OEE | 光电工程 Opto-Electronic Engineering

Article

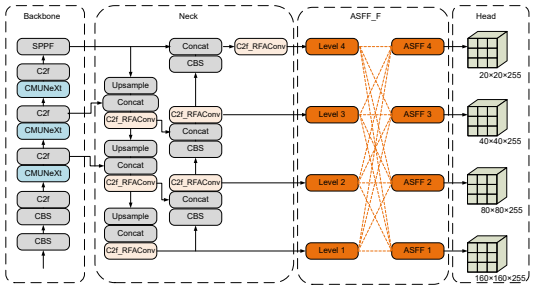
2026 年 1 月, 第 53 卷, 第 1 期

YOLO-SLB：面向结构光成像的电芯外观缺陷检测模型

张　上[1], 熊中越[1*], 张　明[2]

[1] 三峡大学计算机与信息学院，湖北 宜昌 443002；

[2] 三峡大学电气与新能源学院，湖北 宜昌 443002



摘要：针对锂电池电芯外观缺陷检测中目标尺寸小、缺陷类型复杂多样，以及实际生产环境下数据采集困难等问题，提出一种基于结构光成像与改进 YOLOv8 的检测方法，提升检测精度与鲁棒性。首先，在骨干网络中引入 CMUNeXt Block，通过大核卷积增强空间信息交互，提升特征提取能力；其次，在颈部网络中设计 C2f_RFAConv 模块，引入动态参数生成机制，缓解参数共享带来的表达瓶颈，并增强特征传递中的信息保留；最后，改进检测头结构，构建 ASFF_F 模块，融合自适应空间特征融合策略，重构特征金字塔路径，实现新增微小目标检测分支，强化多尺度细节感知能力。实验结果表明，所提模型平均检测精度达 97.98%，准确率和召回率分别为 97.05% 和 92.67%，相较基准模型 YOLOv8 分别提升了 4.29%、1.97% 和 6.46%。在 NEU-DET 公开数据集上的扩展实验亦验证了其良好的泛化性能，对比基准模型，平均检测精度提高了 0.65%，表明该方法具备较高的工程应用价值。

关键词：计算机视觉；缺陷检测；特征提取；YOLO；结构光成像

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250213 | CSTR: 32245.14.oee.2026.250213 | 中图分类号：TP394.1;TH691.9 | 文献标识码：A

张上，熊中越，张明. YOLO-SLB：面向结构光成像的电芯外观缺陷检测模型 [J]. 光电工程，2026，**53**(1): 250213

Zhang S, Xiong Z Y, Zhang M. YOLO-SLB: appearance defect detection model of lithium battery cell oriented to structured light imaging[J]. *Opto-Electron Eng*, 2026, **53**(1): 250213

1　引　言

随着新能源产业的快速发展，锂电池作为核心储能元件，其生产质量直接关系到终端产品的安全性与可靠性。在锂电池的工业化生产过程中，电芯外观缺陷检测是保障产品质量的关键环节。由于电芯表面可能存在划痕、凹陷、褶皱、污染等多种微小且形态复杂的缺陷，传统的人工目检或二维视觉检测方法已难以满足实际生产场景的自动化检测需求。

结构光技术[1]作为一种先进的三维测量方法，在工业应用中占据着重要的位置。通过向目标物体投射特定模式的光线，并分析反射回来的光线变形情况，结构光系统能够快速获取物体表面的三维信息。这种技术因其高精度、非接触性和实时性而广受青睐，成为了实现工业自动化[2]、质量控制[3]、逆向工程[4]等任务的关键工具。近年来，随着深度学习技术的迅猛发展，结构光系统与深度神经网络的融合为复杂表面缺陷的智能识别提供了新的解决方案。例如，深度卷积神经网络 (CNNs)[5]在处理结构光获取的三维点云或形变图像方面展现出强大的特征提取能力，已被成功应用于零部件尺寸测量[6]、表面缺陷检测[7]以及文化遗产的三维建模与状态评估[8]等领域。

收稿日期：2025-07-20；修回日期：2025-09-11；录用日期：2025-09-15；网络出版日期：2026-01-25

基金项目：湖北省数字经济试点示范建设专项 (2312-420625-04-02-996363)；湖北省国家级大学生创新创业训练计划 (S202311075047)

*通信作者：熊中越，eng_xiong@163.com

版权所有©2026 中国科学院光电技术研究所

目标检测作为计算机视觉领域的重要研究方向，近年来在算法架构和性能表现上取得了突破性进展。从早期基于区域提议的 AlexNet[9] 和 R-CNN[10] 及其后续版本 Fast R-CNN[11]、Faster R-CNN[12]，到后来的 YOLO[13] 系列与 SSD[14] 等单阶段检测器，检测速度与精度得以显著提升。与此同时，Transformer 架构的引入，如 DETR[15]、ViT[16] 和 RT-DETR[17]，为目标检测注入了新的建模范式。这些模型不仅刷新了多个权威数据集上的性能指标，也在结构光成像等复杂视觉任务中展现出一定的泛化能力和实用价值。

然而，尽管上述技术在通用工业检测中成效显著，针对锂电池电芯外观缺陷的检测仍面临诸多挑战，比如缺陷尺寸微小、类型多样，对算法的精度与鲁棒性要求高；实际生产中光照变化、反光干扰严重，且数据采集困难，高质量标注样本稀缺等。在现有研究中，Wang 等[18] 提出基于改进 YOLOv5 的智能检测系统，实现了锂电池外观缺陷的高效识别；Liu 等[19] 则利用 3D 点云技术对表面缺陷进行几何量化，弥补了 2D 视觉对高度信息不敏感的不足。尽管这些方法取得了一定成效，但在微小缺陷检测精度、多尺度特征融合效率及模型轻量化方面仍有提升空间。

为此，本文提出一种面向结构光成像的新型目标检测算法——YOLO-SLB，针对锂电池电芯外观微小缺陷检测任务进行优化设计：

1) 将骨干网络的卷积模块替换为 CMUNeXt Block，在保持网络深度的同时增强空间位置信息交互能力，从而提升特征提取效率；

2) 设计 C2f_RFAConv 模块优化颈部网络结构，通过动态参数生成机制缓解卷积核参数共享，有效降低计算冗余并抑制特征传递过程中的信息衰减；

3) 设计改进四头自适应空间特征融合模块 (ASFF_F) 检测头，首先采用自适应空间特征融合策略增强多尺度特征交互能力，其次新增微小目标检测分支并通过重构特征金字塔路径，充分挖掘浅层特征中的细节信息，显著提升对微小缺陷的检测精度。

2 YOLO-SLB 网络模型

YOLOv8 作为一种优秀的单阶段目标检测算法，在检查精度和速度之间取得了较好的平衡。然而，在检测复杂情况和小目标时仍然存在精度有待提升以及误检漏检等情况[20]。

本文为降低锂电池电芯外观缺陷检测在实际应用场景中的误检率及漏检率，提出新的 YOLO-SLB 目标检测模型，模型结构如图 1 所示。针对锂电池电芯外观缺陷检测任务中普遍存在的缺陷尺度微小、分布密集且特征信息复杂等特点，本文从三个层面进行针对性优化。首先，针对微小缺陷易被忽略的问题，传统骨干网络在深层特征中难以保留足够的空间细节。为此，本文在骨干网络中引入 CMUNeXt Block，通过并行多分支结构实现不同感受野下的特征捕获，增强远距离空间位置信息的交互能力，有效提升对微弱、细小缺陷的特征响应，从而增强模型对关键细节的敏感性。针对缺陷密集、边界模糊导致定位不准的问题，现有颈部网络在融合浅层空间细节与深层语义信息时易出现信息损耗，尤其在大核卷积中，参数共享机制易引发计算冗余与特征混淆。为此，本文提出 C2f_RFAConv 模块，引入动态参数生成机制，实现对不同空间位置特征的自适应加权，增强关键区域的特征表达，有效促进高低层特征的高效融合，缓解信息衰减，提升对密集缺陷的分辨能力。最后，针对多尺度特征融合不充分、微小缺陷检测能力弱的问题，传统特征金字塔结构对底层细节利用不足，难以实现跨尺度的有效协同。本文在检测头部改进的 ASFF_F 结构，采用自适应空间特征融合策略，动态调整各尺度特征的贡献权重，增强多尺度间的交互能力；同时，重构特征传递路径，新增微小目标检测分支，直接从浅层特征中提取关键细节，强化对微小缺陷的感知能力，显著提升检测的精度与鲁棒性。

2.1 优化骨干网络

CMUNeXt 是 Tang 等[21] 于 2024 年提出的一种轻量级全卷积医学图像分割网络，其核心采用深度可分离卷积，将标准卷积分解为深度卷积和点卷积。深度卷积在每个通道独立提取空间特征，点卷积通过 1×1 卷积核实现通道间信息融合，显著降低参数量与计算开销，同时保持良好表达能力。如图 2 所示，CMUNeXt Block 采用大核深度卷积捕获全局空间信息，并引入倒置瓶颈结构：通过第一个点卷积将通道数扩展至输入的 4 倍，增强特征融合能力，再通过第二个点卷积压缩回原通道数，形成“先升维、后降维”的非线性变换，进一步提升特征表示能力。

此外，在每一步卷积操作后均引入 Hendrycks 和 Gimpel 提出的一种高斯误差线性单元 (Gaussian error linear unit, GELU) 激活函数[22]，具体形式如式 (1) 所示。

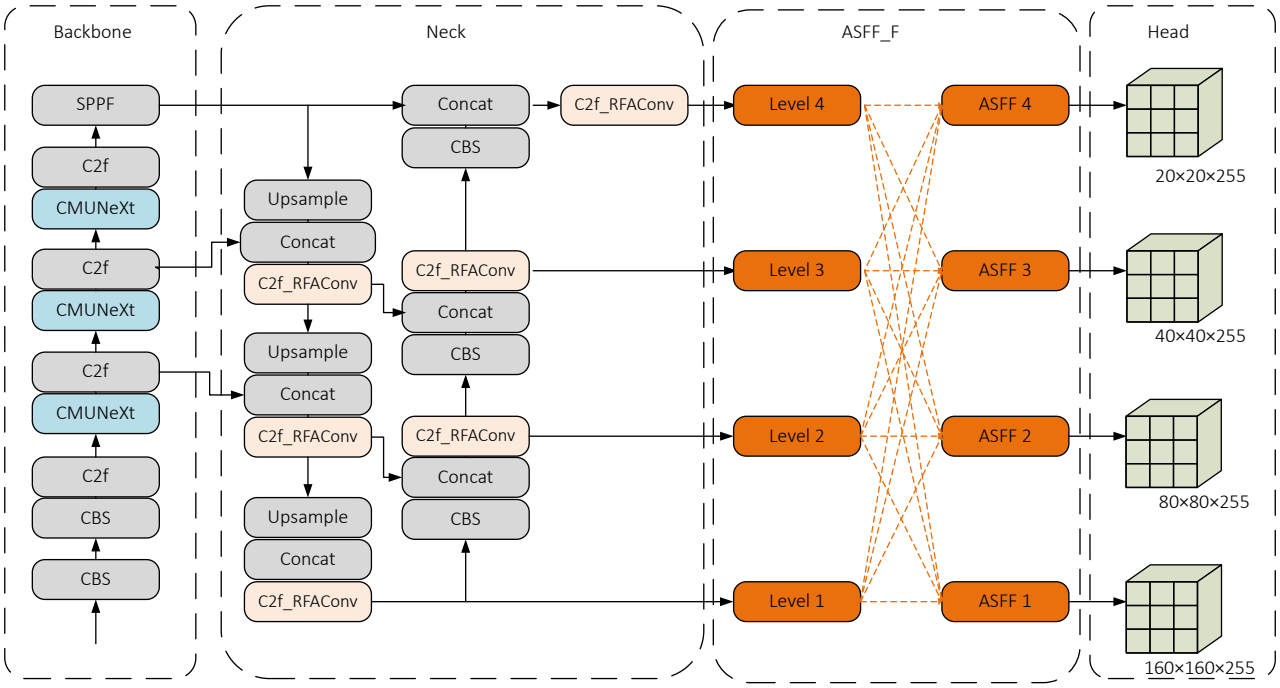


图 1 YOLO-SLB 网络模型

Fig. 1 The network structure of YOLO-SLB

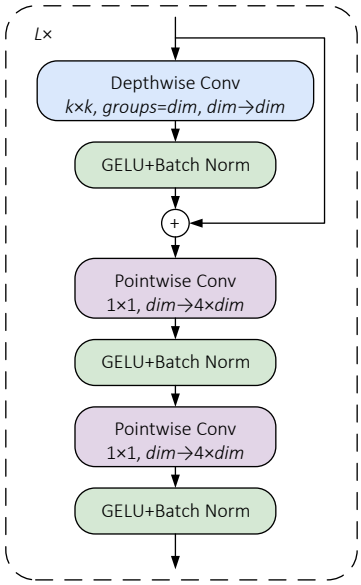


图 2 CMUNeXt 模块

Fig. 2 CMUNeXt Block

$$GELU(x) = x \cdot \Phi(x)\,, \tag{1}$$

式中：$\Phi(x)$是标准正态累积分布函数；x是神经网络中某个神经元的线性输入。GELU 引入了一种门控机制，该机制在 [0,1] 内缩放输入，然后进行批处理归一化以稳定分布。较小的输入可能被抑制，而较大的输入被激活，有效缓解了“死亡神经元”问题，并改善了梯度在反向传播中的流动性，减轻梯度消失现象。近似形式为

$$GELU(x) \approx 0.5x\left(1+\tanh\left(\sqrt{\frac{2}{\pi}}\left(x+0.044715x^3\right)\right)\right). \tag{2}$$

整个 CMUNeXt Block 的数学表达式为

$$f_l' = BN\left(GELU_1\left(DepthwiseConv\left(f_{l-1}\right)\right)\right) + f_{l-1}\,, \tag{3}$$

$$f_l'' = BN\left(GELU\left(PointwiseConv\left(f_l'\right)\right)\right), \tag{4}$$

$$f_l = BN\left(GELU\left(PointwiseConv\left(f_l''\right)\right)\right), \tag{5}$$

式中：f_l表示 CMUNeXt Block 中第 l 层的输出特征图；f_l'表示深度卷积后第l层的输出特征图；f_l''表示逐点卷积后第 l 层的输出特征图；BN 表示 Batch Normalization 层。由于 CMUNeXt 块在输入和输出之间保持一致的分辨率和通道大小，使用标准卷积块来扩展通道维度是原来的 2 倍。

2.2 优化特征提取

YOLOv8n 在颈部网络的 C2f 模块中采用标准卷积提取特征。标准卷积在滑动窗口中对相同位置使用固定参数，难以捕捉输入特征在空间位置上的差异。主流方法通过引入空间注意力机制提升模型性能。

图 3 所示，卷积核 K 仅代表一个参数值。如果把

$A_i \times K$ (注意力权重与卷积核的乘积) 作为一个新的卷积核参数，与 1×1 卷积核相乘，得到的加权卷积核可视为新的非共享参数，从而缓解参数共享问题。通过滑动窗口、注意力权重和 1×1 卷积核的联合作用，每个感受野可获得不同的卷积参数。

然而空间注意力仅作用于空间维度，未能解决大核卷积中的参数共享问题。当其插入至 3×3 卷积前 (见图 4)，虽然注意力图与输入特征同尺寸能够按位置加权，但由于感受野之间存在重叠区域，注意力权重仍被共享，未实现真正意义上的参数去共享。

RFAConv[23] 引入感受野空间注意力机制，通过卷积重排空间特征以缓解权重共享导致的位置不敏感问题，从而提升局部特征建模能力，并融合注意力权重动态整合卷积步长，避免特征重复提取。由于其空间特征由卷积核动态生成，RFA 实质上是卷积操作中一种依赖核大小的固定组合结构。RFAConv 将该机制与标准卷积融合，其结构如图 5 所示，其中 C、H、W 分别表示通道数、高度与宽度。

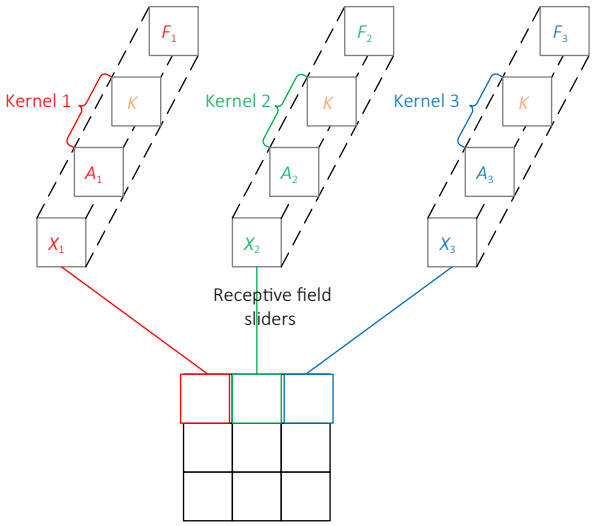


图 3 1×1 空间注意力卷积示图

Fig. 3 Schematic diagram of 1×1 spatial attention convolution

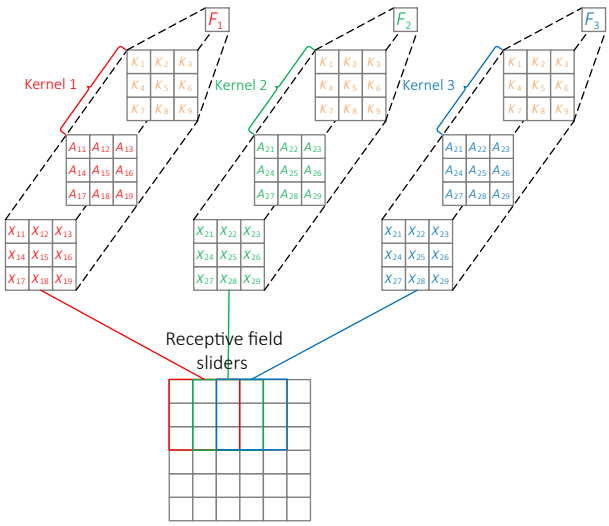


图 4 3×3 空间注意力卷积示图

Fig. 4 Schematic diagram of 3×3 spatial attention convolution

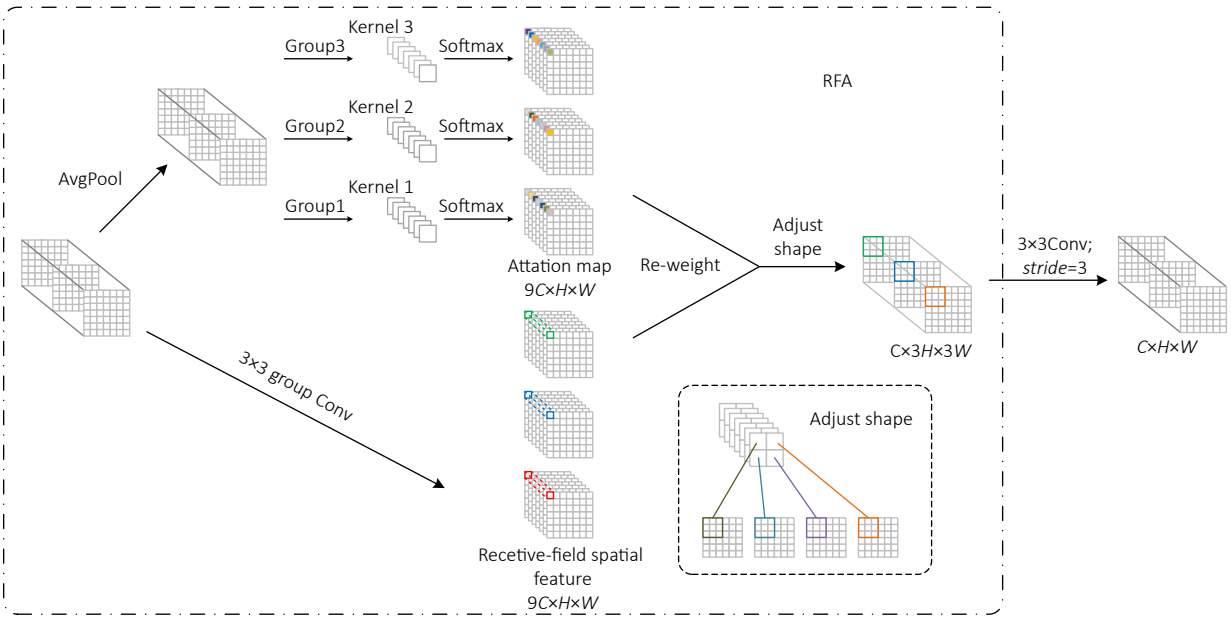


图 5 RFAConv

Fig. 5 RFAConv

RFAConv 通过组卷积 (group convolution) 高效提取感受野空间特征。以 3×3 卷积核为例，输入特征被划分为多个感受野滑块，每个 3×3 区域对应一个滑块。为提升性能，RFAConv 在每个滑块上交互特征以生成注意力图，但逐滑块交互会增加计算开销。为降低计算量与参数数量，RFAConv 先通过$AvgPool$聚合每个滑块的全局信息，再利用 1×1 组卷积实现信息交互，最终通过$Softmax$加权，突出滑块内部关键特征。RFA 的计算过程可表达为

$$F = S \times R = A_{\mathrm{rf}} \times F_{\mathrm{rf}}\,, \tag{6}$$

$$S = Softmax\left(g^{i\times i}(AvgPool(X))\right), \tag{7}$$

$$R = ReLU\left(Norm\left(g^{k\times k}(X)\right)\right), \tag{8}$$

式中：$g^{i\times i}$表示大小$i\times i$的组卷积；k 表示卷积核的大小；$Norm$ 表示归一化；X 表示输入特征图；F 是通过将注意力图 A_{rf} 与变换后的感受野空间特征 F_{rf} 相乘得到的。

RFA 可为每个感受野特征生成专属注意力图，缓解了标准卷积因参数共享而对位置变化不敏感的问题。为实现感受野特征向空间域的准确映射，RFAConv 在注意力加权后引入 Adjust Shape 操作，将通道维度堆叠的滑块重排至其在输入特征图中对应的空间位置，构建结构化的空间特征图。如图 5 所示，不同颜色的滑块被映射至各自区域，实现非重叠感受野的空间分离，有效避免注意力图在滑块间共享，从而保留每个滑块的独立性与判别性。在调整形状后，特征的高度和宽度是 k 倍，需要进行步长为 s=k 的$k\times k$卷积运算来提取特征信息。

考虑本文面向锂电池外观缺陷检测具有高度复杂性与多样性，为增强模型对图像特征的表达能力与多尺度信息的融合效率，本文将 C2f 模块中的标准卷积替换为具备空间感受野注意力机制的 RFAConv，构建改进结构 C2f_RFAConv，以实现更精细的空间特征建模与跨尺度语义增强。

2.3 设计改进检测头

锂电池电芯外观缺陷具有尺度微小、缺陷密集以及特征信息复杂等特点。YOLOv8 颈部的特征金字塔存在尺度不一致问题，直接融合易导致信息丢失。为此，引入自适应空间特征融合模块 (adaptively spatial feature fusion, ASFF)[24] 以增强模型的多尺度特征融合能力。ASFF 结构如图 6 所示，通过自适应学习各层级特征的融合权重，缓解尺度不匹配，提升多尺度特征融合的一致性与模型鲁棒性。

ASFF 通过自适应学习各尺度空间特征的权重，融合多层特征并抑制冗余信息。针对颈部输出的多尺度特征图分辨率不同，采用 1×1 卷积调整通道，上采样用最近邻插值，2 倍下采样用 3×3 步长为 2 的卷积，4 倍下采样先最大池化 (步长为 2)，再接同规格卷积。对齐后进行加权求和，实现跨尺度自适应融合。

$$\boldsymbol{y}_{ij}^{l} = \alpha_{ij}^{l} x_{ij}^{1\to l} + \beta_{ij}^{l} x_{ij}^{2\to l} + \gamma_{ij}^{l} x_{ij}^{3\to l}\,, \tag{9}$$

式中：$\boldsymbol{y}_{ij}^{l}$表示输出特征图在空间位置(i,j)处融合后的特征向量；$x_{ij}^{1\to l}$、$x_{ij}^{2\to l}$、$x_{ij}^{3\to l}$分别表示来自 l 层级不同尺度的 3 个输出特征图；α_{ij}^{l}、β_{ij}^{l}、γ_{ij}^{l}分别表示不同尺度特征在 l 层的空间权重。这些权重由网络自适应地学习，并在所有通道共享，且满足如下条件：

$$\alpha_{ij}^{l} + \beta_{ij}^{l} + \gamma_{ij}^{l} = 1\,, \tag{10}$$

$$\alpha_{ij}^{l},\quad \beta_{ij}^{l},\quad \gamma_{ij}^{l} \in [0,1]\,. \tag{11}$$

权重系数可以通过归一化指数函数来计算，计算公式如表 1 所示，其中τ为调节模型的响应强度。

锂电池电芯外观缺陷中存在大量微小目标。虽然 ASFF 通过动态学习不同层特征的权重，增强了模型对多尺度目标的感知能力，但在融合过程中，不同尺度的特征图需先调整至统一尺寸，通常采用插值等方法，这可能导致微小目标的关键信息部分丢失。

对此，本文设计改进的 ASFF_F，如图 7 所示。ASFF_F 通过增设融合路径，充分挖掘浅层细节特征，

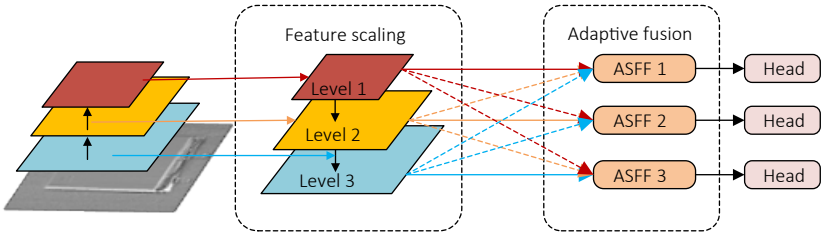


图 6 ASFF 结构

Fig. 6 Structure of ASFF

并基于自适应权重机制动态调节各尺度及路径的贡献，提升微小目标检测精度与特征融合有效性。

表 1 ASFF 权重系数计算公式

Table 1 ASFF weight coefficient calculation formula

ASFF weight	Calculation formula
α_{ij}^{l}	$\frac{e^{\tau\alpha_{ij}^{l}}}{e^{\tau\alpha_{ij}^{l}}+e^{\tau\beta_{ij}^{l}}+e^{\tau\gamma_{ij}^{l}}}$
β_{ij}^{l}	$\frac{e^{\tau\beta_{ij}^{l}}}{e^{\tau\alpha_{ij}^{l}}+e^{\tau\beta_{ij}^{l}}+e^{\tau\gamma_{ij}^{l}}}$
γ_{ij}^{l}	$\frac{e^{\tau\gamma_{ij}^{l}}}{e^{\tau\alpha_{ij}^{l}}+e^{\tau\beta_{ij}^{l}}+e^{\tau\gamma_{ij}^{l}}}$

ASFF_F 在原有的三个特征路径基础上，增加了一个额外的特征路径，这个额外的路径可以从网络的更浅层提取特征图$x_{ij}^{4\to l}$，使得特征尺度更加丰富。不仅增加了模型的表达能力，还提供了更多的信息来源。ASFF_F 使用1×1分别对$x_{ij}^{1\to l}$、$x_{ij}^{2\to l}$、$x_{ij}^{3\to l}$、$x_{ij}^{4\to l}$行卷积操作，生成对应的权重图α_{ij}^{l}、β_{ij}^{l}、γ_{ij}^{l}、θ_{ij}^{l}并进行归一化处理后加权融合，确保权重之和 1，使用归一化后的权重对齐后的特征图进行加权融合，生成最终的融合特征图$\boldsymbol{y}_{ij}^{l}$，公式如下：

$$\boldsymbol{y}_{ij}^{l}=\alpha_{ij}^{l}x_{ij}^{1\to l}+\beta_{ij}^{l}x_{ij}^{2\to l}+\gamma_{ij}^{l}x_{ij}^{3\to l}+\theta_{ij}^{l}x_{ij}^{4\to l}. \tag{12}$$

跨通道空间权重计算需满足表 2 所示公式。

ASFF_F 通过引入四个不同尺度的特征图，并通过自适应权重学习模块动态地融合这些特征图，能够在保持计算效率的同时，显著提升模型对多尺度目标的检测能力。特别是对于微小目标检测，由于$\theta_{ij}^{l}x_{ij}^{4\to l}$特征图包含更多的细节信息，ASFF_F 能够更好地捕捉这些细节，从而提高检测精度。

3 实验与分析

3.1 构建数据集

在锂电池电芯质量检测工作中，图像采集的准确性与清晰度直接影响后续缺陷识别与判断的可靠性。本研究所用图像数据采集自惠州市典名新能源科技有限公司锂电池电芯检修部门，旨在通过科学的视觉采集方案，实现对锂电池电芯表面缺陷的高精度检测与真实还原。

锂电池电芯表面原始图像的获取面临多种挑战。锂电池电芯表面材质为铝塑膜，其颜色、反光率及纹理存在明显差异，同时环境光源的方向性与强度波动也容易引入阴影、眩光等图像伪差。且不同缺陷在固定的光源下，无法同时获取，这些因素会干扰图像的灰度均衡性与边缘信息，影响缺陷的判别。如图 8 所示，锂电池电芯在自然光条件下所获取的原始图像存在明显的反光区

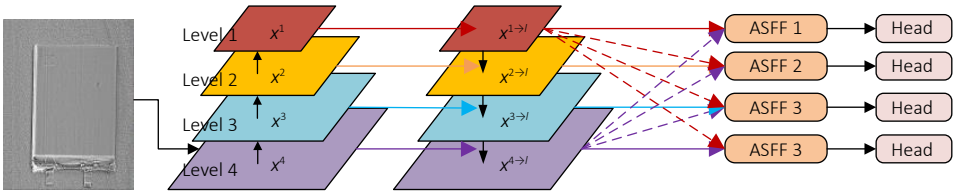


图 7 ASFF_F 结构

Fig. 7 Structure of ASFF_F

表 2 ASFF_F 权重系数计算公式

Table 2 ASFF_F weight coefficient calculation formula

ASFF_F weight	Calculation formula
α_{ij}^{l}	$\frac{e^{\tau\alpha_{ij}^{l}}}{e^{\tau\alpha_{ij}^{l}}+e^{\tau\beta_{ji}^{l}}+e^{\tau\gamma_{jj}^{l}}+e^{\tau\theta_{ij}^{l}}}$
β_{ij}^{l}	$\frac{e^{\tau\beta_{ij}^{l}}}{e^{\tau\alpha_{j}^{l}}+e^{\tau\beta_{j}^{l}}+e^{\tau\gamma_{j}^{l}}+e^{\tau\theta_{j}^{l}}}$
γ_{ij}^{l}	$\frac{e^{\tau\gamma_{j}^{l}}}{e^{\tau\alpha_{j}^{l}}+e^{\tau\beta_{j}^{l}}+e^{\tau\gamma_{j}^{l}}+e^{\tau\theta_{ij}^{l}}}$
θ_{ij}^{l}	$\frac{e^{\tau\theta_{ij}^{l}}}{e^{\tau\alpha_{ij}^{l}}+e^{\tau\beta_{ij}^{l}}+e^{\tau\gamma_{ij}^{l}}+e^{\tau\theta_{ij}^{l}}}$



图 8 锂电池电芯原始图像

Fig. 8 Original image of lithium battery cell

域与暗影过渡，难以呈现表面细节。

为克服上述问题，本文采用线阵相机结合结构光照明的方式进行图像采集。在拍摄系统中，线阵相机以逐行扫描方式获取高分辨率图像，适用于移动中的锂电池电芯目标。而结构光系统通过以固定角度照射的条纹光源对锂电池电芯表面进行主动照明，能够有效抑制环境光变化带来的干扰，提升成像的一致性与稳定性。结构光成像为实际生产场景实现“明场+暗场”的复合照明效果，兼顾表面纹理细节与边缘特征的表达，进一步提升不同类型缺陷的可检测性。整体系统在动态工况下仍能输出高质量、高对比度的图像，为后续高精度缺陷检测提供了可靠的数据支撑。图 9 为系统搭建与拍摄过程示意图，可见锂电池电芯在输送线上匀速通过，条纹光源斜向照射其表面，线阵相机同步完成图像采集。

此外，为进一步增强系统的适应能力，在拍摄装置中对相机焦距、曝光时间及光源强度进行了精确标定，使其适配不同表面颜色与材质的锂电池电芯，确保在各种条件下均可获得对比度充足、细节丰富的图像。

本文最终选取 7 类出现频率次数较高的外观缺陷作为研究对象，包括坑点 (pit point)、坑块 (pit block)、异物 (foreign body)、划痕 (scratch)、褶皱 (wrinkle)、条痕 (stripe)、折叠 (fold)、腐蚀 (corrosion)。原始图像共采集 477 张，图像分辨率为 1520 pixel×2000 pixel，因原始数据难以满足神经网络训练需要，因此标注后使用图像变换方式来对原数据集进行数据增强，变换方式包括旋转、翻转与加噪。图像变化有助于提升检测性能与模型鲁棒性，使其达到工业检测需求。

增强后的数据集训练集达到 3147 张，共有 4063 个待检目标；预测集达到 574 张，共有 700 个待检目标。本文将该数据集命名为 SLBD，数据集的样本按照 7.5:2.5 的比例分为训练集和测试集，分类训练集由 2385 张缺陷图像组成，测试集由 795 张缺陷图像组成。图 10 展示了部分数据集图片。

3.2 训练环境及参数设置

本文对锂电池电芯外观缺陷检测的实验使用深度学习 PyTorch 框架，在 Ubuntu 24.04 操作系统上运行，硬件配置为：CPU 使用 12th Gen Intel (R) Core (TM) i5-12400f，GPU 使用 NVIDIA RTX A4000, 16 G 内存；软件配置为：Python 版本是 3.11，框架使用 Pytorch 2.4.1+CUDA 12.4。

训练轮次设置为 500 轮，patience 设置为 50，batch_size 设置为 16，workers 设置为 4。

3.3 评估指标

在计算机视觉中，评估算法性能的主要指标包括准确率 (precision，P)、召回率 (recall，R)，综合指标平均准确率 (mean average precision，P_{mAP})，以及每秒帧数 (frame per second，FPS)。具体公式如下：

$$P = \frac{TP}{TP+FP}, \tag{13}$$

$$R = \frac{TP}{TP+FN}, \tag{14}$$

$$P_{\mathrm{AP}} = \int_0^1 P(R)\mathrm{d}R, \tag{15}$$

$$P_{\mathrm{mAP}} = \frac{1}{c}\sum_{j=1}^{c} AP_j, \tag{16}$$

$$FPS = \frac{1000}{t_{\mathrm{p}}+t_{\mathrm{i}}+t_{\mathrm{o}}}. \tag{17}$$

式中：TP 表示预测对的正样本、FP 表示预测错的正样本、TN 预测对的负样本和 FN 表示预测错的负样本；c 为总类别数；t_{p}、t_{i} 和 t_{o} 分别表示预处理、推理和后处理阶段的平均耗时。

3.4 实验结果与分析

3.4.1 消融实验

在锂电池电芯外观缺陷检测场景中，缺陷通常表现

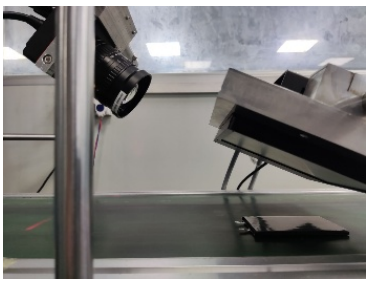


图 9 锂电池电芯结构光成像拍摄过程

Fig. 9 Structured light imaging process of lithium battery cell

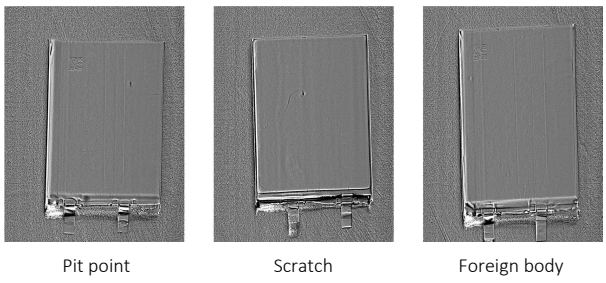


图 10 数据集图像展示

Fig. 10 Dataset images display

为目标尺寸小且类型多样复杂。尽管使用结构光成像来解决光照反射等问题，但复杂的背景噪声显著增加了目标检测任务的难度。为应对上述挑战，本文构建了基于模块化设计的 YOLO-SLB 模型，融合多种结构性改进模块，以提升其在复杂环境下的鲁棒性和精度。

为了验证每一步改动的有效性，对相关创新模块进行消融实验。在每组实验中，首先仅引入一个创新点。然后基于不同改进点进行组合实验。消融实验结果如表 3 所示，其中“√”表示引入该模块，空格则表示未引入。

锂电池电芯外观缺陷检测的核心在于准确识别缺陷，降低误检与漏检率。实际应用中，检测速度已基本满足线阵相机的实时拍摄需求，因而本文重点在于提升模型的平均检测精度 (mAP)。

如表 3 所示，仅引入 CMUNeXt (方案 A) 后，mAP 提升至 95.72%，较基准模型 (方案 Base) 的 93.69% 提高 2.03%。准确率小幅上升至 95.20%，召回率从 86.21% 显著提升至 90.87% (+4.66%)，表明其在增强检出能力方面效果突出。尽管计算量由 8.2 GFLOPs 增至 9.5 GFLOPs，帧率下降至 463 f/s，但仍具较强实时性，实现了精度与速度的良好权衡。

引入 C2f_RFAConv (方案 B) 后，mAP 达到 94.12%，准确率为 93.14%，召回率为 88.71%，分别较基准模型提升了 0.43%、2.93%、1.50%。虽整体优于基准模型，但在各项指标上不及方案 A，这说明该模块对特征增强有效但独立贡献有限，其优势在于后续组合实验中的互补性。

方案 C 仅引入 ASFF_F，mAP 提升至 94.34%，准确率和召回率分别为 94.16% 和 87.73%，提升了 1.08% 和 1.52%。该模块增强了多尺度特征融合能力，但计算量增至 15.4 GFLOPs，帧率降至 224 f/s，更适用于精度优先、资源充足场景。

在多模块组合中，方案 D (CMUNeXt+C2f_RFAConv) 表现最佳，mAP 达 96.20%，准确率为 96.63%，召回率为 92.25%，分别较基准模型提升了 2.51%、1.55%、6.04%。计算量为 9.7 GFLOPs，帧率为 339 f/s，兼具性能与实用性。方案 E (CMUNeXt+ASFF_F) mAP 为 95.89%，准确率为 96.23%，召回率为 91.78%，虽略逊于方案 D，但在特征融合方面仍具优势。方案 F (C2f_RFAConv+ASFF_F) mAP 为 94.75%，虽优于单模块方案，但不及包含 CMUNeXt 的组合。

最终，本文提出的 YOLO-SLB 模型融合了 CMUNeXt、C2f_RFAConv 与 ASFF_F 三模块，性能最优。mAP 达 97.98%，较基准模型提升 4.29%；准确率与召回率为 97.05%、92.67%，提升了 1.97%、6.46%。尽管参数量增至 9.6 MB，计算量为 17.0 GFLOPs，帧率降至 129 f/s，仍满足实时检测需求，具备良好的工程实用性与应用前景。

3.4.2 对比实验

本研究在实验中采用从零训练的方式，未引入迁移学习机制。旨在更客观地评估各网络模型在无预训练支持下的自主学习能力，避免迁移学习所可能带来的预设特征对实验结论造成干扰。训练策略同样适用于本文所提出的方法以及若干主流的基准模型。实验结果如表 4 所示。

实验结果表明，YOLO-SLB 在与各经典模型对比中，

表 3 消融实验结果

Table 3 Results of ablation experiment

Model	Improvement			Params/MB	GFLOPs	P/%	R/%	mAP/%	FPS
	CMUNeXt	C2f_RFAConv	ASFF_F						
Base				**6.12**	**8.2**	95.08	86.21	93.69	**526**
A	√			6.37	9.5	95.20	90.87	95.72	463
B		√		6.4	8.4	93.14	88.71	94.12	228
C			√	9.1	15.4	94.16	87.73	94.34	224
D	√	√		6.45	9.7	96.63	92.25	96.20	339
E	√		√	9.5	16.7	96.23	91.78	95.89	307
F		√	√	9.2	15.6	94.72	87.14	94.75	137
YOLO-SLB	√	√	√	9.6	17.0	**97.05**	**92.67**	**97.98**	129

平均检测精度达到了最高的 97.98%。对比 YOLOv8 原始模型，准确率、召回率和平均检测精度分别提升了 0.97%、4.46% 和 4.29%。与早期的模型 YOLOv5 和 YOLOv7 相比，虽然计算量和参数量有所增加，但是平均检测精度均有大幅提高，分别提高了 6.77% 和 10.05%。与 YOLOv10 原始模型对比，虽然在准确率上下降了 0.31%，但是在召回率和平均检测精度上提高了 5.70% 和 4.89%。对比 YOLOv11 原始模型，准确率、召回率和平均检测精度全部提高，分别为 2.64%、8.91% 和 8.11%，在 FPS 仅下降 2 f/s 的情况下，可以说明 YOLO-SLB 的表现全面优于 YOLOv11 原始模型。在与 RT-DETR 的对比中，各项指标均大幅超过，由于该模型为 Transformer 架构，模型的计算量和参数量过大，导致 FPS 过低，可能无法应用在该场景中，同时平均检测精度过低，为了验证该是否为架构缺陷，为 RT-DETR 模型更换 ResNet50 骨干网络 (RD-ResNet)，实验结果仍然不理想。由对比实验结果可知，综合各项指标 YOLO-SLB 在锂电池外观缺陷检测上有明显优势，符合实际场景需求。

3.4.3 泛化性实验

为了进一步验证本算法的泛化能力，本文使用 NEU-DET[25] 数据集进行实验，该数据集是专为钢材表面缺陷检测设计的一个重要资源，由东北大学宋克臣团队创建，包含了 1800 张图像，覆盖了 6 种常见的钢材表面缺陷类型，分别是：裂纹 (crazing)、夹杂 (inclusion)、斑块 (patches)、凹陷表面 (pitted surface)、压入氧化皮 (rolled-in scale) 以及划痕 (scratches)。这些多样化的缺陷类型使得 NEU-DET 成为训练和评估自动缺陷识别系统性能的理想工具。泛化性实验结果如表 5 所示。

实验结果表明，YOLO-SLB 在各项关键指标上均表现出较强的竞争力。在精确率方面，YOLO-SLB 达到了 79.55 %，显著高于 YOLOv8 的 67.36% 和 YOLOv10 的 69.13；召回率为 73.26%，略低于 YOLOv8 的 80.40%，但优于 YOLOv10 的 70.74%。这表明 YOLO-SLB 在保证较高检出率的同时，具备较强的抗误检能力。此外，在平均检测精度上，YOLO-SLB 取得

表 4 对比实验结果

Table 4 Results of comparative experiment

Model	Params/MB	GFLOPs	P/%	R/%	mAP50/%	FPS
YOLOv5	**1.9**	**4.5**	85.03	84.88	91.21	**750**
YOLOv7	5.73	13.2	91.59	79.89	87.93	679
YOLOv8	6.12	8.2	95.08	86.21	93.69	526
YOLOv10	5.80	8.2	**96.36**	84.97	93.09	667
YOLOv11	5.50	6.3	93.41	81.76	89.87	131
RT-DETR	66.2	103.5	72.38	50.09	58.82	67
RD-ResNet	86.1	125.7	79.60	59.23	69.76	55
YOLO-SLB	9.6	17.0	96.05	**90.67**	**97.98**	129

表 5 泛化性实验结果

Table 5 Results of the generalization experiment

Model	Params/MB	GFLOPs	P/%	R/%	mAP50/%	FPS
YOLOv8	6.12	8.2	67.36	**80.40**	80.66	350
YOLOv10	**5.80**	**8.2**	69.13	70.74	77.75	**356**
文献[26]	19.0	/	/	/	79.90	62.5
文献[27]	/	/	/	/	77.70	17.36
文献[28]	/	/	73.9	76.7	80.7	/
YOLO-SLB	9.6	17.0	**79.55**	73.26	**81.31**	106

了 81.31% 的性能表现，优于 YOLOv8 的 80.66% 和 YOLOv10 的 77.75%，显示出其在目标定位和分类精度方面的综合优势。

尽管文献 [26] 所提方法在 mAP50% 上达到了 79.90%，但由于其参数量高达 19.0 MB，计算资源消耗较大，且未提供 FPS 等实时性指标，限制了其在工业现场部署的可能性。而文献 [27] 虽然强调了模型轻量化设计，但其 FPS 仅为 17.36 f/s，难以满足在线检测的实时需求。文献 [28] 虽然平均检测精度达到 80.7%，但与本文模型相比，仍然低了 0.61%。相比之下，YOLO-SLB 在保持合理模型复杂度 (参数量为 9.6 MB，计算量为 17.0 GFLOPs) 的前提下，实现了 106 f/s 的推理速度，同时达到了最高的平均检测精度 (81.31%)，兼顾了检测精度与效率，更适合于工业环境下的高吞吐量检测任务。

3.4.4 定性实验分析

为进一步验证 YOLO-SLB 算法在实际应用中的有效性与鲁棒性，本文选取了当前主流的 YOLOv8 算法以及改进后的 YOLO-SLB 算法，并在相同的测试集上分别采用各自的最优训练模型进行可视化实验与对比分析。由于本文所使用的锂电池电芯外观缺陷数据集来源于实际生产车间，且检测任务主要针对几类常见且关键的外观缺陷开展无损检测，因此在可视化实验部分，仅选取部分具有代表性的缺陷类型来展示其检测效果，以更贴近实际工业应用中的任务需求与场景特点。

实验结果如图 11 所示，两种检测模型均能够有效识别锂电池电芯的外观缺陷。然而，本文提出的 YOLO-SLB 模型在检测准确率方面表现更为优异，不仅显著提升了检测置信度，在面对复杂环境时也展现出更强的鲁棒性。通过对比图 11(a) 与图 11(b) 的检测结果可以发现，YOLO-SLB 模型在多个方面优于 YOLOv8 模型，其检测置信度更高，同时成功识别出 YOLOv8 模型未能检测到的目标，尤其在小目标检测任务中表现更加稳定和精准。此外，在面对复杂目标分布的情况下，YOLO-SLB 模型不仅保持较高的置信输出，还有效避免了 YOLOv8 模型中存在的漏检问题。综上所述，相较于基准模型 YOLOv8，YOLO-SLB 在整体检测精度方面有明显提升，同时具备更少的误检与漏检情况，展现出更强的实际应用潜力。

4 结 论

本文针对锂电池电芯外观缺陷检测中目标尺寸小、缺陷类型复杂多样以及实际生产环境下数据采集困难等问题，本文提出了一种改进的 YOLO-SLB 检测算法，以提升检测精度与鲁棒性。

通过引入 CMUNeXt Block 增强骨干网络的空间信息交互能力，设计 C2f_RFAConv 模块以缓解卷积参数共享带来的表达瓶颈，并改进融合多尺度细节信息的 ASFF_F 检测头，显著提升了模型对微小缺陷的感知能力与特征表达能力。实验结果表明，YOLO-SLB 在准确率、召回率及平均检测精度上较原始 YOLOv8 模型分别提升了 0.97%、4.46% 和 4.29%，同时在 NEU-DET 工业缺陷数据集上仍保持 81.31% 的检测精度，展现出良好的泛化能力与实用价值。

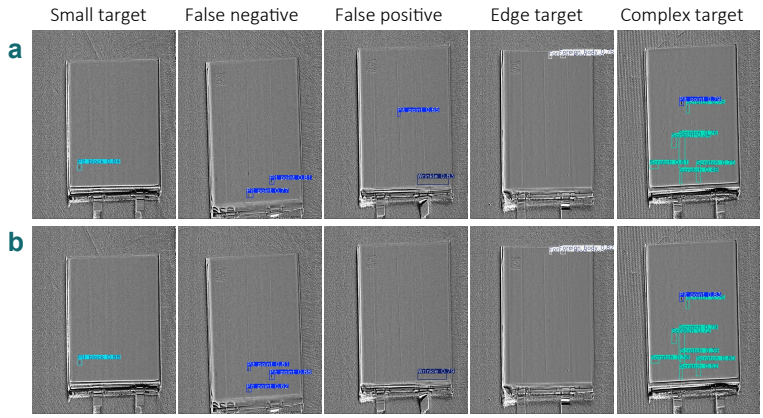


图 11 定性实验分析结果。(a) YOLOv8；(b) YOLO-SLB

Fig. 11 Results of qualitative experimental analysis. (a) YOLOv8; (b) YOLO-SLB

未来将结合半监督学习与自动标注策略，在产线上实现高质量样本的自动积累与模型持续优化，推动锂电池缺陷检测系统向全自动化方向发展。

作者贡献

熊中越：方法的提出，论文构思和撰写；实验设计及数据整理和分析；张上：论文审核与编辑写作；张明：实验数据分析。

利益冲突

所有作者声明无利益冲突

参考文献

1. Geng J. Structured-light 3D surface imaging: a tutorial[J]. *Adv Opt Photonics*, 2011, **3**(2): 128−160.
2. Wang L H, Cai J, Zhang Y N, et al. Application of structured light 3D reconstruction technology in industrial automation scenarios in the context of digital transformation[J]. *SHS Web Conf*, 2024, **181**: 03026.
3. Bieri L S, Jacot J. Three-dimensional vision using structured light applied to quality control in production line[C]//*Proceedings of SPIE 5457, Optical Metrology in Production Engineering*, 2004: 463–471. https://doi.org/10.1117/12.545039.
4. Zhao X S, Lin H Z, Chen H J, et al. Back-propagation-assisted inverse design of structured light fields for given profiles of optical force[J]. *Nanophotonics*, 2023, **12**(11): 2019−2027.
5. Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. *Proc IEEE*, 1998, **86**(11): 2278−2324.
6. Yang G W, Wang Y Z. Three-dimensional measurement of precise shaft parts based on line structured light and deep learning[J]. *Measurement*, 2022, **191**: 110837.
7. Jin L, Li S Y, Qin G, et al. Outer surface defect detection of steel pipes with 3D vision based on multi-line structured lights[J]. *Meas Sci Technol*, 2024, **35**(6): 065203.
8. Pierdicca R, Paolanti M, Matrone F, et al. Point cloud semantic segmentation using a deep learning framework for cultural heritage[J]. *Remote Sens*, 2020, **12**(6): 1005.
9. Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. *Commun ACM*, 2017, **60**(6): 84−90.
10. Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 580–587. https://doi.org/10.1109/CVPR.2014.81.
11. Girshick R. Fast R-CNN[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision (ICCV)*, 2015: 1440–1448. https://doi.org/10.1109/ICCV.2015.169.
12. Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, **39**(6): 1137−1149.
13. Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 779–788. https://doi.org/10.1109/CVPR.2016.91.
14. Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//*Proceedings of the 14th European Conference on Computer Vision*, 2016: 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.
15. Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//*Proceedings of the 16th European Conference on Computer Vision*, 2020: 213–229. https://doi.org/10.1007/978-3-030-58452-8_13.
16. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale[C]//*Proceedings of the 9th International Conference on Learning Representations*, 2021.
17. Zhao Y A, Lv W Y, Xu S L, et al. DETRs beat YOLOs on real-time object detection[C]//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024: 16965–16974. https://doi.org/10.1109/CVPR52733.2024.01605.
18. Wang J, Fan D L, Du J P, et al. Research on artificial intelligence detection method of lithium battery surface defects for production line[J]. *J Comput*, 2023, **34**(2): 203−214.
19. Liu X H, Wu L Q, Guo X Q, et al. A novel approach for surface defect detection of lithium battery based on improved K-nearest neighbor and Euclidean clustering segmentation[J]. *Int J Adv Manuf Technol*, 2023, **127**(1-2): 971−985.
20. Gao L, Zhao J W. YOLOv8 algorithm improvement and its application in small target detection[J]. *Trait Signal*, 2025, **42**(2): 985−994.
21. Tang F H, Ding J R, Quan Q, et al. CMUNEXT: an efficient medical image segmentation network based on large kernel and skip fusion[C]//*Proceedings of 2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, 2024: 1–5. https://doi.org/10.1109/ISBI56570.2024.10635609.
22. Hendrycks D, Gimpel K. Gaussian error linear units (GELUs)[EB/OL]. arXiv:1606.08415, (2023-06-06) [2026-09-15]. https://doi.org/10.48550/arXiv.1606.08415
23. Zhang X, Liu C, Yang D G, et al. RFAConv: innovating spatial attention and standard convolutional operation[Z]. arXiv: 2304.03198, 2023. https://doi.org/10.48550/arXiv.2304.03198.
24. Liu S T, Huang D, Wang Y H. Learning spatial fusion for single-shot object detection[Z]. arXiv: 1911.09516, 2019. https://doi.org/10.48550/arXiv.1911.09516.
25. Song K C, Yan Y H. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects[J]. *Appl Surf Sci*, 2013, **285**: 858−864.
26. Xu W S, Zhang Y F, Jiang X H, et al. An efficient steel defect detection model based on multi-scale information extraction[J]. *Robot Intell Autom*, 2024, **44**(6): 817−829.
27. Liu L J, Zhang Y, Karimi H R. Resilient machine learning for steel surface defect detection based on lightweight convolution[J]. *Int J Adv Manuf Technol*, 2024, **134**(9-10): 4639−4650.
28. Ma Y J, Cao C Q, Zhang J. Steel surface defect detection based on YOLOv8-SOE[J]. *Opto-Electron Eng*, 2025, **52**(5): 250032.
马玉洁, 曹雏清, 张靖. 基于 YOLOv8-SOE 的钢材表面缺陷检测[J]. 光电工程, 2025, **52**(5): 250032.

作者简介



张上 (1979-)，男，博士，副教授，2014 年毕业于中国地质大学，获博士学位。2019 年任三峡大学计算机与信息学院副教授。主要研究方向为计算机视觉、物联网、嵌入式。

E-mail：zhangshang@ctgu.edu.cn



张明 (1999-)，男，硕士研究生，2023 年就读于三峡大学电气与新能源学院，攻读硕士学位，主要研究方向为计算机视觉、嵌入式。

E-mail：z2316428360@163.com



【通信作者】熊中越 (2000-)，男，硕士研究生，2023 年就读于三峡大学计算机与信息学院，攻读硕士学位，主要研究方向为计算机视觉。

E-mail：eng_xiong@163.com

YOLO-SLB: appearance defect detection model of lithium battery cell oriented to structured light imaging

Zhang Shang[1], Xiong Zhongyue[1*], Zhang Ming[2]

[1] College of Computer and Information Technology, China Three Gorges University, Yichang, Hubei 443002, China;
[2] College of Electrical Engineering and New Energy, China Three Gorges University, Yichang, Hubei 443002, China

Abstract

Objective Lithium-ion batteries are critical components in modern energy systems, and their safety and reliability heavily depend on the quality of cell manufacturing. Detecting surface defects on battery cells—such as scratches, pits, wrinkles, and corrosion—is essential yet challenging due to the small size, high diversity, and complex morphology of these defects. Traditional inspection methods, including manual visual checks or conventional 2D imaging, struggle with low accuracy, high false/missed detection rates, and sensitivity to lighting variations in real-world production environments. Moreover, data acquisition under industrial conditions is often hindered by reflective surfaces (e.g., aluminum-plastic film) and inconsistent illumination. To address these issues, this study aims to develop a robust, high-precision defect detection model tailored for structured light imaging, which provides stable, high-contrast 3D-like surface information while mitigating environmental interference.

Methods We propose YOLO-SLB, an enhanced YOLOv8 architecture specifically designed for micro-defect detection on battery cells using structured light images. The model introduces three key innovations. First, the backbone integrates CMUNeXt Blocks featuring large-kernel depthwise convolutions in an inverted bottleneck structure to strengthen spatial feature interaction and preserve fine details critical for detecting minute anomalies. Second, the neck employs a novel C2f_RFAConv module that replaces standard convolutions with receptive field attention convolution (RFAConv), which dynamically generates convolutional weights to overcome the limitations of static weight sharing in large kernels, thereby improving fusion between shallow spatial and deep semantic features. Third, the detection head is redesigned as ASFF_F—a four-branch adaptive spatial feature fusion module that not only applies learnable scale-wise weights but also adds a dedicated micro-target detection path directly utilizing high-resolution shallow features. This reconstruction enhances multi-scale perception and reduces information loss during feature pyramid operations. The model is evaluated on SLBD, a new dataset of 477 original high-resolution (1520 pixel×2000 pixel) battery cell images captured via line-scan cameras under structured light, augmented to 3147

Foundation item: Special Project for Pilot and Demonstration Construction of Digital Economy in Hubei Province (2312-420625-04-02-99663), Hubei Provincial National College Students' Innovation and Entrepreneurship Training Program (S202311075047)

*Correspondence: Xiong Z Y, E-mail: eng_xiong@163.com

samples covering eight common defect types.

Results and Discussions The experimental results show that the proposed model achieves an average detection accuracy of 97.98%, with precision and recall rates of 97.05% and 92.67%, respectively, which are 4.29%, 1.97%, and 6.46% higher than those of the benchmark model YOLOv8. Extensive experiments on the NEU-DET public dataset also verified its good generalization performance, with an average detection accuracy improvement of 0.65% compared to the benchmark model, indicating that this method has high engineering application value.

Conclusions YOLO-SLB effectively addresses the core challenges of micro-defect detection in lithium battery cells by synergizing structured light imaging with architectural innovations in feature extraction, fusion, and detection. The integration of CMUNeXt, C2f_RFAConv, and ASFF_F significantly boosts sensitivity to small anomalies without compromising robustness. The model's high accuracy, generalization capability, and real-time performance underscore its engineering value for automated quality control in battery manufacturing. Future work will explore semi-supervised learning and online model updating to enable continuous improvement using unlabeled production data.

Keywords: computer vision; defect detection; feature extraction; YOLO; structured light imaging

Zhang S, Xiong Z Y, Zhang M. YOLO-SLB: appearance defect detection model of lithium battery cell oriented to structured light imaging[J]. *Opto-Electron Eng*, 2026, **53**(1): 250213; DOI: 10.12086/oee.2026.250213

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文