

基于改进 YOLOX-s 的安全帽检测^①



苏 鹏^{1,2}, 刘 美², 马思群¹

¹(大连交通大学 机车车辆工程学院, 大连 116028)

²(广东石油化工学院 自动化学院, 茂名 525000)

通信作者: 刘 美, E-mail: 879643579@qq.com

摘 要: 在施工现场中, 发生过许多高空坠落事故, 因此在施工现场佩戴安全帽是十分有必要的. 针对安全帽佩戴状况检测中遇到的小目标样本缺检、漏检的情况, 提出一种基于 YOLOX-s 的改进算法. 首先, 在 Neck 层引入主干特征提取网络中的 160×160 特征层进行特征融合, 并且增加了一个针对小目标的检测头; 其次, 采用 SIOU 损失函数计算损失值, 使得网络在训练过程中考虑的损失项更加全面; 并且采用 varifocal loss 函数来计算置信度损失值, 进一步改善训练过程中存在的正样本与困难样本不均衡的问题, 最后, 采用 CA (coordinate attention) 注意力机制来增强模型的特征表达能力. 实验结果表明, 通过对 Neck 层与检测层、损失函数的优化以及引入 CA 注意力机制, 使得网络在训练过程中收敛与回归性能更佳. 改进后的算法的 *mAP* 值为 95.57%, 相较于 YOLOv3 及原 YOLOX-s 算法在 *mAP* 值上分别提高了 17.11%、3.59%. 改进后的算法检测速度为 54.73 帧/s, 符合实时检测速度要求.

关键词: 安全帽检测; YOLOX-s; 小目标检测; SIOU 损失函数; varifocal loss 函数; 注意力机制

引用格式: 苏鹏, 刘美, 马思群. 基于改进 YOLOX-s 的安全帽检测. 计算机系统应用, 2023, 32(7): 145-154. <http://www.c-s-a.org.cn/1003-3254/9127.html>

Safety Helmet Detection Based on Improved YOLOX-s

SU Peng^{1,2}, LIU Mei², MA Si-Qun¹

¹(School of Locomotive and Vehicle Engineering, Dalian Jiaotong University, Dalian 116028, China)

²(School of Automation, Guangdong University of Petrochemical Technology, Maoming 525000, China)

Abstract: In construction sites, many high fall accidents have occurred, so it is necessary to wear helmets. An improved algorithm based on YOLOX-s is proposed to deal with missing and omitted detection of small target samples encountered in helmet-wearing condition detection. First, the 160×160 feature layer in the Neck layer is introduced in the backbone feature extraction network for feature fusion, and a detection head for small targets is added; second, the SIOU loss function is used to calculate the loss value, which makes the loss term considered in the training process of the network more comprehensive, and the varifocal loss function is used to calculate the loss value of the confidence level to further reduce the imbalance of the positive and difficult samples in the training process; finally, coordinate attention (CA) mechanism is used to enhance the feature representation of the model. The experimental results show that the optimization of the Neck layer, detection layer, and loss function and the introduction of the CA mechanism lead to better convergence and regression performance of the network during the training process. The *mAP* value of the improved algorithm is 95.57%, which is 17.11% and 3.59% higher than that of YOLOv3 and the original YOLOX-s algorithm, respectively. The detection speed of the improved algorithm is 54.73 frames/s, which meets the real-time detection speed requirement.

Key words: safety helmet detection; YOLOX-s; small target detection; SIOU loss function; varifocal loss function; attention mechanism

① 基金项目: 国家自然科学基金 (62073091); 广东省普通高校重点领域 (新一代信息技术) 专项 (2020ZDZX3042)

收稿时间: 2022-11-29; 修改时间: 2022-12-23; 采用时间: 2023-01-13; csa 在线出版时间: 2023-04-14

CNKI 网络首发时间: 2023-04-17

安全帽是各行各业安全生产工作者和高空作业人员必不可少的一种安全用具,在施工现场的安全管理中,施工作业人员佩戴安全防护器具可以提高作业过程中的安全性。安全帽呈半球形且表面较为光滑,当有物体砸落在安全帽上时,砸落物体的力不会集中在一点而是会向帽壳周围传递形成一个受力面,这样可以将一个点受到的力缓冲至一个面来承受,避免了冲击过程中的应力集中,从而降低了伤亡事故发生的概率。由于施工作业人员在作业时常常不佩戴安全帽,轻视佩戴安全帽的重要性,导致一些受伤事件。因此可以设计一种安全帽佩戴检测的方法去检测施工现场中人员是否佩戴安全帽,提高作业人员的安全性。传统的人工检查安全帽佩戴情况具有局限性,不能时时刻刻地关注到所有施工作业人员,易造成安全事件的发生。为了解决现有施工场地存在的这一安全管理局限性,出现许多安全帽自动监测的方法,对施工人员是否佩戴安全帽的实时检测具有重大的研究意义。

目前目标检测方法主要可以分为传统检测方法以及基于深度学习的检测方法。传统目标检测方法需要先手工提取对象特征然后再对提取到的特征进行分类处理,传统检测方法主要有 V-J 检测算法^[1]、HOG+SVM 检测算法^[2]、DPM 算法^[3]等,但是传统目标检测方法存在着时间复杂度高、窗口冗余以及手工设计的特征对于多样性的特征变换没有体现出很好的鲁棒性。为了解决以往检测方法存在着的问题,学者们尝试将深度学习的方法运用在目标检测上。基于深度学习的目标检测算法主要分为两类:一类是基于候选区域的两阶段目标检测算法,如 R-CNN^[4]、SPPNet^[5]、Fast R-CNN^[6]、Faster R-CNN 算法^[7]等;一类是基于回归的单阶段目标检测算法,如 SSD^[8]、YOLO 系列^[9-12]、RetinaNet 算法^[13]等。在基于深度学习的安全帽检测算法中,文献[14]中孙勇等人提出一种基于 YOLOv3 的安全帽检测方法,采用密集卷积网络(DenseNet)的方法处理低分辨率的特征层,在保持特征提取充足的情况下降低了计算复杂度。文献[15]中宋晓凤等人针对 YOLOv5s 原主干网对特征融合不充分的问题,将主干特征提取网络中的残差块替换成 ResNetBlock 结构中的残差块,以此提升 YOLOv5s 在细粒度上的融合能力。文献[16]中施辉等人采用图像金字塔结构获取不同尺度大小的特征图,用于位置和类别预测;在模型训练过程中改变输入图像的尺寸大小,提升模型对尺度

大小的鲁棒性,提高了安全帽检测准确率。以上学者通过将原主干特征提取网络替换为更适合当前目标检测任务的特征提取网络、对主干特征提取网络特征提取模块的加强以及在训练过程中的多尺度训练方法,均在一定程度上提升了模型检测性能。本文以密集小目标的检测为总体设计思路,针对在安全帽检测中的密集小目标样本总体存在的缺检、漏检现象提出一种基于改进 YOLOX-s 的安全帽检测方法。

新一代目标检测算法 YOLOX^[17]有着较好检测性能,其算法有着模型大小不同的发行版本如 YOLOX-x、YOLOX-l、YOLOX-m 以及 YOLOX-s。考虑到在施工环境下进行实时检测时需要兼顾检测精度与检测速度,因此本文使用轻量级网络结构的 YOLOX-s 作为基准模型对安全帽进行检测。针对网络对于密集小目标的检测效果不佳的情况,将特征提取网络中含有大量位置信息的浅层特征层经过送入加强特征提取网络与含有大量语义信息的深层特征层进行融合,并增加一个解耦头(decouple head)进行密集小目标检测;对定位损失函数和置信度损失函数进行改进,提升了模型对于密集小目标的检测性能。实验结果表明:本文改进后的算法对安全帽佩戴检测具有较高的检测精度,同时在检测速度上符合实时检测要求。

1 YOLOX 算法原理

YOLOX 算法^[17]是旷视科技在 2021 年提出的新一代 YOLO 系列目标检测算法,相较于在它之前发布的 YOLOv3、YOLOv4、YOLOv5 目标检测算法在性能方面有着显著提升。YOLOX 算法是基于 YOLOv3 进行改进的,YOLOX 算法原作者在对 YOLOv3 算法的检测头进行改进时发现 YOLOv3 原来的耦合检测头在网络结构设置上可能还可以再优化改进,于是进而采用了解耦头来替换原 YOLOv3 中的耦合检测头,改进后发现精度和速度均有所提升。YOLOX 同 YOLOv5 一样,在网络结构上设计了多个深度与宽度不同的版本,用户可根据自身对精度以及检测速度的要求来选择不同版本。本文综合精度以及检测速度两方面的考虑,选择 YOLOX-s 作为本文改进的基准算法,YOLOX-s 算法的深度和宽度分别为 YOLOX 算法的 0.33 和 0.5 倍。YOLOX-s 检测算法同 YOLOX 算法其他版本一样,算法结构由 CSPDarknet 主干特征提取网络、PANet 加强特征提取网络以及解耦头检测模块共同组成,原理图如图 1 所示。

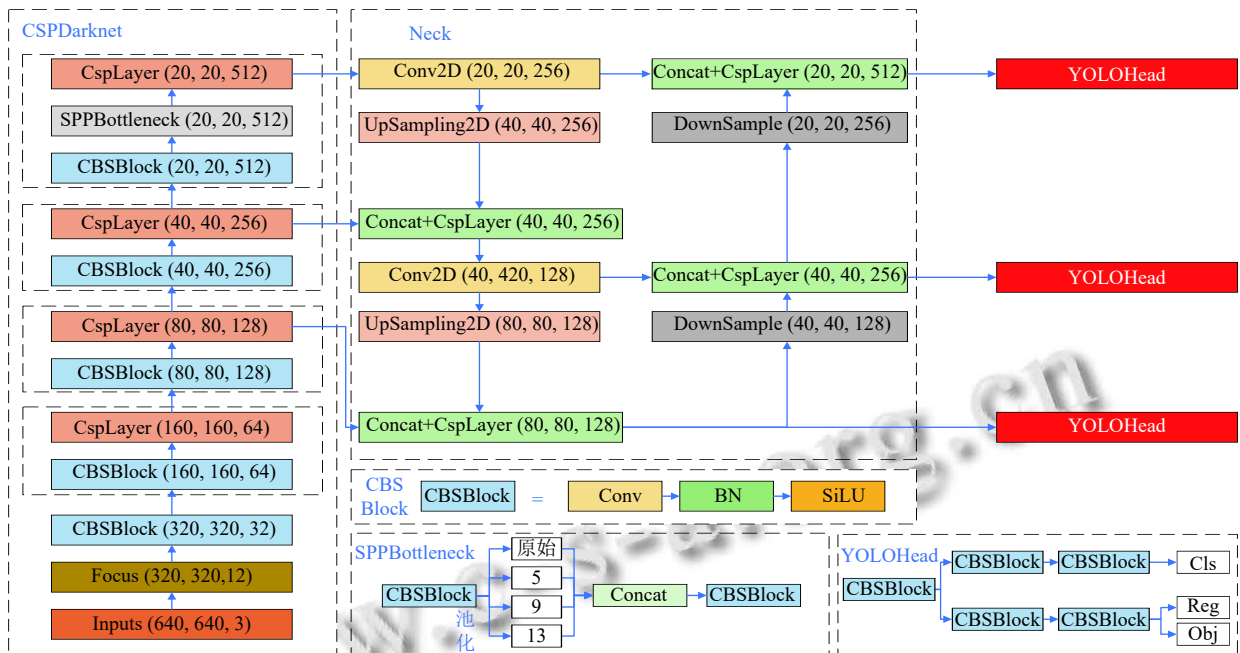


图1 YOLOX-s 算法原理图

1.1 YOLOX-s 算法主要流程

检测图片输入网络后, 首先通过图像预处理统一调整为 640×640 的大小, 之后输入 CSPDarknet 主干特征提取网络进行特征提取, 得到 3 个有效特征层, 有效特征层的大小分别为 $80 \times 80 \times 128$ 、 $40 \times 40 \times 256$ 以及 $20 \times 20 \times 512$. 有效特征层在特征加强网络中依次进行卷积调整通道数、上采样、特征拼接融合, 3 个特征层特征依次叠加融合后进行下采样再次进行特征融合, 加强特征提取能力. 加强特征提取层输出 3 层特征加强融合后的加强特征层, 将 3 层加强特征层分别送入检测小目标、中目标以及大目标的解耦头检测模块进行检测.

1.2 无锚框机制 (anchor-free)

YOLOX-s 虽是由 YOLOv3 改进而来的, 但是在锚框机制上却不同于 YOLOv3 以及由 YOLOv3 改进而来的 YOLOv4、YOLOv5 目标检测算法. YOLOX-s 考虑到有锚框机制 (anchor-based) 需要通过先验知识去手动设计并调整锚框数量及大小, 并且在检测过程中会在特征图上生成大量的锚框, 而这些生成的锚框大多数会在下一阶段通过非极大值抑制 (NMS) 被消除, 这样会造成大量的不必要的计算导致计算资源的浪费, 同时生成大量的锚框中只有少部分置信度大于 IoU 阈值的锚框内的样本会判定为正样本, 而其他置信度低

于 IoU 阈值的锚框则会被判定为负样本, 这样可能会导致网络在训练过程中形成正负样本不均衡的问题, 使得网络收敛性能变差. 采用无锚框机制 (anchor-free) 可以降低不必要的计算开销, 并且减少正负样本不均衡问题的出现, 无锚框机制并不代表是完全取消了锚框, 而是将原来对每个大、中、小网格的 3 个预测框减少为 1 个预先设置好的预测范围 (如 3×3 的网格范围) 如果样本中心点落在了该预测范围内并且置信度大于 IoU 阈值, 则判断该样本为正样本. 无锚框机制 (anchor-free) 减少了其余两个框的计算量以及负样本的数量, 有效地提高了网络在训练过程中的收敛性能.

1.3 解耦头

不同任务的评价体系大多数是不完全相同的, 正如 YOLO 系列目标检测中的分类任务以及定位任务一样, 这两种不同的任务的评价体系一般来说也是不完全相同的. 以往的 YOLO 系列目标检测算法同时也包括近些年的 YOLOv4、YOLOv5 目标检测算法, 采用的都是将分类任务和定位任务融合在一起的耦合头, 这样可能会导致两种不同的任务以及评价体系产生冲突, 从而有可能使得网络的收敛性能达不到最优. 采用解耦头可以使得网络在训练过程中的分类任务与定位任务分别按照各自的评价体系进行损失值计算, 有效

地缓解了任务之间发生冲突的可能性,但是很显然将分类与定位任务进行解耦后不可避免地会增加计算量. YOLOX-s 在权衡检测精度与检测速度后,最终选取了轻量级解耦头的方案(如图1中右下角的YOLOHead模块所示),即特征图由Neck层进入检测层时先经过 1×1 卷积统一调整通道数,而后依据任务的不同,分别设置分类任务通道以及定位任务通道,每个通道按顺序通过两层 3×3 卷积层提取特征后再经 1×1 卷积层调整通道数执行分类及定位任务. YOLOX-s 将耦合头替换成解耦头使得分类及定位两大任务依据各自的评价体系进行评价,使得任务之间发生冲突的可能性降低,提高了网络收敛性能.

2 YOLOX 算法改进

通过对原YOLOX-s算法网络训练完之后得出的模型进行数据分析,发现模型对于本文所选取的数据集的误检及漏检主要是集中在数据集的小目标正样本上,为了提升模型对于数据集中小目标样本的检测效果,本文对YOLOX-s算法改进主要为以下3个方面.

(1) 改进原网络结构,将包含众多位置信息的浅层特征层从主干特征提取网络中提取出来,并在加强特征提取网络中进行特征融合,从而加强网络模型对位置信息的特征提取;将检测层中将原来3个解耦头提升至4个解耦头,提升图像中的小目标样本检测性能.

(2) 改进IoU损失函数,将原IoU损失函数改为SIoU损失函数,SIoU损失函数^[18]考虑到了角度因素并重新定义了相关损失函数,提升了目标定位精度.

(3) 改进置信度损失函数,引入了varifocal loss函数^[19]均衡了正样本与困难样本之间的关系,使网络更加注重困难样本,提升了网络训练效果.

(4) 引入CA(coordinate attention)注意力机制^[20]至Neck层的特征图上采样操作之后,改善原特征层经过上采样后对特征的噪声放大和模糊等影响,增强了网络的特征表达能力,使得网络训练出的模型性能更佳.

2.1 网络结构改进

对原YOLOX-s算法网络训练完之后得出的模型进行推理检测,发现模型对于本文所选取的数据集的误检及漏检主要是集中在数据集的小目标正样本上,为了提升模型对于数据集中的小目标样本的检测效果,将包含众多位置信息的 160×160 浅层特征层从主干特征提取网络中提取出来,将 160×160 浅层特征层在加

强特征提取网络中进行特征融合,从而加强网络模型对位置信息的特征提取.由于本文所采取的数据集中包含有大量微小目标,于是在检测层中将原来3个解耦头提升至4个解耦头,新增加的解耦头输入特征层大小为 160×160 ,该解耦头的感受野可以更好地匹配上数据集中的小目标样本,使得图像中的小目标样本在检测时能够被注意到,而不会因为感受野过大而导致小目标在检测时被忽略,可以让模型对于图像中的小目标有着更好的检测性能.改进后的YOLOX-s原理图如图2所示.

2.2 损失函数改进

2.2.1 SIoU 损失函数改进

针对原定位损失函数在收敛速度上较慢,并且本文数据集中存在着众多小目标样本,在进行定位损失函数计算时,预测框与真实框常是包含与被包含的关系.如果采用原IoU损失函数作为边界框回归损失,那么在出现重叠时,原IoU损失函数并没有考虑到预测框与真实框之间的方向因素,没有能为边界框提供方向上的判断.因此将原IoU损失函数改为SIoU损失函数,SIoU损失函数考虑到了角度因素,重新定义了相关损失函数.损失函数组成包含以下4个部分.

(1) 角度损失,其定义如下:

$$\Lambda = \cos\left(2 \times \left(\arcsin\left(\frac{C_h}{\sigma}\right) - \frac{\pi}{4}\right)\right) \quad (1)$$

其中, σ 为预测框与真实框中心点的距离, C_h 为预测框与真实框中心点的高度差.当 $\frac{C_h}{\sigma}$ 即图3中的 α 的正弦值 $\sin(\alpha)$ 为1或0时,角度损失为0.

(2) 距离损失,如图4,其定义如下:

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma \rho_t}) = 2 - e^{-\gamma \rho_x} - e^{-\gamma \rho_y} \quad (2)$$

其中, ρ_x, ρ_y, γ 分别为:

$$\rho_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{c_w}\right)^2 \quad (3)$$

$$\rho_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{c_h}\right)^2 \quad (4)$$

$$\gamma = 2 - \Lambda \quad (5)$$

其中, $b_{c_x}^{gt}$ 与 $b_{c_y}^{gt}$ 分别为真实框中心点的横坐标与纵坐标, b_{c_x} 与 b_{c_y} 分别为预测框中心点的横坐标与纵坐标, c_w 和 c_h 分别为真实框与预测框最小的外接矩形的宽度与高度.

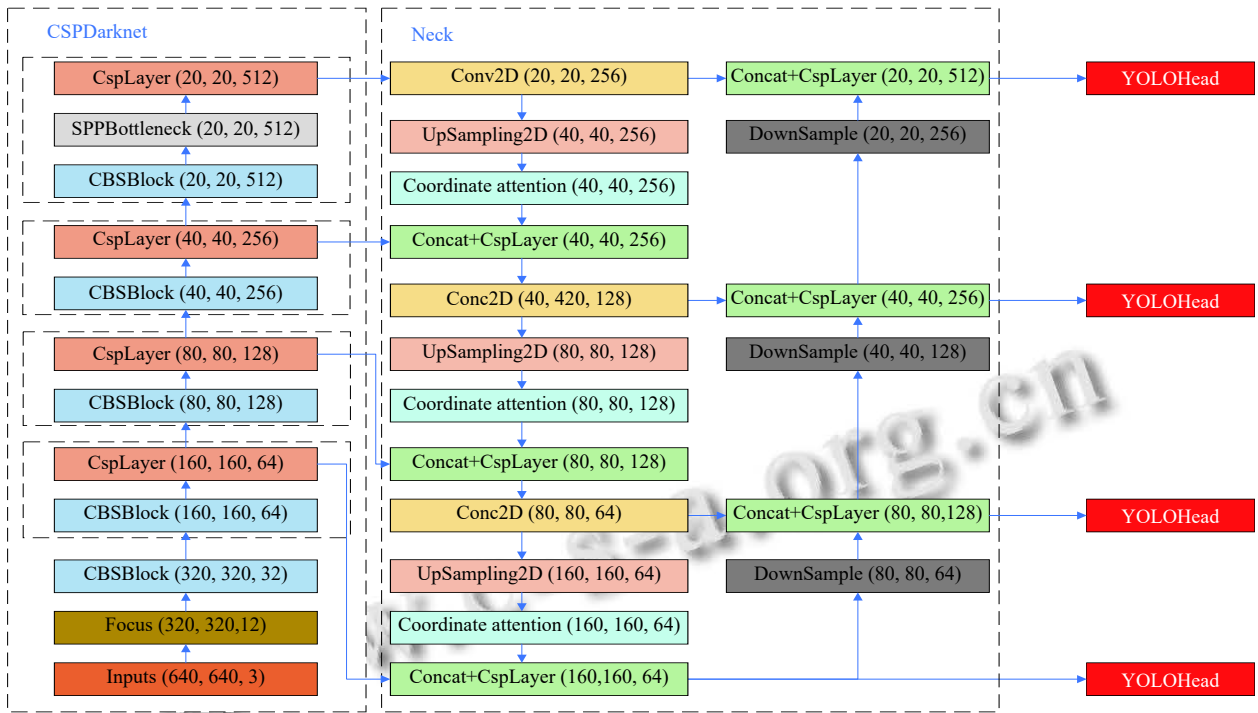


图2 改进后的YOLOX-s原理图

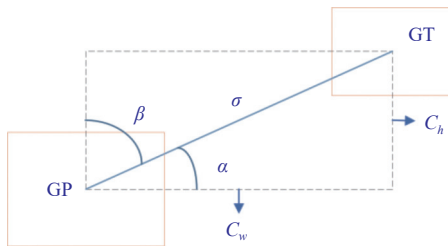


图3 角度损失图

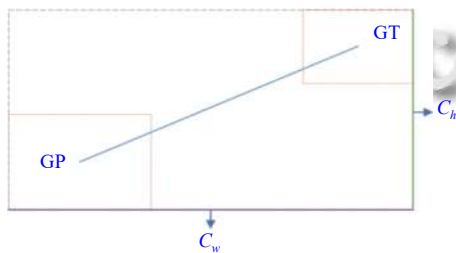


图4 距离损失图

(3) 形状损失, 如图5, 其定义如下:

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega_t})^\theta = (1 - e^{-\omega_w})^\theta + (1 - e^{-\omega_h})^\theta \quad (6)$$

其中,

$$\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})} \quad (7)$$

$$\omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \quad (8)$$

其中, w 和 h 分别为预测框的宽度及高度, w^{gt} 和 h^{gt} 分别为真实框的宽度及高度, 为了避免对于形状损失过度关注, 设置参数 θ 来调整形状损失在总损失计算中的占比.

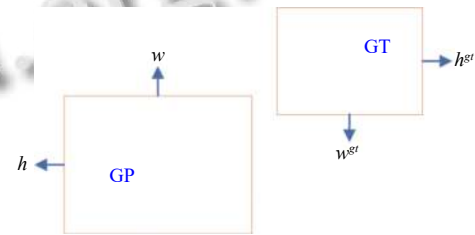


图5 形状损失图

(4) IoU 损失:

$$IoU = \frac{\text{交集}}{\text{并集}} \quad (9)$$

其中, 交集为图6(a)中蓝色区域, 并集为图6(b)中蓝色区域.

最终 SIOU 损失函数由以上 4 部分构成, 公式为:

$$Loss_{SIOU} = 1 - IoU + \frac{\Omega + \Delta}{2} \quad (10)$$

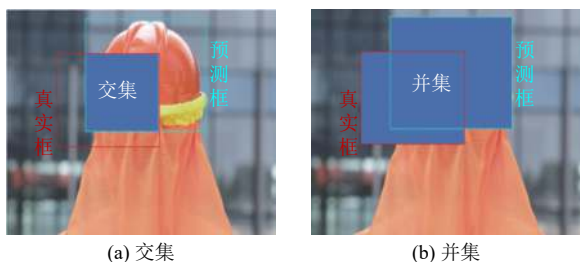


图6 IoU 损失图

2.2.2 Varifocal loss 置信度损失函数改进

因为本文数据集存在着目标密集且正负样本不均匀的问题, 于是引入了 varifocal loss 函数对置信度损失函数进行改进, 平等的对待正负样本, 防止在训练过程中出现类别不平衡的问题, 其定义如下:

$$VFL(p, q) = \begin{cases} -q(q \log(p) + (1 - q) \log(1 - p)), & q > 0 \\ -ap^\gamma \log(1 - p), & q = 0 \end{cases} \quad (11)$$

其中, p 是预测的分类得分, q 是目标分数, 对于一个正样本而言, q 为真实框与预测框的 IoU 值, 否则为 0, 而对于一个负样本而言, 目标分数 q 为 0.

Varifocal loss 函数通过参数 p^γ 来对损失进行缩放控制, 在缩放控制中不降低正样本的损失贡献度, 但会降低负样本的损失贡献度, 如果一个正样本的质量较好 (预测框与真实框拟合得较好), 那么在训练过程中该正样本对于损失的贡献度就更大一些, 使得在训练过程中更加关注高质量的正样本, 而不是在大量的简单正样本中收敛.

2.2.3 改进后的损失函数

本文设计的改进后的损失函数中的定位损失函数为 SIoU 损失函数, 置信度损失函数为 varifocal loss 函数, 分类损失函数采用 Bce loss 函数. 改进后各类损失函数以及总损失函数表达式为:

$$Loss_{IoU} = \sum_{i=1}^N \left(1 - IoU + \frac{\Omega + \Delta}{2} \right) \quad (12)$$

$$Loss_{obj} = \begin{cases} - \sum_{i=1}^N (q(q \log(p) + (1 - q) \log(1 - p))), & q > 0 \\ - \sum_{i=1}^N (ap^\gamma \log(1 - p)), & q = 0 \end{cases} \quad (13)$$

$$Loss_{cls} = - \sum_{i=1}^N (t_i \log(p_i) + (1 - t_i) \log(1 - p_i)) \quad (14)$$

$$Loss_{总} = \frac{1}{N} (Loss_{IoU} + Loss_{obj} + Loss_{cls}) \quad (15)$$

其中, N 为一个 Batchsize 的设置样本数量, $Loss_{IoU}$ 、 $Loss_{obj}$ 、 $Loss_{cls}$ 、 $Loss_{总}$ 分别为定位损失函数、置信度损失函数、分类损失函数以及训练过程中一个 Batchsize 的总损失函数, 通过求得的总损失函数来进行反向传播并更新权重.

2.3 CA 注意力机制的引入

CA 注意力机制不同于通道注意力机制将输入通过空间全局池化将转化为 $C \times 1 \times 1$ 的特征层, CA 是将通道注意力按照空间域的空间分布分解为两个分别沿着垂直方向与水平方向的方向聚合特征的 $C \times H \times 1$ 和 $C \times 1 \times W$ 特征层. 通过对空间域两个方向分别关注, 使其可以沿着一个空间方向获取长距离依赖, 并且保留另一个空间方向精确的位置信息. 然后, 将生成的特征图分别进行特征编码, 形成一对方向感知和位置敏感的注意力特征图. 最后, 将它们乘以输入特征图, 以此来增强感兴趣区域的表达. CA 注意力机制结构示意图如图 7 所示.

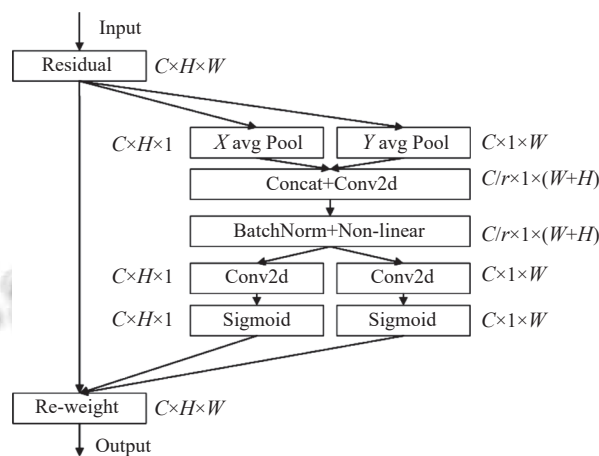


图7 CA 注意力机制结构示意图

CA 注意力机制作用的具体流程如下.

(1) 在空间域的垂直方向与水平方向上分别采用 $1 \times W$ 、 $H \times 1$ 大小的池化核进行平均池化, 将位置信息聚集到各空间方向上. 计算公式如式 (16)、式 (17) 所示:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (16)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (17)$$

其中, $z_c^h(h)$ 和 $z_c^w(w)$ 分别为垂直方向和水平方向位置信息聚合后的输出值。

(2) 在空间位置信息聚合得到输出值后, 将其中一空间方向张量的维度进行维度转换, 使其能与另一空间方向张量进行维度层面的拼接. 将拼接后的张量通过 1×1 卷积核大小的标准卷积模块 (2D 卷积层+Batch-Norm 层+非线性激活函数层) 进行特征提取. 计算公式如式 (18) 所示:

$$f = \delta(F_1[z^h, z^w]) \quad (18)$$

其中, δ 为非线性激活函数 (如 ReLU 激活函数、tanh 激活函数、LeakyReLU 激活函数等), F_1 为 2D 卷积层的卷积操作, $[z^h, z^w]$ 为两个不同空间方向的张量在同一维度进行拼接。

(3) 将经过 1×1 标准卷积模块特征提取后的特征层进行垂直方向和水平方向上的注意力权重重新分配. 计算公式如式 (19)、式 (20) 所示:

$$g^h = \sigma(F_h(f^h)) \quad (19)$$

$$g^w = \sigma(F_w(f^w)) \quad (20)$$

其中, g^h 、 g^w 为空间域上各方向的注意力权重矩阵。

(4) 将水平方向与垂直方向的注意力权重矩阵分别作用于原特征层, 加强原特征层对语义位置信息的表达能力. 计算公式如式 (21) 所示:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (21)$$

其中, $x_c(i, j)$ 、 $y_c(i, j)$ 分别为原特征层特征、注意力权重矩阵对原特征层特征进行加权分配后的具有更好的表达能力的特征层。

本文将 CA 注意力机制引入至 Neck 层的每个上采样操作之后, 将上采样之后的特征层执行 CA 注意力机制, 可以进一步减小上采样过程中可能造成的噪声放大和模糊等影响. 本文考虑到计算开销成本, 以及原网络的下采样操作采用的是卷积核为 3×3 , 步长为 2 的标准卷积模块, 原下采样操作对于特征的提取整合能力相较于上采样操作对于特征更具鲁棒性, 故只将 CA 注意力机制应用于每个下采样操作之后, 在不增加较多计算开销的前提下, 有效增强了网络的特征表达能力。

3 实验

3.1 实验数据集

本文实验所采用的数据集是从开源的 SHWD 数据集选取的 4 290 张图片, 该数据集涵盖了工地以及其

他多种场合下佩戴安全帽的情况, 并将该数据集的类别分为 hat (佩戴安全帽) 类以及 person (未佩戴安全帽) 类, 数据集格式为 Pascal VOC 格式, 标注文件为 XML 格式文件, 在正式开始训练前将数据集按照 8:1:1 的比例划分为训练集、验证集与测试集。

3.2 实验环境

本文所采用的实验环境为: Windows 11 系统, 显卡采用 NVIDIA GeForce RTX 3060 Laptop GPU, 深度学习框架选取 PyTorch. 在网络训练过程中使用 GPU 进行加速, 使得网络训练过程中效率提高, 实验所采用的环境如表 1 所示。

表 1 实验运行环境

配置	版本
CPU	12th Gen Intel(R) Core(TM) i7-12700H
系统	Windows 11
深度学习框架	PyTorch
GPU	NVIDIA GeForce RTX 3060 Laptop

3.3 模型训练

本文使用的训练集的类别与主流开源数据集如 COCO 数据集、VOC2007 数据集以及 VOC2012 数据集在类别上具有一定的相似性, 故采用迁移学习的方式来训练模型, 模型总共训练 100 个 epoch, 总迭代次数为 43 400 次. 网络在开始训练时加载预训练模型并在前 50 个 epoch 阶段进行冻结训练 (此时模型主干特征提取网络参数不发生改变), 在 50 到 100 个 epoch 阶段解冻恢复正常训练. 根据实验环境及数据集情况将网络参数设置为: 冻结训练阶段批次大小为 8, 解冻训练阶段批次大小为 4, 网络训练学习动量为 0.93, 训练时采用的优化器为 SGD 优化器, 网络训练过程如图 8 所示, 相较于改进前整体稳步收敛, 没有发生损失值突变以及过拟合现象。

3.4 评价指标

本实验采用的检测精度评价指标为精确率 (precision, P)、召回率 (recall, R)、平均精度 (average precision, AP) 以及评价精度均值 (mean average precision, mAP); 采用的检测速度评价指标为每秒传输帧数 (frame per second, FPS); 采用网络模型参数量 (Parameter) 来评价网络模型的大小。

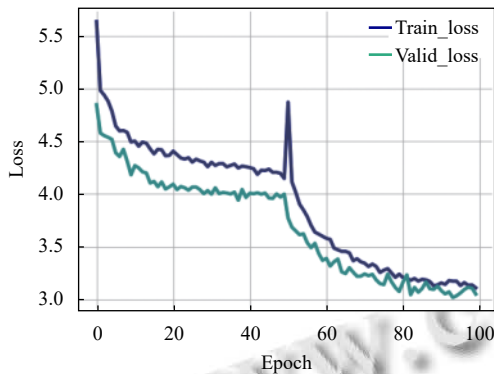
3.4.1 查全率 (R) 与查准率 (P)

查全率与查准率的公式分别为式 (22) 与式 (23):

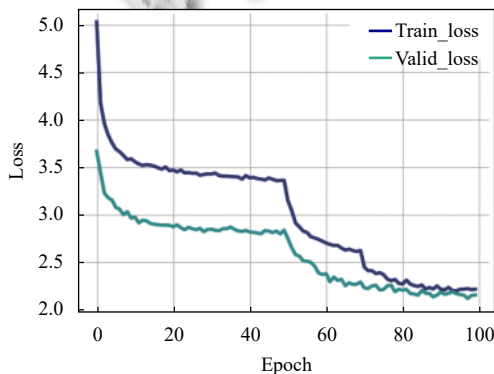
$$R = \frac{TP}{TP + FN} \times 100\% \quad (22)$$

$$P = \frac{TP}{TP+FP} \times 100\% \quad (23)$$

其中, TP 表示正确预测的正样本数量, FN 表示错误预测的负样本数量, FP 表示错误预测的正样本数量, $(TP+FN)$ 表示所有正样本的数量, $(TP+FP)$ 表示所有预测为正确的样本数量 (包含负样本).



(a) 改进前



(b) 改进后

图8 损失值图

3.4.2 平均精度 (AP) 与平均精度均值 (mAP)

平均精度与平均精度均值的公式分别为式 (24)

与式 (25):

$$AP = \int_0^1 P(R)dR \quad (24)$$

$$mAP = \frac{1}{N} \sum_{k=1}^N AP(k) \quad (25)$$

其中, AP 表示单个类别的平均检测精度, 而 mAP 表示的是全局的平均精度均值 (用来评价整体模型的检测精度), 模型检测时可以用 mAP 来衡量检测性能, 其 mAP 越高, 表示模型的检测效果更好.

3.5 实验结果及分析

3.5.1 对比试验

通过对本文改进后的算法进行验证, 并与其他主流检测算法进行性能比较, 对比试验如表 2 所示. 从表 2 中可以看到 Fast R-CNN、RetinaNet、SSD-VGG 算法的 mAP 值均小于 60% 且 person (未佩戴安全帽) 的检测 AP 值均较低不符合检测精度要求, 其中 SSD-VGG 的检测速度 (FPS) 较快, Fast R-CNN 的精度相对较高但检测速度最慢且参数量高达 136.71M. YOLOv3、YOLOv4 的检测精度均小于 80% 且参数量均较大, 不适合在硬件方面部署, 轻量级 YOLOv5-s、原 YOLOX-s 算法的参数量较小且在检测精度与检测速度上均有着良好的表现. 本文改进后的算法相较于主流算法在参数量上较小且检测精度较高, 且相较于原算法改进后的 YOLOX-s 算法 mAP 值提升了 3.59%, 其中 hat (佩戴安全帽) 类别的 AP 值提升了 1.87%, person (未佩戴安全帽) 类别的 AP 值提升了 5.31%, 检测速度为 54.73 帧/s, 能够在大多数情况下及时有效地检测出安全帽佩戴情况. 尽管在检测速度上有所降低, 但符合场景实时检测要求, 能够完成安全帽实时检测任务且适合在硬件方面落地部署.

表2 对比试验

检测算法	Parameters (M)	AP (%)		P (%)		R (%)		mAP (%)	FPS
		hat	person	hat	person	hat	person		
Fast R-CNN	136.71	89.68	20.76	74.70	29.44	89.18	36.12	55.22	11.17
RetinaNet	36.35	87.05	4.20	93.07	96.46	82.24	1.50	45.63	30.02
SSD-VGG	23.75	87.52	16.37	95.94	33.53	77.14	3.17	51.95	82.25
YOLOv3	61.52	91.73	65.19	94.87	70.86	83.06	64.99	78.46	57.28
YOLOv4	63.94	77.02	43.62	80.81	59.88	65.31	45.52	60.32	43.68
YOLOv5-s	7.06	93.96	79.31	94.13	93.26	88.37	62.29	86.63	83.54
YOLOX-s	8.93	95.93	88.03	94.64	92.30	90.00	81.50	91.98	74.48
改进YOLOX-s	9.71	97.80	93.34	93.28	92.54	93.47	89.25	95.57	54.73

3.5.2 消融实验

为了验证本文所采用的各个改进策略的有效性,

将各个改进策略分别加入原模型, 进行消融实验, 如表 3 所示. 可以看到加入 Neck 改进策略、SIoU 损失函数

改进策略、varifocal loss 函数改进策略以及 CA 注意力机制改进策略后, mAP 值总体稳步提升, 验证了每一步改进策略的有效性。

表3 消融实验

改进Neck	SIoU	Varifocal loss	CA	mAP (%)
—	—	—	—	91.98
√	—	—	—	95.04
√	√	—	—	95.28
√	—	√	—	95.31
√	√	√	—	95.42
√	√	√	√	95.57

注: √表示采用该改进策略

3.5.3 检测结果可视化

从检测结果图片可以很清晰地看到, 改进后的算法较改进前检测性能更好。原算法以往对于小目标的漏检以及误检问题在改进后有了很好地改善, 如图9、图10所示, 改进后的算法相较于原算法检测到的目标更多更准, 检测性能更佳。



图9 改进前后检测效果对比图



图10 改进前后检测效果对比图

4 结论与展望

本文提出一种基于YOLOX-s的改进算法, 算法参考多尺度检测的思想针对小目标在实际环境中难以检测的问题, 提出增加一个针对小目标的 160×160 的小目标检测头来提升小目标检测的效果, 并且通过优化损失函数以及引入CA注意力机制, 使得训练过程中网络收敛与回归性能更佳。通过实验可知, 本文改进后的算法在满足现场实时检测速度要求的情况下具有较

高的准确性, 且参数量较小, 在今后的研究中将继续尝试对模型剪枝、量化后移植进硬件后进行现场部署, 实现现场终端实时检测。

参考文献

- Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai: IEEE, 2001. [doi: 10.1109/CVPR.2001.990517.]
- Dalal N, Triggs B. Histograms of oriented gradients for human detection. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego: IEEE, 2005. 886–893. [doi: 10.1109/CVPR.2005.177]
- Felzenszwalb P, McAllester DA, Ramanan D. A discriminatively trained, multiscale, deformable part model. Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage: IEEE, 2008. 1–8. [doi: 10.1109/CVPR.2008.4587597]
- Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 580–587. [doi: 10.1109/CVPR.2014.81]
- He KM, Zhang XY, Ren SQ, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904–1916. [doi: 10.1109/TPAMI.2015.2389824]
- Girshick R. Fast R-CNN. Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1440–1448. [doi: 10.1109/ICCV.2015.169]
- Ren SQ, He KM, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2015. 91–99.
- Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37. [doi: 10.1007/978-3-319-46448-0_2]
- Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788. [doi: 10.1109/CVPR.2016.91]

- 10 Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6517–6525. [doi: [10.1109/CVPR.2017.690](https://doi.org/10.1109/CVPR.2017.690)]
- 11 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv:1804.02767, 2018.
- 12 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. Computer Vision and Pattern Recognition. arXiv:2004.10934, 2020.
- 13 Lin TY, Goyal P, Girshick R, *et al.* Focal loss for dense object detection. Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2999–3007.
- 14 张勇, 吴孔平, 高凯, 等. 基于改进型 YOLOv3 安全帽检测方法的研究. 计算机仿真, 2021, 38(5): 413–417. [doi: [10.3969/j.issn.1006-9348.2021.05.085](https://doi.org/10.3969/j.issn.1006-9348.2021.05.085)]
- 15 宋晓凤, 吴云军, 刘冰冰, 等. 改进 YOLOv5s 算法的安全帽佩戴检测. 计算机工程与应用, 2023, 59(2): 194–201. [doi: [10.3778/j.issn.1002-8331.2206-0468](https://doi.org/10.3778/j.issn.1002-8331.2206-0468)]
- 16 施辉, 陈先桥, 杨英. 改进 YOLOv3 的安全帽佩戴检测方法. 计算机工程与应用, 2019, 55(11): 213–220. [doi: [10.3778/j.issn.1002-8331.1811-0389](https://doi.org/10.3778/j.issn.1002-8331.1811-0389)]
- 17 Ge Z, Liu ST, Wang F, *et al.* YOLOX: Exceeding YOLO series in 2021. arXiv:2107.08430, 2021.
- 18 Gevorgyan Z. SIOU loss: More powerful learning for bounding box regression. arXiv:2202.12740, 2022.
- 19 Zhang HY, Wang Y, Dayoub F, *et al.* VarifocalNet: An IoU-aware dense object detector. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 8510–8519. [doi: [10.1109/CVPR46437.2021.00841](https://doi.org/10.1109/CVPR46437.2021.00841)]
- 20 Hou QB, Zhou DQ, Feng JS. Coordinate attention for efficient mobile network design. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021. 13708–13717.

(校对责编: 孙君艳)