基于改进 YOLOv5 的眼睛及瞳孔检测算法

韩慧妍[1,2,3*]，范鑫茹[1,2,3]
（1. 中北大学 计算机科学与技术学院，山西 太原 030051；
2. 机器视觉与虚拟现实山西省重点实验室，山西 太原 030051；
3. 山西省视觉信息处理及智能机器人工程研究中心，山西 太原 030051）

摘　要：针对眼睛图像易受光照干扰导致的眼睛部位和瞳孔部位检测不准确及误检漏检的问题，提出基于改进 YOLOv5 的眼睛及瞳孔检测算法。首先，进行图像预处理，对比了三种图像增强方法，决定运用效果较好的 CLAHE（限制对比度自适应直方图均衡化）方法进行图像增强，提高对比度；其次，在 YOLOv5 网络中引入 Swin Transformer 模块代替骨干网络的最后一个 C3 模块和三个预测头中的三个 C3 模块，提高网络的特征提取能力，提升眼睛部位的检测精度；最后，在 YOLOv5 网络中通过引入多尺度特征跨层融合机制的方法，增加两个目标预测头，降低网络对眼睛部位和瞳孔部位的漏检率。该文从 ELSE 标准数据集中的 Data setXVIII 中选取了受光照程度不同的眼睛数据集 2 400 张，其中，1 600 张为训练集，800 张为测试集。实验结果表明，改进后的 YOLOv5 网络能检测出眼睛整体部位及完整的瞳孔部位，检测置信度也较高，mAP 提高了 3.2 百分点，Recall 提高了 2.7 百分点，且具有较好的实时性。

关键词：眼睛及瞳孔检测；YOLOv5；CLAHE；Swin Transformer；多尺度特征跨层融合机制
中图分类号：TP391　　**文献标识码**：A　　**文章编号**：1673-629X(2024)04-0076-06
doi：10.20165/j.cnki.ISSN1673-629X.2024.0012

Eye and Pupil Detection Algorithm Based on Improved YOLOv5

HAN Hui-yan[1,2,3*]，FAN Xin-ru[1,2,3]
（1. School of Data Science and Technology，North University of China，Taiyuan 030051，China；
2. Shanxi Key Laboratory of Machine Vision and Virtual Reality，Taiyuan 030051，China；
3. Shanxi Province's Vision Information Processing and Intelligent Robot Engineering Research Center，
Taiyuan 030051，China）

Abstract：To address the issue of inaccurate and missed eye and pupil detection caused by the susceptibility of eye images to light interference，an improved YOLOv5 based eye and pupil detection algorithm is proposed. First of all，image pre-processing is carried out，and three image enhancement methods are compared. It is decided to use CLAHE（limited contrast Adaptive histogram equalization）method with good effect to enhance the image and improve the contrast；Secondly，the Swin Transformer module is introduced into YOLOv5 network to replace the last C3 module of the backbone network and three C3 modules in the three prediction heads，so as to improve the feature extraction ability of the network and improve the detection accuracy of eye parts；Finally，by introducing a multi-scale feature cross layer fusion mechanism in the YOLOv5 network，two target prediction heads are added to reduce the network´s missed detection rate for eye and pupil regions. This article selected 2 400 eye datasets with different levels of illumination from the Data setXVIII in the ELSE standard dataset，of which 1 600 were training sets and 800 were testing sets. The experimental results show that the improved YOLOv5 network can detect the entire part of the eye and the complete pupil，with a high detection confidence. The mAP has increased by 3.2 percentage points，the Recall has increased by 2.7 percentage points，and has good real-time performance.

Key words：eye part detection；YOLOv5；CLAHE；Swin Transformer；Multi scale feature cross layer fusion mechanism

收稿日期：2023-05-16　　**修回日期**：2023-09-19
基金项目：国家自然科学基金（62106238）；山西省科技重大专项计划“揭榜挂帅”项目（202201150401021）；山西省自然科学基金项目（202303021211153）；山西省科技成果转化引导专项（202104021301055）
作者简介：韩慧妍（1980-），女，博士，副教授，CCF 会员（48976M），通讯作者，研究方向为人工智能、虚拟现实；范鑫茹（1998-），女，硕士研究生，研究方向为计算机视觉。

0　引　言

深度学习是机器学习领域中一个新的研究方向,它被引入机器学习使其更接近于最初的目标——人工智能。它的最终目标是让机器能够像人一样具有分析学习能力。深度学习是一个复杂的机器学习算法,在语音和图像识别方面取得的效果远远超过先前相关技术[1]。深度学习在搜索技术、数据挖掘、机器学习以及其他相关领域都取得了很多成果,同时,解决了很多复杂的模式识别难题,使得人工智能相关技术取得了很大进步[2]。

基于深度学习的目标检测算法主要包含两类:双阶段检测算法和单阶段检测算法。双阶段算法主要由提取特征和分类构成。R-CNN 是双阶段算法的典型代表,主要包含三部分:生成独立类别的区域建议候选框、固定长度 CNN 特征提取、基于线性 SVM 的分类[3]。基于 R-CNN,许多越来越多性能更好的双阶段模型被提出[4],如 Fast R-CNN[5] 和 Faster R-CNN[6]。基于 Faster R-CNN,Cascade R-CNN[7] 通过级联多个分类器可以提高对象定位精度;然而,Sun 等人指出,即使面对非常密集的场景,区域提案也大多是多余的。他们提出的 Sparse R-CNN[8] 仅使用少量固定数量的区域提议和稀疏特征就实现了良好的性能。总的来说,双阶段算法的检测精度很好,但速度不能令人满意。

相比之下,单阶段算法直接预测端到端的边界框和类概率。它们比双阶段模型更快,计算成本更低,并且具有实时检测能力。例如,单镜头多盒检测器(SSD)[9] 使用全卷积网络(FCN)进行特征提取,并分别从浅层和高层特征图中检测小对象和大对象。然而,单阶段物体检测器存在类别不平衡问题,故其在检测精度上低于双阶段物体检测器。YOLOv4[10] 设计了基于 Darknet53 的跨阶段部分(CSP)结构,以形成主干网络,从而进一步减少计算工作量并增强梯度性能。此外,强大的 Scaled YOLOv4[11] 为工程应用提供了一系列线性缩放对象检测模型。YOLOv5 继承了上述所有优点[12]。该文使用 YOLOv5 作为检测眼睛图像的基准框架。虽然 YOLOv7 已经面世,但是它的运行速度较慢,为保持检测的实时性,该文使用精度已经足够高的 YOLOv5。

但是,YOLOv5 的四个版本(v5s,v5m,v51 和 v5x)之间的区别在于模型的深度和宽度设置。对于 YOLOv5s 和 YOLOv5m,卷积层相对较少,并且特征层的输出不能很好地提取目标特征[13]。对于 YOLOv5l 和 YOLOv5x,虽然可以通过堆叠更多卷积来提取更深的语义特征,提高模型检测速度,但堆叠卷积层将增加模型的复杂性,从而降低模型检测的速度[14]。YOLOv5 的网络结构简单,但使用卷积来提取特征会导致一些问题,如接收能力有限、特征提取能力差和不可靠的特征集成[15]。于是对 YOLOv5 进行改进,改进 YOLOv5 网络构架的思路如下:

(1)数据预处理:对训练集以及测试集所有的图像进行 CLAHE(限制对比度自适应直方图均衡化)处理,提高图像的黑白对比度,以方便后续的检测及定位。

(2)引入 Swin Transformer 模块代替骨干网络的最后一个 C3 模块和三个预测头中的三个 C3 模块,提高网络的特征提取能力,眼睛部位的检测精度得到提升。

(3)引入多尺度特征跨层融合(M-FCFN),特征融合的同时增加了两个预测头。改善眼睛整体部位的漏检问题,提高眼睛部位的检测精度。

1　眼睛及瞳孔检测方法

1.1　引入 Swin Transformer 的 YOLOv5 方法

结合 YOLOv5 和 Swin Transformer 模块,使新结构可以继承它们的优点,并保留全局和局部特征。此外,使用自注意机制来提高集成模型的检测精度,这种整合对被光照干扰的眼睛图像比较有用。在训练期间采用了 ELSE 数据集的预训练 YOLOv5,以提高网络的泛化能力。

1.1.1　基于移位窗口的自注意力

给定眼睛特征图 $X \in R^{H\times W\times C}$,经过线性投影和重塑操作后,眼睛特征图变为 $Q,K,V \in R^{N\times C}$ 来供给自我注意,其中 $N = H \times W$。自我注意的输出表示如公式 1 和公式 2 所示。

$$Z = AV \tag{1}$$

$$\boldsymbol{A} = \mathrm{softmax}(QK^{\mathrm{T}}) \tag{2}$$

其中,$\boldsymbol{A} \in R^{N\times N}$ 是表示眼睛特征图上所有元素与其他元素之间关系的关注矩阵。输出 Z 聚集全局信息。Transformer 的实际计算是并行进行的,其中输入是单独计算的,然后进行积分。它被称为多头自我关注(MSA)。然而,在实验中发现,由于 SA 的计算复杂性与二次图像大小成正比,Transformer 在处理眼睛图像时消耗了巨大的计算资源。

每个 Swin Transformer 编码器包含两个子层。第一个子层是窗口多头自我关注(W-MSA)。它以非重叠的方式将特征图划分为单独的窗口,然后在这些局部窗口中计算自我关注。W-MSA 就是在一个小窗口内进行 Transformer 的操作。用 W-MSA 而不直接用 MSA 是因为视觉任务本身有局部性,对眼睛图片来说,局部信息足以解决问题。另外一方面,省资源。对于要素地图 $X \in R^{H\times W\times C}$,局部窗口大小为 $m \times m$,基于

具有 $N = h \times w$ 个 patch tokens 的图像窗口的 MSA 模块和基于非重叠局部窗口的 W-MSA 模块的计算复杂度分别如公式 3 和公式 4 所示。

$$\Omega(\mathrm{MSA}) = 4HWC^2 + 2(HW)^2C \tag{3}$$

$$\Omega(W-\mathrm{MSA}) = 4HWC^2 + 2(HW)^2C \tag{4}$$

其中,MSA 关于 patch token 数 $h \times w$ 具有二次复杂度。W-MSA 则当 M 固定时,具有线性复杂度。巨大的 $h \times w$ 对全局自注意力计算而言是难以承受的,而基于窗口的自注意力(W-MSA)则具有良好的扩展性。由于窗口大小比图像小得多,计算复杂性显著降低。W-MSA 和 MLP 之间的残余连接被添加以抵消权重矩阵的梯度消失和退化。

1.1.2 基于 Swin Transformer 的 YOLOv5

基于 Swin Transformer 的 YOLOv5 如图 1 所示。

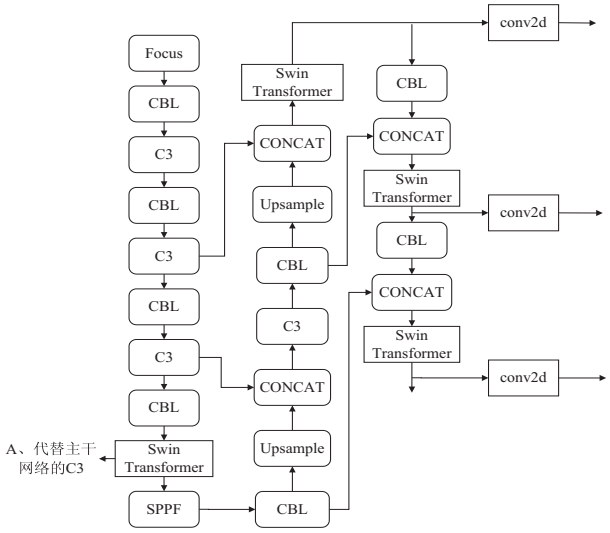


图 1 基于 Swin Transformer 的 YOLOv5

Swin Transformer 模块替换原始 YOLOv5 主干网络中的最后一个 C3 层。骨干网络末端的特征图分辨率为 20 × 20,在低分辨率特征图上应用 Swin Transformer 可以减少计算负载并节省存储空间。Swin Transformer 可用于捕获长距离依赖关系并保留不同的本地信息,并且可以提高眼睛部位的检测精度。

用 Swin Transformer 模块代替 YOLOv5 的三个预测头的三个 C3 模块。YOLOv5 网络中的 C3 模块擅长捕获本地信息而 Transformer 可以补偿全局建模能力,且 Swin Transformer 的移位窗口可以带来更高的效率。该架构具有在各种尺度上建模的灵活性,并且相对于图像大小具有线性计算复杂性。

1.2 引入多尺度特征跨层融合机制的 YOLOv5

引入 M-FCFN 的 YOLOv5 网络结构如图 2 所示。首先,从 PANet 结构中提取浅特征和深特征以进行跨层融合,并获得各种特征尺度作为输出。然后,为了提升眼睛部位检测的概率并显著提高眼睛部位的检测精度,该文提出了通过跨层融合获得的不同尺度特征进行降维,并将其作为预测的另一个输出。

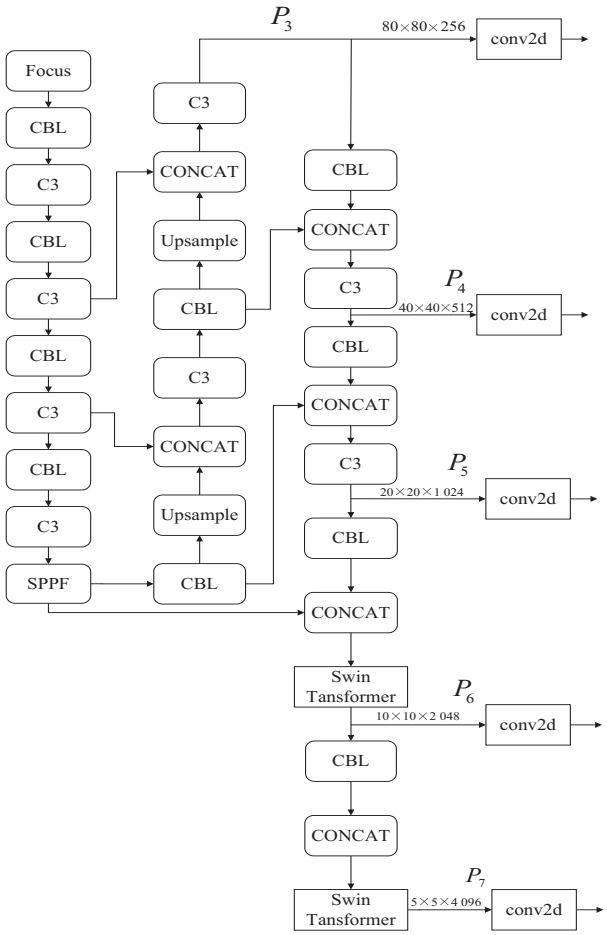


图 2 引入 M-FCFN 的 YOLOv5 网络结构

多尺度特征融合(M-FCFN):PANet 结构是 FPN 结构的优化,该网络模型中有五种特征融合。从网络中提取的中间信息标记为 P3,P4,P5,P6 和 P7,添加模块 P6 和 P7。

预测头:两次上采样和五个特征融合后的特征图为 P3,P4,P5,P6 和 P7。对应的尺度大小为 80×80,40×40,20×20,10×10 和 5×5。这五个不同尺度的特征图被用作检测的输出。这五个不同的尺度对应于网格的大小,该文为每个尺度设计了不同的锚点大小,以达到预测不同大小对象的目的。

改进如下:提取从 SPPF 层获得的特征信息,并将其与从 P5 获得的眼睛特征图像融合,以获得 10×10 大小的特征图像,并将该输出标记为 P6。在 P6 的结构上,添加了另一个 5×5 大小的特征图作为 P7。

网络的浅层包含许多关于眼睛图像的详细特征,但是随着网络结构的加深,信息将逐渐被抽象。而抽象将导致详细特征的丢失,并导致网络学习不到任何有用的东西。考虑到这一点,如果想以某种方式从图像中获得最有用的信息,那么网络的学习就必须是有效的。由于网络的不同层学习不同的特性,只有将这些不同的特征整合起来并再次学习,这个网络的学习

才是有效的。因此,该文从 PANet 结构中提取浅特征和深特征以进行跨层融合,并获得不同于 80×80,40×40 和 20×20 的尺度—10×10 和 5×5 的特征尺度作为输出,将 PANet 结构改进为 YOLOv5 中的 M-FCFN 结构,提取 SPPF 模块后的眼睛特征图,将其与从 P5 获得的特征图融合,并将其输出标记为 P6。基于 P6,建议再次压缩眼睛特征图,并将输出标记为 P7。最终的网络结构具有五个输出尺度。从检测结果和实验结果可以看出,该文的改进是有效的。

不同特征图的特征不同,特征图的比例分别为 80×80,40×40,20×20,10×10 和 5×5。80×80 的特征比例更适合于小目标检测。对于中型目标-瞳孔,40×40 和 20×20 的特征尺度更好,而对于大型目标-眼睛整体部位,10×10 和 5×5 的特征尺度更佳。通过 10×10 和 5×5 的两个特征尺度,网络可以增强对大目标-眼睛整体部位的学习,以提高网络的检测精度。

2　实验和讨论

2.1　数据集及预处理

该文选择标准数据集 ELSE 进行实验。并从数据集中的 Data setXVⅢ中选取了受光照程度不同的眼睛数据集 2 400 张,其中 1 600 张为训练集,800 张为测试集。训练类别包括眼睛整体部位和瞳孔部位。

实验平台为 CPU: Intel(R) Xeon(R) CPU E5-1603 v3 @ 2.80 GHz,GPU 为 GeForce GTX 1070。采用 Pytorch1.7 框架搭建的 YOLOv5 为基础,利用 Python3.7 运行训练程序。对眼睛整体部位和瞳孔部位进行识别,算法中涉及到的参数:Batch size 设置为 4,学习率设置为 0.000 1。

ELSE 存在受不同程度光照影响的情况,该文采用了线性变换、AHE、限制对比度的自适应直方图均衡化(CLAHE)进行预处理,对有光照干扰的眼睛图像进行处理。采用两种性能指标来评估图像的质量,包含 PSNR 和 SSIM。PSNR 是基于对应像素点间的误差计算,数值越大,失真越小。

$$\mathrm{MSE}=\frac{1}{H*W}\sum_{i=1}^{H}\sum_{j=1}^{W}\left(X(i,j)-Y(i,j)\right)^2 \tag{5}$$

$$\mathrm{PSNR}=10\lg\left(\frac{(2^n-1)^2}{\mathrm{MSE}}\right) \tag{6}$$

式中,MSE 表示图像均方误差;H,W 表示图像的宽,高。SSIM 可以对比出两张图像的相似度。计算公式为:

$$\mathrm{SSIM}=\frac{(2x_1x_2+C_1)(2y_{1,2}+C_2)}{(x_1^2+x_2^2+C_1)(y_1^2+y_2^2+C_2)} \tag{7}$$

式中,x_1,y_1 表示输入图像的均值,标准差;x_2,y_2 表示增强后图像的均值,标准差;$y_{1,2}$ 表示输入图像和增强后图像的协方差;C_1,C_2 为常数。SSIM 数值越大,表示输入原图的结构损失越小。

实验定量分析如表 1 所示。CLAHE 算法与线性变换和 AHE 相比,PNSR,SSIM 均最高,分别为 19.392 1,0.887 4。

表 1　对比实验定量分析

方法	PNSR	SSIM
线性变化	13.45	0.610 3
AHE	16.925	0.752 1
CLAHE	19.392	0.887 4

2.2　消融实验

该文选择召回率(R)和平均精度(mAP)作为评估指标。Precision:预测正确的眼睛部位数目占 YOLOv5 预测出来的(不管预测正确与否)所有的眼睛部位数目的比例。Recall:预测正确的眼睛部位数目占实际眼睛部位总数目的比例。P 和 R 可以由真阳性(TP)、假阳性(FP)、真阴性(TN)和假阴性(FN)定义,计算公式如公式 8 ~ 11 所示:

$$P=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{8}$$

$$R=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{9}$$

$$\mathrm{AP}=\sum_{i=1}^{n-1}(r_{i+1}-r_i)P(r_i+1) \tag{10}$$

$$\mathrm{mAP}=\frac{\sum_{i=1}^{k}\mathrm{AP}_i}{k} \tag{11}$$

其中,真阳性(TP)表示 YOLOv5 模型正确预测出眼睛部位的阳性样本,真阴性(TN)表示 YOLOv5 模型错误预测的眼睛部位的阴性样本,假阳性(FP)表示 YOLOv5 模型预测不正确的眼睛部位正样本,假阴性(FN)表示 YOLOv5 模型未正确预测的眼睛部位的负样本。

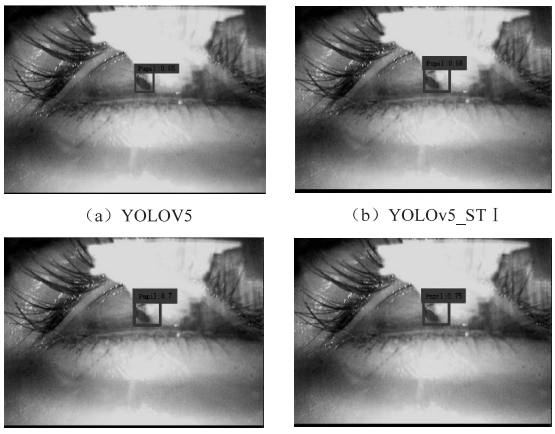
用 Swin Transformer 替换 C3 模块可以分为两部分,第一部分是替换主干网络中最后一个 C3 模块,第二部分为替换预测头中的 C3 模块,分别对第一部分、第二部分以及综合一二部分作对比实验。四个模型的实验定量对比结果如表 2 所示。Recall 方面,YOLOv5_STⅠⅡ相较于传统的 YOLOv5 提高了 1.7 百分点;mAP 方面,YOLOv5_STⅠ、YOLOv5_STⅡ与传统 YOLOv5 模型相比,分别提高了 2.1 百分点和 1.7 百分点,YOLOv5_STⅠⅡ变化最明显,提高了 2.7 百分点。

实验测试如图 3 所示,受光照影响较强的眼睛图像,YOLOv5 模型只能检测出瞳孔的一部分,YOLOv5_STⅠ,YOLOv5_STⅡ,YOLOv5_STⅠⅡ均可检测出

其完整瞳孔,并且 YOLOv5_ST Ⅰ Ⅱ模型置信度最高为 0.75,故 YOLOv5_ST Ⅰ Ⅱ检测受光照影响的瞳孔效果最好。

表 2 引入 Swin Transformer 的三种模型和原模型的性能比较

方法	mAP	Recall
YOLOv5	0.904	0.905
YOLOv5_ST Ⅰ	0.925	0.917
YOLOv5_ST Ⅱ	0.921	0.915
YOLOv5_ST Ⅰ Ⅱ	0.931	0.922



(a) YOLOV5 (b) YOLOv5_ST Ⅰ

(c) YOLOv5_ST Ⅱ (d) YOLOv5_ST Ⅰ Ⅱ

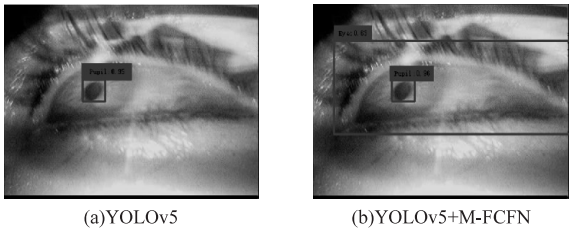
图 3 引入 Swin Transformer 的检测结果对比

引入 M-FCFN 的 YOLOv5 与原模型定量比较结果如表 3 所示,引入多尺度特征跨层融合机制的 YOLOv5 模型在 Recall 方面提升较为明显,提高了 2.3 百分点,从 0.905 提升到 0.928;mAP 从 0.904 提升到 0.922,提升了 1.8 百分点。

表 3 YOLOv5+M-FCFN 模型与 YOLOv5 模型的性能比较

方法	mAP	Recall
YOLOv5	0.904	0.905
YOLOv5+M-FCFN	0.922	0.928

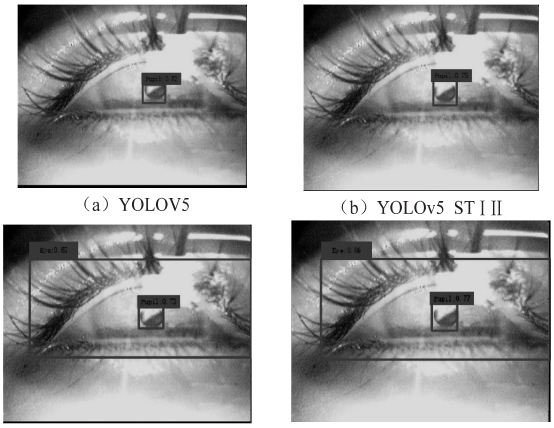
引入 M-FCFN 的 YOLOv5 与原模型的检测结果对比,如图 4 所示,YOLOv5+M-FCFN 可以检测出传统 YOLOv5 漏检的大目标-眼睛整体部位,大幅减少了大目标—眼睛整体部位的漏检率,并且瞳孔的置信度也有所提高。



(a)YOLOv5 (b)YOLOv5+M-FCFN

图 4 引入 M-FCFN 的实验检测结果对比

YOLOv5_ST Ⅰ Ⅱ+M-FCFN 模型、YOLOv5_ST Ⅰ Ⅱ模型、YOLOv5+M-FCFN 模型与 YOLOv5 的实验检测结果如图 5 所示。YOLOv5_ST Ⅰ Ⅱ模型、YOLOv5_ST Ⅰ Ⅱ+M-FCFN 模型均可以检测出完整的瞳孔检测框,但 YOLOv5_ST Ⅰ Ⅱ+M-FCFN 模型的瞳孔置信度最高,为 0.77。YOLOv5+M-FCFN 模型、YOLOv5_ST Ⅰ Ⅱ+M-FCFN 模型均可以检测出受光照影响较强的大目标-眼睛整体部位,但 YOLOv5_ST Ⅰ Ⅱ+M-FCFN 模型的眼睛整体部位置信度最高,为 0.66。综合所得,YOLOv5_ST Ⅰ Ⅱ+M-FCFN 模型检测效果最好。



(a) YOLOV5 (b) YOLOv5 ST Ⅰ Ⅱ

(c) YOLOv5+M-FCFN (d) YOLOv5_ST Ⅰ Ⅱ+M-FCFN

图 5 三种改进模型与 YOLOv5 的实验检测结果对比

四种模型的定量对比结果如表 4 所示,YOLOv5_ST Ⅰ Ⅱ+M-FCFN 模型的 mAP 最高,为 0.936,较原 YOLOv5 模型的 0.904,提高了 3.2 百分点;Recall 方面,YOLOv5_ST Ⅰ Ⅱ+M-FCFN 模型最高,为 0.932,较原 YOLOv5 模型的 0.905 提高了 2.7 百分点。

表 4 三种改进模型与 YOLOv5 性能比较

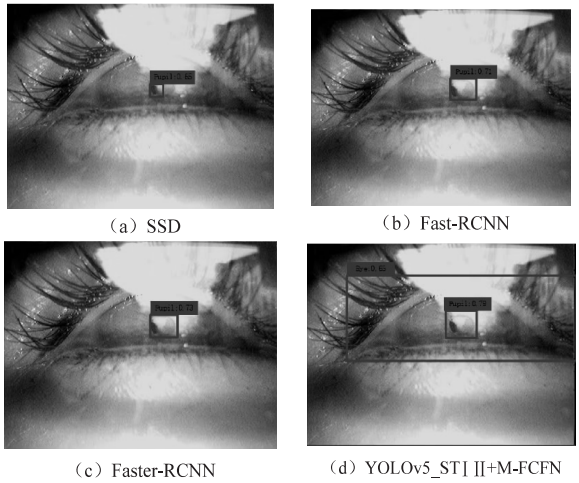
方法	mAP	Recall
YOLOv5	0.904	0.905
YOLOv5_ST Ⅰ Ⅱ	0.931	0.922
YOLOv5+M-FCFN	0.922	0.928
YOLOv5_ST Ⅰ Ⅱ+M-FCFN	0.936	0.932

2.3 与其他模型对比实验

算法在传统的 YOLOv5 网络基础上,通过用 CLAHE 进行图像增强、用 Swin transformer 模块替换 C3 模块、引入多尺度特征跨层机制的方式对 YOLOv5 进行改进,并与当前主流的目标检测网络(SSD,Fast-RCNN,Faster-RCNN)进行对比,使用平均精度均值(mAP)、召回率(Recall)来对比四种算法。

分别用 SSD,Fast-RCNN,Faster-RCNN 和改进后的 YOLOv5_ST Ⅰ Ⅱ+M-FCFN 模型对眼睛数据集进行检测,对比如图 6 所示。SSD 检测出来的瞳孔部分不完整,并且置信度较低,而 Fast-RCNN,Faster-RCNN 均能检测出完整瞳孔,置信度分别为 0.71,

0.73。三个网络均未检测出大目标-眼睛整体部位,而改进后的 YOLOv5_STⅠⅡ+M-FCFN 模型不仅可以检测出完整瞳孔,置信度最高,为 0.79,且可以检测出大目标-眼睛整体部位。



(a) SSD　(b) Fast-RCNN　(c) Faster-RCNN　(d) YOLOv5_STⅠⅡ+M-FCFN

图 6　YOLOv5_STⅠⅡ+M-FCFN 模型与其他模型检测对比

与其他模型的定量对比如表 5 所示。就 mAP 而言,SSD,Fast-RCNN,Faster-RCNN 分别为 0.83,0.862,0.884,YOLOv5_STⅠⅡ+M-FCFN 模型的 mAP 明显最高,为 0.936;就 Recall 而言,不同于 SSD,Fast-RCNN,Faster-RCNN 的 0.843,0.866,0.883,YOLOv5_STⅠⅡ+M-FCFN 模型性能最好,为 0.932。综合考虑,YOLOv5_STⅠⅡ+M-FCFN 模型明显更适合有光照影响的眼睛部位的检测。

表 5　YOLOv5_STⅠⅡ+M-FCFN 与其他模型性能对比

方法	mAP	Recall
SSD	0.83	0.843
Fast-RCNN	0.862	0.866
Faster-RCNN	0.884	0.883
YOLOv5_STⅠⅡ+M-FCFN	0.936	0.932

3　结束语

针对眼睛整体部位和瞳孔部位的检测问题,本文通过改进 YOLOv5 算法,提高了眼睛整体部位和瞳孔部位目标检测的检测精度。用 Swin Transformer 模块替换 YOLOv5 的主干网络和预测头的 C3 模块,提高眼睛部位的检测精确度。用多尺度特征跨层融合方法,促进图像的特征融合,添加两个检测头,降低了眼睛部位的漏检率。

实验结果表明,改进后的 YOLOv5 模型在眼睛部位的目标检测上,比传统的 YOLOv5 模型有更好的效果,可将其应用于眼动判断领域。

参考文献:

[1] 潘雪玲,李国和,郑艺峰. 面向深度网络的小样本学习综述[J/OL]. 计算机应用研究:1-10[2023-05-15].

[2] 游俊哲. ChatGPT 类生成式人工智能在科研场景中的应用风险与控制措施[J/OL]. 情报理与实践:1-11[2023-05-15].

[3] 崔　浩. 基于深度学习的场景分析和行为分类的家居内摔倒识别[D]. 南昌:南昌大学,2019.

[4] 王子琦,管振玉,朱轶昇,等. 基于改进级联 RCNN 的遥感图像目标检测[J]. 计算机工程与设计,2023,44(1):194-202.

[5] 朱德伟. 基于深度学习的钢轨扣件状态检测研究[D]. 南昌:华东交通大学,2022.

[6] 圣文顺,余熊峰,林佳燕,等. 融合注意力与特征金字塔的小尺度目标检测算法[J/OL]. 计算机工程:1-12[2023-05-15].

[7] CAI Z,VASCONCELOS N. Cascade R-CNN:high quality object detection and instance segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2019,43(5):1483-1498.

[8] SUN P,ZHANG R,JIANG Y,et al. Sparse R-CNN:end-to-end object detection with learnable proposals[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Nashville:IEEE,2021:14454-14463.

[9] LIU W,ANGUELOV D,ERHAN D,et al. Ssd:single shot multibox detector[C]//Computer vision - ECCV 2016:14th European conference. Amsterdam:Springer,2016:21-37.

[10] BOCHKOVSKIY A,WANG C Y,LIAO H Y M. Yolov4:optimal speed and accuracy of object detection[J]. arXiv:2004.10934,2020.

[11] WANG C Y,BOCHKOVSKIY A,LIAO H Y M. Scaled-yolov4:scaling cross stage partial network[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Nashville:IEEE,2021:13029-13038.

[12] ZHU X,LYU S,WANG X,et al. TPH-YOLOv5:improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//2021 IEEE/CVF international conference on computer vision workshops (ICCVW). Montreal:IEEE,2021:2778-2788.

[13] GONG H,MU T,LI Q,et al. Swin-transformer-enabled YOLOv5 with attention mechanism for small object detection on satellite images[J]. Remote Sensing,2022,14(12):2861.

[14] LU S,LIU X,HE Z,et al. Swin-transformer-YOLOv5 for real-time wine grape bunch detection[J]. Remote Sensing,2022,14(22):5853.

[15] QU Z,GAO L,WANG S,et al. An improved YOLOv5 method for large objects detection with multi-scale feature cross-layer fusion network[J]. Image and Vision Computing,2022,125:104518.