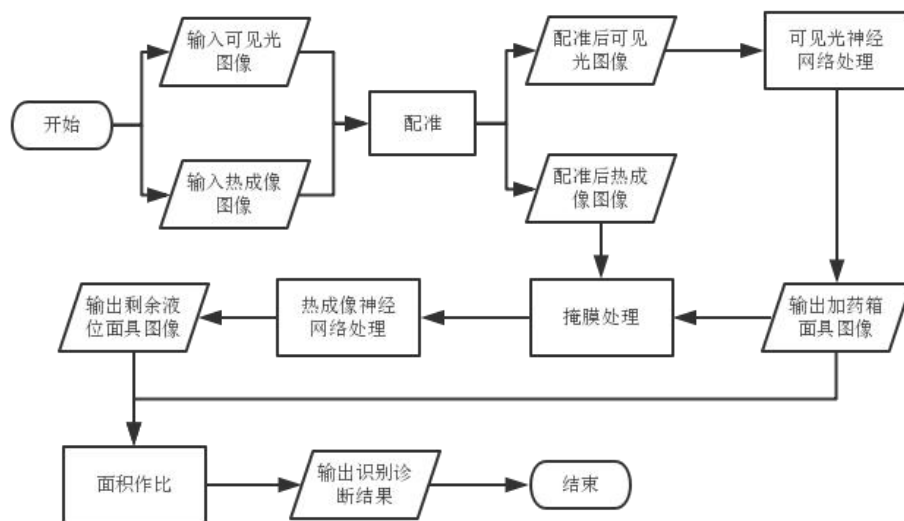


图 3.9 加药箱液位识别诊断总流程

3.3 圆盘式机械表读数

3.3.1 识别对象

如图 3.10，圆盘式机械表广泛存在于机房之中，用于测量介质压力、温度等，常安装在各类阀门前后，其安装位置、角度随管道铺设条件而多变，表面常存在阴影遮挡或光照反射。机房巡检中需记录每个机械表的读数，并判断阀门是否启闭正常、整个管路系统是否正常运行，以避免炸管等重大生产事故的发生。

考虑到现场条件机器人可能无法保证得到机械表的正面图像，且不同机械表盘刻度分布样式不同，因此该部分工作的识别对象设定为表盘平面的八个指定关键点，分别为——指针转轴中心、指针末端、刻度起始点、刻度终止点，以及表盘上下左右四个定位点，算法需定位每个关键点的二维横纵坐标。



图 3.10 圆盘式机械表

装订线

3.3.2 识别方式

对于机械表读数识别，将任务定性为关键点检测，即训练网络识别上述八个关键点的坐标。坐标生成后，通过后续诊断便可推断相应机械表读数。

基于 Mask R-CNN 网络架构，机械表关键点识别网络结构 Backbone 同样采用 Resnet50+FPN，ROI heads 则保留 box_head 分支与 keypoint_head 分支，训练采用迁移学习的方式训练 box 分支的最后两层与 keypoint 分支的最后一层（图 2.2）。

3.3.3 诊断方式

诊断流程分为四个步骤，借助上述识别的关键点坐标，对图像进行矫正并计算关键点之间的相对关系，进而推断机械表读数，具体步骤如下。

A. 透视变换矫正

若拍照时相机镜头平面与机械表表盘平面并非平行，则实际形状为圆形的表盘便会发生透视变换，在实际成像图像中表现为椭圆形。为得到表盘的正面图像，矫正因透视变换带来的读数误差，则需再执行一次透视变换，以抵消成像透视变换的影响（如图 3.11）。具体变换同样遵循式（3.1~3.3）。

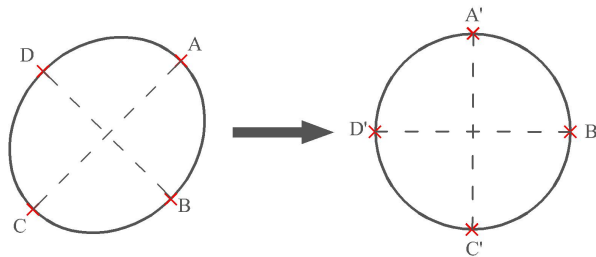


图 3.11 透视变换

不同的是，其中 T 为透视变换矩阵， (U, V) 为变换前点坐标， (X, Y) 为变换后点坐标， $w=w'=a_{33}=1$ ， T 中其余 8 个 a_{ij} 则是通过给出变换前后四个点，联立八个独立方程解出（每一组点可产生两个独立方程）。对于该问题而言，变换前四个点即表盘上下左右四个定位点，变换后四个点取 600×600 正方形四个边的中点。

B. 反三角函数计算夹角

通过以上透视变换，得到矫正后的转轴中心点 O 、指针末端点 P 、刻度起始点 A 与刻度终止点 B 坐标，再利用向量内积与反三角函数，式（3.5~3.6），分别计算指针与起始刻度所成夹角 $\angle POA$ 、起始刻度与终止刻度所成夹角 $\angle AOB$ 。

$$\vec{OP} \cdot \vec{OA} = |\vec{OP}| |\vec{OA}| \cos(\angle POA) \quad (3.5)$$

$$\angle POA = \arccos(\cos(\angle POA)) \quad (3.6)$$

C. 二次投影计算优角

借助向量内积与反三角函数仅能识别劣角（即小于 180° 角），为使模型能够识别优角（即大于 180° 角），本文采用一种二次投影的方法，其原理如图 3.12 所示。

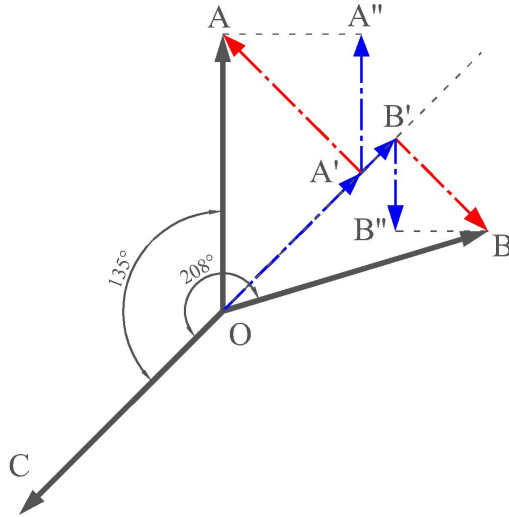


图 3.12 二次投影

$\angle AOC$ 小于 180° ， $\angle BOC$ 大于 180° ，分别将 \vec{OA} 、 \vec{OB} 向 \vec{OC} 方向投影，并得到 \vec{OA} 、 \vec{OB} 垂直 \vec{OC} 方向分量 $\vec{A'A}$ 、 $\vec{B'B}$ ，再将 $\vec{A'A}$ 、 $\vec{B'B}$ 向竖直方向投影得到 $\vec{A''A''}$ 、 $\vec{B''B''}$ ，可以发现当夹角小于 180° 时，二次投影后分量（如 $\vec{A''A''}$ ）竖直向上；当夹角大于 180° 时，二次投影后分量（如 $\vec{B''B''}$ ）竖直向下。

故，若二次投影后向量分量竖直向上则维持上一步计算夹角不变；若二次投影后分量竖直向下则用 360° 减去上一步计算所得夹角。至此，模型可以识别 360° 内所有角度。

D. 角度线性插值

最后，若刻度为均匀分布，给出起始刻度值与终止刻度值，便可利用线性插值，式 (3.7)，计算当前指针所指刻度，其中 A_{pointer} 为逆时针方向指针与起始刻度间夹角， A_{range} 为逆时针方向起始刻度与终止刻度间夹角， max 、 min 指代最大最小刻度值。此外，考虑到实际机械表可能由于故障或其他原因导致指针并不在量程范围内，即指针与起始刻度所成夹角将大于起始刻度与终止刻度所成夹角，此时为突出显示，指针与起始刻度所成夹角将减去 360° ，以使插值结果为负。

$$R = \frac{A_{\text{pointer}}}{A_{\text{range}}}(R_{\text{max}} - R_{\text{min}}) + R_{\text{min}} \quad (3.7)$$

识别诊断总流程如图 3.13 所示。

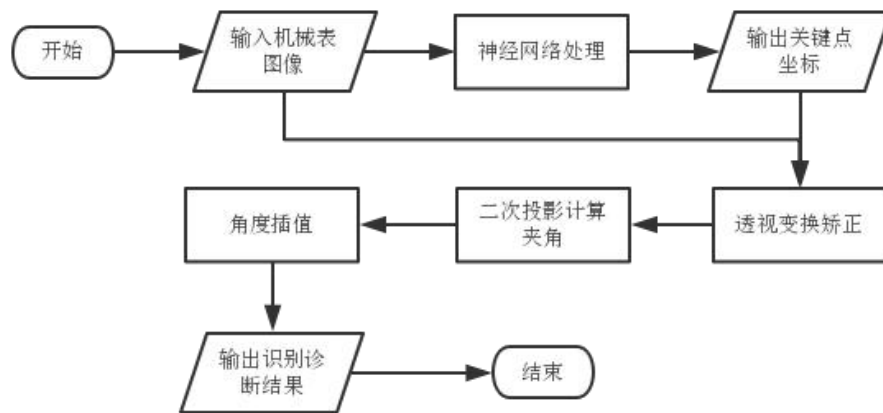


图 3.13 机械表读数识别诊断总流程

装
订
线

4 实验

该部分主要通过实验验证上述提及的方法理论。首先，介绍了各自数据集制作方式，包括数据来源、标注原则、以及数据处理；其次，对相关超参数的设置进行了一定的说明；最后，给出实验结果，包括训练过程中的损失观测、各类型量化评估指标以及识别诊断效果图。

4.1 控制柜指示灯

4.1.1 数据集制作

A. 数据来源

对于 Mask R-CNN 而言，其需要从大量的数据中学习特征，因此原始数据的获取是实验的第一步。完整现场工业数据一般较难获取，本文指示灯数据采用的是提取网络视频的折中手段。首先以一定帧间隔对视频进行图像采样，其次剔除不清晰图像及不符合标注原则图像，最后筛选原始图像共计 79 张。

B. 标注原则

结构化训练数据是监督类学习感知目标的原点，而实现结构化的方法便是对原始数据进行标注。数据标注质量将直接影响网络学习的方向与效果，因此数据标注应根据目标属性与实际应用环境进行统一原则性设定。对于指示灯识别，本文采用的标注原则主要有以下两点：

a. 外观完整度

指示灯遮挡超过 1/4 一律不标注，若指示灯侧面露出，应标注在定位框内。

b. 颜色一致性

值得注意的是，指示灯由于从不同厂家生产，其颜色可能存在一定色差。部分厂家黄灯点亮时颜色可能与其他厂家红灯熄灭时颜色接近，故标注时应注意同一标签对应的颜色基本一致，色差较大样本应直接剔除。也正因此，该实验中数据集并未成功收集到所有颜色指示灯及相应明灭状态信息。

C. 数据增强

普遍认为，训练数据越多模型的准确性与泛化能力越好，但原始数据数量一般受实际条件限制，而数据增强便是通过人工干涉的方式，补充训练数据数量、增加训练数据多样性，进而提升模型表现。

数据增强本文采用的在线增强的方式。相比于离线增强需事先标注好增广数据，所谓在线增强，仅每次读取数据时同时对原始数据及其标注进行指定变换，而无需重复标注。对于该实验而言：

100%概率对原始图像亮度进行 $\pm 20\%$ 之间的变换；

50%概率对图像进行水平翻转；

50%概率对图像进行竖直翻转。

D. 数据一般化

为降低网络对于输入数据多样性的敏感程度，在图像实际送入网络前，还需对其进行一般化处理，本文采用的一般化手段包括图像正规化以及图像尺寸重设：

a. 图像正规化

图像正规化主要具有两点作用：

一是调整数据分布，平衡每张图像更新梯度的大小，优化梯度下降方向，减小损失震动，加快训练速度，示意图如图 4.1，正规化后网络更易达到最优解；

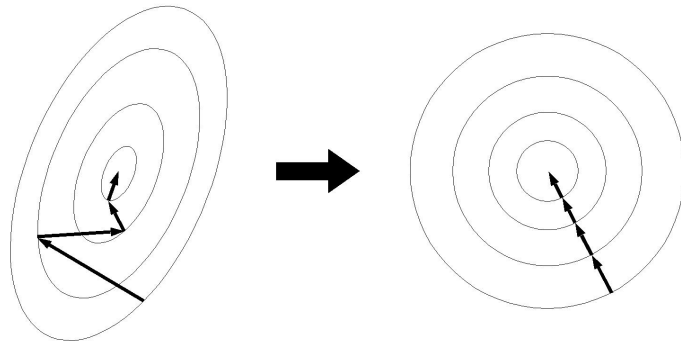


图 4.1 正规化前后数据分布示意图

二是局部细节强化，避免微小细节被淹没在数据噪声中。具体正规化计算如式（4.1），式中脚标 c 表示输入数据点所在图像通道， mean 表示均值， std 表示标准差。

$$\text{output}_c = (\text{input}_c - \text{mean}_c) / \text{std}_c \quad (4.1)$$

当使用预训练参数进行迁移学习时，输入数据的正规化参数应尽量与预训练时参数一致，对于该实验，R、G、B 各颜色通道均值 mean 分别等于 0.485、0.456、0.406，标准差 std 分别等于 0.229、0.224、0.225。该数值也是根据大型图像数据集 Imagenet 计算而得，对于一般训练图像普遍适用。

b. 图像尺寸重设

虽然 Mask R-CNN 由于存在 ROI-Align，不对输入图像大小进行限制，但为了更好识别小尺度目标，一般仍将输入图像的尺寸进行统一重设，使得实际输入网络图像的边长在落在指定范围内。对于该实验而言，重设后最小边长不小于 800，最大边长不大于 1333，具体计算如式(4.2~4.3)。

$$(w', h') = (w \times \text{scale}, h \times \text{scale}) \quad (4.2)$$

$$\text{scale} = \begin{cases} \min_size / \min(w, h) & , \quad \text{scale} \times \max(w, h) \leq \max_size \\ \max_size / \max(w, h) & , \quad \text{else} \end{cases} \quad (4.3)$$

式中 w , h 表示原始图像的宽度与高度， w' , h' 表示重设后的宽度与高度， \min_size 表示重设图像边长下限， \max_size 表示重设图像边长上限， $\min(*, *)$ 表示取较小值， $\max(*, *)$ 表示取较大值。

E. 数据集划分

实际输入模型的数据，根据其作用可划分为训练集与测试集，有时还可以单独划分出验证集。模型通过训练集确定优化方向，直接从训练数据中学习有效特征；通过验证集结果调整网络结构等超参数；通过测试集最终衡量其准确性与泛化能力。而本文中由于原始数据过少，为有效习得一定特征，均没有划分严格的训练集与测试集，同样没有划分验证集。具体划分如下：

a. 训练集

采用全部原始图像，并施加在线数据增强，因每次增强参数为区间内均匀采样，故一定程度上训练集可视为具有无限张图像。

b. 测试集

采用全部图像，不进行数据增强，故虽原始数据与训练集相同，但测试时网络并没有见过完全相同的图像，测试集共有 79 张图像。

4.1.2 超参数设置

超参数（Hyperparameter）是相对于参数（parameter）的概念，在机器学习中，参数可通过模型本身的迭代而更新，而超参数则是在开始训练之前设置的参数，不随模型迭代而改变。设置合适的超参数可提高模型学习效果，加快模型训练效率。

对于该模型而言，参数优化方法采用 SGD，动量参数为 0.9，权重衰减系数为 $5e-4$ ，学习率采用动态规划方法，从 $1e-6$ 增长到 $1e-3$ ，维持一定 Epoch 后最终降到 $1e-4$ ；批大小 $batch_size=4$ 。

下面简单介绍一下 SGD 及各相应超参数的意义：

(1) SGD(Stochastic Gradient Descent)，即随机梯度下降，是一种大规模参数优化算法，其根据损失函数梯度，利用链式法则对相关参数进行更新迭代。对于本实验而言，SGD 具有可以接受的学习效果以及训练速度，故其成为最终选择。其他优化算法，如 Adam 系列等也可以进行尝试，但不意味着一定会有更好的效果。一般而言，Adam 系列具有较高的初始收敛速度，但最终仍需利用 SGD 完成最后的调整；

(2) 动量参数用于减少更新方向，即梯度方向的震荡，使得参数每次更新方向受上一次的方向限制，一般有助于提高学习速率；

(3) 权重衰减系数用于添加 L2 正则项惩罚，如式 (4.4)，其中 C_0 为原始损失函数， λ 为权重衰减系数， ω 为模型权重参数。通过权重衰减，使得参数优化不再以原始损失函数最小化为唯一目标，同时也向使用更小的参数、更少的参数数量方向优化，起到减小过拟合的作用；

$$C = C_0 + \lambda \sum_{\omega} \omega^2 \quad (4.4)$$

(4) 学习率为每次参数更新的步长。起初学习率的“热身”过程可以减少网络学习到过多的无关特征，避免拖慢训练过程；在不产生损失发散或网络过拟合的前提下，维持高学习率可以加快模型训练速度；训练的尾期阶段降低学习率可使网络学习到更多细节特征；

(5) 批大小表示每次迭代同时训练的样本数。一方面可用于加快学习速率，但会牺牲部分细节特征学习能力；另一方面也可用于减缓参数更新震荡。在网络各参数固定的情况下，其数值上限受 GPU 显存限制。

4.1.3 实验结果

A. 损失曲线

损失函数用于度量模型预测值与真实值之间的差距，根据预测任务类型的不同、度量方式的不同，损失函数具有不同的表达形式，而通过“最小化损失函数”便可以进行模型参数估计，损

失曲线即是记录训练过程中损失函数的变化。

训练过程中损失曲线如图 4.2 所示，其中横坐标为遍历数据集次数，纵坐标为网络各部分损失和，红色虚线为训练集损失，蓝色实线为测试集损失。可见，网络在 120 个 Epoch 左右基本达到当前学习率下的收敛。

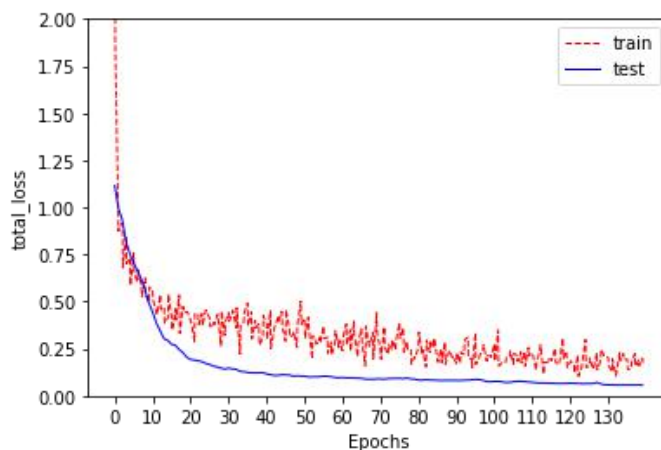


图 4.2 指示灯识别网络训练损失曲线

B. 网络性能评估

网络各方面识别能力如表 4.1 所示。

表 4.1 指示灯识别网络 COCO Metric

| metric | IOU | Area | MaxDets | value |
|--------|-----------|--------|---------|--------|
| mAP | 0.50:0.95 | all | 100 | 0.870 |
| mAP | 0.50 | all | 100 | 0.996 |
| mAP | 0.75 | all | 100 | 0.996 |
| mAP | 0.50:0.95 | small | 100 | 0.866 |
| mAP | 0.50:0.95 | medium | 100 | 0.881 |
| mAP | 0.50:0.95 | large | 100 | -1.000 |
| mAR | 0.50:0.95 | all | 1 | 0.430 |
| mAR | 0.50:0.95 | all | 10 | 0.896 |
| mAR | 0.50:0.95 | all | 100 | 0.896 |
| mAR | 0.50:0.95 | small | 100 | 0.892 |
| mAR | 0.50:0.95 | medium | 100 | 0.900 |
| mAR | 0.50:0.95 | large | 100 | -1.000 |

C. 识别诊断评估

整合算法识别部分与诊断部分，模型全流程检测结果如图 4.3 所示，其中横坐标为图像 ID，

纵坐标为检测正误，1 为正确，0 为错误，当且仅当正确识别出图像中所有指示灯（无遗漏、无多余）且总体状态码排序无误时判定为检测正确。

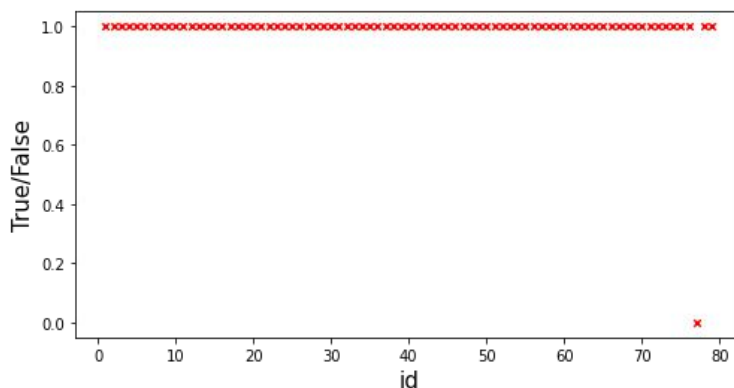


图 4.3 指示灯识别网络识别诊断测试结果

D. 效果图像

模型全流程检测效果如图 4.4 所示，最终检测结果以文字的形式标注在图片的左下方。检测结果由两部分构成：“Detection code”表示模型检测当前图片中的指示灯状态码，其可作为数字信号输出，通过上传服务器或本机就地与预设的正常状态码比较，即可推断设备的工作状况，进而采取相应动作；“Corresponding target”为每一位状态码对应的指示灯类别，通过“/”断开相邻编码。



图 4.4 指示灯识别诊断效果图

4.2 加药箱液位

加药箱液位识别为二阶段任务，相应地，以下实验将包含两部分，一部分为自然图像下加药箱侧表面分割算法验证，另一部分为热成像下剩余液体分割算法验证。

4.2.1 数据集制作

A. 数据来源

a. 自然图像数据来源

数据部分来源于现场机房利用双光摄像机实地拍摄，部分来源于网络图像。

b. 热成像数据来源

数据全部来源于现场机房双光摄像机实地拍摄，但实际使用数据经过配准以及掩膜处理，使每张图像仅保留一个加药箱。

B. 标注原则

a. 自然图像标注原则

对于标注自然图像用于分割加药箱侧表面，其侧表面至少露出三侧边缘，且遮挡超过三分之二不标注；

b. 热成像标注原则

对于标注热成像用于分割剩余液体，由于可见光网络生成的掩膜边缘与实际边缘可能并非十分拟合，故以液位界面所在曲线为分界线，无论边缘效果，标注该分界线以下所有未被掩膜抑制区域。

C. 数据增强

自然图像与热成像数据增强方式相同，均采用在线增强：

100%对图像施加亮度变换，变换亮度在 $\pm 40\%$ 之间均匀采样；

100%对图像施加对比度变换，变换对比度在 $\pm 30\%$ 之间均匀采样；

50%概率对图像施加水平翻转。

D. 数据一般化

a. 图像正规化

自然图像与热成像正规化参数相同，均值均为 0.485、0.456、0.406，标准差均为 0.229、0.224、0.225。

b. 图像尺寸重设

自然图像与热成像尺寸重设参数相同，最小边长下限为 800，最大边长上限为 1333。

E. 数据集划分

a. 训练集

自然图像与热成像训练集与指示灯训练集类似，均采用全部原始图像，并同时施加在线数据增强。

b. 测试集

采用全部原始图像，不进行数据增强，自然图像测试集具有 97 张图像，热成像测试集具有 129 张图像。

4.2.2 超参数设置

自然图像网络与热成像网络参数优化方法均采用 SGD，动量参数、权重衰减系数均分别为 0.9、 $5e-4$ 。而对于学习率，自然图像网络从 $2.5e-6$ 增长到 $2.5e-3$ ，维持若干周期后最终降到 $1e-4$ ；热成像网络则仅从 $2.5e-6$ 增长到 $2.5e-3$ ，然后维持该速率至训练结束。

4.2.3 实验结果

A. 损失曲线

训练过程中可见光网络与热成像网络损失曲线分别如图 4.5、图 4.6 所示。可见，可见光网络在 90 个 Epoch 左右基本达到当前学习率下的收敛、热成像网络在 65 个 Epoch 左右基本达到当前学习率下的收敛。

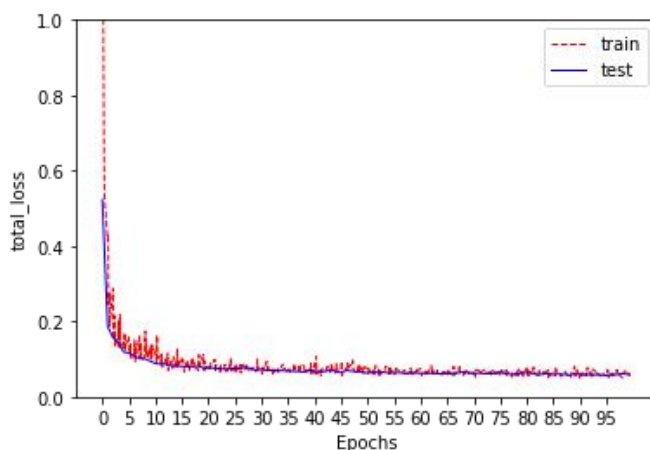


图 4.5 加药箱可见光网络训练损失曲线

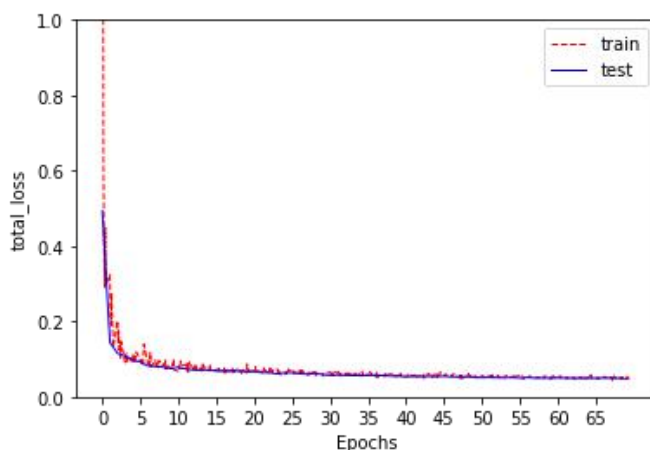


图 4.6 加药箱热成像网络训练损失曲线

B. 网络性能评估

可见光网络与热成像网络各方面识别能力分别如表 4.2、4.3 所示。

装
订
线

表 4.2 加药箱可见光网络 COCO Metric

| metric | IOU | Area | MaxDets | value |
|--------|-----------|--------|---------|--------|
| mAP | 0.50:0.95 | all | 100 | 0.998 |
| mAP | 0.50 | all | 100 | 1.000 |
| mAP | 0.75 | all | 100 | 1.000 |
| mAP | 0.50:0.95 | small | 100 | -1.000 |
| mAP | 0.50:0.95 | medium | 100 | 0.913 |
| mAP | 0.50:0.95 | large | 100 | 0.998 |
| mAR | 0.50:0.95 | all | 1 | 0.386 |
| mAR | 0.50:0.95 | all | 10 | 0.998 |
| mAR | 0.50:0.95 | all | 100 | 0.998 |
| mAR | 0.50:0.95 | small | 100 | -1.000 |
| mAR | 0.50:0.95 | medium | 100 | 0.950 |
| mAR | 0.50:0.95 | large | 100 | 0.999 |

表 4.3 加药箱红外光网络 COCO Metric

| metric | IOU | Area | MaxDets | value |
|--------|-----------|--------|---------|--------|
| mAP | 0.50:0.95 | all | 100 | 0.989 |
| mAP | 0.50 | all | 100 | 1.000 |
| mAP | 0.75 | all | 100 | 1.000 |
| mAP | 0.50:0.95 | small | 100 | -1.000 |
| mAP | 0.50:0.95 | medium | 100 | 0.944 |
| mAP | 0.50:0.95 | large | 100 | 0.997 |
| mAR | 0.50:0.95 | all | 1 | 0.992 |
| mAR | 0.50:0.95 | all | 10 | 0.992 |
| mAR | 0.50:0.95 | all | 100 | 0.992 |
| mAR | 0.50:0.95 | small | 100 | -1.000 |
| mAR | 0.50:0.95 | medium | 100 | 0.967 |
| mAR | 0.50:0.95 | large | 100 | 0.997 |

C. 识别诊断评估

整合算法识别部分与诊断部分，模型全流程检测结果如图 4.7 所示，其中横坐标为配准后每张自然图像中每个加药箱的 ID，纵坐标为模型检测值相对于标注真实值的偏差，偏差计算基于相对液位，即剩余液体掩膜面积与加药箱侧表面掩膜面积比值。

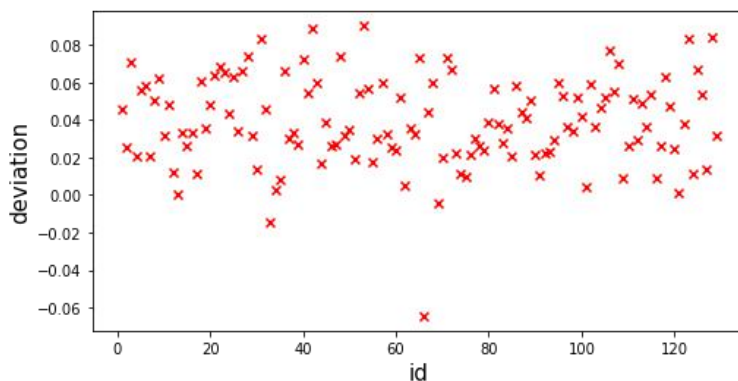


图 4.7 加药箱液位识别诊断测试结果

D. 效果图像

模型识别诊断效果如图 4.8 所示，第一行为加药箱分割效果，第二行为剩余液体分割效果，模型最终检测结果以文字的形式标注在剩余液体分割图片的左上方。“remain”表示剩余液体百分比。

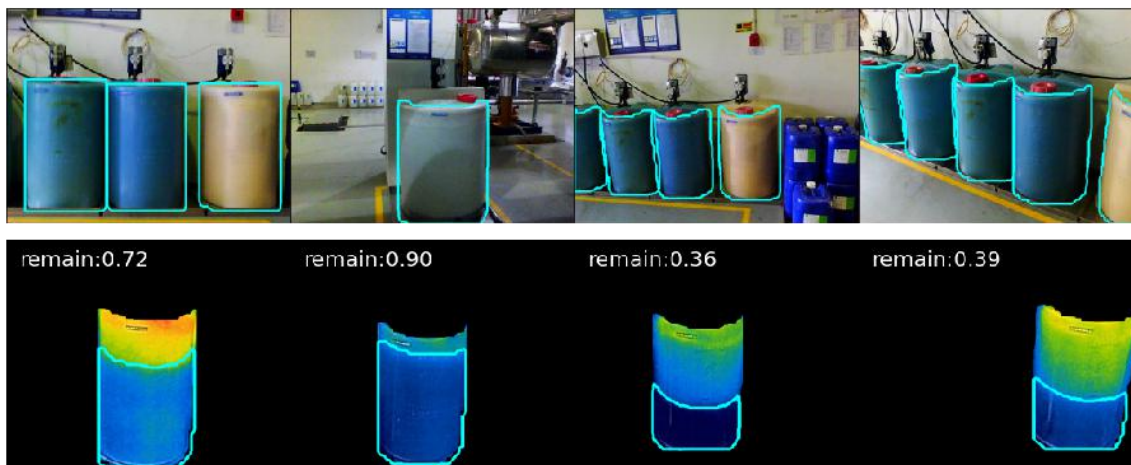


图 4.8 加药箱液位识别诊断效果图

4.3 圆盘式机械表读数

4.3.1 数据集制作

A. 数据来源

数据来源共有三部分：其中一部分数据来源于网络视频，与指示灯数据集类似，同样按照一定帧间隔采样后筛选合适图像；另外一部分数据来源于机房现场实地拍摄；其余部分数据则来源于网络图像。

B. 标注原则

a. 轮廓完整性

由于表盘边缘可能存在密封胶，导致表盘露出部分并非为严格圆形，故表盘上下左右定位四

装 订 线

点应取自表盘自中心向外扩展接触到的第一个完整圆形轮廓。

b. 直线不变性

根据透视变换前后直线关系不变的特性，表盘上下定位点与左右定位点应分别与表盘中心点尽量保持共线。

C. 数据增强

数据增强采用在线增强方式，具体内容如下：

100%对图像施加透视变换，扭曲率 $distortion_scale$ 为 0.4，透视变换矩阵由起始点组 $startpoints$ 与终止点组 $endpoints$ 构成，具体计算遵从式（4.5~4.10），其中 $width$ 、 $height$ 分别表示输入图像的宽度与高度， $randint(a,b)$ 表示在 a ， b 之间随机选取整数， $int(*)$ 表示向下取整；

$$startpoints = [(0, 0), (width - 1, 0), (width - 1, height - 1), (0, height - 1)] \quad (4.5)$$

$$endpoints = [topleft, topright, botright, botleft] \quad (4.6)$$

$$\begin{aligned} topleft = (randint(0, int(distortion_scale * 0.5 * width)), \\ randint(0, int(distortion_scale * 0.5 * height))) \end{aligned} \quad (4.7)$$

$$\begin{aligned} topright = (randint(width - int(distortion_scale * 0.5 * width) - 1, width - 1), \\ randint(0, int(distortion_scale * 0.5 * height))) \end{aligned} \quad (4.8)$$

$$\begin{aligned} botright = (randint(width - int(distortion_scale * 0.5 * width) - 1, width - 1), \\ randint(height - int(distortion_scale * 0.5 * height) - 1, height - 1)) \end{aligned} \quad (4.9)$$

$$\begin{aligned} botleft = (random.randint(0, int(distortion_scale * 0.5 * width)), \\ random.randint(height - int(distortion_scale * 0.5 * height) - 1, height - 1)) \end{aligned} \quad (4.10)$$

100%对图像施加旋转变换，旋转中心为表盘中心点，旋转角度在 $\pm 46^\circ$ 之间均匀采样；

100%概率对图像施加对比度变换，变换对比度在 $\pm 30\%$ 之间均匀采样。

D. 数据一般化

a. 图像正规化

均值等于 0.485、0.456、0.406，标准差等于 0.229、0.224、0.225。

b. 图像尺寸重设

最小边长下限从 640、672、704、736、768、800 中等概率选取，最大边长上限等于 1333。

E. 数据集划分

a. 训练集

训练集采用全部原始图像，并施加在线数据增强。

b. 测试集

测试集采用全部原始图像，不进行数据增强，共有 224 张。

4.3.2 超参数设置

a. 优化器

参数优化方法依然采用 SGD，动量参数为 0.9，权重衰减系数为 $5e-4$ ，学习率采用动态规划

方法，从 $1e-6$ 增长到 $1e-3$ ，维持若干周期后最终降到 $5e-4$ 。

b. OKS sigma

该参数并非严格属于模型超参数，但考虑到该参数同样是为预先设定，用于计算相关评价指标，所以本文且将其归在“超参数”中。由于本文数据集并无横向比较对象，且该超参数计算需大量人工标注成本，故本文并未实际赘余标注，而是采取类比的方法，选用 COCO 发布的人体关键点超参数：中心点、指针末端、刻度起始点、刻度终止点、表盘上定位点、表盘下定位点、表盘左定位点、表盘右定位点分别取 0.026, 0.025, 0.025, 0.025, 0.035, 0.035, 0.072, 0.072，依次对应鼻子、眼睛、眼睛、眼睛、耳朵、耳朵、手肘。

4.3.3 实验结果

A. 损失曲线

训练过程网络损失曲线如图 4.9 所示。可见，网络在 120 个 Epoch 左右基本达到当前学习率下的收敛。

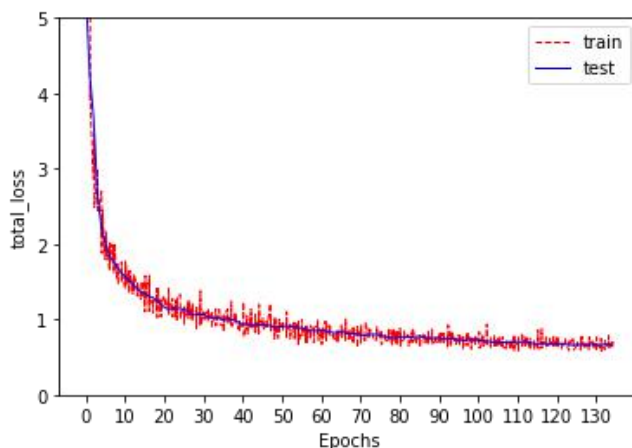


图 4.9 机械表识别网络训练损失曲线

B. 网络性能评估

网络各方面识别能力如表 4.4 所示。

表 4.4 机械表识别网络 COCO Metric

| metric | OKS | Area | MaxDets | value |
|--------|-----------|--------|---------|--------|
| mAP | 0.50:0.95 | all | 20 | 1.000 |
| mAP | 0.50 | all | 20 | 1.000 |
| mAP | 0.75 | all | 20 | 1.000 |
| mAP | 0.50:0.95 | medium | 20 | -1.000 |
| mAP | 0.50:0.95 | large | 20 | 1.000 |

续表 4.4

| metric | IOU | Area | MaxDets | value |
|--------|-----------|--------|---------|--------|
| mAR | 0.50:0.95 | all | 20 | 1.000 |
| mAR | 0.50:0.95 | all | 20 | 1.000 |
| mAR | 0.50:0.95 | all | 20 | 1.000 |
| mAR | 0.50:0.95 | medium | 20 | -1.000 |
| mAR | 0.50:0.95 | large | 20 | 1.000 |

C. 识别诊断评估

整合算法识别部分与诊断部分，模型全流程检测结果如图 4.10 所示，其中横坐标为图像 ID，纵坐标为模型检测值与真实值之间的偏差，其中偏差基于逆时针方向下指针旋转角与最大刻度旋转角间比值计算。

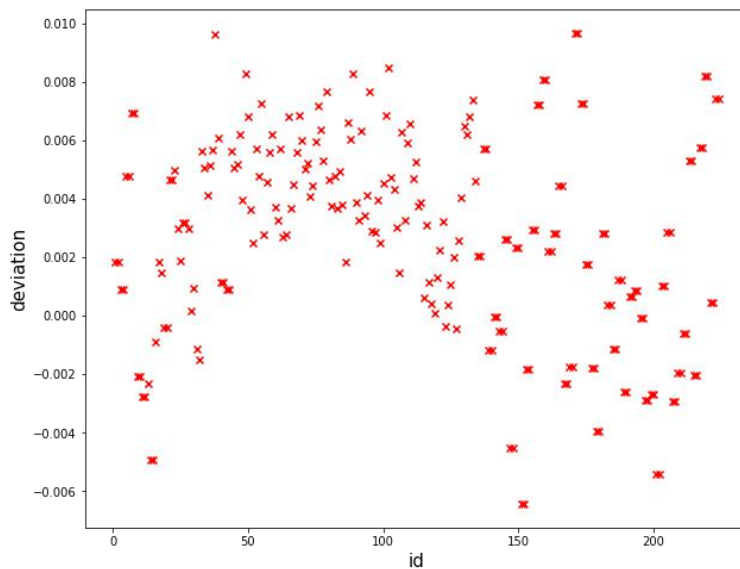


图 4.10 机械表识别诊断测试结果

D. 效果图像

模型识别诊断效果如图 4.11 所示，每行有两组图片，每组图片中左侧为原始图片，右侧为透视矫正后图片，检测结果以文字的形式标注在矫正后图片的下方，“metric”即为模型检测读数。

装 订 线

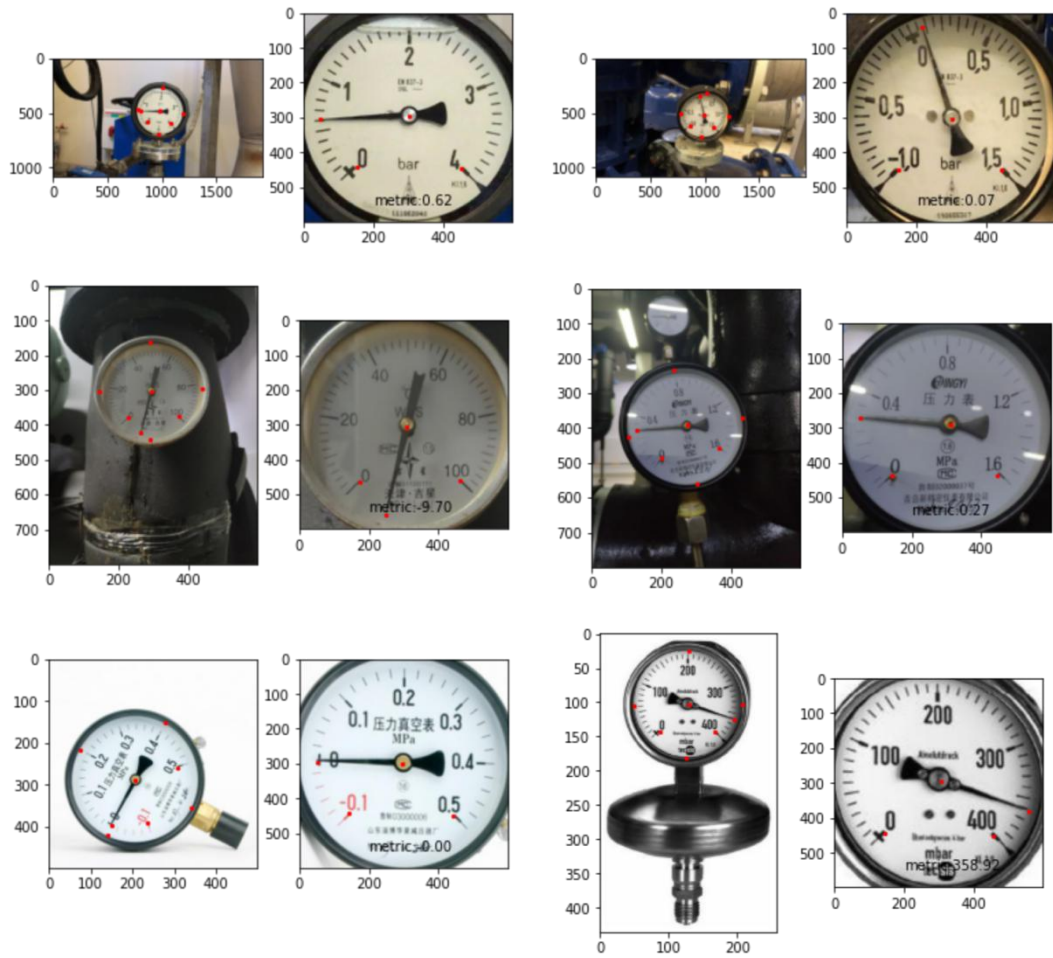


图 4.11 机械表识别诊断效果图

装
订
线

5 分析与讨论

该部分内容首先针对上述实验结果进行了分析，借助总体识别效果与网络量化评估指标，评价了模型的优势与不足，并研讨了不足产生的原因及相应的优化方向；其次，讨论了本文在方法理论、实验过程中的部分细节，以供相关研究人员进行参考；最后，对巡检机器人的技术效益进行了一定的评价。

5.1 结果分析

5.1.1 指示灯识别结果分析

A. 整体效果分析

从效果图来看，该模型可以基本准确定位、识别不同状态颜色指示灯；

从识别诊断评估结果来看，该模型大体表现良好，但存在一张错误识别样本——模型在正确识别了图上所有的指示灯外，多识别出一例 FP。进一步观察发现：一，此类样本在整体数据集中数量偏少；二，该样本为此数据集中唯一一张旋转 90° 、周围存在黑色填充的图像，且训练时并没有采取数据旋转增强。而卷积神经网络在设计之初并没有刻意考虑旋转不变性，其无法有效提取旋转不变特征，即生成特征图对旋转变换敏感^[29]。以上两点原因均可导致该样本识别失败，但有待进一步实验验证。

B. 网络性能分析

从网络性能评估结果来看，当 IOU 等于 0.5 或 0.75 时，mAP 均接近于 1，说明网络确实可以基本定位并识别不同类别指示灯，未完全等于 1 是因为存在一例错误识别样本；

但当 IOU 为 0.5:0.95 时，mAP 与 mAR 可以看到有一定下降，说明大部分网络识别定位框与真实标注定位框之间的交并比不大于 0.75，当阈值提高时，总体 TP 数目会产生一定下降，即定位精度相对不足。但对于该实际问题而言，因无需更高的定位精度，故此类指标降低影响不是很大；

当 MaxDets=1 时，mAR 再次产生显著下降，从另一个角度上也说明了此网络的高置信度识别目标交并比较低的问题；

该数据集中无大尺度目标样本，中等尺度目标与小尺度目标识别效果基本一致，中等尺度目标识别效果略微较好。

C. 优化方向

结合以上分析及本文技术路线，可以大致得出以下几点优化方向：

a. 样本均衡性

补充数据均衡性，尤其是补充完全各色指示灯及其明暗状态数据，使各类样本分布尽量均匀，可考虑人工合成数据；在获得所有类别样本的前提下，若样本均衡性难以满足，可考虑使用其他对于样本均衡性不敏感的损失函数。

b. 数据增强

补充数据多样性，虽实际中一般不会出现控制柜旋转的情况，但训练数据增加旋转变换样本，可使网络对于旋转变换不敏感，提高网络表现。

c. 提高定位精度

增加原始图像数量、进一步降低学习率、训练更多层参数、调整预设 Anchor 大小、更换性能更为良好的 Backbone 与 box_head、或直接换用其他网络结构等。

d. 诊断流程优化

本文诊断由聚类 and 排序构成，仍需人为提供一定聚类信息。未来可考虑简化该流程，取消人为干预，如从二维坐标空间稳定地映射至某一维距离空间，可以直接根据该距离对坐标进行排序，或考虑其他聚类手段、神经网络手段。

5.1.2 液位识别结果分析

A. 整体效果分析

从效果图来看，该模型可以基本准确分割加药箱及加药箱内剩余液体。可见光下加药箱边缘并非十分准确，不过由于最终诊断结果为百分比，故此影响尚可接受；

从识别诊断评估结果来看，模型检测加药箱内剩余液体百分比与人为观测结果基本一致，最大误差约为 9%，基本满足一般巡检需求。

B. 网络性能分析

a. 可见光网络

从可见光网络性能评估结果来看，当 IOU 等于 0.5 或 0.75 时，mAP 均等于 1，说明网络在该测试集中可以正确识别并找出所有加药箱；

与指示灯识别类似，当 IOU 为 0.5:0.95 时，各项指标均有所下降，但大多数指标降幅不大，说明当阈值提高时，网络识别结果的 TP 数目下降并不大；

当 MaxDets=1 时，mAR 下降显著，说明每张图像网络识别置信度最高的结果与真实结果的交并比不高、边缘并非十分精确，置信度稍微低一点的识别结果交并比反而较高；

该数据集中无小尺度目标样本，大尺度目标识别效果略优于中等尺度目标，边缘识别更加准确。

b. 热成像网络

从热成像网络性能评估结果来看，当 IOU 等于 0.5 或 0.75 时，mAP 均等于 1，说明在该测试集下此网络可以正确分割剩余药液；

当 IOU 为 0.5:0.95 时，所有相应指标仅略微下降，说明该网络识别结果的交并比较为稳定且处于高位，同时这也是热成像网络在训练的尾声阶段并没有再次减小学习率的原因；

该数据集中无小尺度目标样本，大尺度目标识别效果同样略优于中等尺度目标。

C. 优化方向

根据以上分析，可以得出优化方向主要在可见光下边缘分割准确性上。值得一提的是，该网络并非完整分割加药箱，而是仅分割加药箱侧面，一定程度上确实会提高分割难度。如若进一步优化网络的边缘分割效果，更换性能更好的 Backbone 与 mask_head 或直接选择更优的网络构架效果将较为显著，大量增加训练数据以训练全部网络参数、更换边缘敏感的损失函数或引入辅助损失函数也不失为一种有效方案，进一步降低学习率训练更多回合也可以作为进一步的尝试。当然，也可以引入形态学操作，对识别结果进行一定修补。

5.1.3 机械表识别结果分析

A. 整体效果分析

从效果图像来看，该模型可以基本准确定位所有关键点，并识别当前机械表读数；

从识别诊断评估结果来看，该模型输出结果与人为观测结果基本吻合。若实际生产中指针存在晃动，可取多次识别的平均值作为最终诊断结果，一方面可抵消指针晃动影响，另一方面可进一步减小识别误差，以满足一般观测需求。

B. 网络性能分析

从网络性能评估结果来看，所有参与计算指标均为 1，说明网络在此数据集下表现确实良好、定位精度较高。但仍存在几点需要说明：此数据集中仅包含大尺度目标，尺度变化较为单一；OKS 是由所有关键点的相似度共同计算而得，因此在体现单个关键点的偏离程度上略显乏力；本文计算 OKS 所需的经验常数 k 为类比所得，并非实际统计。

C. 优化方向

结合以上分析及本文技术路线，可以大致得出以下几点优化方向：

a. 尺度变换

若实际中需对其他尺度目标识别，或仅为了增强网络识别能力，可在训练时向网络输入不同尺度的图像，或通过增加不同尺度的原始图像、或通过数据增强进行多尺度变换。

b. 统计 k 值

雇佣大量人员标注同一数据集，生成大量冗余标注，以计算更为合理的 k 值。

c. 改善评估指标

单独分离各个关键点的相似性或寻找其他评估指标，以获得更加全面深入的评估结果。

d. 起止刻度数值识别

目前算法虽已减少大量需人工提供信息，但仍需给出起止刻度的数值大小。该算法目前已具有定位起止刻度位置的能力，因此未来可考虑根据定位点所在区域进行起止刻度识别，进一步减少人工干预。

5.1.4 综合结果分析

横向比较指示灯状态识别、加药箱液位识别、机械表读数识别，可以发现其识别效果大致呈递增趋势。从其各自数据集角度来看，指示灯数据集最小，样本均匀性、多样性最差；液位数据集中等；机械表数据集最大，样本多样性最强。由此可见，数据集质量确实对于网络的识别效果有着重要影响，在进行相关研究时应着重关注。

5.2 其他问题讨论

5.2.1 关于网络选择：为什么选择 Mask R-CNN？

(1) 因为 Mask R-CNN 提出至今已有三年左右时间，基本吻合学术界提出到工业界落地所需时间。该模型较为成熟，相关研究较多，性能良好，能够胜任本文任务；

(2) Mask R-CNN 具有在大型数据集上的预训练参数，通过迁移学习可降低数据需求量。实际任务中原始工业图像十分难以获得，在数据严重不足的情况下训练网络全部参数，一方面网络

参数将不稳定，无法提取有效特征，导致识别结果较差；另一方面，又将花费大量算力更新网络参数，在设备要求以及时间花费等方面极其不友好；

(3) Mask R-CNN 为多功能任务网络，刚好对应本文多个问题，一定程度上简化了本文的工作；

(4) 条件允许情况下，选择更好的网络结构将有更好的模型表现。

5.2.2 关于数据集划分：为什么没有严格意义的测试集？

(1) 主要因为工业图像数据获取较难，本文原始数据较少，单独分割一部分数据做验证集或测试集过于奢侈，将导致网络训练不足，且较少的测试样本统计意义也较弱；

(2) 此外，因为本文并非大型公开数据集，存在样本分布不均等特殊特性，即使划分出验证集、测试集，其评估的具体数字结果对于其他研究人员的参考价值也较为有限，而其数字结果的相对趋势价值相比较较高；

(3) 再者，在实际生产生活中，巡检机器人仅在某一特定环境中工作，且目前仍然以定点拍照为主，其可获取图像的多样性有限。因此，在数据不足的情况下适当牺牲网络泛化能力，使其捕捉该环境下更多的细节特征，将有助于提高在网络在该环境下识别的准确率；

(4) 本文主要目的在于验证能否借助卷积神经网络捕获指定特征，如若为测试、增强网络泛化能力，则需进一步补充数据、调试网络；

(5) 当可获取数据足够时，应当进行严格的训练集、验证集、测试集划分。

故训练集使用所有图像并采用在线增强，训练集使用所有图像但不采用在线增强。

5.2.3 关于指示灯：为什么人为给定聚类数？

(1) 一方面，当模型诊断不同控制柜时，模型需了解相应的不同正常状态与故障状态，即模型需要一定的先验信息，聚类数便可以包含在内，若自动确定聚类数则可能增加误差与时间损耗；

(2) 另一方面，K-means 聚类受初始化影响较大，聚类数以及初始位置均可能影响聚类效果，因此这里介入人为监督具有一定必要性。

5.2.4 关于加药箱：热成像效果是否对识别造成影响？

(1) 热成像效果会对识别产生影响，在热成像网络训练过程中，液面上下颜色对比较为明显的样本训练识别较快，颜色较为接近的、液面分界线不完整的样本训练识别较慢。因此，提高加药箱药液覆盖侧表面与未覆盖侧表面温差，如将加药箱置于周期性热源附近，将有助于提高识别效果。同时，这也是本文没有采用二值化的方法从热成像中提取剩余液位的原因之一——因为难以寻得某一维度空间，使得在此空间下确定的阈值可以稳定准确分割剩余液体；

(2) 此外，因红外测温仅能测量表面温度，若加药箱内药液下降速度过快以至热量传递不及时、或红外测量时具有其他干扰，均将产生识别误差。

5.2.5 关于机械表：为什么选择 8 个关键点？

(1) 表盘上下左右四个定位点可以带来一定的矫正；

(2) 训练过程中发现，即使表盘上下左右定位点部分产生偏移，但因诊断结果采用的是角度比值的方式，故最终结果差别并不大。实际中若拍摄时没有过大的透视角度、读数精度要求不高，也可选择仅识别其余四个关键点。

5.3 技术效益评价

5.3.1 经济效益

巡检机器人带来的经济效益难以准确衡量，直观上来看，其可直接节约的便是巡检人力成本。虽受地域性、工作场景等影响，人工成本波动较大，但基本上相关企业每年的人工成本在数十万元左右。根据功能不同，当前巡检机器人的购置费用从几万元到几十万元不等，日常运行电力费用可忽略不计。因此单独凭借人力成本，企业便可在一年左右收回初投资。

而巡检机器人的间接经济效益也表现在多方面。最突出的便是事故维修费用，借助机器人巡检，可以及时发现故障设施，及时止损，降低重大生产事故风险，节约潜在风险成本；除此之外，其经济效益仍可辐射到人力资源部门，减少其招聘成本以及管理成本等。

5.3.2 社会效益

巡检机器人的社会效益则主要表现在促使相关产业进行产业升级。一方面，其落地应用推进了科技发展与生产力解放，加速了建筑信息化与智能化基础建设；另一方面，其为当前世界普遍面临的人口问题之一，即适龄劳动人口数降低，提供了一定的解决方案。

5.3.3 环境效益

巡检机器人的环境效益至少也表现在两方面。对于建筑内部，其更高频次的获取了环境参数与设备参数，有助于室内环境控制，以获得更为舒适的室内空间；对于建筑外部，则受益于室内精细化管理，降低了日常污染排放，同时也减小了重大事故带来严重环境污染的风险。

6 结论与展望

6.1 结论

(1) 本文基于机器视觉，借助 Mask R-CNN 神经网络，对机房巡检机器人的部分检测模块进行了开发，其中包括控制柜指示灯识别模块、加药箱液位识别模块以及圆盘式机械表识别模块。通过分析目标特性、确定检测特征、现场或网络获取相关数据、确立数据标注原则、选择对应神经网络结构、调试超参数以及损失观测等完成了对指定目标的识别，并根据识别特征提出了相应的诊断方式，最终整体评估结果表现良好，可基本满足机房相关巡检需求。

(2) 本文采用的方法相比于传统识别方式克服了诸多缺点，分别表现在：①本文对于指示灯的识别不依赖于参考图像、可对多种角度、尺度下的指示灯进行识别，并给出了较为通用的识别结果诊断方式；②本文对于液位的识别，可在非纯净背景中同时对多个非透明容器进行液位判断，并降低了红外热像图诊断中对于色彩差异的敏感；③本文对于机械表读数的识别更为简洁，不依赖于参考图像、起止刻度旋转角，或表盘区域划分，同时也提出了一种较为通用的优、劣角计算方式。

(3) 实验表明，借助卷积神经网络对机房特定目标进行识别诊断具有其独特的优越性：相比建立传统特征工程需要的大量人力以及专业知识，其优越性主要体现在表征学习能力上，这意味着①它可以不断扩展信息维度，提取到人类难以感知的、稳定的、具有不变性的高维特征，生成并组合更多的特征检测器，以包含更丰富的边缘信息与语义信息，最终在复杂多变的环境中表现出更高的准确性与鲁棒性。②它可以实现端到端的学习，节省大量人力劳动成本，如反复设计特征、检验特征效果等。

(4) 实验表明，目前图像识别的神经网络方法受数据质量限制较大：这是因为目前其实现仍以强监督学习为主，而强监督学习的目的便是在高维空间中的对训练数据特征分布进行拟合。但当训练数据无明显的分布模式，或测试数据与训练数据分布模式显著不同时，模型表现将大打折扣。

(5) 本文将各巡检模型均划分为识别部分与诊断部分，其原因在于——为使机器视觉算法真正落地产业应用，离不开对图像特征的识别与整体理解。而就目前而言，深度学习的强势之处更多体现在“识别”方面，其“理解”能力较为薄弱。因此，仍需人工根据“识别”特征，设计“理解”算法。这时，良好的特征选择将有助于后续算法设计及总体表现，而不良的特征选择将复杂化后续设计，甚至难以为继。

6.2 展望

机房巡检机器人结合神经网络进行目标识别诊断具有较为广阔的发展前景，多样的采集数据结合不同用途的神经网络，将碰撞出众多的可能性，其可能的发展方向包括但不限于以下几点。

A. 多维数据输入

一般用于机器视觉的图像数据输入仅具有可见光下 R-G-B 三个通道的信息，但其仍可以结合更多维度的信息：如将温度的伪彩色热像图信息叠加到可见光 R-G-B 通道上，或直接并列出一条温度 T 通道；利用深度摄像机，检测不同像素点深度信息，并列 D 通道；纳入时间 t 维度，对

序列图像进行识别等。

B. 数据处理

正如前文所说，该模式对于数据要求较高，因此对于原始数据也存在着两种相互关联的发展方向：

a. 补充原始数据数量

对于补充原始数据，一种方案是生成合成数据，如建立三维实物仿真模型、搭建自编码器（Auto-Encoder）、搭建生成对抗网络（GAN）等；另一种方案是收集更多真实数据，如建立网络平台，为使用者提供预训练网络，使用者再补充自身数据进行微调（fine-tune）等。该方向旨在解决原始数据数量需求，不涉及数据标注问题。

b. 降低原始数据处理需求

减少数据处理需求则是指发展弱监督或无监督学习方法，该方法的突破可本质上拓宽该模式应用边界。此外，值得一提的是，弱监督、无监督方法仅是降低了数据标注成本，但仍然需要大量数据。

C. 任务定性

不同的任务性质代表了不同的技术路线，对于同一任务可能具有单一或多种技术路线，而这种路线便是巡检机器人与神经网络的结合纽带。如本文中任务性质可包含多目标识别、实例分割、语义分割与关键点检测。其他性质任务亦存在其潜在应用空间，如图像描述生成（Image Caption）可能用于简化部分任务的诊断环节设计、语音识别（Automatic Speech Recognition）可能用于处理声音信号或其他序列数据、姿态估计（Pose Estimation）可能用于机械机构的状态判定等。随着深度学习发展出更多的任务性质，其二者的结合方式也将变得愈加多元化。

参考文献

- [1] 刘鹤,刘芬香,伍林.水电站智能巡检系统设计[J].水电站机电技术,2019,42(12):24-27.
- [2] 刘孝颂,黄挺.巡检机器人在互联网数据中心智能运维中的应用探索与研究[J].电信科学,2018,34(12):132-137.
- [3] 张俊,马迅飞,张进铎,陈翔,李华松.基于人工智能学习的巡检监控机器人系统的探索与研究[J].计算机时代,2019(11):15-17+22.
- [4] 袁晓东.数据中心高压配电机房巡检机器人应用及能源互联网解决方案[J].通信电源技术,2019,36(S1):214-221.
- [5] 刘杨,刘俊,柯奕辰.基于变电站巡检机器人的指针式仪表读数识别[J].化工自动化及仪表,2019,46(8):636-639,658.
- [6] 鲍友革,黄宗碧,姬晓辉.基于强人工智能的水电厂智能巡检机器人研究与应用[J].水电与抽水蓄能,2019,5(01):39-43+92.
- [7] Merriaux P, Rossi R, Bouteau R, et al. The VIKINGS Autonomous Inspection Robot: Competing in the ARGOS Challenge[J]. IEEE Robotics & Automation Magazine, 2018, 26(1): 21-34.
- [8] 王泽民,林晓焕,宋扬.铁路中继站室内巡检机器人的设计[J].国外电子测量技术,2019,38(12):97-102.
- [9] 陈卓雄,于晓丽.智能巡检机器人在地铁通信、信号机房的应用研究[J].中国新通信,2019,21(11):7-8.
- [10] 张楠.中央泵房自动化及机器人巡检系统的研究[J].机械管理开发,2019,34(09):241-242.
- [11] Liu L, Yan R J, Maruvanchery V, et al. Transfer learning on convolutional activation feature as applied to a building quality assessment robot[J].
- [12] 徐波,张立群,刘朝欣.变电站巡检机器人保护装置识别关键技术研究[J].山东电力技术,2018,45(08):18-23.
- [13] 陈卓雄,于晓丽.智能巡检机器人在地铁通信、信号机房的应用研究[J].中国新通信,2019,21(11):7-8.
- [14] 胡灿林,肖尚华.基于颜色属性的信号指示灯检测识别[J].现代计算机(专业版),2018(07):75-78+89.
- [15] Lv L, Tan Y. Detection of cabinet in equipment floor based on AlexNet and SSD model[J]. The Journal of Engineering, 2019, 2019(15): 605-608.
- [16] Hao J, Dai M, Zong X. Design of inspection system for railway engine room based on visual detection[C]//2018 WRC Symposium on Advanced Robotics and Automation (WRC SARA). IEEE, 2018: 1-6.
- [17] 张宁,田慕琴,宋建成,郑丽君,江湘津,刘成林.基于图像识别的吸量管液位检测方法研究[J].现代电子技术,2019,42(23):40-43.
- [18] 李永,东琦,宋晓川,李惠琴.基于数字处理的电厂水位计液位测量系统误差分析[J].自动化技术与应用,2019,38(07):111-114.

- [19] 刘晓洲, 龚演平. 基于图像处理的断路器压力表计识别技术研究[J]. 技术与市场, 2019, 26(11): 102+104.
- [20] Liu Y, Liu J, Ke Y. A detection and recognition system of pointer meters in substations based on computer vision[J]. Measurement, 2020, 152: 107333.
- [21] 黄嘉明, 胡欣欣, 韦亦龙, 史筱川, 周灏, 李剑. 一种基于红外图像的电缆终端异常发热自动诊断方法[J]. 广东电力, 2020, 33(03): 103-110.
- [22] 李文璞, 谢可, 廖道, 李小宁, 王皓. 基于 Faster RCNN 变电设备红外图像缺陷识别方法[J]. 南方电网技术, 2019, 13(12): 79-84.
- [23] 刘小波, 徐波, 宋爱国, 田定胜, 吴道平. 基于变电站巡检机器人数字仪表识别算法[J]. 供用电, 2019, 36(12): 80-86.
- [24] Dumoulin V, Visin F. A guide to convolution arithmetic for deep learning[J]. arXiv preprint arXiv:1603.07285, 2016.
- [25] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.
- [26] Hosang J, Benenson R, Dollár P, et al. What makes for effective detection proposals?[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 38(4): 814-830.
- [27] 李玉浩, 曹学文, 梁法春, 等. 湿气集输管线非介入式积液检测技术研究[J]. 科学技术与工程, 2014, 14(30): 18-22. DOI:10.3969/j.issn.1671-1815.2014.30.004.
- [28] 罗为. 基于图像的空调系统故障诊断算法研究[D]. 上海: 同济大学, 2019.
- [29] Cohen T, Welling M. Group equivariant convolutional networks[C]//International conference on machine learning. 2016: 2990-2999.

谢辞

转眼间本科四年已经过去，从未想到会在这样一个特殊的时期结束了我的本科生涯，昨日往往显得有些模糊与不真切，但毫无疑问的是日子仍将继续、未来依旧充满可能。首先，感到十分幸运的是，能够找到许鹏教授作为我本次论文指导教师以及未来的研究生导师。非常感谢许老师交予我的论文选题以及在此期间的细心指导。虽论文写作方向几经波折，内容不断调整，但许老师依然给予了我充分的信任与空间，我才得已顺利完成这篇论文。他严肃的科学态度、精益求精的工作作风、和蔼可亲的言谈举止，深深地感染和激励着我。

其次我也要感谢课题组的诸多师兄、师姐在生活学习上对我的帮助，尤其是李鸿名师兄和陈智博师兄。他们帮助我收集数据、教我和企业打交道、指导我学习新知识、给予我参考建议、与我分享写作经验等，有效推进了课题进度。

当然，还有我可爱的室友们，他们营造了良好和谐的宿舍氛围，促使了我们共同成长、共同进步，很荣幸能够在学习生涯的下一阶段与你们继续同行！

最后，我要感谢我的母亲。一直以来，虽独自顶着生活的各方面压力，却始终如一地、毫无保留地支持我、鼓励我。感谢您的付出，令我在人生旅程中看到更多的风采！