

基于 Mask R-CNN 和改进 OCRNet 的冷热设备结霜检测的研究

黄炳睿¹, 张恒文¹, 王宝友², 邱文瀛²

(1. 北京邮电大学智能工程与自动化学院北京 100086 ;

2. 中国工业互联网研究院智能化研究所, 北京 100000)

摘要: 冷热设备结霜检测对保障设备高效运行和降低能耗具有重要意义。传统的结霜检测方法存在准确性不足和适应性差的问题,无法实时监测结霜程度,从而影响除霜策略的有效性。为此,本研究提出了一种基于 Mask R-CNN 和改进 OCRNet 的冷热设备结霜检测方法。首先利用 Mask R-CNN 对不同角度图像中散热器底板的位置进行定位,然后利用改进的 OCRNet 对结霜区域进行精细分割,计算结霜面积占比,为除霜控制提供依据,从而实现按需除霜,有效减少能源浪费,提高设备运行效率和延长设备寿命。实验表明,所提方法在结霜检测的准确性和适应性方面优于传统方法,为智能除霜提供了可靠的技术支持。

关键词: 模式识别与智能系统; Mask R-CNN; OCRNet; 结霜检测

中图分类号: TP391

The study on frost detection of cold and hot equipment based on Mask R-CNN and improved OCRNet

HUANG Bingrui¹, ZHANG Hengwen¹, WANG Baoyou², QIU Wenying²

(1. School of Intelligent Engineering and Automation, Beijing University of Posts and Telecommunications, Beijing 100086;

2. Intelligent Research Institute, China Academy of Industrial Internet, Beijing 100000)

Abstract: Frost detection on the heat exchange surfaces of heat and cold source equipment is crucial for improving operational efficiency and energy conservation. Traditional frost detection methods mostly rely on manual observation, which is time-consuming, inefficient, and lacks consistency and precision. This paper proposes a method based on an improved Mask R-CNN and OCRNet for frost detection. First, Mask R-CNN is employed to identify and extract the heat exchange surface region, followed by OCRNet for frost area segmentation and frost coverage calculation. Meanwhile, relevant frost removal data is provided to support practical applications. The proposed method demonstrates strong robustness and accuracy, effectively enhancing operational efficiency and reliability. This approach provides a theoretical and technical basis for automatic frost detection and removal systems, serving the development of intelligent equipment.

Keywords: Object Recognition and Intelligent Analysis; Mask R-CNN; OCRNet; Frost Detection

0 引言

冷链物流是传统物流体系中的重要组成部分,它使产品处于符合规定的温度环境、保障产品质量、减少损耗,是一项涉及多环节的系统工程^[1],自 21 世纪初以来,我国冷链物流迎来了快速发展,尤其是在冷库建设方面。目前,冷库在食品、医疗卫生等多个领域得到广泛应用,市场规模巨大,并且每年保持近 5% 的增长速度。然而,由于冷库结构性问题及智

作者简介: 黄炳睿 (1999-), 男, 智能装备及制造技术

通信联系人: 张恒文, 男, 硕导, 计算机集成制造系统. E-mail: zhanghengwen@bupt.edu.cn

能化、数字化、自动化技术的不足，导致能源浪费现象严重，如何减少冷库中的能源浪费已成为冷链物流行业的关键问题^[2]。

制冷系统的核心包括蒸发器、冷凝器、压缩机和节流装置四个基本部件，其中蒸发器起着至关重要的作用。它通过制冷剂的蒸发过程吸收热量，以降低流经物料的温度。然而，当蒸发温度低于空气露点且低于 0℃ 时，蒸发器表面不可避免地会出现结霜现象^[3]。结霜初期，霜层的形成能够增加蒸发器的换热面积，起到强化换热的作用^[4]，但随着霜层不断增厚，翅片间距减少、气流通道被堵塞，换热性能大幅下降，系统的性能系数（COP）可能下降高达 35% 至 60%^[5]。因此，为了确保制冷设备的稳定运行，减小能耗，提高冷库性能，对蒸发器进行周期性除霜是必不可少的。

在数字化和自动化技术不断发展的今天，人们愈加追求高效、便捷的自动化控制和数据采集系统。针对冷库能源浪费问题，利用自动化控制和数据信息技术实现蒸发器的按需除霜、远程监控，显得尤为重要。这不仅能降低能源浪费，还能提升冷库运维的便捷性和高效性。然而，结霜现象复杂多变，霜层的形成具有显著的随机性，使得传统的除霜控制方法常常难以做到精准除霜，容易出现过度除霜或除霜不彻底的现象，最终导致系统性能下降、能耗增加及库温波动等问题^[6]。

基于此背景，开发一种合适的除霜控制方法和监控系统，是制冷系统研究中的重要课题之一。本文提出了一种精准、按需的除霜控制方法，利用先进的图像处理技术，对蒸发器的结霜程度进行实时判断，并据此执行适时适度的除霜操作，从而减少除霜能耗，确保设备的高效运行，延长机组的使用寿命，提高能源利用效率。

1 冷库散热底板图像特点及识别需求分析

根据某集团冷库摄像头实拍图片，冷库的散热板底板图像具有一定特点，对图像的识别也提出了更高要求。

1.1 冷库散热底板图像特点分析

冷库内的散热底板图像具有显著的夜间低光照、反射严重、摄像头畸变和拍摄角度不一致等特点，如图 1、图 2 所示。这些因素给结霜检测带来了诸多挑战：

（1）光照条件差：夜间冷库内光线较弱，图像对比度低，导致结霜区域与底板背景难以区分，影响分割精度。

（2）反射干扰：由于冷库内金属结构反射强烈，图像中易产生光斑和高亮区域，增加了检测算法对霜层识别的难度。

（3）摄像头畸变：受制于摄像头的拍摄位置和透镜特性，图像存在一定的畸变，导致结霜区域在图像中发生形变，进而影响结霜率的计算准确性。

（4）拍摄角度不一：冷库内的摄像头多以倾斜角度安装，且各摄像头视角不同，导致图像中底板的形状和比例变化，给统一的结霜检测带来了挑战。



图 1 白天冷库散热板图像



图 2 夜冷库散热板图像

1.2 识别需求分析

- 结合冷库环境下散热底板图像的特点，自动化结霜检测系统需要满足以下需求：
- （1）高精度图像分割：能够在低光照、反射严重的环境中准确分割散热底板区域，排除背景干扰。
 - （2）抗畸变与多角度适应性：应对摄像头畸变带来的图像形变，并在不同摄像头角度下识别并分割底板区域。
 - （3）结霜率精确计算：能够精确识别结霜区域从而计算结霜率，以便为按需除霜提供可靠数据支持。

2 结霜检测流程

为了适用不同角度不同规格的底板，以及后续多个散热器底板的处理，首先对采集到的图像进行预处理，其次进行摄像头畸变矫正，然后采取 MaskR-CNN 对底板区域进行分割，从而排除画面中其他物体的干扰，提升后续结霜率的计算准确度，并且通过透视变换的方法对裁剪到的图像进行矫正，排除距离因素的干扰，最后使用改进的 OCRNet 对底板区域进行结霜检测，计算结霜率。结霜检测流程图见图 3。



2.1 图像预处理

在冷库环境下采集的图像往往具有光照不均、对比度低的问题，尤其是在夜间或低光条

件下, 图像的亮度分布集中在灰度值较低的范围内, 导致底板结霜区域与背景之间的区分度降低。为了提升图像的对比度和细节可见性, 本文在图像预处理中引入了直方图均衡化技术。直方图均衡化通过重新分配图像灰度值的概率密度, 使灰度值在整个范围内分布更加均匀, 从而增强图像的对比度。在具体实现中, 首先统计输入图像的灰度直方图, 并计算 CDF (Cumulative Distribution Function, 累积分布函数), 然后依据 CDF 对每个像素点的灰度值进行映射, 生成增强后的图像。直方图均衡化不仅能够有效提升低亮度区域的细节表现, 还能强化结霜区域与背景之间的对比, 为后续的分割和检测算法提供更优质的输入数据。

考虑离散灰度级的情况, 设一幅图像的像素总数为 n , 分为 L 个灰度级, 其中 n_k 表示第 k 个灰度级出现的个数, $p(r_k) = \frac{n_k}{n}$ 表示第 k 个灰度级出现的概率, ($0 \leq r_k \leq 1, k = 0, 1, 2, \dots$), 公式见式 (1):

$$S_k = T(r_k) = \sum_{j=0}^k P_r(r_j) = \sum_{j=0}^k \frac{n_j}{n} \quad (1)$$

S_k 即为使直方图变平直的一种变换, 处理效果如图 4 所示:

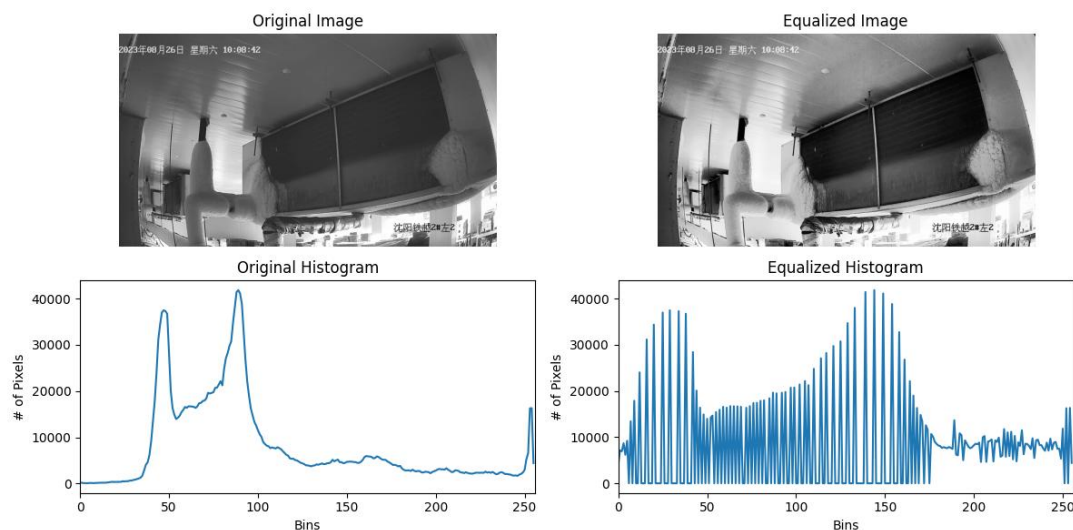


图 4 直方图均衡化效果对比

同时, 在冷库环境中, 摄像头倾斜安装或因镜头特性导致的几何畸变会对图像的准确性产生显著影响。尤其是桶形畸变 (barreldistortion) 和枕形畸变 (pincushiondistortion), 会使图像中的直线出现弯曲, 目标区域的形状和比例发生变化, 进而影响结霜检测的精度和霜层覆盖率的计算。为解决这个问题, 本文在图像预处理中加入了畸变矫正模块。首先, 通过摄像头标定获取内参矩阵和畸变系数, 利用 OpenCV 中的畸变矫正方法对图像进行校正。标定过程通过对标准棋盘格图案的多视角拍摄, 计算镜头的畸变参数 (径向畸变和切向畸变系数), 并生成矫正所需的内参矩阵。校正过程使用反向映射方法, 将图像中每个像素点重新映射到真实位置, 恢复图像中的直线结构。处理前后对比如图 5、图 6 所示



图5 畸变校正前



图6 畸变校正后

2.2 基于 Mask R-CNN 的底板分割

FasterR-CNN 是一种广泛使用的目标检测框架，MaskR-CNN 是基于 FasterR-CNN 发展而来的，将 FasterR-CNN 扩展为一个实例分割框架。MaskR-CNN 采用了两阶段的处理流程：第一阶段，框架扫描图像并生成可能包含目标的区域；第二阶段，进行目标分类，并生成目标的边界框和分割掩码。如图所示，MaskR-CNN 中的特征图提取骨干网络采用的是 ResNet-101^[6]，残差网络对输入图像进行特征提取，并通过特征金字塔网络（FPN）进行特征融合，得到的特征图被送入区域候选网络（RPN）。

RPN 利用滑动窗口机制扫描特征图，生成候选框并识别潜在目标区域。对于每个候选框，RPN 会输出两个结果：一是用于区分前景与背景的分类标签；二是用于优化目标拟合的边界框预测。如果存在多个候选框重叠，系统会通过非极大值抑制（NMS）方法进行筛选。接下来，ROI 池化技术^[7]被用于解决感兴趣区域（RoI）大小不一致的问题，利用 ROI-Align 方法^[8]通过双线性插值来精确计算每个感兴趣区域的四个边界点的像素值，从而减少采样误差。

在分类过程中，这些 ROI 区域会通过全连接网络进行处理，预测目标的类别和边界框。同时，掩码分割子网络使用前述的 ROI-Align 定位信息，对每个候选区域进行逐像素分类，从而得到目标的最终掩码。这个掩码用于精确地匹配和分割目标。

在获得散热板的掩码后，通过掩码的轮廓提取出凸包，从而实现对整个底板区域的完整包裹。凸包是计算机图形学中的常用概念，指的是将二维平面中点集的最外层点连接形成的凸多边形。凸包可以有效地进行多边形近似，对于规则和不规则形状均具有良好的描述效果。在获得凸包后，使用霍夫变换检测边界直线，如图 7 所示，并通过延长相邻直线的交点，即图 8 中白色的四个点来定位四边形，从而实现对底板的拟合。

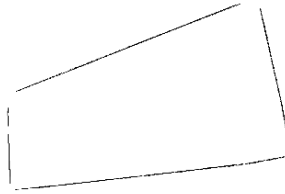


图 7 霍夫变换检测直线

150



图 8 掩膜处理示意图

155

最后，使用透视变换将底板转换为标准的正视图，为后续的结霜检测提供统一的视角。透视校正是一种将图像投影到新的视平面上的变换技术，其基本原理是通过四对对应点（分别来自原始图像和目标图像）来计算一个 3×3 的变换矩阵。透视校正的变换矩阵可以通过式（2）来求得。

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2)$$

最终底板被变换为图 9 状态。



图 9 透视变换后散热板图片

160

2.3 基于改进 OCRNet 语义分割模型的结霜检测算法

165

在当前实际应用中，大多数场合仍依赖人工观察散热板的结霜程度来决定何时进行除霜。传统的卷积神经网络，如 Mask-RCNN 和 U-Net，通过堆叠卷积层和池化层进行逐层特征提取，这种方法仅考虑了像素局部区域的上下文信息，忽略了全局图像中目标像素之间的关联，从而影响了分割的准确性。为了解决这个问题，本文通过显式建模全局图像中目标间的上下文关系，采用了目标上下文卷积神经网络（Object-Contextual Representations Network，

OCRNet) 来进行结霜检测。OCRNet 的主干网络使用了 HRNet^[9]。

OCRNet 是一种专为语义分割设计的网络结构，能够有效地利用对象上下文信息，从而提升分割的精度。OCRNet 的工作流程分为三个阶段，如图 10 所示。在第一阶段，网络根据骨干网输出的特征图生成初步的语义分割结果，这些结果作为软目标区域 (SoftObjectRegions)，每个区域与一个特定的语义类别相对应；第二阶段，利用网络深层的像素特征表示 (PixelRepresentations) 与软目标区域，生成 K 组向量，作为目标区域表示 (ObjectRegionRepresentations)，每组向量代表一个语义类别的特征；在第三阶段，采用自注意力机制^[10]，计算像素表示与对象区域表示之间的关系矩阵，并根据矩阵中的值加权求和各对象区域特征，从而得到对象上下文表示 (ObjectContextualRepresentations)。最终，将像素表示与对象上下文表示在通道维度上进行拼接，生成增强的特征表示 (AugmentedRepresentations)，用于预测每个像素的语义类别。

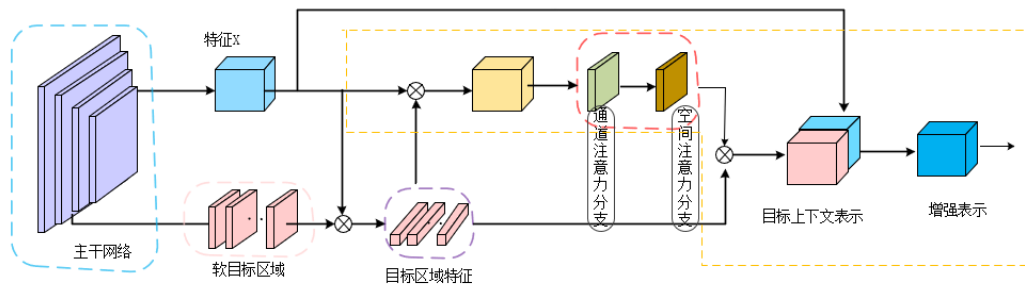


图 10 改进 OCRNet 网络结构

为了提升模型在处理结霜图像时的表现，本文对基于 OCRNet 的原始结构进行了改进。经典的 OCRNet 通过像素级的类别标签进行像素分类，并在网络结构中引入了多头注意力机制，以增强特征的表达能力。然而，这种结构仅能捕捉空间维度上的上下文信息，无法充分考虑通道维度的上下文信息。为了解决这个问题，本文引入了极化注意力模块 (polarizedself-attention, PSA)^[11]，如图 10 红色区域所示，该模块包含通道注意力机制，从而进一步提升了模型的性能。

极化注意力模块由通道注意力和空间注意力两个部分组成。通道注意力模块通过计算每个通道的权重并将其应用到相应的通道上，进而增强重要特征。空间注意力模块通过建立像素间的远程依赖关系，显著提升了 OCRNet 中目标区域的判别能力。

设输入到极化注意力模块的特征张量为 $X \in R^{C \times H \times W}$ ，增强后的特征张量为 $Y \in R^{C \times H \times W}$ 。本文采用了串联结构的极化注意力模块，公式(3)所示。

$$Z = A_{sp} \left[A_{ch}(X) \odot_{ch} X \right] \odot_{sp} A_{ch}(X) \odot_{ch} X \quad (3)$$

其中: A_{ch} 代表通道注意力; \odot_{ch} 表示通道维度上的逐元素乘法; A_{sp} 表示空间注意力; \odot_{sp} 表示空间维度上的逐元素乘法。

195 通道注意力部分计算公式为式(4):

$$A_{ch}(X) = \sigma \left(W_z \left\{ \text{Softmax} \left[W_q(X) \right] \otimes \sigma \left[W_v(X) \right] \right\} \right) \quad (4)$$

其中: W_Q 、 W_V 、 W_z 表示 1×1 卷积; σ 表示激活函数; \otimes 表示矩阵点乘。

空间注意力部分计算公式为式(5):

$$A_{sp}(X) = \sigma \left(\text{Softmax} \left\{ \sigma \left[\text{AvgPool} \left(W_q(X) \right) \right] \otimes \sigma \left[W_v(X) \right] \right\} \right) \quad (5)$$

200 其中 AvgPool 表示全局平均池化。

为了提升边界区域的分割精度, 本文采用了 OhemCrossEntropyLoss 损失函数^[12]。OhemCrossEntropyLoss 是专为深度学习中的目标检测任务设计的, 它在传统交叉熵损失函数的基础上进行了改进, 旨在应对数据分布不平衡和难样本训练问题。“Ohem”代表
205 “OnlineHardExampleMining”, 即在线困难样本挖掘。在目标检测任务中, 背景类的样本通常远多于目标类样本, 这种不平衡的分布可能会影响模型的训练效果。而一些难以分类的样本对于网络的训练来说, 常常是最具挑战性的。为了解决这些问题, OhemCrossEntropyLoss 损失函数在训练过程中重点关注那些损失值较大的困难样本, 帮助模型更好地适应这些样本, 从而提升整体性能。其具体计算公式如(6)所示:

$$OhemCrossEntropyLoss = -\frac{1}{N} \sum_{i=1}^N \begin{cases} \log(p_{target}) & \text{if } y_{target}=1(\text{目标类样本}) \\ \log(1-p_{target}) & \text{if } y_{target}=0(\text{背景类样本且损失高于阈值}) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

210 其中, N 是 Batch 中样本的数量, p_{target} 是模型预测目标类的概率, y_{target} 是真实标签 (1 表示目标类, 0 表示背景类), 损失计算根据标签的情况进行不同的处理。背景类样本中损失值高于一个预定义的阈值的样本会被选中进行梯度更新, 这样网络更关注于难以分类的样本, 有助于提高性能。

最终底板图像及结霜检测效果分别如图 11 和图 12 所示。



图 11 原始图像

215



图 12 结霜检测结果

3 结霜检测实验及分析

本文所用的数据集来自某集团冷库实拍图片，包括训练集 588 张，验证集 72 张，测试集 72 张。

3.1 实验环境

(1) 操作系统与硬件信息：

操作系统：Windows64 位

GPU: NVIDIA GeForce RTX3060 Laptop GPU

CUDA 版本：12.0，支持 CUDA 加速

Python 版本：3.8.20

(2) 软件库与框架：

PyTorch 版本：1.11.0

CUDARuntime: 11.3

编译工具：MSVC192829337 (Microsoft C/C++ 优化编译器)

Intel Math Kernel Library (MKL)：版本 2020.0.2，支持 AVX512 指令集

3.2 评价指标

(1) IoU: IoU 指标即交并比，在语义分割中作为标准度量一直被人使用。交并比不仅在语义分割中使用，在目标检测等方向也是常用的指标之一。计算公式为：

$$IoU = \frac{|P \cap G|}{|P \cup G|} \quad (7)$$

$|P \cap G|$ 表示的是所预测的区域与真实区域的交集的像素个数， $|P \cup G|$ 表示预测的区域与真实区域的并集的像素个数。

(2) F1-score

表示 *Precision* 和 *Recall* 的调和平均数，计算公式为：

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (8)$$

其中 *Precision* 为查准率，表示模型所预测的正例样本中，预测正确样本的占比，

Recall 表示所有真实标签为正的样本，被预测出来的占比。

3.3 实验结果

表 1 各模型 IoU 与 F1-score 对比

	IoU	F1-score
Ocrnet_hr18	83. 58	90. 64
Ocrnet_psa	85. 28	91. 73
OCRNet_ohem	85. 99	92. 17
OCRNet_psa_ohem	86. 93	93. 62

可以看到，本文提出的改进 OCRNet 对比原始 OCRNet 均有改进，并且两种改进方式结合会有更好的效果。

4 计算结霜率

为定量评估图像中结霜区域的覆盖程度，本研究采用了基于 OCRNet 的语义分割模型对图像进行处理，提取结霜区域的掩膜。通过 OCRNet 对输入图像进行推理，得到一个包含结霜区域的掩膜。掩膜中的非零部分对应于结霜区域，而零值部分则表示背景或非结霜区域。随后，将掩膜图像被转换为二值图像，其中结霜区域被标记为白色（255），其他部分标记为黑色（0），如图 13 所示。



图 13 二值化掩膜图

在得到二值化掩膜后，就可以计算结霜区域在整张图像中的面积占比，首先将掩膜数据转化为 NumPy 数组，然后统计二值掩膜中所有非零像素的数量，即结霜区域的像素数，结霜率即为：

结霜率 = (结霜区域像素数 / 图像总像素数) × 100% (9)

与真实标注做对比，并计算误差：

表 2 计算所得结霜率与真实结霜率误差比较（部分）

真实值	测量值	误差
0.4019	0.4066	0.011694
0.3819	0.3848	0.007594
0.4163	0.3948	0.05165
0.4143	0.4135	0.000242
0.1890	0.18	0.004762
0.3582	0.3673	0.025405
0.2041	0.1914	0.06222
0.0993	0.0896	0.09768
0.346	0.3179	0.08121
0.2491	0.2618	0.050984
0.2224	0.2034	0.08543
0.0225	0.0223	0.00889
0.2102	0.213	0.013321

由表 2 可见，在测试集上平均误差为 4.32%，满足实际需求。

5 结论

本文提出了一种基于改进 MaskR-CNN 和 OCRNet 的冷热设备结霜检测方法，旨在解决传统检测方法在低光照、反射干扰、摄像头畸变以及多角度拍摄条件下的不足。通过 MaskR-CNN 进行散热器底板区域的精确分割，并通过改进的 OCRNet 进行结霜区域的精确分割和结霜率计算，从而实现了高效且精准的按需除霜控制。通过对比实验证明，改进后的方法在 IoU、F1-score 等指标上均表现出优异的性能，进一步验证了该方法的有效性。结霜率的计算误差保持在 4.32% 的范围内，满足某集团实际应用的精度要求，为智能除霜控制系统的实现提供了可靠的技术支持。

[参考文献] (References)

[1] 姚建凤, "冷链物流发展思考," 合作经济与科技, no. 16, pp. 76-77, 2021, doi: 10.13665/j.cnki.hzjjykj.2021.16.031.

[2] J. A. Dopazo, J. Fernandez-Seara, F. J. Uhía, and R. Diz, "Modelling and experimental validation of the hot-gas defrost process of an air-cooled evaporator," International journal of refrigeration, vol. 33, no. 4, pp. 829-839, 2010.

[3] J.-M. Huang, W.-C. Hsieh, X.-J. Ke, and C.-C. Wang, "The effects of frost thickness on the heat transfer of finned tube heat exchanger subject to the combined influence of fan types," Applied Thermal Engineering, vol. 28, no. 7, pp. 728-737, 2008

[4] H. Cho, Y. Kim, and I. Jang, "Performance of a showcase refrigeration system with multi-evaporator during on-off cycling and hot-gas bypass defrost," Energy, vol. 30, no. 10, pp. 1915-1930, 2005.

[5] Y. Jiang, J. Dong, M. Qu, S. Deng, and Y. Yao, "A novel defrosting control method based on the degree of

refrigerant superheat for air source heat pumps," International journal of refrigeration, vol. 36, no. 8, pp. 2278-2288, 2013.

290 [6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[7] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 6, pp. 1137-1149, 2016.

295 [8] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961-2969.

[9] K. Sun et al., "High-resolution representations for labeling pixels and regions," arXiv preprint arXiv:1904.04514, 2019.

[10] A. Vaswani, "Attention is all you need," Advances in Neural Information Processing Systems, 2017.

300 [11] H. Liu, F. Liu, X. Fan, and D. Huang, "Polarized self-attention: Towards high-quality pixel-wise mapping," Neurocomputing, vol. 506, pp. 158-167, 2022.

[12] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 761-769.