

DOI:10.19651/j.cnki.emt.2211203

改进YOLOX的轻量级人体跌倒检测算法*

龙艺文¹ 闵宣霖¹ 陈奕兆¹ 罗欢¹ 刘洪² 易军¹

(1.重庆科技学院智能技术与工程学院 重庆 401331; 2.重庆科技学院安全工程学院 重庆 401331)

摘要: 针对边缘计算设备算力和存储空间有限的问题,提出了一种基于YOLOX改进的轻量级人体跌倒检测算法。首先采用GhostNet中的Ghost模块降低YOLOX中Neck和Prediction层的卷积参数冗余;其次,在Neck层中添加坐标注意力机制,增强关键信息提取能力,减少背景带来的噪音影响;最后,针对轻量级模型检测头检测能力不足问题,引入辅助头模块以加强轻量检测头的学习能力。通过算法检测性能以及在边缘计算端NVIDIA Jetson Xavier NX运行实验,结果显示,所提模型的mAP@0.5达到84.9%,且模型大小为25.6 MB。相较于YOLOX模型,仅以牺牲少量推理速度提升了4.6%的检测精度,减少了25.4%的模型大小,另外与一些主流目标检测算法相比,也具有一定的优越性。这些结果表明所提模型能够更好地满足边缘计算设备在人体跌倒检测中对轻量化和准确性的需求。

关键词: 跌倒检测;边缘计算;YOLOX;鬼影模块;坐标注意力;辅助头模块

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 520.20

Lightweight human fall detection algorithm of improved YOLOX

Long Yiwen¹ Min Xuanlin¹ Chen Yizhao¹ Luo Huan¹ Liu Hong² Yi Jun¹

(1. College of Intelligent Technology and Engineering, Chongqing University of Science and Technology, Chongqing 401331, China;

2. College of Safety Engineering, Chongqing University of Science and Technology, Chongqing 401331, China)

Abstract: Aiming at the problem of limited computing power and storage space of edge computing devices, a lightweight human fall detection algorithm based on YOLOX was proposed. Firstly, Ghost module in GhostNet is used to reduce the redundancy of convolution parameters in Neck and Prediction layers in YOLOX. Secondly, the coordinate attention mechanism is added to the Neck layer to enhance the ability of key information extraction and reduce the influence of background noise. Finally, aiming at the problem of insufficient detection ability of lightweight model detection head, an auxiliary head module is introduced to strengthen the learning ability of lightweight detection head. The performance of the proposed model is tested by the algorithm and the experimental results are run on the edge computing end of NVIDIA Jetson Xavier NX. The results show that the mAP@0.5 of the proposed model reaches 84.9%, and the model size is 25.6 MB. Compared with the YOLOX model, only a small amount of reasoning speed is sacrificed to improve the detection accuracy of 4.6% and reduce the model size of 25.4%. In addition, compared with some mainstream object detection algorithms, the proposed model also has certain advantages. These results show that the proposed model can better meet the requirements of lightweight and accuracy of the model for edge computing equipment in human fall detection.

Keywords: fall detection; edge computing; YOLOX; Ghost module; coordinate attention; auxiliary head module

0 引言

据统计,我国每年至少有两千万以上的老年人发生跌倒^[1],而跌倒已在全世界范围内成为影响老年人健康的主要原因,其会导致老年人在生理与心理上产生不同程度的

损伤,甚至会导致死亡^[2]。如何对老年人跌倒进行及时和准确的检测,成为目前社会广泛关注的问题。

当前对人体跌倒检测的方法主要分为3类:基于环境设备的检测方法^[3]、基于穿戴式传感器的检测方法^[4-5]以及基于计算机视觉识别的检测方法。前两类方法通过检测

收稿日期:2022-08-28

* 基金项目:油气生产安全与风险控制重庆市重点实验室开放基金(CQSRC202115)、重庆科技学院硕士研究生创新计划项目(YKJXC2120809)资助

人体加速度或跌倒产生的异常声波等信息识别跌倒行为,但这些监测方法存在着设备复杂、成本高昂和误报率较高等诸多问题。

随着深度学习技术的不断发展,基于计算机视觉的跌倒检测方法得到了越来越多的关注,尤其在 Krizhevsky 等^[6]提出 AlexNet 模型后,使用卷积神经网络(convolutional neural network, CNN)进行人体跌倒检测的方法得到广泛应用。例如 Wang 等^[7]基于 CNN 提出了一种线核卷积神经网络,用于提高跌倒检测算法的准确性; Panahi 等^[8]利用机器视觉技术,分析具备深度信息的图像实现人体跌倒检测,但它们无法同时兼顾准确性和实时性。而伴随着目标检测算法的相继提出,使用这类算法进行跌倒检测,成为了解决先前模型缺陷的方案之一,例如 Min 等^[9]提出了一种基于 Faster R-CNN 来检测检测人与家具之间的空间关系并追踪人体特征,根据关系和特征的变化从而判断人体是否从家具上跌倒; Wang 等^[10]使用 YOLOv3 算法来检测人体的跌倒; Cheng 等^[11]利用 YOLOv5 并改进其网络的损失函数和非极大值抑制,实现对人体跌倒的高效检测。综合文献[9-11]使用目标检测算法进行跌倒检测是目前较好的方法之一,因此本文也将使用目标检测方法,选取当前性能更优异的 YOLOX^[12]来实现人体跌倒检测。然而对于边缘计算设备来说,想直接应用 YOLOX 算法于设备中仍一定的问题。

边缘计算是继云计算后的一种新型计算模式,它把计算深入到用户和数据源的边缘,并提供数据缓存和处理的能力,具有低延迟、安全、节约成本、高可靠性 5 大优势,应用于医疗健康、工业生产、智慧城市等诸多领域^[13]。目前的边缘计算设备要求被移植模型具有体积小、计算复杂度低等特点,而常见的目标检测模型并不适用于边缘计算设备,因此模型轻量化就成为解决这一问题的关键手段之一。随之而来的各种轻量级检测模型相继被提出,例如 MobileNet^[14]、ShuffleNet^[15]和 GhostNet^[16]等,但这些轻量级模型很难在节省计算资源和检测精度方面达到平衡。

针对上述文献中所存在的问题,本文提出了一种基于 YOLOX 改进的轻量级目标检测模型,该模型在保证准确性和实时性的前提下,实现了轻量化。其主要改进如下:首先,使用鬼影模块(Ghost module)替换 YOLOX 网络中颈层和预测层的部分卷积模块,在小幅度降低模型精度的情况下减少模型参数量和计算复杂度。其次,为弥补因缩小模型体积而造成的检测精度损失,在颈层加入坐标注意力机制(coordinate attention module, CA)^[17],使模型更加关注重要特征。然后,为进一步提高模型的精度,在预测层添加辅助头模块(AuxHead module),增强轻量检测头的检测能力。最后,针对各种复杂场景搭建了基于 NVIDIA Jetson Xavier NX 人体跌倒检测平台,并在平台上进行模型移植与测试验证。实验结果表明,所提模型在提高检测精确度的情况下,压缩了模型体积,能够有效地应用于边缘

计算设备中。

1 YOLOX 检测算法

YOLOX 作为 YOLO 系列检测算法中的又一新作,在精度和复杂性方面都是一个非常不错的目标检测模型。YOLOX 采用了 YOLOv3^[18]的 Darknet53 作为基线,改进了 YOLOv4^[19]和 YOLOv5 的过度优化问题,其主要结构分为主干网络层(Backbone),颈层(Neck)以及预测层(Prediction),网络结构如图 1 所示。YOLOX 相较于先前的目标检测算法,主要的改进体现在:1)在标签分配策略中,使用了 SimOTA 技术为不同标签动态分配正样本数量,其优势在于不仅减少训练时间,而且避免额外的参数;2)锚框生成采用 Anchor Free 方式,不需要使用聚类算法获取先验框,其主要优点在于不使用先验框从而减少网络的参数量和浮点数;3)在 Prediction 层中,创新使用预测分支解耦作为检测头,该检测头被分为分类和回归两个部分,其分别负责获取类别信息和获取检测框信息以及置信度信息,最后在预测时进行整合,这种方式提高了模型的收敛速度,还优化了模型的性能。

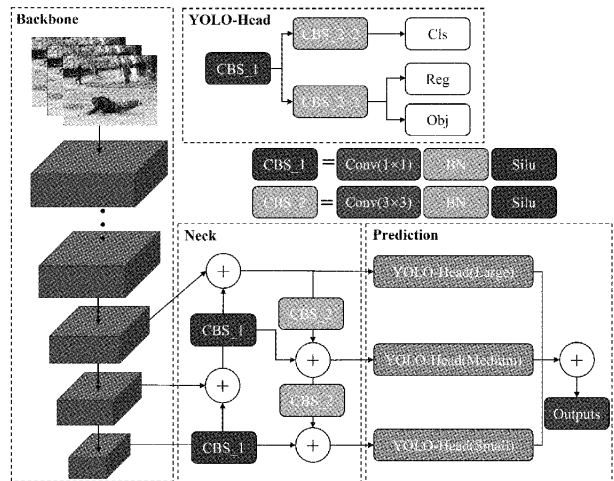


图 1 YOLOX 网络结构

2 改进 YOLOX 模型

本文为了满足边缘计算设备对算法模型的要求,同时达到更好的检测效果。因此本文对 YOLOX 进行了 3 个方面的改进,1)YOLOX 的 Neck 层和 Prediction 层中的卷积过于繁重,对此本文采用 Ghost 模块替换两层中的 3×3 卷积模块(CBS_2),减少参数冗余,缩小模型体积;2)考虑到卷积模块替换后特征提取能力下降导致模型精度降低的问题,本文在 Neck 层加入了 CA 注意力机制,其不仅能减少背景带来的噪音影响,加强对关键信息的提取能力,而且损耗的成本低,满足轻量级模型需求;3)为进一步实现模型轻量化,在 Prediction 层中采用了轻量检测头,但轻量检测头的最优标签匹配能力不足,导致模型检测能力低,对此本文

设计出了一种基于知识蒸馏技术的 AuxHead 模块,该模块主要进行辅助任务以加强轻量检测头的性能,在不损耗推理的情况下提升整体泛化性。改进 YOLOX 的网络结构如图 2 所示。

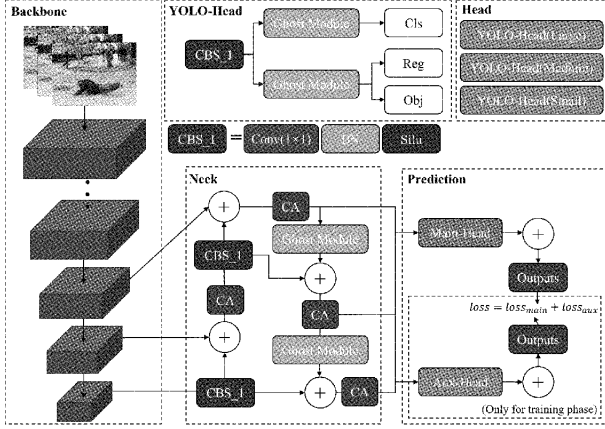


图 2 改进 YOLOX 网络结构

2.1 引入鬼影模块

本文为了优化 YOLOX 模型的参数,降低计算复杂度,采用了 Ghost 模块替换 Neck 和 Prediction 层中的 CBS_2 模块(CBS_2 模块是由普通卷积(Conv2D)、批量归一化(BatchNorm)以及 SiLU 激活函数组成)。Ghost 模块是由华为诺亚方舟实验室发表的端侧神经网络架构 GhostNet 网络所提出的轻量级模块,该模块利用了线性运算来扩充特征,增加通道数。本文将 Ghost 模块引用在 YOLOX 网络中,不仅可以使使用更少的参数来生成更多的特征图,而且在不改变输出特征图大小的条件下,所需的参数总量和浮点数均低于普通卷积。另外在维持相似识别性能的同时降低卷积层的计算成本,提升算力。

在图 3 中,Ghost 模块先通过 1×1 卷积模块,将输入的通道数减少,再通过廉价操作生成相似特征图,其目的在于用成本更低的方式保留有效信息。图中的 ϕ_i 即为廉价操作,使用的是深度可分离卷积(deep-width convolution, DW Conv)^[14]。最后,将 1×1 卷积模块得到的特征图通过恒等变换(Identity)与廉价操作生成的特征图拼接生成最终特征图。

在实际操作中,设输入特征图的尺寸是 $h_x \times w_x \times c$,输出特征图的尺寸是 $h_y \times w_y \times n$,普通卷积核大小为 $k \times k$,深度卷积核大小为 $d \times d$,则普通卷积的参数量 P_m 和计算量 F_m 分别为:

$$P_m = n \times c \times k \times k \quad (1)$$

$$F_m = n \times h_y \times w_y \times c \times k \times k \quad (2)$$

而在 Ghost 模块中,将输入特征图分为 s 份,则参数量 P_g 和计算量 F_g 分别为:

$$P_g = \frac{n}{s} \times c \times k \times k + (s-1) \times \frac{n}{s} \times d \times d \quad (3)$$

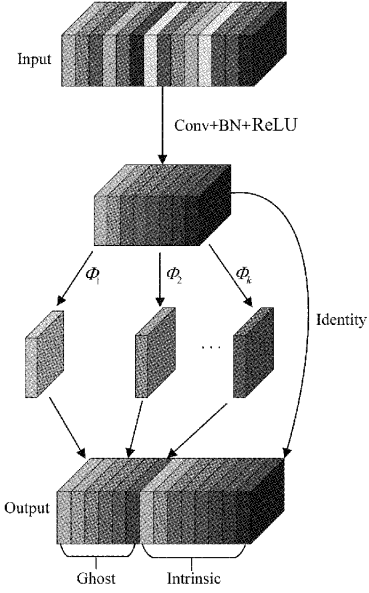


图 3 鬼影模块结构

$$F_g = \frac{n}{s} \times (h_y \times w_y) \times [c \times k \times k + (s-1) \times d \times d] \quad (4)$$

Ghost 模块与普通卷积操作的参数量和计算量的比值为:

$$R_p = \frac{P_m}{P_g} \approx s \quad (5)$$

$$R_f = \frac{F_m}{F_g} \approx s \quad (6)$$

由上述公式可知,采用 Ghost 模块相较于普通卷积可以减少 s 倍的计算量和缩减 s 倍的参数量。

2.2 嵌入坐标注意力机制

在 YOLOX 中,Neck 层使用了 PAFPN 结构,其融合了特征金字塔网络(feature pyramid networks,FPN)^[20]和路径聚合网络(path aggregation network,PAN)^[21],该结构通过自上而下路径和横向连接的方式,将分辨率低、语义强的特征和分辨率高、语义弱的特征相结合。PAFPN 用于构建所有尺度的高级语义特征,融合多尺度特征、扩大感受野;另外还可以缓解 Anchor Free 带来的正负样本极端不平衡,语义模糊性等问题,因此 PAFPN 对整个网络而言至关重要。

然而本文在实验中发现,使用 Ghost 模块替换 Neck 层的 CBS_2 模块会在一定程度上影响 PAFPN 对关键特征的提取能力,为了降低其带来的负面影响,本文将注意力机制嵌入到 Neck 层中来增强 PAFPN。由于各级特征融合的过程中容易造成特征混合,因此将注意力嵌入至如图 2 所示的位置,其优势在于该处添加注意力能够更有效增强各级特征对人体行为的关注度,筛选出融合特征的关键区域,并去除其中的背景噪音。

目前较主流的通道注意力机制(squeeze and excite,

SE)^[22]只考虑了内部通道信息而忽略了位置信息的重要性,仅能够小幅度提升网络的性能。而本文所使用的 CA 注意力机制不仅能跨通道捕获信息,还能捕获方向感知和位置感知的信息,这使得模型能够更准确的定位并识别到目标区域。该注意力机制的主要思想在于通过精确的位置信息对通道关系和长期依赖进行编码,分为坐标信息嵌入和坐标注意力生成两个步骤,其结构如图 4 所示。

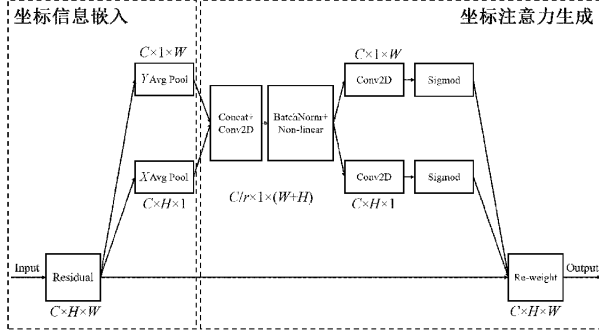


图 4 坐标注意力结构

首先在坐标信息嵌入操作中,设输入尺寸为 (C, H, W) 特征图,利用尺寸为 $(H, 1)$ 和 $(1, W)$ 的池化核沿着水平坐标(X Avg Pool)方向和竖直坐标(Y Avg Pool)方向分别做平均池化来对每个通道进行编码,然后将这两个变换沿着两个空间方向进行特征聚合,得到一对尺寸分别为 $C \times H \times 1$ 和 $C \times 1 \times W$ 的感知注意力特征图 z^H 和 z^W 。然后在坐标注意力生成操作中,先拼接坐标信息嵌入模块生成的两个特征图,接着使用 1×1 卷积(Conv2D)进行通道变换,表达式如下:

$$f = \delta[F_1(\text{cat}(z^H, z^W))] \quad (7)$$

在式(7)中, F_1 表示 1×1 卷积, δ 表示非线性激活函数,最后得到的 $f \in \mathbf{R}^{(\frac{C}{r} \times H \times W)}$ 为中间特征图,其含有水平方向和竖直方向的空间信息, r 表示下采样比例。将得到的 f 沿空间维度进行拆分生成张量 f^h 和 f^w ,使用 1×1 卷积将特征图 $f^h \in \mathbf{R}^{(C \times H \times W)}$ 和 $f^w \in \mathbf{R}^{(C \times H \times W)}$ 变换到和输入特征图相同的通道数,再 sigmoid 函数进行激活,得到分别利用于输入特征图的水平 and 竖直方向的 g^h 和 g^w 注意力权重,最后可表示为:

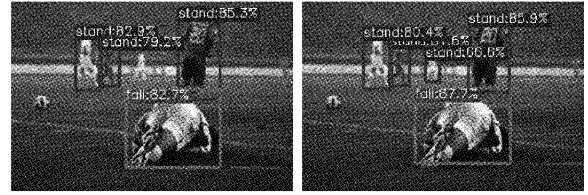
$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (8)$$

$x_c(i, j)$ 是输入特征图在通道 c 一个高度为 i , 宽度为 j 的特征, $y_c(i, j)$ 是经过注意力模块得到的输出结果。

使用 CA 注意力机制前后检测效果对比如图 5 所示,从图中可以看出,原始 YOLOX 检测图 5(a)中出现了行为漏检,而在使用 CA 注意力机制后检测图 5(b)中准确的检测出了该行为,由此可见 CA 注意力机制能够提高模型对人体行为的检测能力。

2.3 加入辅助头模块

YOLOX 的 Prediction 层分为解耦、锚框生成、标签匹配以及损失计算 4 个步骤,其中标签分配主要进行初步筛



(a) YOLOX 检测结果 (b) 添加 CA 检测结果

图 5 添加 CA 注意力机制效果对比

选和精细化筛选。初步筛选可根据中心点和目标框来判断。而精细化筛选使用的是 SimOTA 技术,该技术不仅减少训练时间,而且避免额外的参数。SimOTA 是在 OTA^[23]基础上提出的改进方法,其作用是针对不同标签动态分配正样本数量。SimOTA 首先会计算一个代价矩阵,代表每个真实框和网络解耦特征点之间的代价关系,该矩阵的计算公式为:

$$c_{ij} = L_{ij}^{cls} + \lambda L_{ij}^{reg} \quad (9)$$

式中: λ 为平衡系数, L_{ij}^{cls} 和 L_{ij}^{reg} 分别为真实框与预测框之间的分类损失和回归损失。代价矩阵的目的是为真实框动态分配候选框,其由 3 个部分组成:每个真实框和当前特征点的预测框的重合程度、每个真实框和当前特征点的预测框的分类精度、每个真实框中心的几何约束(落在特征点的一定范围内)。然后将代价矩阵的值按从小到大的顺序排列,每个真实框取 k 个最小代价的预测框作为正样本,其余的预测框被指定为负样本。最后对正负样本进行分类损失和回归损失计算,筛选出目标所对应的预测框。

本文在 Prediction 层中采用轻量检测头作为原检测头(MainHead),然而使用轻量检测头会出现最优标签匹配能力不足的问题,从而导致检测精度下降。对此本文借鉴 NanoDet-Plus 的训练辅助模块^[24],设计了一种可用 YOLOX 网络的 AuxHead 模块,该模块采用知识蒸馏技术,在网络训练时辅助 MainHead 缓解其能力不足的问题,并且在不损耗推理的情况下,加强了模型的整体泛化性。

图 6 中,输入的 3 个尺寸为 (80×80) 、 (40×40) 和 (20×20) 的特征图是经 PAFPN 提取所得。首先将特征图输入到 MainHead 和 AuxHead 中,MainHead 采用 YOLOX-S 网络的检测头作为轻量检测头,而 AuxHead 采用 YOLOX-L 网络的检测头作为重型检测头。重型检测头的参数量大,特征提取能力更强,因而在标签匹配时匹配效果更好,可以很好弥补 MainHead 在标签匹配上的不足。然后把 AuxHead 的检测结果进行 SimOTA 动态匹配正样本,得到的匹配结果应用于 MainHead 与 AuxHead 检测结果中,并分别计算损失。最后将两者的损失相加得到模型总损失,其表达式如下:

$$loss_{iou} = loss_{iou}^{main} + loss_{iou}^{aux} \quad (10)$$

$$loss_{obj} = loss_{obj}^{main} + loss_{obj}^{aux} \quad (11)$$

$$loss_{cls} = loss_{cls}^{main} + loss_{cls}^{aux} \quad (12)$$

$$loss_{total} = loss_{iou} + loss_{cls} + loss_{obj} \quad (13)$$

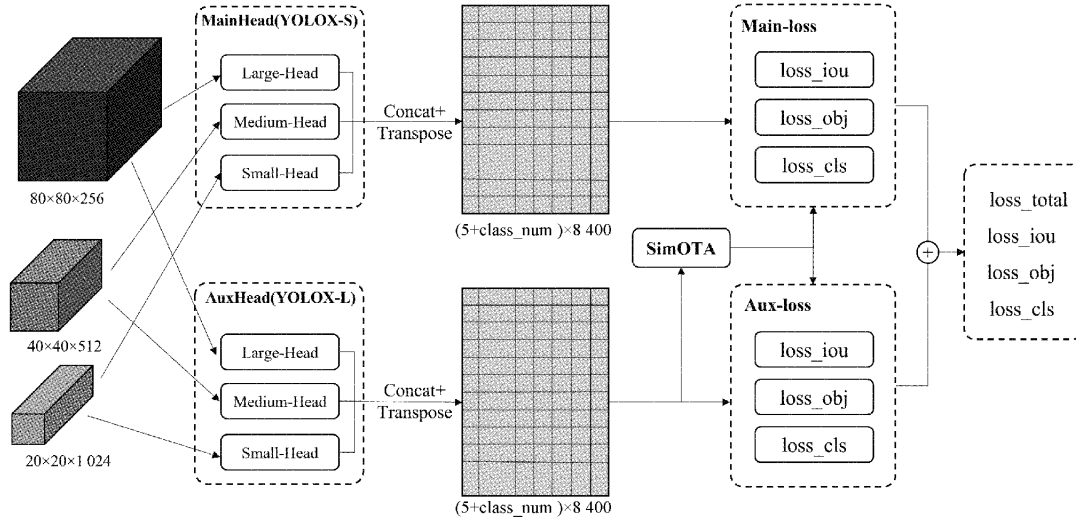


图6 辅助头模块结构

式中： $loss_{total}$ 是模型的总损失、 $loss_{iou}$ 是模型的检测框损失、 $loss_{obj}$ 是模型的置信度损失、 $loss_{cls}$ 是模型的分类损失。

本文所提出的辅助头模块属于在线知识蒸馏模式，该模式属于知识蒸馏^[25]的分支。知识蒸馏就是将重型网络作为教师，轻量网络作为学生，其主要操作是先让教师网络进行训练，然后把得到的知识用于学生网络的训练，这样的好处在于不仅保留了教师网络的性能，又有学生网络的轻量。而作为知识蒸馏分支的在线知识蒸馏模式和离线知识蒸馏模式，区别在于前者是一边训练教师网络一边训练学生网络，将教师网络训练出来的结果直接转移给学生网络训练；而后者是先训练教师网络，训练完毕后再选择最好的结果用于学生网络训练。两者本质差异在于在线知识蒸馏模式无需提前训练好一个教师网络，而是在学生模型学习时，实时进行“指导”，所以其训练模型消耗的资源远不小于离线知识蒸馏模式中的教师网络。此外，辅助头模块仅使用于模型的训练阶段，训练完则去掉辅助头模块分支进行推理，并不在测试阶段使用，因此辅助头模块不会增加模型的参数数量和计算复杂度，也不影响模型的推理速度。

3 实验结果与分析

3.1 实验数据集

本文的实验数据集来自于公开图像和自制图像一共有1046张如图7所示，其涵盖多种室内和室外场景。数据集的标签分为3类，其中跌倒(fall)类图像有350张，站立(stand)类图像有348张，下蹲(squat)类图像有348张。数据集通过LabelImg工具标注目标，标注完成后保存为xml文件，然后将xml文件转化为VOC2007的数据格式。实验将90%的图像用于模型训练，剩余10%数据用于验证模型性能。



图7 实验数据集

3.2 实验环境及参数设置

实验训练所使用操作系统为Window10，CPU为AMD Ryzen 5 5600X 6-Core Processor 3.70 GHz，GPU为NVIDIA RTX 3060 12 GB显卡，内存为16 GB，在消融实验和对比实验中均使用相同的硬件配置环境。由于修改了YOLOX网络结构的原因，无法使用迁移学习，所以模型参数都为随机初始化。训练轮次设置为300，批处理大小设置为8，动量设置为0.9，初始学习率为 1.5×10^{-4} 。此外辅助头模块在15轮后开启，250轮后关闭；马赛克数据增强在第275轮后关闭。训练完毕后，验证集的平均检测精度均值(mAP@0.5)如图8所示。本文实验会选取各轮次中损失最低和精度最高的一组权重文件进行对比。

3.3 消融实验及结果分析

本文通过消融实验来验证各改进模块的有效性，实验将比较各模块组合下，模型的平均检测精度均值(mAP@0.5和mAP@0.5:0.95)、参数数量以及模型大小。

表1为消融实验结果，表中Ghost代表用Ghost模块

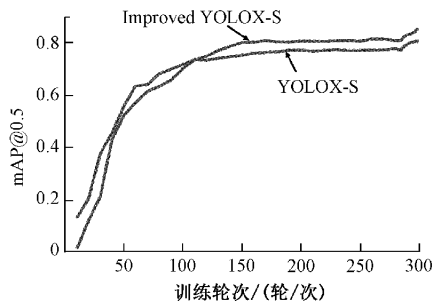


图 8 验证集 mAP@0.5

替换 Neck 和 Prediction 层中的 CBS_2 模块、CA 代表在

Neck 层中添加 CA 注意力机制、AuxHead 代表在 Prediction 层引用 AuxHead 模块。根据表中所示,改进后的 YOLOX-S 相比于 YOLOX-S, mAP@0.5 和 mAP@0.5:0.95 分别提高 4.6%、2.2%,而参数量和模型大小都分别减少 2.3 M、8.7 MB。此外对于 YOLOX-S+Ghost 模型而言,只添加 CA 注意力机制, mAP@0.5 和 mAP@0.5:0.95 会分别增加 0.9% 和 1.8%;只引用 AuxHead 模块也会分别增加 3.1% 和 4.3%;同时使用 CA 注意力机制和 AuxHead 模块则会分别增加 5.9% 和 4.6%,且这两处改进对模型的参数量和模型大小没有较大影响,综上所述结果表明 3 种改进方案有效。

表 1 改进 YOLOX 消融实验

| 算法模型 | mAP@0.5/% | mAP@0.5:0.95/% | 参数量/M | 模型大小/MB |
|--------------------------|-----------|----------------|-------|---------|
| YOLOX-S | 80.3 | 44.6 | 8.94 | 34.3 |
| YOLOX-S+Ghost | 79.0 | 42.2 | 6.59 | 25.4 |
| YOLOX-S+Ghost+CA | 79.9 | 44.0 | 6.64 | 25.6 |
| YOLOX-S+Ghost+AuxHead | 82.1 | 46.5 | 6.59 | 25.4 |
| YOLOX-S+Ghost+CA+AuxHead | 84.9 | 46.8 | 6.64 | 25.6 |

此外,为了验证本文所提改进点在实际检测中是否有效,本文选取了 5 张拥有复杂背景且包含跌倒、站立以及蹲坐 3 种行为的图片进行检测,检测结果如图 9 所示,图中各矩形框内标记了模型检测后的行为类别,框中的数字代表了检测框的置信度。置信度表示每个检测框的准确度,其计算的是真实框与预测框的重合程度,计算公式如式(14)所示。

$$C_{obj} = P_{obj} \times IOU \quad (14)$$

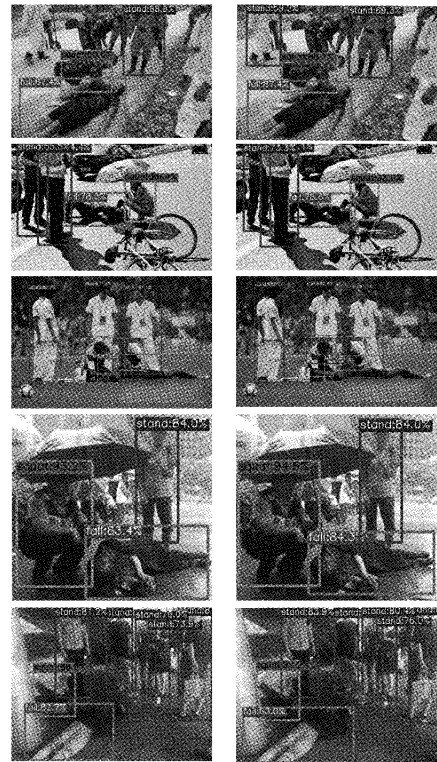
式中: C_{obj} 表示置信度, P_{obj} 表示检测到某类行为的概率, IOU 表示真实框与检测框的交并比。

对比图 9 的检测结果,在引用 CA 注意力机制和 AuxHead 模块之后的模型检测效果要优于原始 YOLOX。原始模型出现的漏检或误检行为,在改进后能够被准确检测出来,且在置信度方面表现更好。

3.4 算法对比

为进一步验证改进 YOLOX 算法的性能,本文采用相同数据集来训练当前主流的轻量级目标检测算法 YOLOX-S、YOLOv5-S、YOLOv4-Tiny 以及 YOLOX-Tiny,并进行对比实验。首先将各检测算法对每个行为类别的平均精度(AP)进行对比,实验结果如表 2 所示。另外对比各检测算法分别在 mAP@0.5、mAP@0.5:0.95、参数量、计算量和模型大小指标上的优劣,实验结果如表 3 所示。

由表 2 结果可知,在 3 类行为识别检测中,Improved YOLOX-S 的 AP 值均高于 YOLOX-S、YOLOX-Tiny 以及 YOLOv4-Tiny,尤其蹲坐行为的识别精度,AP 值分别比 YOLOX-S、YOLOv5-S、YOLOv4-Tiny 和 YOLOX-Tiny 高 7.3%、9.3%、26.1%、20.9%。虽然在跌倒行为和



(a) 原始 YOLOX (b) 改进 YOLOX

图 9 模型效果对比

站立行为 AP 值略低于 YOLOv5-S,但从检测的综合性能来看,效果要好于 YOLOv5-S。

表 3 中模型权重均以 32 位精度保存,实验结果表明,相较于表中的主流轻量算法,改进后的算法更加轻量,精

表2 行为类别的AP值对比

%

| 行为类别 | Improved YOLOX-S | YOLOX-S | YOLOv5-S | YOLOv4-Tiny | YOLOX-Tiny |
|------|------------------|---------|----------|-------------|------------|
| 跌倒 | 85.7 | 81.7 | 86.2 | 67.5 | 83.0 |
| 站立 | 82.6 | 80.0 | 86.8 | 65.3 | 66.4 |
| 下蹲 | 86.5 | 79.2 | 77.2 | 60.4 | 65.6 |

表3 改进YOLOX和主流轻量算法对比

| 算法模型 | 输入尺寸 | mAP@0.5/ % | mAP@0.5:0.95/ % | 参数量/ M | 计算量/ Gflops | 模型大小/ MB |
|---------------------|---------|---------------|--------------------|-----------|----------------|-------------|
| YOLOX-S | 640×640 | 80.3 | 44.6 | 8.94 | 26.64 | 34.3 |
| YOLOv5-S | 640×640 | 83.4 | 47.7 | 7.20 | 16.52 | 27.7 |
| Improved YOLOX-S | 640×640 | 84.9 | 46.8 | 6.64 | 16.54 | 25.6 |
| YOLOv4-Tiny | 416×416 | 64.4 | 34.4 | 6.05 | 16.43 | 22.5 |
| YOLOX-Tiny | 416×416 | 71.6 | 35.7 | 5.03 | 6.40 | 19.4 |
| Improved YOLOX-Tiny | 416×416 | 74.1 | 39.6 | 3.75 | 4.01 | 14.5 |

度表现更好。与YOLOv5-S相比,Improved YOLOX-S的mAP@0.5提高1.5%,且参数量和模型大小分别减少7.8%和7.6%。另外,对比YOLOv4-Tiny和YOLOX-Tiny,Improved YOLOX-Tiny的mAP@0.5分别比它们高9.7%和2.5%,同时参数量、计算量和模型大小分别只有3.75M、4.01Gflops、14.5MB,其远低于YOLOv4-Tiny和YOLOX-Tiny。因此综合各项指标,改进YOLOX确实拥有较好性能,占用更少存储资源和计算资源,优于当前主流轻量算法。

4 模型移植与验证

最后为进一步验证本文所提出的改进YOLOX模型在边缘计算设备上的实际效果,将在PC端上训练完成的

模型移植到高性能NVIDIA Jetson Xavier NX边缘计算设备上并连接网络摄像头进行实验。该边缘计算设备安装JPack4.6.1系统,系统镜像带有Ubuntu18.04操作系统、python3.6.9、TensorRT8深度学习推理框架等。

将模型移植到边缘计算设备上后,采用TensorRT加速模型检测。模型的mAP@0.5与检测速度对比如表4所示,从表中数据可以看出,Improved YOLOX-S和Improved YOLOX-Tiny模型移植到NVIDIA Jetson Xavier NX后FPS虽略低于YOLOX-S和YOLOX-Tiny,但mAP@0.5仍高于它们,且模型大小更小。总体而言,在边缘计算设备中,本文所提出的改进YOLOX模型占用存储空间更少,行为检测更加准确,而检测速度也能够满足设备的要求。

表4 在NVIDIA Jetson Xavier NX上的算法对比

| 算法模型 | mAP@0.5/% | 检测速度/(fps) | 模型大小/MB |
|---------------------|-----------|------------|---------|
| YOLOX-S | 80.3 | 67.8 | 34.3 |
| YOLOX-Tiny | 71.6 | 82.7 | 19.4 |
| Improved YOLOX-S | 84.9 | 62.5 | 25.6 |
| Improved YOLOX-Tiny | 74.1 | 79.3 | 14.5 |

5 结论

本文提出了一种改进的YOLOX跌倒检测模型,它不仅满足了边缘计算设备对模型轻量化和实时性的要求,而且保证了在复杂环境下的检测精度。所提模型在YOLOX的Neck和Prediction层采用Ghost模块替换CBS₂模块,减少参数冗余,完成了模型轻量化改造;其次,为弥补因缩小模型体积牺牲的精度,在Neck层中添加CA注意力机制,增强了PAFPN对关键特征的提取能力;此外,在Prediction层中引入AuxHead模块,缓解了轻量检测头最

优标签匹配能力不足的问题,检测精度也得到了进一步的提升。根据实验结果表明,本文所提出的改进YOLOX不仅比原始模型更加轻量,而且检测精度更高,能够更好的搭载于边缘计算设备中。在后续的工作中,一方面需要进一步提高模型的精度和检测速度,另一方面巩固提升模型的泛化性,将应用场景扩展到建筑工地、大型商场、旅游景点和医院等更多实际环境中,进行实时的跌倒检测。

参考文献

- [1] 师昉,李福亮,张思佳,等.中国老年跌倒研究的现状与对策[J].中国康复,2018,33(3):246-248.

- [2] 赵珍珍,董彦如,曹慧,等.老年人跌倒检测算法的研究现状[J].计算机工程与应用,2022,58(5):50-65.
- [3] MINVIELLE L, ATIQ M, SERRA R, et al. Fall detection using smart floor sensor and supervised learning [C]. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society(EMBC), IEEE, 2017: 3445-3448.
- [4] 彭亚平,贺乾格,柯希垚,等.一种基于加速度传感器的摔倒检测腰带[J].电子测量技术,2018,41(11):117-120.
- [5] 任晶秋,蒋杨,张光华.AHRS的老人跌倒检测算法[J].电子测量与仪器学报,2020,34(12):190-196.
- [6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [7] WANG B, GUO L, ZHANG H, et al. A millimetre-wave radar-based fall detection method using line kernel convolutional neural network[J]. IEEE Sensors Journal, 2020, 20(22): 13364-13370.
- [8] PANAHIL, GHODS V. Human fall detection using machine vision techniques on RGB-D images [J]. Biomedical Signal Processing and Control, 2018, 44: 146-153.
- [9] MIN W, CUI H, RAO H, et al. Detection of human falls on furniture using scene analysis based on deep learning and activity characteristics[J]. IEEE Access, 2018, 6: 9324-9335.
- [10] WANG X, JIA K. Human fall detection algorithm based on YOLOv3[C]. 2020 IEEE 5th International Conference on Image, Vision and Computing(ICIVC), IEEE, 2020: 50-54.
- [11] CHENG S C, ZHANG H K. YOLOv5-based fall detection method[J]. Academic Journal of Computing & Information Science,2022,5(4):29-34.
- [12] GE Z, LIU S, WANG F, et al. YOLOx: Exceeding YOLO series in 2021 [J]. ArXiv Preprint, 2021, ArXiv:2107.08430.
- [13] 赵明.边缘计算技术及应用综述[J].计算机科学,2020,47(S1):268-272,282.
- [14] HOWARD A, SANDLER M, CHEN B, et al. Searching for mobileNetv3 [C]. 2019 IEEE/CVF International Conference on Computer Vision(ICCV), October 27-November 2, New York: IEEE Press, 2019: 1314-1324.
- [15] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6848-6856.
- [16] HAN K, WANG Y, TIAN Q, et al. GhostNet: More features from cheap operations [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE,2020:1577-1586.
- [17] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [18] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. ArXiv Preprint, 2018, ArXiv:1804.02767.
- [19] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. ArXiv Preprint, 2020, ArXiv: 2004.10934.
- [20] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [21] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [22] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [23] GE Z, LIU S, LI Z, et al. Ota: Optimal transport assignment for object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 303-312.
- [24] RANGILYU. NanoDet-Plus: Super fast and high accuracy lightweight anchor-free object detection model[EB/OL]. (2021)[2022-5-7], <https://github.com/RangiLyu/nanodet>.
- [25] CHEN P, LIU S, ZHAO H, et al. Distilling knowledge via knowledge review[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 5008-5017.

作者简介

龙艺文,硕士研究生,主要研究方向为目标检测与边缘计算。

E-mail:longyi1998@qq.com

易军(通信作者),博士,教授,主要研究方向为目标跟踪、计算机视觉和机器学习。

E-mail:laoyifrcq@163.com