



## 特征差异增强与残差蒸馏网络结合的医药可见光图像异物检测

刘优武, 张辉, 孔森林, 陶岩, 李冲

引用本文:

刘优武, 张辉, 孔森林, 等. 特征差异增强与残差蒸馏网络结合的医药可见光图像异物检测[J]. 智能系统学报, 2025, 20(1): 118-127.

LIU Youwu, ZHANG Hui, KONG Senlin, et al. Foreign object detection in pharmaceutical visible-light images using feature difference enhancement and residual distillation network[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(1): 118-127.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202311023>

## 您可能感兴趣的其他文章

### 多视图主动学习的多样性样本选择方法研究

Diversity sample selection method of multiview active learning classification

智能系统学报. 2021, 16(6): 1007-1014 <https://dx.doi.org/10.11992/tis.202007037>

### 自步稀疏最优均值主成分分析

Sparse optimal mean principal component analysis based on self-paced learning

智能系统学报. 2021, 16(3): 416-424 <https://dx.doi.org/10.11992/tis.201911028>

### 基于孪生变分自编码器的小样本图像分类方法

A small-sample image classification method based on a Siamese variational auto-encoder

智能系统学报. 2021, 16(2): 254-262 <https://dx.doi.org/10.11992/tis.201906022>

### 面向自动驾驶目标检测的深度多模态融合技术

Deep multi-modal fusion in object detection for autonomous driving

智能系统学报. 2020, 15(4): 758-771 <https://dx.doi.org/10.11992/tis.202002010>

### 多层卷积特征的真实场景下行人检测研究

Research on pedestrian detection based on multi-layer convolution feature in real scene

智能系统学报. 2019, 14(2): 306-315 <https://dx.doi.org/10.11992/tis.201710019>

### 基于医学征象和卷积神经网络的肺结节CT图像哈希检索

Hashing retrieval for CT images of pulmonary nodules based on medical signs and convolutional neural networks

智能系统学报. 2017, 12(6): 857-864 <https://dx.doi.org/10.11992/tis.201706035>

DOI: 10.11992/tis.202311023

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20241223.1250.010>

# 特征差异增强与残差蒸馏网络结合的 医药可见光图像异物检测

刘优武<sup>1</sup>, 张辉<sup>2</sup>, 孔森林<sup>1</sup>, 陶岩<sup>1</sup>, 李冲<sup>3</sup>

(1. 长沙理工大学 电气与信息工程学院, 湖南 长沙 410114; 2. 湖南大学 机器人学院, 湖南 长沙 410012; 3. 楚天科技股份有限公司, 湖南 长沙 410600)

**摘要:** 医药中的异物通常形态微弱, 导致轻量化算法无法准确检测, 而高精度算法通常实时性差。为兼顾医药异物检测的实时性与准确性, 提出了一种深度学习蒸馏算法, 能够快速、准确地检测药液图像中的异物。首先, 在教师网络中引入基于语义特征的上采样方法, 增强了教师网络与学生网络之间的特征差异。同时, 在学生网络的训练图像中加入随机噪声, 提高了在高干扰场景下的鲁棒性。为验证算法的有效性, 在灯检设备采集了药液异物数据集并进行了对比实验, 蒸馏后平均精度提升了 4.1 百分点, 每秒帧数达到了 65, 优于目前已有的先进方法。最后, 在天池酒液数据集进行拓展实验, 检测的平均精度提升了 3.9 百分点, 验证了模型在类似场景中的适用性。

**关键词:** 医药; 异物; 轻量化; 深度学习; 蒸馏; 特征差异; 上采样; 灯检**中图分类号:** TP391 **文献标志码:** A **文章编号:** 1673-4785(2025)01-0118-10

中文引用格式: 刘优武, 张辉, 孔森林, 等. 特征差异增强与残差蒸馏网络结合的医药可见光图像异物检测 [J]. 智能系统学报, 2025, 20(1): 118-127.

英文引用格式: LIU Youwu, ZHANG Hui, KONG Senlin, et al. Foreign object detection in pharmaceutical visible-light images using feature difference enhancement and residual distillation network[J]. CAAI transactions on intelligent systems, 2025, 20(1): 118-127.

## Foreign object detection in pharmaceutical visible-light images using feature difference enhancement and residual distillation network

LIU Youwu<sup>1</sup>, ZHANG Hui<sup>2</sup>, KONG Senlin<sup>1</sup>, TAO Yan<sup>1</sup>, LI Chong<sup>3</sup>

(1. School of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha 410114, China; 2. School of Robotics, Hunan University, Changsha 410012, China; 3. Truking Technology Limited, Changsha 410600, China)

**Abstract:** Foreign objects in pharmaceuticals are typically small, which causes difficulty for lightweight algorithms to detect them accurately, while high-performance algorithms often struggle with real-time capability. To balance real-time performance and accuracy, a deep learning distillation algorithm is proposed for the precise and rapid detection of foreign objects in pharmaceutical liquid images. The teacher network incorporates a semantic feature-based upsampling method to enhance the feature disparity between teacher and student networks. In addition, random noise is added to the training images of the student network to improve robustness in high-noise detection scenarios. To validate the effectiveness of the algorithm, a pharmaceutical liquid foreign-object dataset is collected using lamp inspection equipment, and comparative experiments are conducted. After distillation, the average precision improves by 4.1%, and the model achieves 65 frames per second, which surpasses current state-of-the-art methods. Extended experiments on the Tianchi liquor dataset show a 3.9% improvement in detection accuracy, which demonstrates the applicability of the model in similar scenarios.

**Keywords:** pharmaceutical; foreign objects; lightweight; deep learning; distillation; feature disparity; upsampling; lamp inspection

收稿日期: 2023-11-17. 网络出版日期: 2024-12-24.

基金项目: 科技创新 2030—“新一代人工智能”重大项目 (2021ZD0114503); 国家自然科学基金重大研究计划项目 (92148204); 国家自然科学基金项目 (62027810); 湖南省科技创新领军人才项目 (2022RC3063); 湖南省十大技术攻关项目 (2024GK1010); 湖南省重点研发计划项目 (2023GK2068, 2022GK2011).

通信作者: 张辉. E-mail: [zhanghuihy@126.com](mailto:zhanghuihy@126.com).

医药异物通常指的是药品生产中非有意添加的、不溶于药物的、非气泡的微粒, 常见的异物包括橡胶、玻璃屑、纤维、毛发等可见的杂质<sup>[1]</sup>。由于药液大多以口服或注射方式作用于人体, 药液中的异物可能会降低药效, 或者导致口腔或食道划

伤,甚至在严重情况下可能阻塞血管,对人体的健康和生命构成威胁<sup>[2]</sup>。然而,在药液的加工或灌装过程中,不可避免地会产生带有异物的次品<sup>[3]</sup>。因此,药液异物检测成为了生产过程中至关重要的环节,研究一种高效的药液异物检测技术对保障药品质量与安全具有极其重要的社会意义。

在实际的工业生产中,医药检测通常划分为人工灯检和机器灯检2类。然而,人工灯检存在着效率低、判断主观性强等诸多问题,难以满足高速准确检测异物的需求。随着设备算力的不断突破,越来越多的领域开始使用机器视觉算法代替人工进行检测<sup>[4-8]</sup>,基于机器视觉的灯检算法也持续演进,其检测精度逐渐上升,已成为目前主流的工业医药检测方式。在早期阶段,学者们借助传统的机器学习算法来实现对较大异物的检测,例如,Wang等<sup>[9]</sup>利用机械臂按一定力度摇晃药瓶,然后使用相机获取药液图像,通过基于边缘检测和形态学处理的图像分割技术将注射器图像中的液体和玻璃碎片分离;Dai等<sup>[10]</sup>提出了基于差异的细节保留自适应滤波算法,利用基于三帧差相位的方法检测图像中的异物;Zhang等<sup>[11]</sup>提出了基于空间信息的在线顺序极限学习机,在检测过程中利用面积、平均灰度值等特征,区分了气泡和异物。然而,这些方法存在泛化性差的问题,特别在微小异物的检测方面容易产生漏检。一些学者为了提升检测准确度,尝试使用更为强大的卷积神经网络进行检测。例如,Zhang等<sup>[12]</sup>将药液检测任务分为2部分,第1部分利用卷积神经网络定位异物,第2部分则利用轨迹信息剔除误检的定位结果。然而,由于选用的卷积神经网络并未针对小目标检测进行改进,导致定位效果并不理想。随后,Yi等<sup>[13]</sup>提出了一种基于多特征融合的无锚框卷积神经网络,从而提升了微小异物的定位精度,但实时性难以满足生产需求。由此可见,在异物检测领域仍存在难点,本文将难点总结如下:1)相对于其他工业检测项目,药液异物检测对检测速度要求更高,现有的深度学习异物检测算法难以在保证准确性的情况下提高速度。2)由于异物微小,其在药液图像中所占像素比例极小,容易在特征提取过程中丢失信息,从而导致检测算法定位不精确或者漏检。3)药液检测场景中存在许多干扰因素,例如反光、气泡等,需要检测算法具备强大的辨别能力。

针对这些难点,本文提出了一种基于卷积神经网络的异物检测方法,能够对药液中的异物进行精确的检测定位,该方法的主要贡献包括:1)使用改进后的交叉蒸馏方法,利用大网络指导小网络的训练,令小网络兼备较高的检测能力和

检测速度。2)设计实验探究在训练图像中引入不同类型的噪声对网络在高干扰药液检测场景下检测性能的影响。3)在教师网络的特征融合阶段中引入了基于语义特征的上采样方法,能够有效增强微小异物的特征信息,增强了教师网络和学生网络的特征差异性,使得学生网络得到更好的训练。4)在实际的灯检设备中采集并制作了药液异物数据集,对所提出的网络进行训练和测试对比实验,以证明其先进性,同时在天池酒液异物数据集也进行了拓展实验,证明算法在相似场景下的有效性。

## 1 研究基础与相关技术

### 1.1 灯检机

在药液生产中,灯检机的主要任务是判定生产的药品是否符合标准,并将合格品和次品进行分离。通常情况下,灯检机需要进行瓶身缺陷、瓶盖缺陷以及药液异物等方面的检测工作。图1给出了本文所使用的灯检机,其工作流程如下:首先,通过进料口的进瓶螺杆将瓶子输送至灯检系统中。接着,旋转轮组推动瓶子进入检测区域,定位机构将瓶子牢固地固定在指定位置。随后,旋转机构以高速旋转瓶子,以便于将瓶底可能存在的异物旋起,从而方便拍摄。接下来,相机会连续拍摄每个瓶子中的药液图像。然后,采集到的图像将被发送至工业服务器进行检测,系统通过串行端口将服务器有关药液鉴定的数据传输至上位机。最终,转盘组会根据服务器作出的判断,将合格品和次品分别送往不同的出料口。

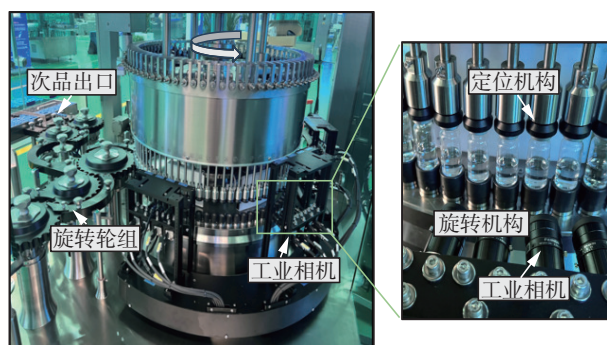


图1 灯检机实物

Fig. 1 Physical of the inspection machine

### 1.2 药液异物数据集

由于实际生产中次品出现的概率相对较低,因此获取足够数量的带有异物的药液图像比较困难。为此,本文将装有毛发、橡胶屑、玻璃碎片以及纤维等异物的药液送入上述的药液异物灯检机中,并利用其中的工业相机拍摄了共3 300张药液图像。随后,按照视觉目标类(visual object



classes, VOC)<sup>[14]</sup> 数据集的格式, 对每一张图像中的异物进行了标注, 以用于训练和验证算法。其中, 训练集与测试集的比例为 8:2, 图像的大小统一设定为 1 088 像素×460 像素, 所有类别的异物在标签中均统一归为一类。图 2 给出了数据集的示例, 可以明显看出, 检测场景中存在较多的干扰, 药液中的异物十分微小, 并且不同类别的异物形态差异较大。

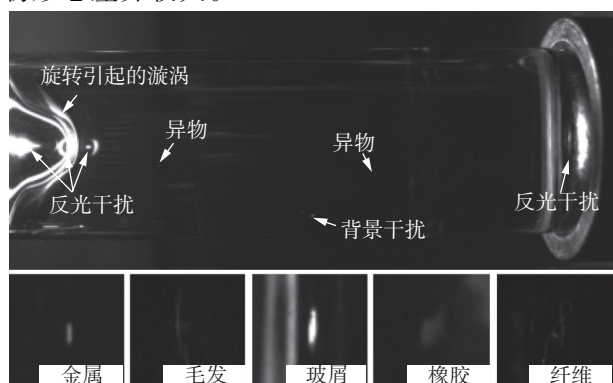


图 2 药液异物数据集示例

Fig. 2 Example of foreign particulates in pharmaceutical liquid

### 1.3 交叉蒸馏方法

在异物微小且干扰繁多的药液异物检测任务中, 即使采用大型深度网络, 也很容易出现误检漏检的问题。这导致了直接应用传统的知识蒸馏方法来训练学生网络时, 大型教师网络提供的指导信息与数据集真实标签的指导信息存在显著差异, 反而降低了学生网络的训练效果。为了解决这一问题, 本文采用了交叉蒸馏方法<sup>[15]</sup>。该方法通过将学生模型检测头的中间特征传递至教师模型的检测头进行预测, 并计算蒸馏损失, 使得蒸馏指导和正常的数据集标签指导相互独立, 能够

有效缓解上述问题。

### 1.4 注意力机制

注意力机制是一种可以通过训练自动学习数据权重的神经网络模块, 其作用是使网络关注特定的重点区域。具体地说, 注意力模块会分析输入的语义信息, 并生成一个与输入大小一致的权重矩阵将输入中被检测对象(如异物)所在位置的关注度提升, 即在该位置分配更高的权重, 而背景的位置则分配低权重, 以实现注意力的效果。目前的主流注意力方法主要分为 3 种, 第 1 种是基于通道的注意力机制<sup>[16]</sup>, 通过通道之间的语义关系生成注意力权重; 第 2 种是基于像素和通道的注意力机制<sup>[17-18]</sup>, 通过组合分析通道和像素的语义信息生成注意力权重; 最后一种通过能量函数实现注意力<sup>[19]</sup>, 本文利用注意力机制提高教师网络提取的特征图质量, 使学生网络有更好的学习效果。

## 2 特征差异增强的残差蒸馏方法

本文主要围绕医药检测对速度和精度的高要求, 以及根据异物微小的特征, 进行模型的搭建和改进。为了获取兼顾高速和高精度的网络, 采用交叉蒸馏作为整体框架进行算法搭建, 同时为了让学生网络在蒸馏时获得更好的学习效果, 在教师网络中加入了基于语义的特征融合模块以及注意力模块, 提升了教师网络的特征质量, 并在学生网络的训练图像中随机加入噪声, 加强了学生网络的鲁棒性, 也扩大了学生网络和教师网络之间的特征差异, 下面详细介绍本文的具体架构和改进部分。

### 2.1 整体框架

本文算法的整体框架如图 3 所示。

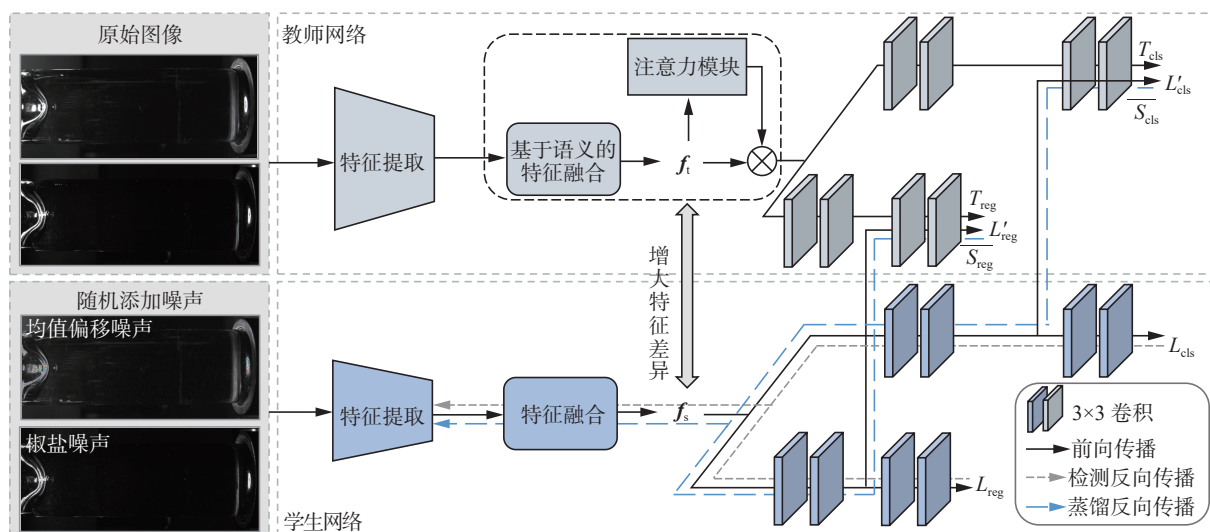


图 3 基于特征差异增强的交叉蒸馏药液异物检测方法整体结构

Fig. 3 Overall of feature-difference-enhanced cross-distillation method for pharmaceutical liquid foreign particulates

算法主要由 2 个部分构成, 分别是教师网络和学生网络。其中, 学生网络部署于实际的药液异物检测任务中, 因此为了保证其实时性, 特征提取网络采用了深度为 18 层的残差网络 (residual network18, ResNet-18)<sup>[20]</sup>, 特征融合网络采用了特征金字塔网络 (feature pyramid networks, FPN)<sup>[21]</sup>。而教师网络作为指导网络, 需要考虑到检测精度, 因此选用了 ResNet-50 作为特征提取网络, 特征融合网络结构采用了 FPN 的结构, 但是融合过程中的上采样方法替换为了基于语义特征的上采样方法, 并在特征融合部分的输出后加入了注意力模块, 以增强教师网络对微小异物信息的提取能力。最后, 教师网络和学生网络的检测头均选用边界分布<sup>[22]</sup>检测头。

## 2.2 基于语义特征的上采样方法

传统的上采样大多采用插值法将低分辨率的图像放大, 插入的值通常以固定的权重进行采样, 并没有考虑到像素之间的语义信息, 文献 [23] 利用卷积分析语义信息生成上采样核, 在开放世

界数据集中取得了很好的效果。但是其考虑到网络部署时的轻量化和实时性, 没有设计合适的深度卷积网络提取像素间的语义信息, 导致微小异物信息不能很好地被还原。本文在此基础上, 提出了基于语义特征的上采样方法, 具体的结构图如图 4 所示, 该方法包括上采样核预测模块和上采样重组模块, 前者目的是根据语义信息生成  $\alpha H \times \alpha W$  个大小为  $k_{up} \times k_{up}$  的上采样核, 后者是利用上采样核和原特征图  $F$  进行运算, 得到最终的上采样特征图  $F''$ 。以  $\alpha$  倍上采样为例, 网络首先将长为  $H$ 、宽为  $W$ 、通道数为  $C$  的特征图  $F$ , 送入上采样核预测模块, 为了充分提取  $F$  中的语义信息, 语义分析模块采用了深度阶梯式提取方法, 首先用  $1 \times 1$  卷积提升特征通道数, 然后按通道将  $F$  平均分为  $n$  份, 再利用  $3 \times 3$  卷积  $\text{Conv}_i(\cdot)$  进行语义特征提取, 分别得到  $n$  份特征  $F'_i$ , 公式为

$$F'_i = \begin{cases} \text{Conv}_i(F), i = 1 \\ \text{Conv}_i(F_i + F'_{i-1}), 1 < i \leq n \end{cases}$$

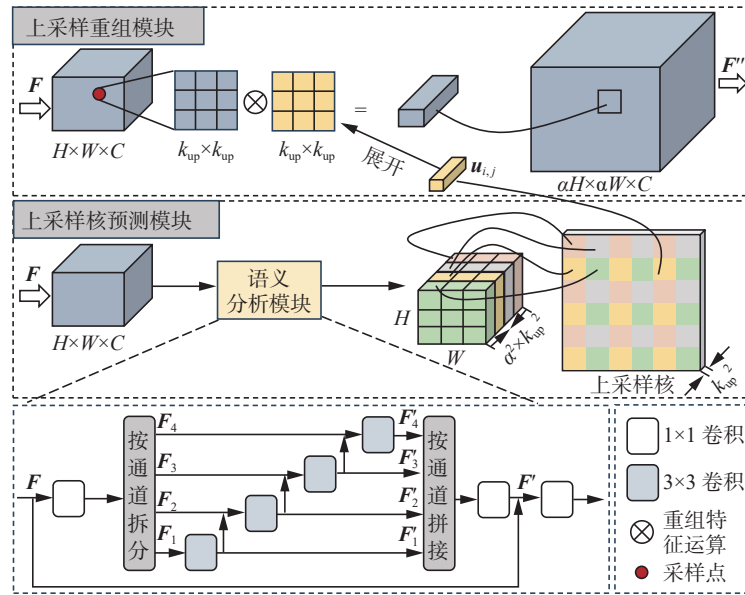


图 4 基于语义特征的上采样方法结构

Fig. 4 Structural of upsampling method based on semantic features

提取的  $F'_i$  中包含了不同深度的语义信息, 本文通过通道拼接  $\text{concat}(\cdot)$  融合信息, 并通过多个  $1 \times 1$  卷积  $L(\cdot)$  对齐通道, 公式为

$$F' = L(\text{concat}(F'_1, F'_2, \dots, F'_n)) + F$$

融合特征后, 再将  $F'$  的通道压缩为  $\alpha^2 \times k_{up}^2$ , 然后将压缩后的通道均分为  $\alpha^2$  部分, 即每部分通道数为  $k_{up}^2$  (图 4 中给出  $\alpha = 3$  的例子)。按照图中对应关系, 使大小  $H \times W \times \alpha^2 \times k_{up}^2$  的  $F'$  变为  $\alpha H \times \alpha W \times k_{up}^2$  的上采样核矩阵  $U_k$ , 其中每一个像素的

通道展开后对应一个大小为  $k_{up} \times k_{up}$  的上采样核  $u_{i,j}$ 。

得到  $\alpha H \times \alpha W$  个上采样核后, 上采样重组模块进行重组操作, 重组操作得到的上采样图像  $F''$  大小为  $\alpha H \times \alpha W \times C$ ,  $F''$  中每一个点  $(i', j')$  与原特征图  $F$  点  $(i, j)$  的对应关系为

$$\begin{cases} i = [i' / \alpha] \\ j = [j' / \alpha] \end{cases}$$

式中  $[\cdot]$  表示取整。  $F''$  每个点  $(i', j')$  对应一个上采样核  $u_{i,j}$ 。重组时, 从  $i' = 0, j' = 0$  为起点依次求取

$F''_{i,j}$  对应的值, 特征图  $F$  中以点  $(i, j)$  为中心取出  $k_{up} \times k_{up}$  大小的区域记为  $F_{i,j}(k_{up})$ , 与不同的上采样核运算, 得到上采样图像  $F''$ , 该计算过程称为重组特征运算,  $F_{i,j}(k_{up})$  与上采样核运算的公式为

$$F''_{i,j} = \sum_{n=-k_{up}/2}^{k_{up}/2} \sum_{m=-k_{up}/2}^{k_{up}/2} u_{i,j} \cdot F_{i+n,j+m}(k_{up})$$

### 2.3 特征差异增强

为了使学生网络有更好的学习效果, 本文在特征融合的输出端加入了注意力模块  $\phi$ , 目的是利用注意力模块生成输入特征  $f$  的权重矩阵  $\phi(f)$ , 增强特征图中异物位置的值, 使异物定位更加准确, 提升特征图的质量, 注意力模块按公式接入网络:

$$F = \phi(f) \times f$$

除此之外, 为了增加网络在高干扰场景下的检测能力, 本文还在学生网络的训练输入图像中增加了随机噪声。网络单独训练时, 由于只有检测框标签的指导, 此时在训练集中添加噪声对网络的提升很有限。但是在交叉蒸馏中, 除了检测的反向传播外, 还存在蒸馏的反向传播, 此时添加噪声, 会对学生网络提取的特征产生一定影响, 这增大了与教师网络特征之间的差异, 通过与教师网络特征对比, 能够区分图像中的干扰, 并从干扰中提取有效的信息, 有效地增强高干扰场景下的检测能力。具体的操作如下:

在每一个训练轮次中, 训练集共有  $n$  张输入图像  $P_i$  以及对应的标签, 输入到教师网络的图像为原始图像  $P_i$ , 以保证所得特征图的准确性。而学生网络的输入图像  $P'_i$  会在整张图像中随机加入椒盐噪声  $\gamma$ 、均值偏移噪声  $\delta$ , 具体公式为

$$P'_i = P + \gamma \times \text{rand}(1) + \delta \times \text{rand}(1)$$

式中:  $\text{rand}(1)$  表示随机从 0 和 1 两者中选取一个数, 并且每个  $\text{rand}(1)$  相互独立。这种添加噪声的方式能够提升数据集的多样性, 提升学生网络的训练效果。

### 2.4 训练流程和损失函数

在训练学生网络之前, 需要在药液数据集中单独训练教师网络, 并将训练后的权重保存。

训练学生网络的具体流程如图 3 所示, 此时教师网络导入并冻结了所有的权重, 不受反向传播的影响。训练开始后, 在前向传播阶段, 教师网络输入原始图像  $P_i$ , 学生网络则同步输入带有噪声的对应图像  $P'_i$ , 做出对应的预测。反向传播阶段, 学生网络则通过与标签  $l_i$  进行对比得到的分类损失  $L_{cls}$  和回归损失  $L_{reg}$  以及与教师网络预测

结果的蒸馏损失  $L'_{cls}$  和  $L'_{reg}$  进行反向优化。训练完成后, 部署时单独使用学生网络进行异物检测。

考虑到药液数据集存在正负样本不平衡的问题, 采用了质量焦点损失 (quality focal loss, QFL)<sup>[22]</sup> 损失作为分类损失  $L_{cls}$ , 同时选择了广义交并比 (generalized intersection over union, GIoU)<sup>[24]</sup> 损失作为回归损失  $L_{reg}$ , 并引入了分布焦点损失 (distribution focal loss, DFL)<sup>[22]</sup> 损失  $L_D$  加速回归损失的收敛。通过应用系数  $\lambda_0$  和  $\lambda_1$ , 模型平衡了  $L_{reg}$  和  $L_D$  2 个损失, 由于这 2 个损失仅适用于包含异物的预测框, 数量为  $N_{pos}$ , 因此引入了一个指示函数  $K(z)$  以明确其应用范围, 具体而言, 如果预测框中有可见的异物, 该函数将返回 1; 否则, 返回 0。将这些损失综合起来, 得到了检测损失  $L_{label}$ , 其具体表达式为

$$L_{label} = \frac{1}{N_{pos}} \sum_z L_{ds} + \frac{1}{N_{pos}} \sum_z K(z) (\lambda_0 L_{reg} + \lambda_1 L_D)$$

在蒸馏反向传播部分, 蒸馏损失  $L'_{cls}$  和  $L'_{reg}$  均采用了交叉蒸馏损失<sup>[15]</sup>, 两者计算的目标是学生网络交叉到教师网络检测头部分的回归预测  $\overline{S}_{reg}$ 、分类预测  $\overline{S}_{cls}$  与教师网络的回归预测  $T_{reg}$ 、分类预测  $T_{cls}$  之间的损失, 蒸馏损失的公式为

$$L_{KD} = L'_{cls}(\overline{S}_{reg}, T_{reg}) + L'_{reg}(\overline{S}_{cls}, T_{cls})$$

最终, 模型的总损失  $L$  为

$$L = L_{label} + L_{KD}$$

## 3 实验与结果分析

### 3.1 实验配置和评价指标

实验均在 Windows10 系统下进行, 中央处理器为 i5-13400F, 显卡为 NVIDIA GeForce RTX 4070 Ti, 使用 PyTorch1.8.1 版本, 批量大小为 4, 训练总轮次为 48, 学习率在 32 轮次前设置为 0.001, 在 32~48 轮次设为 0.000 1, 采用随机梯度下降方法进行网络优化。在该实验场景的显卡上训练本文提出的蒸馏算法需要耗费 3 h, 单独训练教师网络则需要 2.5 h。

实验主要使用的数据集为 1.2 节介绍的药液异物数据集, 拓展实验使用了 2020 年天池大赛复赛的酒液异物数据集, 经过数据清洗后, 选择出 1 522 张带有异物的图像, 大小为 4 096 像素  $\times$  3 000 像素按 8:2 的比例划分训练集和数据集。

为了对比模型的精度、复杂度以及速度, 实验使用的评价指标包括: 平均精度 (average precision, AP)、在交并比 (intersection over union, IoU) 值为 0.5 时的平均精度  $AP_{50}$ 、平均召回率 (average



recall, AR)、每秒帧数 (frames per second, FPS)、浮点运算次数 (floating point operations, FLOPs) 以及参数量 (params)。由于药液异物中大多数都是像素面积小于 32 像素 $\times$ 32 像素的小目标, 因此指标中还加入了小目标预测的平均精度  $AP_{small}$  和平均召回率  $AR_{small}$ 。

### 3.2 对比实验

为了探究在训练图像中引入不同类型的噪声对网络检测性能的影响, 本节设计了加入噪声的对比实验。选择了椒盐噪声、高斯噪声和均匀噪声, 椒盐噪声具有随机性且突出的黑白点, 与异物相似; 高斯噪声模拟药液中的传感器噪声; 均匀噪声具有均匀分布的随机性, 用于模拟药液中的背景噪声。单一噪声作用时网络性能变化不明显, 因此探究组合使用噪声对性能的影响, 实验时其他模块均不作用, 结果如表 1 所示。实验表明, 均匀噪声和椒盐噪声结合使用能最有效地提高召回率, 而 3 个噪声均作用时影响了对正常样本的学习, 导致性能最差。

表 1 加入不同噪声的对比实验结果  
Table 1 Results of experiments on different noise %

添加噪声	$AP_{50}$	$AP_{small}$	AR	$AR_{small}$
不添加噪声	52.2	18.2	26.7	26.1
高斯+椒盐	52.1	17.9	26.9	26.2
均匀+高斯	52.2	18.0	27.1	26.5
均匀+椒盐	<b>52.3</b>	<b>18.4</b>	<b>27.5</b>	<b>26.8</b>
三者同时作用	51.6	17.5	26.3	25.8

为验证本文提出的蒸馏算法的有效性, 本节进行了一系列对比实验, 涉及多个主流蒸馏算法, 包括定位蒸馏网络 (localization distillation, LD)<sup>[25]</sup>、耐心蒸馏网络 (patient knowledge distillation, PKD)<sup>[26]</sup> 以及交叉蒸馏网络 (cross-head knowledge distillation, CKD)<sup>[15]</sup>。为确保实验的公平性, 本文统一了各蒸馏网络所采用的基础网络。学生网络选用了 ResNet-18 作为特征提取网络的广义焦点损失 (generalized focal loss, GFL)<sup>[22]</sup> 网络, 而教师网络则采用了 ResNet-50 的 GFL 网络, 各蒸馏算法在此基础上进行各自的改进 (表 2 中 ours 包含了本文所有模块改进, 包括添加随机噪声)。此外, 为验证添加注意力算法的有效性, 本文在实验中选择了较为主流的注意力算法压缩和激励 (squeeze and excitation, SE) 网络<sup>[16]</sup>、卷积块注意力模块 (convolutional block attention module, CDAM)<sup>[17]</sup>、简单无参数注意力模块 (simple, para-

meter-free attention module, simAM)<sup>[19]</sup> 作为插入的注意力算法, 并进行了详细对比。最终的对比结果整理如表 2 所示。从表中可以看出本文提出的蒸馏算法选用 SE 作为注意力机制时在各项指标上均为最优, 蒸馏效果最好。在此基础上与其他蒸馏算法相比, 本文蒸馏得到的学生网络在  $AP_{50}$ 、AP、AR 分别领先了第 2 名 1.4、1.0 和 1.9 百分点, 由于均没有对学生网络进行改动, 各蒸馏网络在 FPS、params 和 FLOPs 保持了一致。值得注意的是, 经过 LD 蒸馏后学生网络的  $AP_{50}$ 、AP、 $AP_{small}$  这 3 个指标分别降低了 0.9、0.6 和 0.8 百分点, 证明了在药液异物检测任务中, 教师网络大量的错误检测结果会降低蒸馏效果。

表 2 蒸馏算法实验结果  
Table 2 Results of distillation algorithms %

检测模型	$AP_{50}$	AP	$AP_{small}$	AR	$AR_{small}$
不蒸馏	43.3	16.1	15.2	23.5	22.8
PKD <sup>[26]</sup>	45.0	15.8	14.8	23.1	22.4
LD <sup>[25]</sup>	42.4	15.5	14.4	23.8	23.1
CKD <sup>[15]</sup>	52.2	19.2	18.2	26.7	26.1
ours+simAM <sup>[19]</sup>	50.9	18.3	17.2	25.9	25.2
ours+CDAM <sup>[17]</sup>	53.2	20.0	19.1	27.3	26.6
ours+SE <sup>[16]</sup>	<b>53.6</b>	<b>20.2</b>	<b>19.6</b>	<b>27.8</b>	<b>27.2</b>

为了更直观地对比蒸馏网络的效果, 提取了各网络的特征图, 并将其转换为热力图进行可视化对比, 具体结果见图 5。

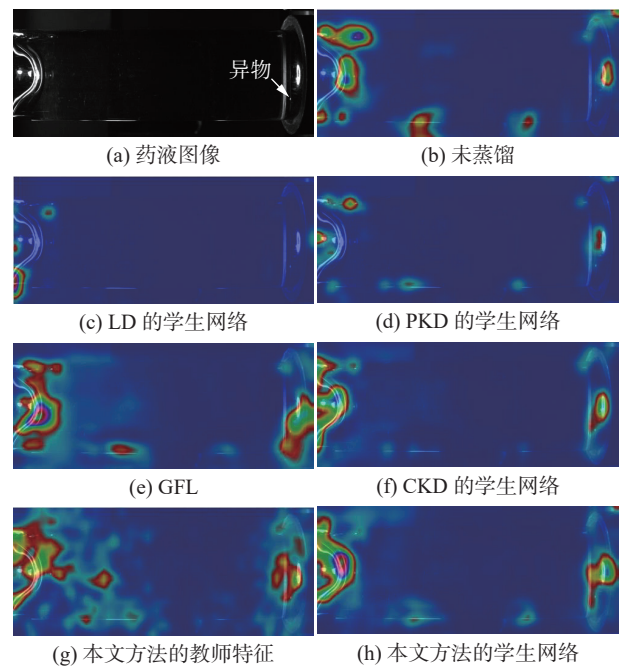


图 5 各蒸馏方法的热力图

Fig. 5 Heatmaps of various distillation methods

图中热力值的大小通过可见光谱映射到图像中, 高热值区域呈现红色, 低热值区域呈现紫色, 热力值越高, 表示网络对相应区域更感兴趣。对比教师特征和 GFL 特征, 可以看出, 本文方法有效增强了特征差异。本文方法的学生网络的特征图与未蒸馏以及 GFL 网络的特征图相比, 背景区域的热力值明显降低, 同时也没有学习到教师网络中错误的特征信息, 能够有效降低误检的产生; 与 LD、PKD 和 CKD 的学生网络相比, 本文的学生网络异物处的热力值更高, 定位更为准确。上述实验验证了本文蒸馏网络的有效性。

为了验证本文在教师网络中所提出模块的有效性, 本节设计了教师网络与主流检测算法的对比实验, 教师网络的基准网络为 GFL 算法, 因此该算法作为主要的对比对象。除此之外, 本文将算法与 Zhang 等<sup>[12]</sup>使用的定位网络——更快的区域卷积神经网络 (faster-region convolutional neural network, Faster-RCNN)、Yi 等<sup>[13]</sup>针对药液异物设计的自适应卷积和多尺度注意力网络 (adapt-

ive convolution and multiscale attention network, ACMA) 以及在公共数据集有较高检测精度的变焦网络 (varifocal net, VFNet)<sup>[27]</sup>、自适应训练样本选择网络 (adaptive training sample selection, ATSS)<sup>[28]</sup>、任务对齐单阶段目标检测 (task-aligned one-stage object, TOOD)<sup>[29]</sup>、只看一层特征网络 (you only look one-level feature, YOLOF)<sup>[30]</sup> 和选择性查询回忆网络 (selective query recollection, SQR)<sup>[31]</sup> 进行对比, 为了对比的严谨性, 这些算法的特征提取网络与教师网络一样均采用 ResNet-50。在药液异物数据集训练后, 进行对比实验, 最终结果如表 3 所示。以  $AP_{50}$  为主要精度指标、AR 作为召回率指标进行对比, 改进后的教师网络相较于其它算法均有较大的提升, 精度指标均为最高,  $AP_{50}$  和 AR 相较于第 2 名的 ACMA 分别高出了 3.1 和 1.0 百分点, 验证了其在蒸馏网络中的检测指导优势。同时与特征提取网络为 ResNet-50 的 GFL 算法相比,  $AP_{50}$  和 AR 分别提升 6.4 和 2.5 百分点, 证明了本文基于语义的上采样模块和加入注意力模块的有效性。

表 3 主流目标检测算法的实验结果  
Table 3 Results of state-of-the-art object detection algorithms

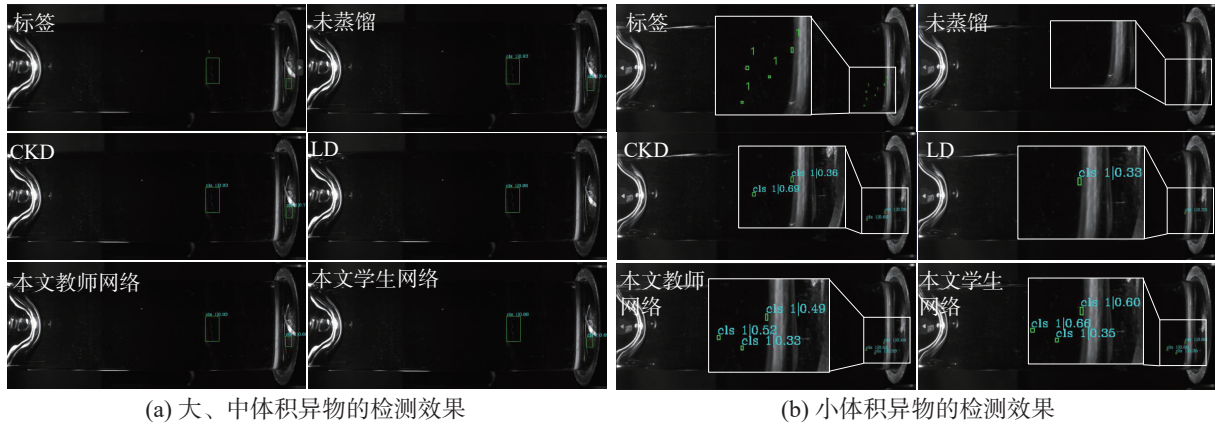
方法	$AP_{50}/\%$	$AP/\%$	$AP_{small}/\%$	$AR/\%$	$AR_{small}/\%$	FLOPs/ $10^9$	params/ $10^3$	FPS
Faster-RCNN <sup>[12]</sup>	28.5	6.8	6.5	18.1	17.9	193.8	41.1	43
ACMA <sup>[13]</sup>	54.3	18.8	17.9	28.9	28.4	197.9	32.6	41
ATSS <sup>[28]</sup>	44.4	15.6	7.1	24.0	23.4	201.4	31.9	47
YOLOF <sup>[30]</sup>	33.6	9.2	8.9	16.9	16.5	<b>98.2</b>	42.1	43
VFNet <sup>[27]</sup>	30.2	13.2	8.3	15.7	14.8	190.0	32.5	27
TOOD <sup>[29]</sup>	54.0	21.4	20.5	28.6	27.9	149.2	32.0	39
GFL <sup>[22]</sup>	51.0	19.8	18.9	27.4	26.7	204.5	32.0	48
SQR <sup>[31]</sup>	52.0	20.4	20.3	28.4	28.1	102.9	134.6	13
本文教师网络	<b>57.4</b>	<b>22.9</b>	<b>22.1</b>	<b>29.9</b>	<b>29.3</b>	212.1	46.4	19
本文学生网络	53.6	20.2	19.6	27.8	27.2	154.9	<b>19.1</b>	<b>65</b>

综合表 2 和表 3, 可以看出不使用蒸馏方法时, 采用 ResNet-18 和采用 ResNet-50 的 GFL 算法相比,  $AP_{50}$  和 AR 分别损失了 7.7 和 3.5 百分点。而经过本文蒸馏后的学生网络相较于教师网络  $AP_{50}$  和 AR 只分别下降了 3.8 和 2.1 百分点, 且精度仍然高于大部分使用 ResNet-50 的主流检测算法, 同时 FPS 达到了 65, 相较于教师网络, FPS 提升了 46, 并且 params 和 FLOPs 大大低于其他算法, 兼顾了实时性和准确性。

除此之外, 本文还在图 6 中可视化了部分网

络的检测结果。图 6(a) 给出了较大体积异物的检测场景, 图中只有 LD 蒸馏后的学生网络出现了漏检, 其余算法均能正常检测该类型异物, 在该场景下, 本文方法对异物的定位精度更高。图 6(b) 给出了多个微小异物的检测场景, 由于异物过于微小, 图中网络均出现了漏检。其中, 未蒸馏的网络完全没有检测出异物, 而 CKD 和 LD 的学生网络分别漏检了 2 个和 3 个异物, 而本文方法的学生网络仅漏检了 1 个异物, 在置信度方面甚至超过了教师网络, 验证了其有效性。





(a) 大、中体积异物的检测效果

(b) 小体积异物的检测效果

图 6 检测结果可视化

Fig. 6 Detection results visualization

### 3.3 消融实验

为评估本文蒸馏算法中各个组成部分对蒸馏效果的影响, 本节设计了消融实验进行验证。

首先将本文的改进划分为 3 部分, 分别是基于语义特征的上采样方法 (记为语义上采样)、在蒸馏前插入注意力机制 (记为注意力)、在学生网络训练图像中加入噪声 (记为随机噪声), 随后将它们单独作用于蒸馏网络进行测试, 最终蒸馏后学生网络的结果如表 4 所示。

表 4 不同模块消融实验结果

模块	AP <sub>50</sub>	AP <sub>small</sub>	AR	AR <sub>small</sub>
均不作用	52.2	18.2	26.7	26.1
语义上采样	52.9	19.0	27.4	27.1
注意力	52.7	18.9	26.8	26.1
随机噪声	52.3	18.4	27.5	26.8
语义上采样+注意力	53.4	19.3	27.5	27.1
随机噪声+注意力	52.9	18.9	27.2	26.7
随机噪声+语义上采样	53.1	19.1	27.7	27.0
同时作用	<b>53.6</b>	<b>19.6</b>	<b>27.8</b>	<b>27.2</b>

当添加随机噪声单独作用时, 网络的精度指标没有明显变化, 但是召回率指标 AR 和 AR<sub>small</sub> 分别提升了 0.8 和 0.7 百分点; 当基于语义特征的上采样方法单独作用时, 所有指标均有明显提升; 当注意力机制单独作用时, 在精度指标上有一定提升; 上述结果验证了每个模块单独作用的有效性。接着, 本节对每个模块进行了组合验证, 除了“随机噪声+注意力”的组合相较于单独使用“随机噪声”的 AR 和 AR<sub>small</sub> 分别下降了 0.3 和 0.1 百分点, 其余的组合使用均提升了网络性能。最后, 将这 3 个模块同时应用于网络, 可以使其性能达到最佳水平。

### 3.4 天池酒液数据集的拓展实验

为了证明所提方法在相似场景下的有效性, 本文在公共的天池酒液数据集上设计了拓展实验, 包括与蒸馏网络 LD、PKD 和 CKD 以及主流检测算法对比, 结果如表 5 所示, 本文方法在各项指标中均为最优。表 6 给出了与主流检测算法的对比结果, 蒸馏后的学生网络的精度和召回率仍然领先大多数算法, 证明了其在相似场景下仍然有最佳的检测效果和蒸馏效果。

表 5 天池数据集蒸馏网络实验结果

检测模型	AP <sub>50</sub>	AP	AP <sub>small</sub>	AR	AR <sub>small</sub>
不蒸馏	75.4	40.7	31.6	51.6	46.3
PKD <sup>[26]</sup>	77.6	39.8	30.1	52.1	48.3
LD <sup>[25]</sup>	75.1	40.5	32.1	51.2	46.3
CKD <sup>[15]</sup>	78.2	43.4	34.4	54.2	49.3
本文方法	<b>80.9</b>	<b>44.6</b>	<b>36.4</b>	<b>55.9</b>	<b>52.1</b>

表 6 主流目标检测算法在天池数据集的实验结果

方法	AP <sub>50</sub>	AP	AP <sub>small</sub>	AR	AR <sub>small</sub>
Faster-RCNN <sup>[12]</sup>	38.2	19.7	5.2	23.8	5.1
ACMA <sup>[13]</sup>	77.0	44.4	34.7	56.1	53.0
ATSS <sup>[28]</sup>	68.1	37.0	26.7	47.7	40.2
VNet <sup>[27]</sup>	78.7	41.9	34.4	50.5	43.4
TOOD <sup>[29]</sup>	80.2	46.6	36.9	59.4	57.1
GFL <sup>[22]</sup>	77.9	45.0	33.2	58.6	56.5
SQR <sup>[31]</sup>	79.4	45.9	35.4	52.6	47.1
本文教师网络	<b>84.2</b>	<b>47.9</b>	<b>38.5</b>	<b>59.9</b>	<b>57.6</b>
本文学生网络	80.9	44.6	36.4	55.9	52.1

## 4 结束语

针对医药生产中的药液检测,本文提出了一种基于知识蒸馏的深度学习异物检测方法,利用改进后的交叉蒸馏模型,让精度高的大网络指导速度快的小网络,得到兼顾速度和精度的网络,有效提高了异物检测任务的效率和准确度。本文提出的基于语义特征的上采样和随机添加噪声的方法是根据药液异物检测场景和异物特性设计的,能够有效提升微小异物检测的精度以及抗干扰性能。本文通过实验证明了所提出算法在药液异物检测以及相似场景下的有效性,所得的学生网络在精度方面能够与大网络媲美,同时网络的 FPS 也能够达到 65。然而在实验中发现由于单帧定位任务缺乏序列帧特有的异物轨迹信息,导致即使是性能最好的教师网络也仍然会出现漏检和误检的问题。因此,如何利用连续帧的异物轨迹信息来滤除误检和修正漏检是下一步的研究内容。本论文代码与数据集开源,开源地址为:<https://github.com/YouWLi/DT>。

## 参考文献:

- [1] 国家药典委员会. 中华人民共和国药典[S]. 2020 年版第三部. 北京: 中国医药科技出版社, 2020: 481.  
Chinese Pharmacopoeia Commission. Pharmacopoeia of People's Republic of China[S]. Part III, 2020 Edition. Beijing: China medical science press, 2020: 481.
- [2] PEREZ M, MAIGUY-FOINARD A, BARTHÉLÉMY C, et al. Particulate matter in injectable drugs: evaluation of risks to patients[J]. *Pharmaceutical technology in hospital pharmacy*, 2016, 1(2): 91–103.
- [3] 徐帆, 杨伶俐, 毛盼盼, 等. 大输液包装材质与注射用药物相容性研究进展与思考[J]. *中国药业*, 2014, 23(20): 91–94.  
XU Fan, YANG Lingli, MAO Panpan, et al. Research advance and investigation on compatibility of large volume parenteral solution package material with injectable drugs[J]. *China pharmaceuticals*, 2014, 23(20): 91–94.
- [4] 冯晗, 姜勇. 使用改进 YOLOv5 的变电站绝缘子串检测方法[J]. *智能系统学报*, 2023, 18(2): 325–332.  
FENG Han, JIANG Yong. A substation insulator string detection method based on an improved YOLOv5[J]. *CAAI transactions on intelligent systems*, 2023, 18(2): 325–332.
- [5] 刘嘉轩, 胡非易, 张辉, 等. 上下文空间与实例信息的皮肤镜图像自监督分类[J]. *智能系统学报*, 2023, 18(4): 783–792.  
LIU Jiaxuan, HU Feiyi, ZHANG Hui, et al. Dermoscopic images classification based on context and instance-level feature of self-supervised learning[J]. *CAAI transactions on intelligent systems*, 2023, 18(4): 783–792.
- [6] 郑广智, 彭添强, 肖计春, 等. 基于语义信息增强的化纤丝线网络度检测方法[J]. *自动化学报*, 2024, 50(10): 1963–1976.  
ZHENG Guangzhi, PENG Tianqiang, XIAO Jichun, et al. A detection method for the interlacing degree of filament yarn based on semantic information enhancement[J]. *Acta automatica sinica*, 2024, 50(10): 1963–1976.
- [7] 康守强, 邢颖怡, 王玉静, 等. 基于无监督深度模型迁移的滚动轴承寿命预测方法[J]. *自动化学报*, 2023, 49(12): 2627–2638.  
KANG Shouqiang, XING Yingyi, WANG Yujing, et al. Rolling bearing life prediction based on unsupervised deep model transfer[J]. *Acta automatica sinica*, 2023, 49(12): 2627–2638.
- [8] 闫涵, 卢伟, 吴玉虎. 基于轻量化卷积神经网络的金属断口图像识别[J]. *控制与决策*, 2024, 39(9): 2858–2866.  
YAN Han, LU Wei, WU Yuhu. Metal fracture recognition based on lightweight convolutional neural network[J]. *Control and decision*, 2024, 39(9): 2858–2866.
- [9] WANG Sirui, ZHUO Qing, XIA Ji'an. Detection of glass chips in liquid injection based on computer vision[C]//2015 International Conference on Computational Intelligence and Communication Networks. Jabalpur: IEEE, 2015: 329–331.
- [10] DAI Jinzhi. Retracted: drug residue detection system for infusion based on machine vision[C]//2016 International Conference on Smart City and Systems Engineering. Hu'nan: IEEE, 2016: 571–574.
- [11] ZHANG Hui, LI Xuanlun, ZHONG Hang, et al. Automated machine vision system for liquid particle inspection of pharmaceutical injection[J]. *IEEE transactions on instrumentation and measurement*, 2018, 67(6): 1278–1297.
- [12] ZHANG Hui, ZHAO Miao, LIU Li, et al. Deep multimodel cascade method based on CNN and random forest for pharmaceutical particle detection[J]. *IEEE transactions on instrumentation and measurement*, 2020, 69(9): 7028–7042.
- [13] YI Junfei, ZHANG Hui, MAO Jianxu, et al. Pharmaceutical foreign particle detection: an efficient method based on adaptive convolution and multiscale attention[J]. *IEEE transactions on emerging topics in computational intelligence*, 2022, 6(6): 1302–1313.
- [14] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge[J]. *International journal of computer vision*, 2010, 88(2): 303–338.
- [15] WANG Jiabao, CHEN Yuming, ZHENG Zhaohui, et al. CrossKD: cross-head knowledge distillation for object detection[EB/OL]. (2023–06–20)[2023–11–17]. <https://arxiv>.

- [org/abs/2306.11369](https://arxiv.org/abs/2306.11369).
- [16] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7132–7141.
- [17] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018: 3–19.
- [18] PARK J, WOO S, LEE J Y, et al. BAM: bottleneck attention module[EB/OL]. (2018–07–17)[2023–11–17]. <https://arxiv.org/abs/1807.06514>.
- [19] YANG Lingxiao, ZHANG Ruyuan, LI Lida, et al. Simam: a simple, parameter-free attention module for convolutional neural networks[C]//Proceedings of the 38th International Conference on Machine Learning. Online: PMLR, 2021: 11863–11874.
- [20] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770–778.
- [21] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 936–944.
- [22] LI Xiang, WANG Wenhai, WU Lijun, et al. Generalized focal loss: learning qualified and distributed bounding boxes for dense object detection[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2020, 33: 21002–21012.
- [23] WANG Jiaqi, CHEN Kai, XU Rui, et al. CARAFE: content-aware reassembly of features[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 3007–3016.
- [24] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 658–666.
- [25] ZHENG Zhaohui, YE Rongguang, WANG Ping, et al. Localization distillation for dense object detection[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 9397–9406.
- [26] SUN Siqi, CHENG Yu, GAN Zhe, et al. Patient knowledge distillation for BERT model compression[EB/OL]. (2019–08–25)[2023–11–17]. <https://arxiv.org/abs/1908.09355>.
- [27] ZHANG Haoyang, WANG Ying, DAYOUB F, et al. VarifocalNet: an IoU-aware dense object detector[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 8510–8519.
- [28] ZHANG Shifeng, CHI Cheng, YAO Yongqiang, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 9756–9765.
- [29] FENG Chengjian, ZHONG Yujie, GAO Yu, et al. TOOD: task-aligned one-stage object detection[C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 3490–3499.
- [30] CHEN Qiang, WANG Yingming, YANG Tong, et al. You only look one-level feature[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 13034–13043.
- [31] CHEN Fangyi, ZHANG Han, HU Kai, et al. Enhanced training of query-based object detection via selective query recollection[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 23756–23765.

#### 作者简介:



刘优武, 硕士研究生, 主要研究方向为深度学习、医药异物检测。  
E-mail: [liuyouwu1999@163.com](mailto:liuyouwu1999@163.com)。



张辉, 教授, 博士生导师, 主要研究方向为计算机视觉。主持科技创新 2030—新一代人工智能重大项目、国家自然科学基金共融机器人重大研究计划重点项目、国家重点研发计划子课题、国家科技支撑计划项目子课题等 20 余项, 获省部级科学技术奖励一等奖 8 项, 获 2022 年湖南省第十三届教学成果特等奖等, 获发明专利授权 38 项, 发表学术论文 50 余篇。E-mail: [zhanghuihy@126.com](mailto:zhanghuihy@126.com)。



孔森林, 硕士研究生, 主要研究方向为无监督学习和工业图像缺陷检测。E-mail: [986735244@qq.com](mailto:986735244@qq.com)。

[ 责任编辑: 丁钰 ]