

基于改进 YOLOv7 的手套佩戴检测算法

敖良忠¹ 黄浩宇¹

AO Liangzhong HUANG Haoyu

摘要

维修培训人员在进行培训时需要佩戴手套对飞机进行维修工作,若未佩戴手套进行操作,可能会出现安全问题。针对机务维修培训中,传统人工对培训人员是否佩戴手套的检测效率低下、容易漏检和误检,无法保障培训人员在培训过程中的安全问题,提出了基于改进 YOLOv7 的检测方法。在 YOLOv7 中引入 SPDconv 模块,提高模型针对小目标和低分辨率图片的检测性能。同时,在模型的 Backbone 中加入 CBAM 注意力机制,以提高模型的精度。改进后的 YOLOv7 手套佩戴检测算法平均精度均值达到 86.5%,相较于原本的 YOLOv7 算法,精度提高了 4.3%。实验表明,改进后的算法能有效地检测出维修培训人员的手套佩戴情况。

关键词

目标检测;佩戴检测;YOLOv7 算法;小目标

doi: 10.3969/j.issn.1672-9528.2024.06.004

0 引言

飞机维修人员的培养要求必须在保障人员和设备安全的前提下进行^[1]。由于飞机的机械复杂性,飞机维修工作通常涉及与机械部件、工具和设备的接触。若未佩戴手套,培训人员的手可能会被切割、挤压或造成其他手部伤害。同时,飞机维修涉及使用各种液体、油漆、溶剂等化学物质,若未佩戴手套时,培训人员的皮肤容易受到这些化学物质的刺激,可能引发过敏反应或其他皮肤问题。正确佩戴手套能有效避免上述问题,目前针对维修培训人员是否佩戴手套,采用的是人工检测方法。传统人工检测存在效率低、无法保证实时性、难以有效检测等问题,在实际培训过程中佩戴手套主要依靠培训人员的安全意识和自觉性,因此无法准确有效地确定培训人员是否佩戴手套。针对这种情况,提出一种新型检测方式来改变传统的人工检查。随着计算机技术的发展,基于深度学习的目标检测技术得到广泛应用。利用深度学习能实现对培训人员手套佩戴情况进行快速高效的检测,保护培训人员在操作过程中的安全。

目前对于防护用具的佩戴检测大多是基于 YOLO 目标检测算法进行改进。例如,2021 年,任大丛^[2]提出了一种基于 YOLOv4 改进的手套佩戴检测算法 YOLOv4-DN-CM。该算法以密集型网络 DenseNet121 为主干网络,提升模型检测速度,引入 CBAM 注意力模块提升模型检测精度。2022 年,

肖体刚等人^[3]在 YOLOv3 算法的基础上改进网络结构,使用深度可分离卷积结构替换 Darknet-53 提升了安全帽佩戴检测速率和准确性。2023 年,王彦生等人^[4]在 YOLOv5s 引入自适应卷积和 SK 注意力机制的同时,将 YOLOv5s 的损失函数替换为 EIoU,提高了对绝缘手套佩戴的检测精度。吴昊祺等人^[5]通过 SE 注意力机制和小样本处理技术修改 YOLOv7 主干网络,改进 IoU 损失函数,提高模型的准确性,改进后的 YOLOv7 模型在多目标和小目标场景下具有较好的适应性。2024 年,聂壮壮等人^[6]在 YOLOv5 中添加 SPD 小目标检测模块,接着引入坐标注意力机制,解决了因添加 SPD 所带来的相邻位置关系无法有效提取的问题。同时,将原本的损失函数替换为 SIoU 损失函数,实现了对复杂环境下护目镜佩戴情况的基本检测。

1 YOLOv7 目标检测算法

1.1 YOLOv7 算法概况

YOLO 系列算法在不断发展和改进的过程中,逐渐实现了在保持较高速度的同时提高检测精度,成为目标检测领域的代表性算法^[7]。Alexey Bochkovskiy 团队提出的 YOLOv7 算法^[8]凭借检测精度相对较高、检测速度快等特点,被广泛应用在各个领域。

1.2 YOLOv7 网络结构

YOLOv7 在 YOLOv5 算法的基础上进行了改进和优化,拥有更好的性能。YOLOv7 网络结构主要由 Input、Backbone、Head 3 个部分组成,如图 1 所示。

1. 中国民用航空飞行学院 四川广汉 618307

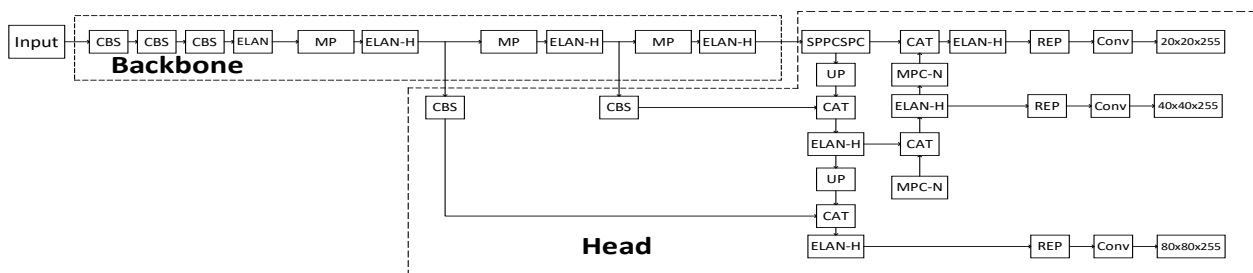


图 1 YOLOv7 网络结构图

1.2.1 Input

Input 对输入图片进行预处理，主要包括 Mosaic 数据增强、自适应锚框计算、自适应图片缩放三部分。采用 Mosaic 数据增强利用了四张图片对其随机缩放、随机裁剪、随机排布的方式进行拼接^[9]，增加了样本的多样性，加快了模型的训练速度。由于原始图像和 Input 要求输入图像大小可能存在不同，会将原始图像自适应图片放缩处理，使原始图像调整到 640×640 大小，再将图像输入，从而提高目标检测速度。

1.2.2 Backbone

Backbone 对 Input 输出的特征进行提取。Backbone 包含 CBS 模块、ELAN 模块、MP 模块。其中，CBS 模块主要由 Conv、BN、SiLU 构成，如图 2 所示。ELAN 模块由多个 CBS 拼接而成，通过控制最短最长的梯度路径，使网络能够有效学习和收敛，具有更强的鲁棒性，如图 3 所示。MP 模块由 CBS 和 Maxpool 组成，作用是进行下采样，减少网络计算量，如图 4 所示。

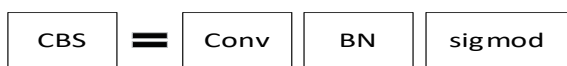


图 2 CBS 模块结构图

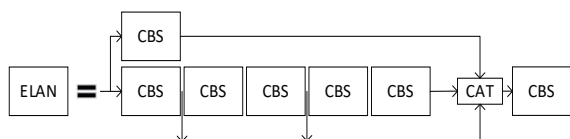


图 3 ELAN 模块结构图

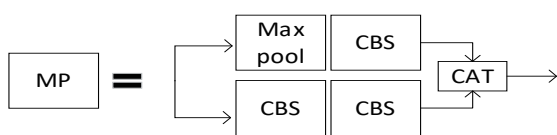


图 4 MP 模块结构图

1.2.3 Head

Head 对特征图进行目标检测，输出目标位置和目标种类。Head 包含 MP-2 模块、SPPCSPC 模块、ELAN-H 模块、RepConv 模块。SPPCSPC 模块得到不同的感受野，适应不同分辨率图像，如图 5 所示。RepConv 结构重参数化进行特征通道数调整，并在三个不同尺度特征图上进行结果预测，输出对应的目标位置与类别信息。ELAN-H 模块与 ELAN 模块的功能和结构相类似。

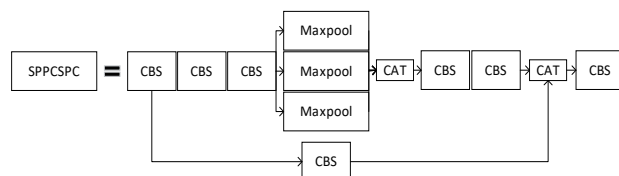


图 5 SPPCSPC 模块结构图

1.3 YOLOv7 改进

1.3.1 SPDconv 模块

数据集中检测目标较小，原 YOLOv7 检测算法对小目标检测精度较低。同时，数据集中部分图片分辨率低。为了提高小目标和低分辨率检测性能，在 YOLOv7 的 Head 中引入 SPDconv（space-to-depthconvolution）^[10] 模块。

SPDconv 由空间到深度（space-to-depth，SPD）层和非跨步卷积层（non-strided convolution）组成。SPDconv 模块对特征图进行下采样，对大小为 $S \times S \times C$ 的特征图 X ，可以将子特征映射序列切片为：

$$\begin{aligned} f_{0,0} &= X[0:S:scale, 0:S:scale], f_{1,0} = X[1:S:scale, 0:S:scale], \dots, \\ f_{scale-1,0} &= X[scale-1:S:scale, 0:S:scale]; \\ f_{0,1} &= X[0:S:scale, 1:S:scale], f_{1,1}, \dots, \\ f_{scale-1,1} &= X[scale-1:S:scale, 1:S:scale]; \\ &\vdots \\ f_{0,scale-1} &= X[0:S:scale, scale-1:S:scale], f_{1,scale-1}, \dots, \\ f_{scale-1,scale-1} &= X[scale-1:S:scale, scale-1:S:scale]; \end{aligned} \quad (1)$$

通常来讲，给定任何特征图 X 进行切片，得到 $scale^2$ 个子映射，其形状为 $(\frac{s}{scale}, \frac{s}{scale}, C_1)$ ，同时对特征图 $scale$ 倍下采样。

把 $scale^2$ 个子映射 $f_{0,0}, f_{0,1}, f_{1,0}, f_{1,1}, \dots, f_{scale-1,scale-1}$ 沿着通道维度拼接起来，得到特征图 $X'(\frac{s}{scale}, \frac{s}{scale}, C_1 scale^2)$ 。相对于 X ， X' 的空间维度减少 $scale$ ，通道维度增加 $scale^2$ 。为了保留特征信息，非跨步卷积层使用 C_2 过滤器将 X' 的通道数 C_1 转化为后续需要的通道数 C_2 ，即 $X'(\frac{s}{scale}, \frac{s}{scale}, C_1 scale^2)$ 转换为 $X''(\frac{s}{scale}, \frac{s}{scale}, C_2 scale^2)$ 。经过上述处理后，大部分小目标特征能够被保留，方便模型检测，提高模型对小目标检测效果。

1.3.2 CBAM 注意力机制

CBAM^[11] 是一种轻量级注意力机制，由通道注意力机制（channel attention module，CAM）和空间注意力机制（spatial attention module，SAM）串联组成^[12]。引入 CBAM 注意力机制，能够在不增加参数数量的情况下，提高模型特征提取能力，提高模型准确度。本文在 Backbone 中加入 CBAM 注意力机制。CBAM 注意力机制结构图如图 6 所示。

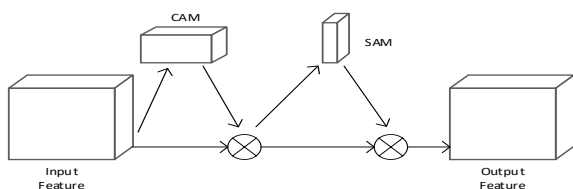


图6 CBAM 结构图

CAM 模块关注输入图片中有意义的信息^[13]。首先，对输入特征图 $F(H \times W \times C)$ 进行基于宽度和高度的全局最大池化和全局平均池化操作，得到两个 $1 \times 1 \times C$ 的特征图。然后，将这两个特征图分别送入一个共享的双层神经网络（MLP）中，其中第一层的神经元数量为 C/r （ r 为减少率），使用 $ReLU$ 激活函数，第二层的神经元数量为 C 。接下来，对 MLP 输出的特征进行逐元素加和操作，然后经过 Sigmoid 激活函数，生成最终的通道注意力特征。最后，将通道注意力特征与输入特征图 F 进行逐元素乘法操作，生成空间注意力模块所需的输入特征。CAM 模块如图 7 所示。

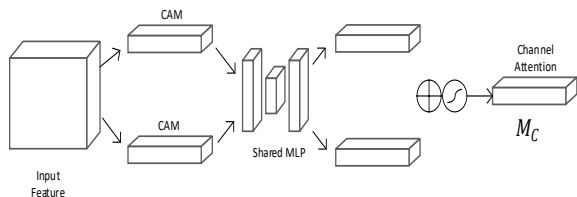


图7 CAM 模块图

SAM 模块关注的是目标的位置信息^[13]。首先，利用基于通道的全局最大池化和全局平均池化操作，对 CAM 模块中的 Channel attention 模块输出的特征图 F' 处理，分别得到两个 $H \times W \times 1$ 的特征图。然后，对这两个特征图按通道进行 concat 操作（通道拼接）。随后，经过一个 7×7 卷积操作（相对于 3×3 卷积， 7×7 卷积效果更佳），将特征维度降至 1 个 channel，即 $H \times W \times 1$ 。接着，经过 Sigmoid 函数生成空间注意力特征，即 M_s 。最后，将该特征与本模块的输入特征进行逐元素乘法操作，得到最终生成的特征。SAM 模块如图 8 所示。

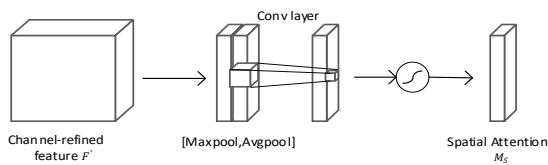


图8 SAM 模块图

2 实验与结果分析

2.1 数据集及实验环境

本文实验所用的手套数据集主要来自网络上收集的图片 and 现场真实拍摄图片，数据集中总共 3516 张图片，其中包含在正面和侧面条件下，单人未佩戴手套、单人佩戴手套、多人佩戴手套、多人佩戴手套和部分未佩戴手套、部分佩戴手

套和多人未佩戴手套等情况。通过 Labeling 对图片进行标记，将佩戴手套、未佩戴手套两种情况标记为 gloves、no_gloves 两个类别。数据集按 8:2 的比例划分为训练集和验证集，验证集用于评价训练过程中模型的状态和收敛性^[14]。

本文实验在 PyCharm 软件中进行，采用 PyTorch 作为深度学习框架对 YOLOv7 算法进行训练。具体环境是：Windows10 操作系统，CPU 型号 i9-11900K，GPU 型号 NVIDIA GeForce RTX 3080ti，显存 12 GB，编程语言 Python，CUDA 版本为 12.2。实验选用 adam 优化器，训练轮次为 100 轮，图片尺寸 640×640 ，每次处理图片数量 16 张，学习率为 0.01，学习率动量 0.937。

2.2 评价指标

本实验模型从精确度（precision, P）、平均精度均值（mean average precision, mAP）、召回率（recall, R）三个指标评价。

精确度能够反映模型正确预测正样本精度的能力，值越接近 1，证明模型效果越好。精确度公式为：

$$P = \frac{TP}{TP + FP} \quad (2)$$

召回率能够反映模型正确预测正样本全度的能力，值越接近 1 越好。召回率公式为：

$$R = \frac{TP}{TP + FN} \quad (3)$$

式中：TP 为正确识别出佩戴手套目标数量，FP 为错误识别出佩戴手套目标数量，FN 为错误识别出未佩戴手套目标数量。

2.3 结果与分析

为验证本文基于 YOLOv7 改进算法的检测效果，在保证实验数据集及实验环境完全相同的情况下，在 YOLOv7 算法上分别加入 SPD 模块、CBAM 注意力机制、SPD 模块和 CBAM 注意力机制，并且与目前较为流行的目标检测算法进行对比实验。实验结果如表 1 所示。

表 1 实验结果

模型	precision	recall	mAP	权重
YOLOv5s	0.833	0.654	0.719	13.6
YOLOv7	0.916	0.684	0.822	71.3
YOLOv7-SPD	0.86	0.836	0.879	73.3
YOLOv7CBAM	0.919	0.723	0.861	71.6
YOLOv7-SC	0.923	0.788	0.865	73.6

根据表 1 中的实验结果看出，相比于 YOLOv5s，YOLOv7 的 precision 提高了 8.3%，recall 提高了 3%，mAP 提高了 10.3%。在 YOLOv7 基础上引入小目标检测模块 SPD，precision 降低了 5.6%，recall 提高了 15.2%，mAP 提高了 5.7%。在 YOLOv7 基础上引入 CBAM 注意力机制，precision 提高了 0.3，recall 提高了 3.9%，mAP 提高了 3.9%。同时，引入 SPD 模块和 CBAM 注意力机制，precision 提高了 0.7%，recall 提高

了 10.4%，mAP 提高了 4.3%。改进后模型 YOLOv7-SC 相比于 YOLOv5、YOLOv7 模型，其精确度、召回率、平均精度均值都有所提高，能够有效检测出手套佩戴情况。

为了检测算法在实际运用中的效果，拍摄了某航空发动机培训过程中的人员图像进行测试，各模型效果图如图 8~10 所示，图 8 为 YOLOv5s 检测结果，图 9 为原 YOLOv7 检测结果，图 10 为 YOLOv7-SC 检测结果。可以看到，在实际场景下，改进后的 YOLOv7-SC 模型能够更准确地检测出培训人员手套佩戴情况。本文改进的算法在测试中综合性能优异，能满足对培训人员手套佩戴情况进行快速高效检测的要求。

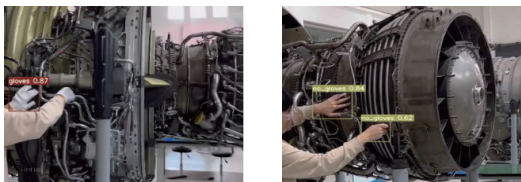


图 8 YOLOv5s 佩戴手套检测结果

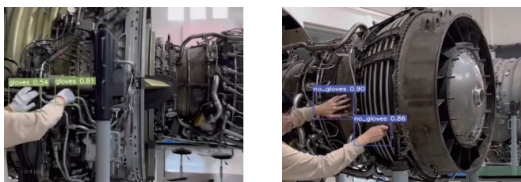


图 9 YOLOv7 佩戴手套检测结果

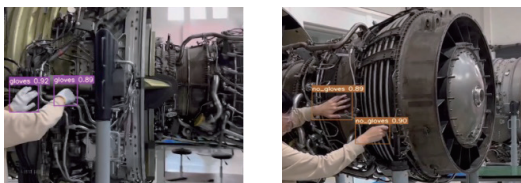


图 10 YOLOv7-SC 佩戴手套检测结果

3 总结

为保障维修培训人员的手部安全，本文提出基于 YOLOv7 改进的手套佩戴检测方法。首先在 YOLOv7 的 Head 中引入 SPDconv 模块，提高了模型对小目标和低分辨率目标的检测性能，再在 Backbone 引入 CBAM 注意力机制，提高模型提取特征的能力。本文改进后的 YOLOv7 手套佩戴检测算法精确率达到 92.3%，相较于其他主流算法模型，精确率、召回率、平均精度均值得到了较大的提升，同时改进后的算法能够满足在实际场景下正确检测手套佩戴情况的要求，能够有效保障维修人员在培训中的安全。但对于手部有一定遮挡以及某些特定角度下，本文改进的算法会存在检测效果不佳的情况，后续针对这一问题进一步改进算法，以提高检测手套佩戴效果。

参考文献：

- [1] 杨莹. 虚拟现实技术在飞机维修人员培养中的应用探讨[J]. 中国航务周刊, 2022(31):61-63.
- [2] 任大丛. 面向工业现场的手套佩戴检测系统开发[D]. 南京: 南京理工大学, 2021.
- [3] 肖体刚, 蔡乐才, 高祥, 等. 改进 YOLOv3 的安全帽佩戴检测方法[J]. 计算机工程与应用, 2021,57(12):216-223.
- [4] 王彦生, 朱佳佳, 王紫仪, 等. 基于改进 YOLOv5 的电厂人员绝缘手套佩戴检测[J]. 计算机测量与控制, 2023, 31(11): 60-65+73.
- [5] 吴昊祺, 刘小军, 周倩文, 等. 基于改进 YOLOv7 的骑行头盔佩戴识别方法[J]. 自动化应用, 2023,64(21):1-4.
- [6] 聂壮壮, 汪军, 黄翔翔. 基于改进 YOLOv5 的护目镜佩戴检测算法[J/OL]. 重庆工商大学学报(自然科学版):1-9 [2024-02-22]. <http://kns.cnki.net/kcms/detail/50.1155.N.20240112.1019.002.html>.
- [7] 苗茹, 岳明, 周珂, 等. 基于改进 YOLOv7 的遥感图像小目标检测方法[J/OL]. 计算机工程与应用:1-14 [2024-02-22]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20240130.1559.016.html>.
- [8] WANG C, BOCHKOVSKIY A, LIAO H. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[EB/OL]. (2022-07-06)[2024-01-01]. <https://doi.org/10.48550/arXiv.2207.02696>.
- [9] 陈嘉靖, 李龙飞. 基于 YOLOv5 神经网络模型的违采视频检测[J]. 水利技术监督, 2021(8):61-63+119+124.
- [10] RAJA S, LUO T. No more strided convolutions or pooling: a new CNN building block for low-resolution images and small objects[C]//Machine Learning and Knowledge Discovery in Databases, part 3. Cham: Springer Nature Switzerland AG, 2023: 443-459.
- [11] WOO S H, PARK J C, LEE J Y, et al. CBAM: convolutional block attention module[C]//Computer Vision-ECCV 2018, p. VII. Berlin: Springer, 2018:3-19.
- [12] 张文锐, 刘佳. 基于改进 YOLOv7 的口罩佩戴检测算法[J]. 北京信息科技大学学报(自然科学版), 2024,39(1):55-60.
- [13] 王宸, 杨帅, 周林, 等. 基于自适应多尺度特征融合网络的金属齿轮端面缺陷检测方法研究[J]. 电子测量与仪器学报, 2023, 37(10):153-163.
- [14] 李胜利, 刘忆宁, 高谭芮. 基于改进 YOLOv7-tiny 的安全帽佩戴检测算法[J/OL]. 电子设计工程:1-7 [2024-04-30]. <http://kns.cnki.net/kcms/detail/61.1477.TN.20240111.1143.004.html>.

【作者简介】

敖良忠(1971—)，男，重庆人，教授，研究方向：航空发动机维护、状态监控和故障诊断。

黄浩宇(2000—)，男，四川成都人，硕士研究生，研究方向：航空发动机维修。

(收稿日期：2024-03-19)