文章编号:1671-4598(2025)11-0083-14 DOI:10.16526/j.cnki.11-4762/tp.2025.11.010 中图分类号:TP391.41 文献标识码:A

基于改进 YOLOv5s 和 DeepLabV3+的摄像头模组瑕疵检测

张 满[1],李璟文[1,2],毕杰方[1],张金莹[1]
(1.江南大学 理学院,江苏 无锡 214122;
2.江西盛泰精密光学有限公司,江西 新余 336600)

摘要: 为了克服摄像头模组现有检测方法的局限性,提出了一种基于改进 YOLOv5s 和 DeepLabV3+的摄像头模组瑕疵检测方法,以满足摄像头模组工业生产过程中外观与功能检测的需求;针对摄像头模组表面瑕疵检测中存在难度大、目标小、种类多的问题,采用了基于改进 YOLOv5s 的外观定性检测方法;该方法引入多维协作注意力机制,并结合基于 NWD 的损失函数优化,有效提高了模型对小目标的检测能力;测试结果表明,改进后 YOLOv5s 的平均精度 *mAP* 达 95.7%,相比原始模型提升了 8.5%,同时,每秒帧率 FPS 为 38.5,基本满足工业实时检测的要求;此外,针对需要进行定量检测的组件(如脖子胶区域),进一步研究了一种基于 DeepLabV3+语义分割的脖子胶定量分析方法;通过提取区域边界,并计算其面积与长宽比特征,有效评估模组的组装质量并识别潜在功能瑕疵;相比传统方法,该方法能够同时实现摄像头模组的外观与功能检测,同时保障检测的精度与速度,并为其他工业领域的质量控制和瑕疵检测提供了有益借鉴与参考,具有较高的应用价值。

关键词: 摄像头模组;瑕疵检测;YOLOv5s;小目标优化;DeepLabV3+

Defect Detection of Camera Modules Based on Improved YOLOv5s and DeepLabV3+

ZHANG Man[1], LI Jingwen[1,2], BI Jiefang[1], ZHANG Jinying[1]
(1. School of Science, Jiangnan University, Wuxi 214122, China;
2. Jiangxi Shengtai Precision Optics Co., Ltd., Xinyu 336600, China)

Abstract: To overcome the limitations of existing inspection methods for camera modules, a defect detection method for camera modules based on improved YOLOv5s and DeepLabV3+ is proposed, which meets the requirements of appearance and functional inspection in the industrial production process of camera modules. In the detection of surface defect in camera modules, there are problems such as high complexity, small targets, and diverse defect types, and an improved YOLOv5s-based qualitative appearance inspection method is presented. A multimodal co-attention (MCA) mechanism is introduced to combine with an optimized loss function based on a normalized wasserstein distance (NWD), significantly enhancing the model's ability to detect small targets. Experimental results show that the improved YOLOv5s achieves a mean average precision (*mAP*) of 95.7%, an 8.5% improvement over the original model, with a frames per Second (FPS) of 38.5, meeting the requirements for real-time industrial inspection. Additionally, for the components requiring quantitative analysis, such as the neck glue area, a DeepLabV3+based quantitative analysis method for the segmented neck glue is proposed. By extracting the region boundaries and calculating the area and length to width ratio, this method effectively evaluates the assembly quality and identifies the potential functional defect. Compared with traditional methods, the proposed approach can simultaneous-

收稿日期:2024-10-14; **修回日期:**2024-11-21。
基金项目:江西省 03 专项及 5G 项目(S2023ZXXM C0049);中国博士后科学基金第 70 批面上资助一等(2021M700039);国家自然科学基金项目(11904135)。
作者简介:张 满(1998-),男,硕士。
通讯作者:李璟文(1985-),男,博士,副教授。
引用格式:张 满,李璟文,毕杰方,等.基于改进 YOLOv5s 和 DeepLabV3+的摄像头模组瑕疵检测[J].计算机测量与控制,2025,33(11):83-96.

ly detect the appearance and function of camera modules while ensuring the accuracy and speed of detection. It provides a valuable reference for quality control and defect detection in other industrial fields, demonstrating a significant application value.

Keywords: camera module; defect detection; YOLOv5s; small object optimization; DeepLabV3+

0 引言

摄像头模组作为智能手机、个人电脑等数码产品中的核心组成部分，其质量直接影响产品的成像效果和用户体验[1]。因此，如何在生产过程中有效检测摄像头模组表面的瑕疵，成为确保高质量、高分辨率摄像头模组生产的关键步骤。

一般而言，摄像头模组主要由镜头、柔性电路板（FPC，flexible printed circuit）和连接器三部分构成。在生产过程中，往往涉及诸多工序，如镜头表面清洗、FPC 板烘烤、晶圆固定与烘烤、装订、连接器点胶等[2]。然而，由于工业生产环境并非完全无尘，模组在加工安装过程中常受到灰尘、棉絮等异物的影响，导致镜头表面出现黑点、白点、白斑、脏污、划痕等瑕疵，严重影响成像质量[3]。此外，在模组点胶和组装过程中，当点胶量不足或比例失调时，往往会出现组件松动或脱落。上述外观瑕疵和质量瑕疵不可避免的会影响产品的外观和质量。因此，在生产和组装过程中实时检测和识别这些瑕疵，从而确保产品质量是非常必要的。

目前，最常见的检测方法往往通过人工检测和传统视觉检测，然而，这些检测手段存在多种弊端。首先，人眼容易疲劳且具有不稳定性，难以保证检测精度，容易出现漏检和误检。其次，由于人工检测在精度与速度上的限制，复杂精细的检测任务难以完成，导致效率低且成本高[4]。基于传统视觉的检测方法通过图像采集与处理，可以对摄像头模组瑕疵进行在线检测。这种方法尽管在某些程度上能够弥补人工检测的不足，但由于其主要依赖于浅层特征的提取，难以应对摄像头模组中复杂多样的小目标瑕疵特征[5]。因此，如何快速、准确地检测摄像头模组的外观瑕疵及功能瑕疵，并对其质量进行评估，是摄像头模组生产线面临的重大挑战。

随着计算机视觉的快速发展，深度学习模型开始被应用于工业产品检测。深度学习模型具有优秀的特征学习和特征表达能力，可以逐层提取特征，弥补传统方法的不足，为模组瑕疵和质量的定性检测提供了一种有效途径[6-7]。许多基于不同网络架构和算法的目标检测模型被广泛应用于瑕疵检测领域，这些模型一般可分为双阶段和单阶段检测算法。

其中，双阶段检测算法主要包括 Region-CNN（R-CNN）[8]、Fast Region-based CNN（Fast R-CNN）[9]和 Faster R-CNN（Faster Region-based CNN）[10]。由于这类算法涉及两个阶段，即候选区域生成和区域分类，处理时间较长。相较之下，单阶段检测模型如 You Only Look Once（YOLO）[11-15]系列、Single Shot MultiBox Detector（SSD）[16]和 CornerNet[17]在实时性上表现更优。其中，YOLO 系列由于其在检测精度和速度上的优势，已被广泛应用于工业生产中的瑕疵检测。例如，文献［18］描述了了一种基于 YOLOv3 的端到端缺陷检测模型，文献［19］描述了一种多注意力深度学习网络，用于解决纺织品图像中多尺度缺陷的共存问题。同年，文献［20］描述了改进的 MS-YOLOv5 网络，成功识别了铝表面上随机分布的 7 种缺陷，取得了 87.4% 的总体准确率。文献［21］描述了将基于 YOLO 的目标检测模型用于摄像头模组的检测，检测精度达到了 90.5%。

尽管深度学习算法在摄像头模组质量检测领域具有巨大潜力，但当前的研究和应用仍存在以下局限：小目标检测能力有限：首先，主流深度学习模型在大中型目标检测任务中表现出色，但工业生产中的检测对象多为小目标，这要求对现有模型进行优化以适应小目标检测任务。例如，模型结构中的主干网络需针对性地优化，以提升其在工业生产环境中的适应性。因此，针对摄像头模组的外观质量检测，需要进一步优化现有的深度学习模型，以满足工业生产中对小目标瑕疵检测的实时性、精度和效率的要求。其次，难以对瑕疵直接定量分析：目标检测模型往往仅能对待检瑕疵进行定性分类，当需要对瑕疵或特定区域进行定量分析时，较为常规的做法是首先利用目标检测模型识别瑕疵区域，然后采用基于传统图像处理算法的边界提取方法，从而对其进行定量分析与评估。这无疑增加了算法的复杂性，而且模型的通用性和可重复性有待提高。

因此，仅依靠目标检测模型难以判断脖子胶区域是否异常，无法评估模组组装质量和潜在异常。语义分割技术可以对图像进行像素级别的分类，能够精确区分瑕疵像素与正常像素，并提供瑕疵的几何形状、面积等相关信息。这对于工业生产线上的多类别瑕疵的定量检测尤其重要。因此，基于语义分割的特征提取方法可以快速识别特征边界，有望直接对特定功能区域（如脖子胶）进行定量分析（如计算面积与长宽特征），通过与相关标准进行对比，从而评估模组是否存在潜在的功能瑕疵，可以实现功能瑕疵的高效检测[22]。

在相关研究中，文献［23］描述了通过卷积神经网络中的类激活映射技术解决了传统网络无法定位缺陷的问题。文献［24］描述了利用改进的 SegNet 网络，在

轮胎缺陷检测中实现了像素级的缺陷分割。文献［25］描述了将 Inception 模块嵌入 U-Net 中，用于磁瓦表面缺陷的准确分割。类似地，文献［26］描述了通过 U-Net 分割模型实现了对磁粉缺陷的准确分割，并结合骨干网络和优化损失函数来提升模型性能。文献［27］描述了利用双注意力网络检测钢材表面缺陷，并精准定位瑕疵形状。文献［28］描述了通过非对称卷积技术检测瓷砖表面的裂纹瑕疵，进一步降低了模型复杂度。

DeepLabV3＋是目前广泛应用于图像分割任务中的先进语义分割模型。它的编码器－解码器结构、空洞空间卷积金字塔池化（ASPP，atrous spatial pyramid pooling）模块以及深度可分离卷积使其在捕捉全局上下文信息的同时有效保留细节，并显著降低计算开销，适合工业场景中的应用[29]。例如，文献［30］描述了将密集卷积模块引入 DeepLabV3＋以提升高光谱图像分割性能，文献［31］描述了 DeepLabV3＋在自动驾驶场景中的实时分割性能的优化。文献［32］描述了通过 MobileNetV2 优化了 DeepLabV3＋，实现了水利工程施工中的骨料粒径检测。文献［33］描述引入注意力机制，设计了多尺度特征融合的 DeepLabV3＋，有效提升了钢板表面瑕疵的分割精度。尽管基于语义分割的特征提取与分割在工业瑕疵检测中具有很好的应用前景，但目前的研究工作主要集中于瑕疵区域的分割与识别，很少对分割区域进行定量分析，尚未发现有研究人员将语义分割模型用于摄像头模组中特定瑕疵的分割、定量分析与潜在功能评估。

基于上述研究与应用背景，本文研究了一种基于目标检测与语义分割融合的摄像头模组瑕疵检测方法。

首先，针对镜头表面瑕疵，采用基于目标检测的定性分类方法，选择了在实时性以及精度上都表现优秀的 YOLOv5s 模型，并在此基础上，进一步优化其小目标检测能力。通过引入多维协作注意力机制（MCA，multidimensional collaborative attention），对注意力计算中的特征进行压缩和优化，从而提高小目标的检测精度。在降低计算复杂度的同时，提升了模型的性能和实时性。此外，在后处理阶段，采用了基于归一化高斯瓦斯坦距离（NWD，normalized gaussian wasserstein distance）的损失函数，通过优化 Wasserstein 距离，实现了更平滑、稳定的训练过程，进一步提升了小目标的检测能力。实验结果表明，改进后的 YOLOv5s 算法平均精度均值（mAP，mean average precision）可达 95.7%，相比初始的 YOLOv5 算法提升了 8.5%，每秒帧率（FPS，frames per Second）达到 38.5。

其次，针对需要定量检测的功能组件（如脖子胶），采用 DeepLabV3＋语义分割模型，首先对点胶区域进行分割识别，然后提取区域的面积与长宽特征，通过将其与标准区间进行对比，有效评估模组的组装质量并识别潜在的功能瑕疵。实验结果表明，采用 DeepLabV3＋模型之后的平均像素准确率（mPA，mean pixel accuracy）达到了 97.10%。论文研究的检测方法能够同时实现摄像头模组外观瑕疵的定性分类和部分功能组件的定量评估，为摄像头模组的整体质量检测提供了一种更全面的方法。该方法具备良好的推广性，可应用于其它工业质量控制和瑕疵检测场景。

1　模型与方法

1.1　基于 YOLOv5s 模型的目标检测

自 2016 年首次问世以来，YOLO 系列已经历多次升级与优化，成为目标检测领域中的佼佼者。其中，YOLOv5 作为最经典且广受欢迎的版本，通过卷积神经网络的最佳优化策略，在精度、效率和识别能力等方面表现出色。YOLOv5 官方提供了 4 种不同的模型版本，分别为 YOLOv5s、YOLOv5m、YOLOv5l 和 YOLOv5x。考虑到工业生产中的实际应用环境及硬件设备的计算能力限制，本文选用了轻量级的 YOLOv5s 模型作为摄像头模组镜头表面瑕疵检测的基础模型，使得在保持较高检测精度的同时，降低了计算复杂度与开销，保障实时性和效率要求。

YOLOv5s 模型由输入层、骨干网络、颈部网络和输出层四部分构成，各部分的功能如下。

输入层：负责接收并处理输入图像，将其转换为网络所需的格式。YOLOv5s 采用了马赛克数据增强、自适应锚点计算和自适应图像缩放三大策略[34]。这些策略在训练过程中发挥了重要作用，尤其是在处理不同数据集时。马赛克数据增强通过结合多张图像，提高了数据多样性，并改善了模型的泛化能力；自适应锚点计算根据目标物体的大小动态调整锚点框尺寸，提升了对不同尺度目标的定位精度；自适应图像缩放则在保证输入尺寸灵活性的同时，最大限度保留了图像细节，优化了检测效率。

骨干网络：用于从图像中提取特征信息。YOLOv5s 采用了高效的骨干架构，结合了 Focus 结构与 Cross-Stage Partial Network（CSP）结构。Focus 结构通过将图像分块后进行拼接，增强了细节捕捉能力，尤其适合高分辨率图像处理；CSP 结构则通过特征图的分离与逐步融合，减少了重复计算，提高了推理速度与准确性。这两者结合，使得 YOLOv5s 在特征提取方面既高效又精准，同时降低计算资源消耗[35-37]。

颈部网络：连接骨干网络与输出层，通过融合不同层次的特征信息，进一步提升检测性能。YOLOv5s 采用了特征金字塔网络（FPN，feature pyramid network）[38]和像素聚合网络（PAN，pixel aggregation net-

work)[39]进行特征融合。FPN 通过自上而下的金字塔结构，将高级语义信息传递至低层特征图，从而增强了对小目标的检测能力；PAN 则优化了自下而上的路径聚合，进一步加强了特征的多尺度融合，提升了对不同尺度目标的检测精度。这种架构组合在复杂场景下，能够高效应对多尺度目标。

输出层：生成最终的检测结果，包括目标的边界框位置、类别概率和置信度分数。YOLOv5s 的输出层经过优化设计，能够高效输出精准的检测和分类结果。通过边界框的精确回归和置信度计算，模型在不同场景下能够快速识别并生成高质量的检测结果。此外，输出层的优化进一步提升了推理速度，使得 YOLOv5s 既能保证精度，又能满足实时检测的需求。

1.2 基于 YOLOv5s 的小目标检测的性能优化

1.2.1 基于多维协作注意力模块 MCA 的小目标检测性能优化

深度学习中的注意力机制是受人类视觉和认知系统启发而设计的一种机制。通过在机器学习目标检测任务中引入注意力机制模块，能够显著增强模型对特定目标特征的提取能力，从而提高检测性能和效率。常见的注意力机制模块包括 Efficient Channel Attention（ECA）模块、Selective Residual Module（SRM）模块、Convolutional Block Attention Module（CBAM）模块[40]，这些常见的机制都各有优缺点，其中，ECA 和 SRM 模块主要关注通道和空间维度的建模，但往往忽视了两者之间的交互关系；CBAM 模块虽然提供了更全面的建模能力，但引入了较高的模型复杂度和计算负担。

针对上述注意力机制的不足，本文采用了一种轻量级且高效的注意力机制——多维协作注意力模块（MCA）[40]。MCA 模块通过创新性地采用三分支架构，分别在通道、高度和宽度 3 个维度上进行注意力计算。该模块的设计在保持计算效率的同时，有效提升了检测性能。其主要特点如下。

三分支架构：MCA 模块在不同分支中独立且并行地对高度、宽度和通道维度的信息进行建模。这使得模型能够充分考虑每个维度的特征差异，并避免信息在某个维度上的遗漏。

高效性与轻量化：相较于 CBAM 较为复杂的模块，MCA 在不显著增加计算成本的前提下，提供了更全面的特征权重分配。其计算复杂度大幅降低，更适用于工业生产中的实时检测需求。

通过 MCA 模块，模型能够在处理特征图时更全面地聚焦于重要特征，实现高精度的小目标检测。同时，MCA 通过并行处理和结果整合，有效提高了特征提取的效率和实时性。

图 1 展示了 MCA 注意力机制模块的具体结构。模块从顶部到底部分别为用于捕获空间维度宽度（W）中特征之间的交互分支、高度（H）中特征之间的交互分支以及用于捕获通道（C）之间的交互分支。其三个分支相互独立但又彼此联系，其中前两个分支会使用置换操作来捕获任一空间维度与通道维度之间的远程依赖关系，此操作可以使模型再计算某个维度的注意力权重时，同时考虑到其它维度的信息，以捕获更丰富的特征信息，此方法在处理小目标时效果更为显著；最终在 3 个通道完成计算后得到的注意力权重进行简单的加权平均之后应用到特征图上，并对其进行融合。

其中 3 个分支的具体操作过程以及计算公式[41]如下所示。

1）高度注意力分支（Height Attention Branch）：通过对特征图的高度维度进行压缩，计算出该维度的注意力权重，从而强化特征图中的重要高度信息，计算公式如下：

$$A_h = \sigma(W_{1h} \cdot \mathrm{Avg}_A(X)) \tag{1}$$

其中：X 表示输入的特征张量，Avg_h 表示高度维

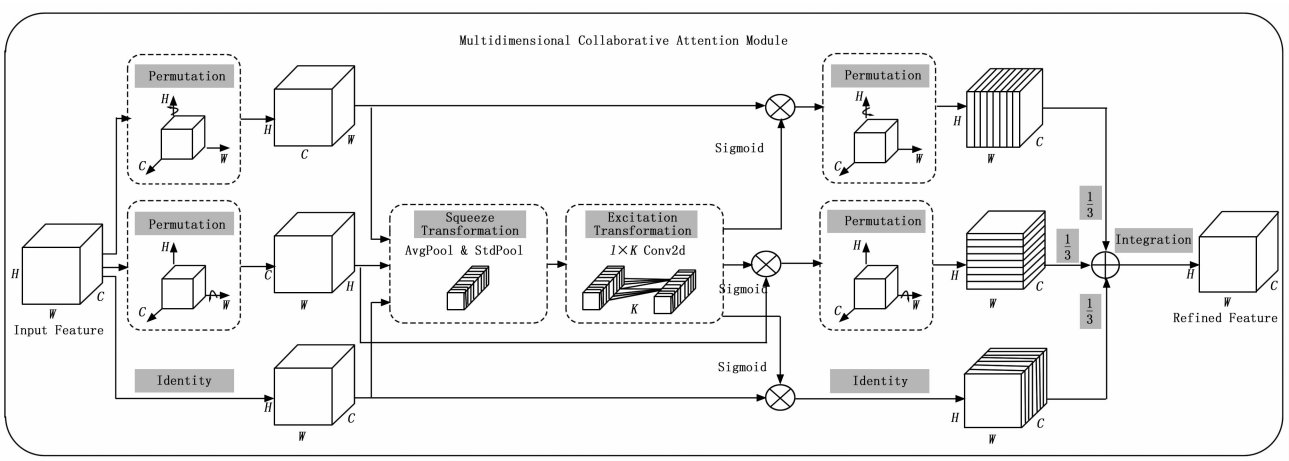


图 1 MCA 注意力机制模块结构

度上的平均池化操作，σ 表示激活函数（softmax），W_{1h}表示不同的高度上的特征向量，A_h 表示高度分支上的注意力权重向量；

2）宽度注意力分支（Width Attention Branch）：通过对特征图的宽度维度进行压缩，计算出该维度的注意力权重，从而强化特征图中的重要宽度信息，计算公式如下：

$$A_w = \sigma(W_{1w} \cdot Avg_w(X)) \tag{2}$$

其中：X 表示输入的特征张量，Avg_w表示宽度维度上的平均池化操作，σ 表示激活函数（softmax），W_{1w}表示不同的宽度上的特征向量，A_w 表示宽度分支上的注意力权重向量；

3）通道注意力分支（Channel Attention Branch）：通过对特征图的通道维度进行压缩，计算出该维度的注意力权重，从而强化特征图中的重要通道信息，计算公式如下：

$$A_c = \sigma[W_{1c} \cdot \mathrm{Avg}_c(X)] \tag{3}$$

其中：X 表示输入的特征张量，Avg_c表示通道维度上的平均池化操作，σ表示激活函数（sigmoid），W_{1c}表示不同的通道上的特征向量，A_c 表示通道分支上的注意力权重向量；

在 3 个分支计算完成后，用逐元素加和法将各自的注意力权重分别应用到原始特征图的相应维度，计算公式为：

$$Y_h = \boldsymbol{A}_h \odot X, Y_w = \boldsymbol{A}_w \odot X, Y_c = \boldsymbol{A}_c \odot X \tag{4}$$

其中：$\boldsymbol{A}_h$、$\boldsymbol{A}_w$、$\boldsymbol{A}_c$ 分别表示高度、宽度和通道分支上的注意力权重向量，Y_h、Y_w、Y_c 分别表示高度、宽度和通道分支上加权后的特征张量，⊙表示逐元素乘法，X 表示输入张量。

最后将不同分支上加权后的特征张量进行融合，生成最终的特征张量即特征图用于后续的目标检测。经过 MCA 注意力模块这种特殊的并行处理和加权融合的方式，使得每个维度的信息都能被充分挖掘和利用，显著增强了特征图的表达能力，大大提高了模型的整体性能。

1.2.2　基于 NWD 的 Loss 优化

在目标检测任务的后处理阶段，大多数模型都是使用 IoU 以及其变种（如 GIoU、DIoU、CIoU）等传统的评估指标，通常关注的是预测框与真实框的重叠区域。然而，在边界框位置差异较大或者存在部分重叠的时候，往往难以提供精确的度量。为了解决这些问题，本文引入了一种全新的度量方式——归一化高斯瓦斯坦距离（NWD）[42]，这种方式主要是通过计算预测边界框与真实边界框之间的 Wasserstein 距离来评估其相似性，Wasserstein 距离是一种度量分布之间差异的距离指标。相较于传统的 IoU，它能够更精确地反映边界框的重叠情况及其位置差异，特别是在处理小目标时，Wasserstein 距离具有更显著的优势，因为它能够更好地捕捉小目标的特征，从而提高检测精度。

本文中所使用的 NWD Loss 其核心是 Wasserstein 距离，介绍此概念之前，首先需要说明一下高斯分布。在 NWD Loss 中，预测边界框和真实边界框都用二维高斯分布来表示，这种方式可以更准确地捕捉边界框的空间分布特征。不同像素的权重，可以将边界框建模为二维高斯分布，其中边界框的中心像素最高，像素的重要性从中心向边界递减。具体说，对于水平边界框 $R=(c_x, c_y, w, h)$，其中 (c_x, c_y)，w 和 h 分别表示中心坐标、宽度和高度。其内切椭圆方程可以表示为[42-43]：

$$\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} = 1 \tag{5}$$

其中：(μ_x, μ_y) 为椭圆的中心坐标，σ_x，σ_y为沿 x、y 轴的半轴长度，因此 $\mu_x=c_x$，$\mu_y=c_y$，$\sigma_x=\frac{w}{2}$，$\sigma_y=\frac{h}{2}$；

二维高斯分布的概率密度函数如下所示：

$$f(X \mid \mu, \sum) = \frac{\exp\left[-\frac{1}{2}(X-\mu)^T \sum^{-1} (X-\mu)\right]}{2\pi \mid \sum \mid^{1/2}} \tag{6}$$

其中：X、μ 和 $\sum$ 表示高斯分布的坐标（x，y）、均值向量和协方差矩阵：

$$(X-\mu)^T \sum^{-1} (X-\mu) = 1 \tag{7}$$

当公式（7）成立时，公式（5）中的椭圆将是二维高斯分布的密度轮廓，因此，水平边界框 $R=(c_x, c_y, w, h)$ 可以被建模成一个二维高斯分布：

$$N(\mu, \sum):$$

$$\mu = [(c_x \cdot c_y)], \sum = \begin{bmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{bmatrix} \tag{8}$$

得到每个预测边界框和真实边界框对应的二维高斯分布参数后，将进行 Wasserstein 距离的计算，对于两个二维高斯分布 $\mu_1 = N(m_1, \sum_1)$ 和 $\mu_2 = N(m_2, \sum_2)$，μ_1 和 μ_2 之间的二阶 Wasserstein 距离定义为：

$$W_2^2(\mu_1, \mu_2) \| m_1 - m_2 \|_2^2 + Tr\left[\sum_1 + \sum_2 - 2\left(\sum_2^{1/2}\right)\sum_1 \sum_2^{1/2}\right)^{1/2}\right] \tag{9}$$

上式可以简化为：

$$W_2^2(\mu_1, \mu_2) = \| m_1 - m_2 \|_2^2 + \| \sum_1^{1/2} - \sum_2^{1/2} \|_F^2 \tag{10}$$

其中：$\| \cdot \|_F$ 是 Frobenius 函数。

此外，对于高斯分布 Na 和 Nb，它们是根据边界

框 $A=(cx_a, cy_a, w_a, h_a)$ 和 $B=(cx_b, cy_b, w_b, h_b)$ 建模的，公式（10）可以进一步化为：

$$W_2^2(N_a,N_b)=\left\|\left\{\left(cx_a,cy_a,\frac{w_a}{2},\frac{h_a}{2}\right)^T,\left[\left(cx_b,cy_b,\frac{w_b}{2},\frac{h_b}{2}\right)\right]\right\}^T\right\|_2^2 \tag{11}$$

但是此时得到的结果 $W_2^2(N_a, N_b)$ 是距离度量，并不能直接用作相似性度量，因此为了使其适用于损失函数，通常要对其进行归一化处理，归一化的处理公式如下所示：

$$NWD(N_a,N_b)=\exp\left(-\frac{\sqrt{W_2^2(N_a,N_b)}}{C}\right) \tag{12}$$

与传统的 IoU 相比，归一化后的 Wasserstein 距离在作为损失函数时展现出多个优势。首先，它能够更准确地反映边界框在空间上的分布差异；其次，由于 Wasserstein 距离的梯度更加平滑和稳定，训练过程中的收敛性和模型稳定性得到了显著改善；此外，该度量方式特别适用于检测细小瑕疵，如摄像头模组镜头表面的细小瑕疵，从而有效提升小目标的检测精度和模型的训练效果。通过上述改进，YOLOv5s 模型在处理目标检测任务时，能够充分利用归一化高斯 Wasserstein 距离的优越特性，进一步提升了检测精度和训练稳定性。

本文改进后的 YOLOv5s 网络结构如图 2 所示，展示了在经过改进之后的模型网络架构。

1.3 基于语义分割的脖子胶区域定量检测与功能分析

1.3.1 基于 DeepLabV3＋的特定区域瑕疵的分割

语义分割是计算机视觉中与目标检测不同的一项任务。与目标检测定位物体并确定其类别不同，语义分割在识别物体的同时，还会精确定位每个物体的边界，其核心是对图像中的每个像素进行分类，使同类像素具有相同的语义标签，从而提供更加细粒度的图像信息[44]。在本研究中，针对摄像头模组功能部分组装点胶过程中脖子胶区域的特殊瑕疵，传统的目标检测方法无法满足精确定位和区域提取的需求。因此，本文采用基于语义分割的区域特征提取与定量分析方法为特定功能瑕疵的判别提供了一种可能。

DeepLabV3＋ 是一种功能强大的语义分割模型，具有高精度、出色的细节保留能力以及多尺度信息捕捉能力。其架构主要包括两个部分：编码器（Encoder）和解码器（Decoder）[44]。编码器负责提取图像的高级特征，解码器则将这些特征逐步上采样至原始图像的分辨率，恢复并细化分割结果。具体的网络结构如图 3 所示，其中，编码器部分包括主干网络（图 3 中 DCNN 部分）和空洞空间金字塔池化（ASPP）模块[45]。输入的图像经过主干网络处理，得到两个输出：一个是高级特征，另一个是低级特征。高级特征通过编码器的 ASPP 模块，进行 5 种不同操作（1×1 卷积、3 个不同膨胀率的空洞卷积、1 个全局平均池化），然后将这些输出通过拼接操作和 1×1 卷积结合，得到多尺度的特征表示。

解码器部分则处理主干网络的低级特征图和 ASPP 输出。首先，对低级特征图使用 1×1 卷积进行通道降维，实验表明将通道数从 256 降至 48 效果最佳（较多的通道会掩盖 ASPP 输出特征的重要性）。然后，ASPP 输出的特征图进行插值上采样，得到与低级特征图相同尺寸的特征图。接着，将降维后的低级特征图与上采样后的 ASPP 特征图进行拼接，并送入一组 3×3 卷积块进行处理；最后，再次进行线性插值上采样，得到与原

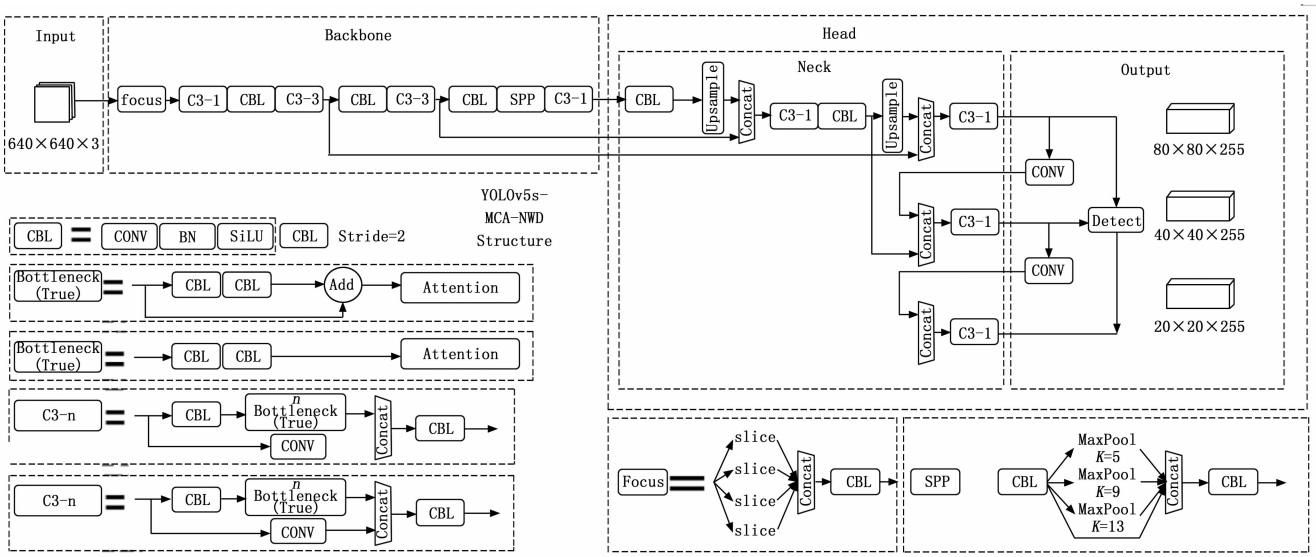


图 2 改进后的 YOLOv5s-MCA-NWD 模型网络结构

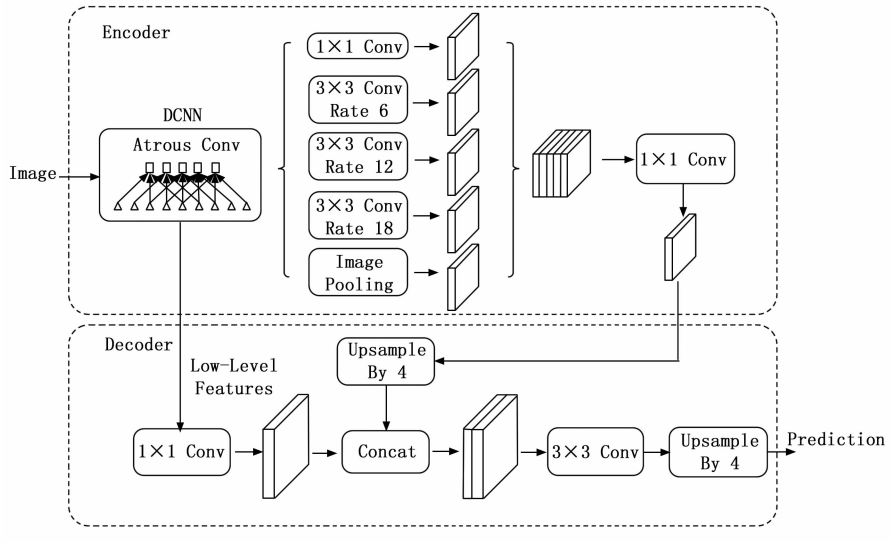


图 3　DeepLabV3＋网络模型结构

始图像相同分辨率的预测图。

通过上述操作（结合编码器－解码器结构、空洞卷积和 ASPP 模块），DeepLabV3＋在多尺度特征捕捉、细节保留以及处理小目标等方面表现优异，能够在语义分割任务中取得高精度结果，适用于各种工业生产场景。此外，DeepLabV3＋采用深度可分离卷积技术，显著降低了计算复杂度和参数量，同时保持较高的特征提取能力和分割精度，从而一定程度上避免了本研究中由于计算资源有限带来的挑战。

1.3.2　基于语义分割输出的特定区域定量分析

基于前文提取的脖子胶区域分割结果，本文采用了一种基于深度学习的图像处理技术，提出了一种用于判断点胶是否符合标准的检测方案。该方案主要通过计算分割区域的面积来判断瑕疵情况，依托开源的计算机视觉库 OpenCV 实现对瑕疵区域面积、长宽的准确计算和可视化。具体的方案流程如下。

分割区域提取：首先，通过 1.3.1 小节中所述的分割模型及方法，对脖子胶区域进行分割提取，得到完整的点胶区域图像。

轮廓提取与面积、长宽计算：接下来，使用 OpenCV 库中的多个函数对瑕疵区域进行轮廓提取和面积计算。主要函数包括 cv2.imread（读取图像）、cv2.cvtColor（灰度转换）、cv2.findContours（检测轮廓）等，通过这些函数完成瑕疵区域的轮廓提取，并计算其具体面积、长度、宽度，从而得到面积和长宽比大小。

面积标准比较：在完成瑕疵区域面积和长宽比计算后，将计算所得面积、长宽比与设定的标准区间进行对比。一般而言，行业设定的标准为：点胶区域面积在 69 000 到 76 000 像素平方之间且同时满足区域的长宽比在 8∶1～9∶1 之间的才为合格产品，不满足这两项中的任一项均视为不合格产品。该标准基于实际生产要求和质量控制需求，能够有效评估点胶过程的合格性。

该检测方案中面积的单位为像素平方，计算的是轮廓内部的像素数。具体的计算公式如下所示：

$$S=\frac{\sum_{i=0}^{(n-1)}(x_i\cdot y_{i+1}-y_i\cdot x_{i+1}}{2} \tag{13}$$

其中：$(x_i,\ y_i)$ 为轮廓上的第 i 个点，n 表示轮廓上的点的数量。

通过上述方法，能够对脖子胶区域的瑕疵面积、长宽特征进行定量分析。通过将其与预定标准进行实时对比，可以有效实现工业生产中摄像头模组功能部分组装点胶部位的自动化实时检测。

1.4　模型评价指标

在本节中，将简要描述用于评估模型性能的指标。我们采用精度 P（定义为正确分类的瑕疵产品占分类器所划分的所有瑕疵产品的比例）和召回率 R（定义为正确分类的瑕疵产品占瑕疵产品数量的比例）来计算模型的 mAP：

$$mAP=\frac{1}{N}\sum\nolimits_1^N AP \tag{14}$$

式中，N 表示 N 个分类，AP 表示平均准确率，AP 的计算方法如下：

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

$$P=\frac{TP}{TP+FP} \tag{16}$$

$$R=\frac{TP}{TP+FN} \tag{17}$$

其中：采用了真正例（TP，True Positive）、假正例（FP，False Positive）和假反例（FN，False Negative）评估模型的性能。其中，真正例表示模型正确识别的瑕疵，与人工标注的瑕疵相符；假正例表示模型错误地将非瑕疵标记为瑕疵；而假反例则表示模型未能正确识别存在的瑕疵。除了这些指标，还利用精准率－召回率（PR，precision-recall）曲线和 $F1$ 曲线来综合评估模型性能。PR 曲线展示了精准率和召回率之间的关系，这两者是一对相互矛盾的指标。通常情况下，高精准率往往伴随着低召回率，反之亦然。为了全面评估模型，还需要综合考虑精准率和召回率。因此，曲线越接近图的右上角，说明模型性能越好。然而，不同算法的 PR 曲线常常交叉，难以单凭曲线形态来判断模型的优

劣。因此，通常会借助 $F1$ 曲线来进行衡量。$F1$ 分数是精准率和召回率的调和平均，能够更综合地反映模型的性能表现，定义为：

$$F_1 = 2 \times \frac{P \times R}{P + R} \tag{18}$$

同时采用 mPA 和平均交并比（mIoU，mean Intersection over union）两个评价指标来衡量语义分割任务中的性能。

mPA 的计算公式如下：

$$PA = \frac{TP}{TP + FP} \tag{19}$$

$$mPA = \frac{1}{N} \sum_{i=1}^{N} \frac{TP_i}{TP_i + FN_i} \tag{20}$$

其中：PA 表示每个类别的像素准确率，TP、FP、FN 分别代表真正例、假正例、假反例，i 代表第 i 个类别，N 代表类别的总数。IoU 的计算公式如下所示：

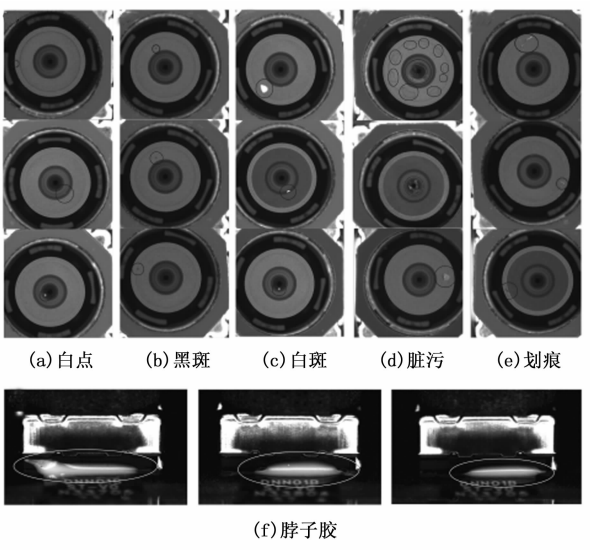
$$\mathrm{IoU} = \frac{TP}{TP + FP + FN} \tag{21}$$

$$\mathrm{mIoU} = \frac{1}{N} \sum_{i=1}^{N} \frac{TP_i}{TP_i + FN_i + FN_i} \tag{22}$$

其中：TP 表示正确分割的像素数，FP 表示错误分割为该类的像素数，FN 表示该类中被错误分割为其他类的像素数，i 代表第 i 个类别，N 代表类别的总数。

2 摄像头模组瑕疵数据集准备

本文所使用的摄像头模组数据集主要采集自摄像头模组工业生产线。在本文的研究中，主要针对摄像头模组镜头表面的多种外观瑕疵（即白点、黑斑、白斑、脏污和划痕），以及一种组装点胶过程中容易出现的功能瑕疵（脖子胶区域异常）。正如图 4 中，展示了上述 5 种需要定性分析的外观瑕疵和 1 种需要定量检测的功能瑕疵。



(a) 白点 (b) 黑斑 (c) 白斑 (d) 脏污 (e) 划痕

(f) 脖子胶

图 4 摄像头模组瑕疵外观

由于工业生产中的瑕疵样本数据集规模较小且分布不均匀，本文首先采用了图像增强的方法对数据集进行了扩展。图像增强是一种常用且有效的数据预处理技术，通过对原始图像进行一系列变换来生成新的图像，从而增加数据集的多样性，提高模型的泛化能力和鲁棒性。常见的图像增强方法包括旋转、翻转、裁剪、缩放、色彩调整以及添加去除噪声等[46]。

在本研究中，使用基于 OpenCV 和 PyQt5 的图像增强工具对瑕疵样本图像进行数据扩展。原始数据集中，镜头表面的瑕疵样本图像共计 640 张，脖子胶区域的瑕疵样本图像共计 270 张。通过图像增强处理，将数据集扩展为：镜头表面瑕疵样本增加至 9 600 张，脖子胶区域瑕疵样本增加至 4 050 张。

扩展后的数据集按照 7∶2∶1 的比例被划分为训练集、验证集和测试集。其中，镜头表面瑕疵数据集训练集 6 720 张图像、验证集 1 920 张、测试集 960 张；脖子胶瑕疵数据集训练集 2 835 张图像、验证集 810 张、测试集 405 张。经过扩展和随机划分后对各类瑕疵的具体图像数量进行汇总，分别统计了原始数据集中包含各类瑕疵样本的数量以及经过扩增之后的各类瑕疵样本数量，其中镜头表面瑕疵数据集存在单张图像上包含多种瑕疵的情况，而脖子胶瑕疵数据集单张图像只存在一种瑕疵，最终汇总统计结果如表 1 所示。

表 1 镜头表面各种类别瑕疵的数量 张

部位	瑕疵类型	原始数量	扩增数量	训练集	验证集	测试集	总数
镜头表面	白点	443	6 645	6 720	1 920	960	9 600
	黑斑	223	3 345				
	白斑	212	3 180				
	脏污	198	2 970				
	瑕疵	142	2 130				
功能部位	脖子胶	270	4 050	2 835	810	405	4 050

在数据增强的过程中，不仅注重了图像的多样性，还特别关注了增强方法的合理性，确保每种变换都能真实地反映实际生产中的各种情况。例如，通过适当的旋转和翻转，可以模拟生产过程中摄像头模组在不同角度下的瑕疵表现；通过色彩调整和噪声添加，可以反映出生产环境光照变化和其他干扰因素对瑕疵检测的影响。这些增强方法使得数据集更加全面和具备代表性，有助于提高模型在实际应用中的可靠性和准确性。通过图像增强处理，有效地丰富了数据集样本，提升了深度学习模型在不同环境和条件下的鲁棒性和准确性[46]。这一改进有助于提升工业生产中瑕疵检测的可靠性。

3 结果与讨论

3.1 基于 YOLOv5s 的小目标检测优化

基于 YOLOv5s 在保证检测速度的同时还能提供良

好的检测精度，因此本研究选择 Yolov5s 作为摄像头模组瑕疵检测的基础模型。实际上，尽管 YOLOv5s 具有较高的检测性能，但离实际工业应用仍有一定的差距。为了进一步提升检测性能，首先列出并分析了 YOLOv5s 在对不同瑕疵的检测性能。表 2 提供了 YOLOv5s 对摄像头模组数据集中观察到的 5 种主要瑕疵类型的详细检查度量（如 Recall，F_1，Precision，Average Precision 等），这些信息有助于在后续部分中进一步进行数据分析和优化。

表 2　5 种瑕疵类型在 YOLOv5s 模型下的详细检查度量

检测器	瑕疵类型	Precision /%	Recall /%	F_1 /%	AP /%	mAP /%	FPS
YOLOv5s	白点	85.25	71.24	0.75	0.82	87.20	39.90
	黑斑	96.00	77.42	0.86	0.88		
	白斑	83.68	71.60	0.72	0.79		
	脏污	86.76	94.19	0.89	0.95		
	划痕	96.30	89.66	0.93	0.92		

之后，为了更加全面且准确地验证本文中所采用的两种改进方法的有效性，我们在最终制作的镜头表面瑕疵样本数据集上进行了烧蚀实验。该实验的目的是测试模型在不同超参数设置下的表现，并确保避免模型过度拟合，同时兼顾计算资源的使用。

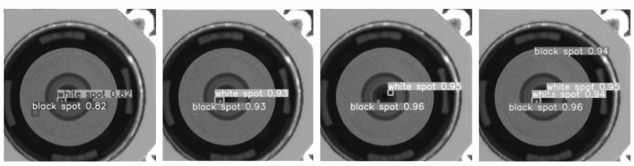
在烧蚀实验中，经过多次实验和调整，最终确定了以下的超参设置：Epochs 设置为 200 轮、Batch Size 大小设置为 16、图像大小设置为 640×640、初始学习率 lr 设置为 0.01、优化器种类设置为随机梯度下降（SGD）、动量参数设置为 0.9、Num _ workers 设置为 16，这些设置能够在保证模型训练稳定性的同时，优化网络性能，最终得到了最优的网络模型。烧蚀实验的结果已在下表 3 中详细列出，以便对改进方法的有效性进行深入分析和讨论。

表 3　YOLOv5s 的烧蚀实验结果

MCA	NWD	Precision /%	Recall /%	$mAP@0.5$ /%	$mAP@0.5:0.95$ /%
×	×	89.60	88.82	87.20	83.64
√	×	93.34	91.68	90.82	86.82
×	√	93.91	92.81	92.43	90.36
√	√	96.83	94.77	95.70	93.81

从表中可以看出，第一行展示了 YOLOv5s 在未进行任何改进时在数据集上的原始性能，其 mAP 结果为 87.20%；第二行和第三行展示了引入了 MCA 注意力机制和引入了 NWD 模块之后的性能表现，结果表明 MCA 机制和 NWD 模块在 Precision、Recall 和 mAP 方面对比原始模型均有明显改善，其中 NWD 模块相较于 MCA 模块在优化效果上更具优势。这是因为在本文的数据集中，瑕疵的形状复杂且具有多维特征；以及小目标之间普遍存在重叠情况；而 NWD 还能够提供更精准的边界框定位；此外，由于 MCA 模块相较于 NWD 模块会增加更多额外的计算量。综合这些因素，最终导致了 NWD 模块的优化效果更为显著。当同时引入 MCA 和 NWD 模块之后，模型的检测效果达到了最佳状态，平均精度（mAP）比原始网络模型提高了 8.50%，达到了 95.70%。

为了更直观地展示不同改进方案对检测结果的影响，选择了其中一张图像在不同模型下的检测结果示意图，如图 5（a）～（d）所示。从图中可以看出，原始模型以及仅添加 MCA 模块和仅添加 NWD 模块的模型均存在漏检情况；然而，添加 MCA 和只添加 NWD 的模型相比原始模型精度均有所提升，其中 NWD 模型的提升效果略优于 MCA 模块。最终，当两者同时引入时，模型不仅弥补了漏检问题，还实现了最佳的检测精度。



(a)原始　(b)+MCA　(c)+NWD　(d)+MCA+NWD

图 5　原始模型及不同优化模型的检测结果对比

同时在表 4 中提供了使用原始 YOLOv5s 网络模型和改进之后 YOLOv5s-MCA-NWD 的相同样本图像的检查细节。

表 4　改进后的 YOLOv5s 和原始的 YOLOv5s 图像检查细节

%

瑕疵	原始	＋MCA	＋NWD	＋MCA＋NWD
白斑 1	0.82	0.93	漏检	0.94
白斑 2	漏检	漏检	0.95	0.95
黑斑 1	0.82	0.93	0.96	0.96
黑斑 1	漏检	漏检	漏检	0.94

此外，在工业生产中，对模型的 FPS 有一定的需求，因此对 3 种模型的检测精度和 FPS 结果进行了可视化对比，如图 6 所示。结果表明，引入了 MCA 机制和 NWD 模块之后确实有助于检测精度的提升，然而，引入 MCA 机制时，由于增加了额外的计算量，导致模型的 FPS 有所下降。相对而言，NWD 模块由于其在梯度信息上的稳定性，使其在检测精度上提升的同时并未显著影响 FPS。特别地，在同时引入 MCA 和 NWD 模块之后，检测精度得到了显著提升，而 FPS 的变化幅度较小，表明这些改进并未对实时性能造成显著影响。

综合考虑检测精度的实质性提高，计算需求的增加是合理的，同时改进后的 FPS 并没有发生太大的改变，仍然能符合工业生产的需求。这表明，对模型的改进并没有显著影响其整体效率和实时性能。

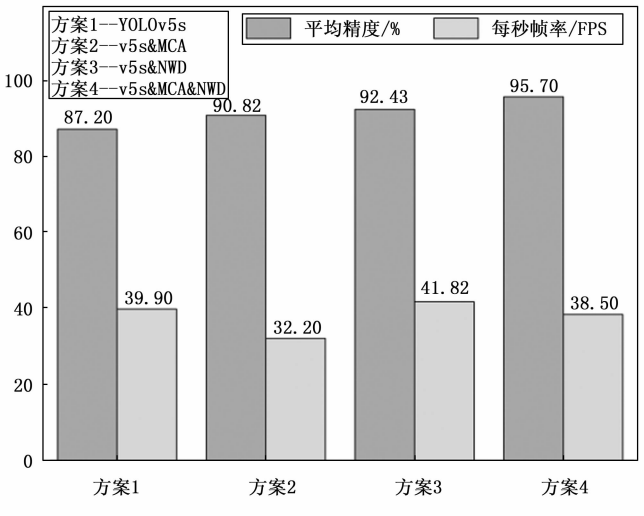


图 6 原始模型与改进后模型 mAP 与 FPS 结果对比

在本文中，我们基于 YOLOv5s 网络模型针对摄像头模组生产过程中小目标瑕疵的检测精度做出了两项关键改进：引入了 MCA 注意力模块和 NWD-Loss 模块，改进后的 YOLOv5s-MCA-NWD 网络模型在训练和验证过程种的各项输出如图 7 所示。图中详细展示了模型训练完成后的损失函数下降曲线、精度（Precision）、召回率（Recall）以及平均精度（$mAP\_0.5$、$mAP\_0.5:0.95$）的变化趋势。从图中可以看出，随着训练轮次的增加，模型的损失值逐渐降低，验证集的精度和召回率逐步提升，最终趋于稳定。特别是在后期训练阶段，模型的 $mAP\_0.5$ 达到较高水平，表明模型在检测小目标和复杂背景下具备良好的泛化能力与稳定性。

最终得到的最佳模型应用于摄像头模组镜头表面瑕疵数据集的部分检测效果如图 8（a）～（f）所示。从图中可以看出，模型对各类瑕疵的检测均获得了较高的精度。其中，对白斑的检测精度最高且最稳定，基本保持在 95％以上；对于其它的瑕疵类型如黑斑、白点、脏污、划痕的检测精度在 90％～97％之间波动，但大多数检测精度结果还是能够保持在 94％以上；此外，检测结果中未发现漏检或误检现象。

3.2 基于 DeepLabV3＋语义分割模型的脖子胶区域定量分析与功能判别

在前文的研究中，我们研究了基于目标检测的摄像头模组外观瑕疵检测。然而，在实际应用场景中，还需对特定区域进行定量分析。例如，在摄像头模组组装过程中，脖子胶区域的长宽和面积特征与模组的组装性能息息相关，为了避免模组部件的松动与脱落，工业中往往对脖子胶区域的面积和长宽比做了明确的限制。只有当其满足特定标准时，才可以被认为是正常样本。如何对该特定区域进行快速提取与定量分析，是摄像头模组质量控制的另一个挑战。

在本节中，将研究基于语义分割的脖子胶区域提取及潜在功能异常检测。首先，展示了基于 DeepLabV3＋语义分割模型的主要评价指标和性能，如图 9（a）和（b）所示。可以看出，模型的 mPrecision 和 mRecall 都能稳定达到 97％以上，同时每个类别的平均像素准确率（mPA）能达到 97.10％，平均交并比（mIoU）

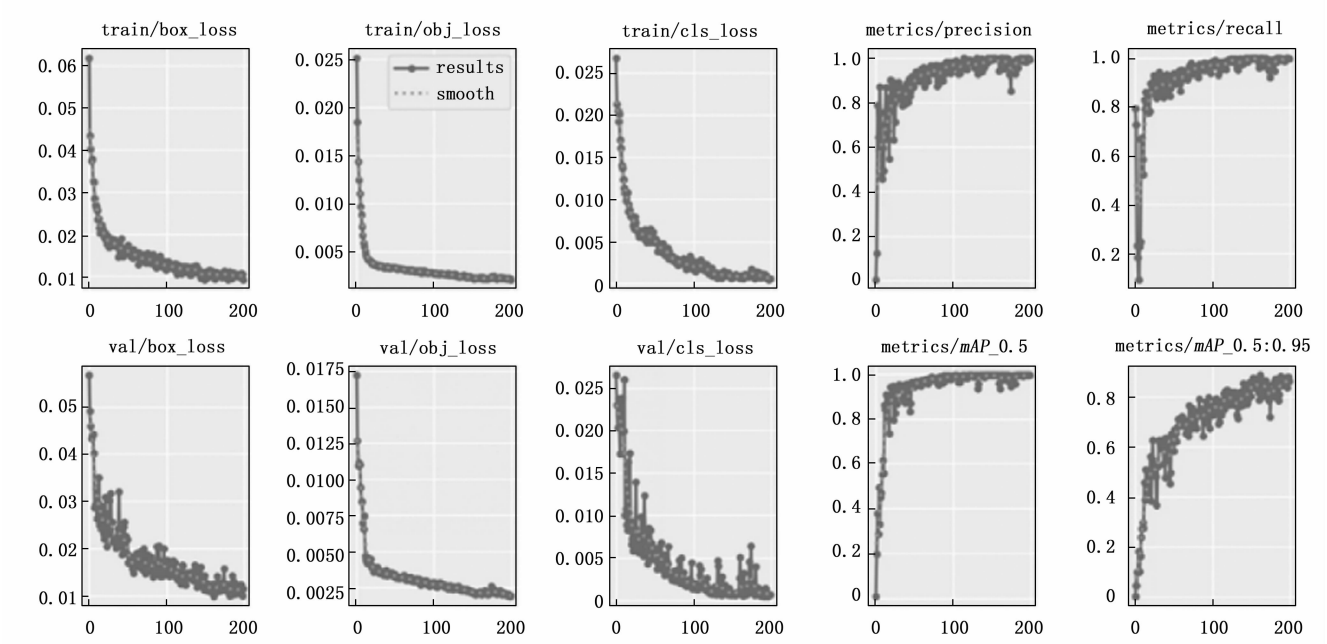
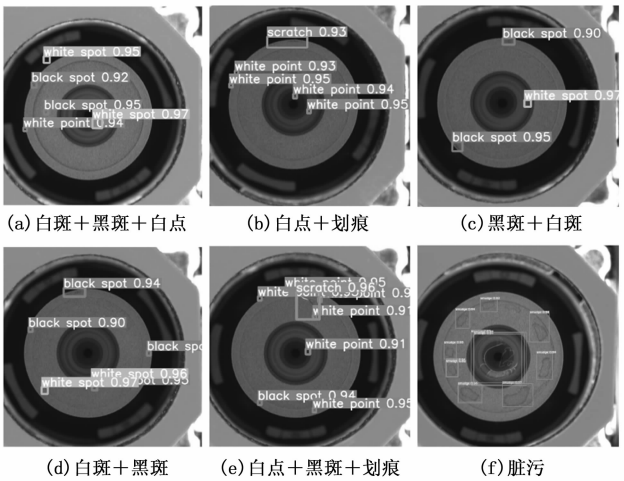


图 7 模型训练和验证过程的输出

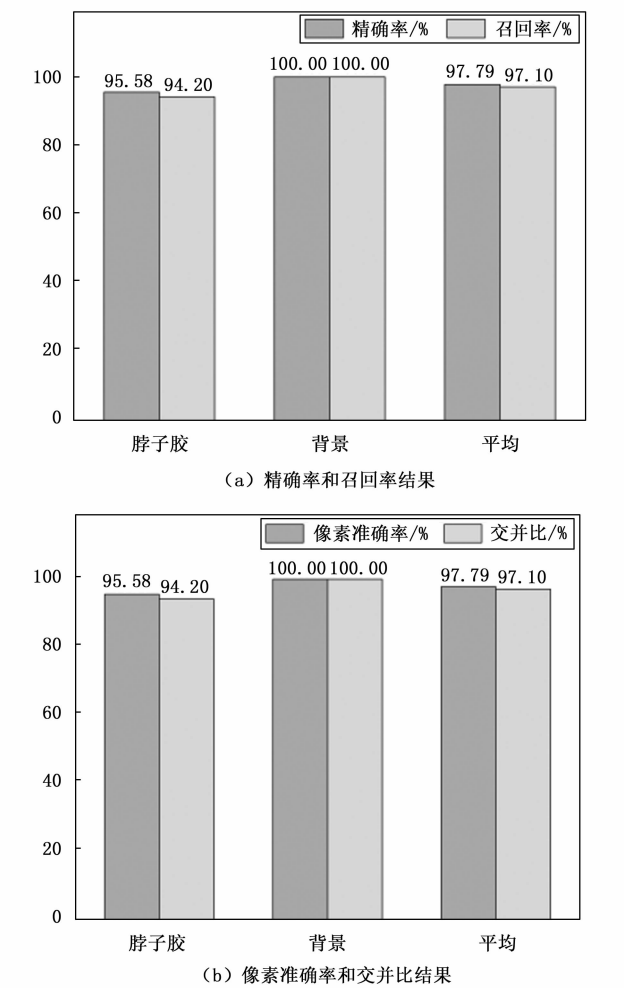
能达到 95.11%，基本满足工业生产上的需求，也符合实验预期。



(a) 白斑＋黑斑＋白点　(b) 白点＋划痕　(c) 黑斑＋白斑

(d) 白斑＋黑斑　(e) 白点＋黑斑＋划痕　(f) 脏污

图 8　不同类型瑕疵最终检测结果



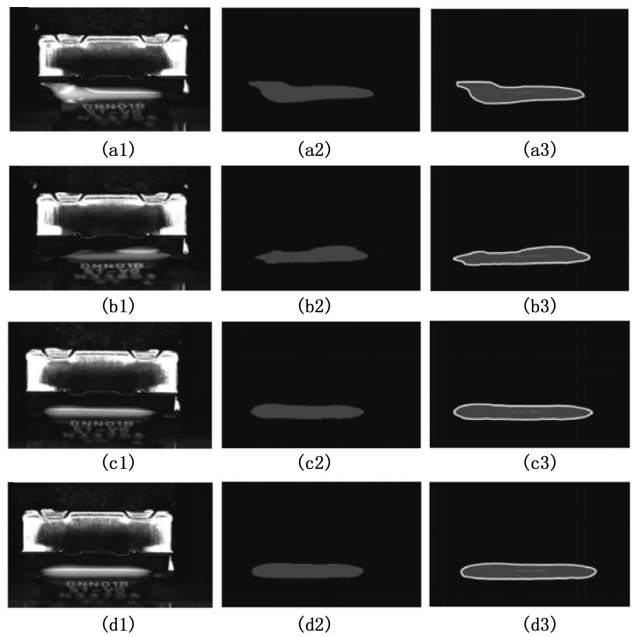
(a) 精确率和召回率结果

(b) 像素准确率和交并比结果

图 9　DeepLabV3＋模型性能评价结果

在得到瑕疵区域的分割结果后，进一步对这些区域的面积和长宽比进行计算，通过与行业标准对比从而判断点胶是否符合生产标准。

图 10 所示了四组脖子胶区域的边界提取过程。原始图像（a1）～（d1）经过 DeepLabV3＋模型处理后，成功提取出所需的脖子胶区域图（a2）～（d2）；提取完成后，对提取到的区域进行面积和长宽比的计算，并在图中使用白线描绘出了瑕疵区域的轮廓，同时会在区域中心显示具体的面积（Contour Area）和长宽比（Aspect ratio）数值，结果如图（a3）～（d3）所示。根据合作企业提供的脖子胶判别标准，如表 5 所示。根据规定，必须面积和长宽比均符合标准才能判定为 OK 产品。图像样本 a 的面积大小为 90 164.5，长宽比为 8∶3，图像样本 b 的面积大小为 67 091，长宽比为 4∶1，根据与标准对比判定其不符合标准为 NG 产品；图像样本 c 的面积大小为 70 496，长宽比为 9∶1，图像样本 d 的面积大小为 75 193.5，长宽比为 8∶1，根据与标准对比，其面积大小和长宽比均符合标准，因此判定其为 OK 产品。



(a1)　(a2)　(a3)

(b1)　(b2)　(b3)

(c1)　(c2)　(c3)

(d1)　(d2)　(d3)

图 10　脖子胶区域面积计算流程

表 5　脖子胶区域组装点胶比例合格标准（盛泰光学提供）

合格标准	长度/mm	宽度/mm	长宽比	面积/像素
最大	6.1	0.75	9∶1	76 000
最小	6	0.65	8∶1	69 000

最后，我们对比了测试集样本在本方案和合作企业传统人工检测下的 *mAP* 和 *FPS* 结果，并进行了可视化对比，如图 11 所示。结果显示，采用本方案进行检测，不仅显著提高了检测的精度，*mAP* 提升至 97.10%，同时还大幅提升了检测速度，*FPS* 提高至 32.71。相比之下，传统人工检测步骤繁琐且耗时，而本方案有效地减少了时间成本，大大提高了工业生产的效率。

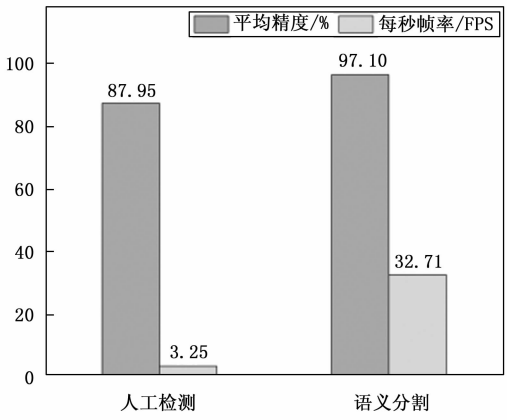


图 11 人工检测和 DeepLabV3+的 *mAP* 与 *FPS* 结果对比

4 结束语

综上所述，本文研究了基于目标检测和语义分割模型检测摄像头模组镜头表面瑕疵以及特殊部位瑕疵检测中的应用。首先，基于 YOLOv5 在检测精度以及检测速度上的优异表现，选择其轻量化版本 YOLOv5s 作为基准模型。在此基础上，针对小目标瑕疵检测精度较低的问题，对原来的 YOLOv5s 算法进行了以下两个关键改进：（1）在骨干网络中引入了 MCA 注意力机制模块；（2）在后处理阶段，采用了基于 NMD 的 Loss 函数来代替传统的 IoU。实验结果表明，本文研究的 YOLOv5s-MCA-NWD 算法的平均精度 *mAP* _ 0.5 为 95.7%，比原来的 YOLOv5s 算法提升了 8.5%，检测速度也达到了 38.5 FPS，显著优于原始模型，能够更快、更准确地检测出摄像头模组镜头表面的细小瑕疵，更好地满足了工业生产的需求。此外，针对摄像头模组组装点胶工艺中的特殊脖子胶瑕疵，研究了基于 DeepLabV3+的语义分割模型的特定区域提取与定量分析方法，并通过与行业标准进行对比来判断是否存在组装瑕疵和功能异常，从而实现了摄像头模组功能瑕疵的自动化检测。

尽管本研究在精度和速度上取得了显著提升，但仍然存在很多改进空间，未来我们的工作将从以下几个方面展开：首先，在模型网络结构方面，将探索更轻量化的设计，以进一步提升检测精度和速度，尤其是在实时检测和大规模工业生产环境中具有更高的实用性。并计划引入更多先进的深度学习技术，如 Transformer 架构和自适应卷积神经网络，以增强模型的特征提取能力。其次，为了应对数据集样本不足的问题，将重点研究基于生成对抗网络（GAN）的数据扩增方法，特别是 DCGAN、CGAN 和 WGAN 等模型，通过生成高质量的瑕疵图像来扩展训练集，从而提升模型的泛化能力。除了数据扩增，还将探索如何引入多任务学习机制，使模型能够同时完成不同瑕疵类型的检测与分割任务，以实现更高的效率。此外，为了进一步简化检测流程，还将研究开发集成模型，使其能够一次性完成摄像头模组不同部位的瑕疵检测，简化工业生产中的操作步骤，增强模型的应用适应性。这些改进有望显著提升模型的实用性和通用性，为未来的智能制造提供更强有力的技术支持。

参考文献：

[1] SONG W T，LIU X M，CHENG Q J，et al. Design of a 360-deg panoramic capture system based on a smart phone [J]. Optical Engineering，2020，59 (1)：015101.

[2] CHANG Y H，LU C J，LIU C S，et al；Design of miniaturized optical image stabilization and autofocusing camera module for cellphones [J]. Sensors and Materials，2017，29 (7)：989－995.

[3] HUANG H X，HU C，WANG T，et al. Surface defects detection for mobilephone panel workpieces based on machine vision and machine learning [C] // In Proceedings of the ICIA，Macao，China：IEEE，2017：370－375.

[4] LV Y F，MA L，JIANG H Q. A mobile phone screen cover glass defect detection model based on small samples learning [C] // In Proceedings of the 4th IEEE International Conference on Signal and Image Processing，Wuxi，China：IEEE，2019：1055－1059.

[5] CHANG C F，WU J L，CHEN K J，et al. M. C. A hybrid defect detection method for compact camera lens [J]. Advances in Mechanical Engineering，2017，9 (1)：16878-14017722949.

[6] YANG J，LI S B，WANG Z，et al. Real-Time tiny part defect detection system in manufacturing using deep learning [J]. IEEE Access，2019，7：89278－89291.

[7] MA L，LU Y，JIANG H Q，et al. An automatic small sample Learning-Based detection method for LCD product defects [J]. CAAI Transactions on Intelligent Systems，2020，15：560－567.

[8] GIRSHICK R，DONAHUE J，DARRELL T，et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // In Proceedings of the CVPR，Columbus，OH，USA：IEEE，2014：580－587.

[9] GIRSHICK R. Fast R-CNN [C] // In Proceedings of the CVPR，Santiago，Chile：IEEE，2017：1440－1448.

[10] REN S Q，HE K M，GIRSHICK R，et al. Faster R-CNN：Towards Real-Time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2017，39 (6)：1137－1149.

[11] ZHU X K，LYU S C，WANG X，et al. TPH-YOLOv5：Improved YOLOv5 based on transformer prediction head for object detection on Drone-Captured scenarios [C] // Proceedings of the CVPR，Honolulu，HI，USA：IEEE，

2017：6517－6525.

[12] REDMON J，FARHADI A. YOLO9000：better，faster，stronger［C］// Proceedings of the CVPR，Salt Lake City，UT，USA：IEEE，2018：895－900.

[13] REDMON J，FARHADI A. YOLOv3：an incremental improvement［C］// In Proceedings of the CVPR，Seattle，WA，USA：IEEE，2020：7263－7274.

[14] BOCHKOVSKIY A，WANG C Y，LIAO H Y M. YOLOv4：optimal speed and accuracy of object detection［C］// Proceedings of the CVPR，Seattle，WA，USA：IEEE，2020：7263－7274.

[15] REDMON J，DIVVALA S，GIRSHICK R，et al. You only look once：unified，Real-Time object detection［C］// Proceedings of the CVPR，Las Vegas，NV，USA：IEEE，2016：779－788.

[16] LIU W，ANGUELOV D，ERHAN D，et al. SSD：single shot MultiBox detector［C］// Proceedings of the ECCV，Amsterdam，Netherlands：Springer，2016：21－37.

[17] LAW H，DENG J. CornerNet：detecting objects as paired keypoints［J］. International Journal of Computer Vision，2019，128（3）：642－656.

[18] SALEH R A，KONYAR M Z，KAPLAN K，et al. End-to-end tire defect detection model based on transfer learning techniques［J］. Neural Computing and Applications. 2024，36（20）：12483－12503.

[19] XIANG X Y，LIU M Q，ZHANG S L，et al. Multi-scale attention and dilation network for small defect detection［J］. Pattern Recognition Letters，2023，172：82－88.

[20] WANG T，SU J H，XU C，et al. An intelligent method for detecting surface defects in aluminium profiles based on the improved YOLOv5 algorithm［J］. Electronics，2022，11（15）：2304.

[21] ZOU H T，HE G，YAO Y，et al. YOLOv7-EAS：a small target detection of camera module surface based on improved YOLOv7［J］. Advanced Theory and Simulations，2023，6（11）：2300397.

[22] TABERNIK D，ŠELA S，SKVARC J，et al. Segmentation-based deep-learning approach for surface-defect detection［J］. Journal of Intelligent Manufacturing，2020，31（3）：759－776.

[23] LEE S Y，TAMA B A，MOON S J，et al. Steel surface defect diagnostics using deep convolutional neural network and class activation map［J］. Applied Sciences，2019，9（24）：5449.

[24] ZHENG Z Z，ZHANG S，YU B，et al. Defect inspection in tire radiographic image using concise semantic segmentation［J］. Ieee Access，2020，8：112674－112687.

[25] CAO X C，CHEN B Q，HE W P. Unsupervised defect segmentation of magnetic tile based on attention enhanced flexible U-Net［J］. IEEE Transactions on Instrumentation and Measurement，2022，71：1－10.

[26] SUGISAKA A，TANAKA K，NAKAMURA T. Magnetic particle defect segmentation using U-Net with optimized backbone and loss function［J］. NDT & E International，2021，120：102448.

[27] PAN Y，ZHANG L M. Dual attention deep learning network for automatic steel surface defect segmentation［J］. Computer-Aided Civil and Infrastructure Engineering，2022，37（11）：1468－1487.

[28] BIVALKAR M，AGARWAL S，SINGH D. Development of an efficient approach for detection and measurement of crack length in ceramic tile manufacturing using millimeter-wave imaging［J］. NDT & E International，2022，129：102656.

[29] CHEN L C，ZHU Y K，PAPANDREOU G，et al. Encoder-decoder with atrous separable convolution for semantic image segmentation［C］// Proceedings of the European Conference on Computer Vision（ECCV），2018：801－818.

[30] CHEN H，QIN Y S，LIU X Y，et al. An improved DeepLabv3＋ lightweight network for remote－sensing image semantic segmentation［J］. Complex & Intelligent Systems，2024，10（2）：2839－2849.

[31] SIAM M，GAMAL M，ABDEL-RAZEK M，et al. A comparative study of real-time semantic segmentation for autonomous driving［C］// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops，2018：587－597.

[32] 张社荣，欧阳乐颖，王　超，等. 基于 DeepLabV3＋的骨料图像自动分割算法［J］. 水利水电科技进展，2022，42（6）：28－32.

[33] 范瑶瑶，王兴芬，刘亚辉. 改进 DeepLabV3＋网络的钢板表面缺陷检测研究［J］. 计算机工程与应用，2023，59（16）：150－158.

[34] YUN S D，HAN D Y，CHUN S，et al. CutMix：regularization strategy to train strong classifiers with locatable features［C］// Proceedings of the ICCV，Seoul，Korea：IEEE，2019：6023－6032.

[35] GUO S Y，LI L L，GUO T Y，et al. Research on Mask-Wearing detection algorithm based on improved YOLOv5［J］. Sensors（Basel），2022，22（13）：1－16.

[36] LIU H Y，SUN F Q，GU J，et al. SF-YOLOv5：a lightweight small object detection algorithm based on improved feature fusion mode［J］. Sensors（Basel），2022，22（15）：1－14.

[37] LI S，WANG X Q. YOLOv5-based defect detection model for hot rolled strip steel［J］. Journal of Physics：Conference Series，2022，2171（1）：1－6.

[38] LIN T Y，DOLLAR P，GIRSHICK R，et al. Feature pyramid networks for object detection［C］// In Proceedings of the CVPR，Honolulu，HI，USA：IEEE，

2017：2117－2125.
[39] LIU S，QI L，QIN H F，et al. Path aggregation network for instance segmentation [C] // In Proceedings of the CVPR，Salt Lake City，UT，USA：IEEE，2018：8759－8768.
[40] GUO Z H，HAN D Z. Multi-modal co-attention relation networks for visual question answering [J]. The Visual Computer，2023，39 (11)：5783－5795.
[41] PASZKE A，GROSS S，MASSA F，et al. PyTorch：an imperative style，high-performance deep learning library [C] // In Proceedings of the 33rd International Conference on Neural Information Processing Systems，Vancouver，Canada，2019：8026－8037.
[42] WANG J W，YANG W，LI H C，et al. Learning center probability map for detecting objects in aerial images [J]. IEEE Transactions on Geoscience and Remote Sensing，2020，59 (5)：4307－4323.
[43] WANG J W，CHANG X，WEN Y，et al. A normalized gaussian wasserstein distance for tiny object detection [J]. ArXiv Preprint ArXiv：2110.13389 (2021).
[44] NEWELL A，YANG K Y，DENG J. Stacked hourglass networks for human pose estimation [C] // In Proceedings of the ECCV，Amsterdam，Netherlands：Springer，2016：483－499.
[45] CHEN L C，PAPANDREOU G，SCHROFF F，et al. Rethinking atrous convolution for semantic image segmentation [C] // Ithaca，NY：ArXiv，2017. ArXiv：1706.05587.
[46] ZHANG H Y，WNAG X Q，CAO L C，et al. Advanced Retinex-Net image enhancement method based on value component processing [J]. Acta Physica Sinica，2022，71 (11)：110701.

(上接第 57 页)

[5] JIN K H，MCCANN M T，et al. Deep convolutional neural network for inverse problems in imaging [J]. IEEE Transactions on Image Processing，2017，26：4509－4522.
[6] GAO G，XU Z，LI J，et al. CTCNet：a CNN-Transformer cooperation network for face image super resolution [J]. IEEE Transactions on Image Processing，2023，32：1978－1991.
[7] SINGH S，ANAND R S. Multimodal medical image fusion using hybrid layer decomposition with CNN based feature mapping and structural clustering [J]. IEEE Transactions on Instru-mentation and Measurement，2020，69：3855－3865.
[8] ZHENG Z，ZHAO J，LI Y. Research on detecting bearing cover defects based on improved YOLOv3 [J]. IEEE Access，2021，9：10304－10315.
[9] REDMON J，DIVVALA S，GIRSHICK R，et al. You only look once：Unif-ied，real-time object detection [C] // Proc-eedings of the IEEE Conference on Computer Vision and Pattern Recognition，2016：779－788.
[10] 胡明合. 多摄像机下模糊图像细节特征目标快速检测研究 [J]. 现代电子技术，2019，42 (13)：76－80.
[11] 范 琳，王亚超. 基于人工智能算法的机器人目标检测技术研究 [J]. 电视技术，2024，48 (6)：205－207.
[12] KIM V H，CHOI K K. A reconfigur-able CNN based accelerator design for fast and energy efficient object detection system on mobile FPGA [J]. IEEE Access，2023，11：59438－59445.
[13] CAO J，YANG Z，LU J，et al. A high performance YOLOV5 accelerator for object detection with near sensor intelligence [C] // 2023 IEEE 15th International Conference on ASIC (ASICON)，2023：1－4.
[14] RUAN Z，XIAO S，DENG Q，et al. FPGA-Based implementation of the quantized neural network for image super resolution [C] // 2023 8th International Conference on Image，Vision and Computing (ICIVC)，2023：374－379.
[15] BAI L，ZHAO Y，et al. A CNN accelerator on FPGA using depthwise separable convolution [J]. IEEE Transactions on Circuits and Systems II：Express Briefs，2018，65：1415－1419.
[16] JAMEIL A K，Al-RAWESHIDY H. Efficient CNN architecture on FPGA using high level module for healthcare devices [J]. IEEE Access，2022，10：60486－60495.
[17] WANG Z，XU K，WU S，LIU L，et al. Sparse-YOLO：Hardware/Software C-o-Design of an FPGA accelerator for YOLOv2 [J]. IEEE Access，2020，8：116569－116585.
[18] ZHANG L，PAN Z. Design implementation of FPGA-based neural network acceleration [C] // 2024 4th International Conference on Consumer Electronics and Computer Engineering (ICCECE)，2024：163－166.
[19] JACOB B et al. Quantization and training of neural networks for efficient integer-arithmetic-only inference [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition，2018：2704－2713.
[20] ADIONO T，PUTRA A，SUTISNA N，et al. Low latency YOLOv3-Tiny accelerator for Low-Cost FPGA using general matrix multiplication principle [J]. IEEE Access，2021，9：141890－141913.
[21] YU Z，BOUGANIS C S，et al. A parameterisable FPGA-tailored architecture for YOLOv3-tiny [C] // Applied Reconfigurable Computing，2020：330－344.