文章编号:1671-4598(2025)05-0053-09 DOI:10.16526/j.cnki.11-4762/tp.2025.05.007 中图分类号:TP183;TP391.4 文献标识码:A

用于导览机器人的轻量化行人姿态检测算法

徐启航[1]，李 迎[1]，刘雪凯[1]，孟祥印[1]，任永川[2]
(1. 西南交通大学 机械工程学院，成都 610031；
2. 民航成都物流技术有限公司，成都 611435)

摘要：针对搭载嵌入式设备的智能导览机器人算力较低，分析参观者行为困难的问题，提出一种基于轻量化YOLOv5的姿态检测算法；在YOLOv5的Backbone部分采用基于GhostNet设计的C3Ghost并在Bottleneck部分采用GSConvns和VoV-GSCSP设计细颈结构，在此基础上使用Slim算法和PAGCP算法分别对模型进行进一步压缩；基于YOLOv5提供的目标检测结果使用DeepSORT算法和SlowFast算法实现对视频中连续的人类目标进行跟踪和行人的姿态检测；经实验验证轻量化改进的YOLOv5其FLOPs下降71%，Parameters下降68%，配合行人姿态检测算法能够准确地识别人类目标并对行人的姿态进行分类。

关键词：导览机器人；模型轻量化；模型剪枝；目标检测；YOLOv5；目标跟踪；人体姿态检测

Lightweight Pedestrian Attitude Detection Algorithm for navigation Robots

XU Qihang[1], LI Ying[1], LIU Xuekai[1], MENG Xiangying[1], REN Yongchuan[2]
(1. School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China;
2. Civil Aviation Logistics Technology Co., Ltd., Chengdu 611435, China)

Abstract: Aiming at the problems of low computing power for intelligent navigation robots equipped with embedded devices and difficulty in analyzing visitor behavior, a posture detection algorithm based on lightweight YOLOv5 is proposed. The Backbone part of YOLOv5 adopts the C3Ghost based on the GhostNet, and in the Bottleneck part, the GSConvns and VoV-GSCSP are used to design a thin-neck structure. Based on this, the Slim and PAGCP algorithms are used to further compress the model, respectively. Based on the target detection results provided by YOLOv5, the DeepSORT and SlowFast algorithms are used to track continuous human targets in the video and detect pedestrian postures. Experimental results show that the lightweight improved YOLOv5 reduces the FLOPs by 71%, and the parameters are reduced by 68%. The detection algorithm of pedestrian postures can accurately recognize human targets and classify pedestrian postures.

Keywords: navigation robot; model lightweighting; model pruning; target detection; YOLOv5; target tracking; human posture detection

0 引言

随着我国经济和科技的发展，服务类机器人正在快速地进入到人们的日常生活中。其中，导览机器人是一类重要的服务类机器人。对于导览机器人而言，各类传感器模块作为决策模块、行为模块的前提和基础，对增强其为用户提供服务的能力起着至关重要的作用。在各类机器人所能获得的环境信息中，图像信息包含的内容是最为丰富且最有助于机器人理解环境现状的。然而，图像信息作为环境中的原始信息，需要进一步的加工和分析。在一系列的图像分析方法中，深度学习正在以其卓越的性能成为图像分析的主要方式。然而嵌入式设备有限的计算资源往往难以满足深度学习模型的需求。且很多的场景下服务机器人无法联网需要在本地的嵌入式系统中运行某些算法并执行预先设定的任务。因此为了既能够利用深度学习网络高质量的处理性能又能保证其能够在有限算力的嵌入式设备上运行有必要对深度学习模型进行轻量化改进以匹配低算力设备的需求。

在解决了嵌入式设备算力不足的问题后，还需要进

收稿日期:2024-03-11； **修回日期**:2024-05-15。

作者简介:徐启航(1998-),男,硕士研究生。

引用格式:徐启航,李 迎,刘雪凯,等.用于导览机器人的轻量化行人姿态检测算法[J].计算机测量与控制,2025,33(5):53-61.

一步考虑如何使机器人正确理解环境中用户的行为。当前大多数服务机器人在与人类进行交互时都是等待人类向机器人发起互动请求机器人提供响应式的服务。而为了服务机器人能够更好地融入到人类社会中去，其有必要具备准确判别用户参观意图并向其发起主动交互的能力。为了解决上述问题，需要机器人具备处理环境中连续时空信息的能力，即机器人要能够在一个连续的时间段内准确捕捉到当前的关键用户并能够对关键用户的行为进行正确的分析并加以理解。

在导览机器人的工作进程内，目标检测是实现其功能的基础。近年来，基于深度学习方法的目标检测取得重要突破。其中，YOLOv5 作为一种单阶段的目标检测网络，被大量地用于工程实践中。但随着手机、智能穿戴设备、各类服务机器人的普及，有限的计算资源对深度学习模型提出了轻量化的需求。于是，近年来许多工作集中于压缩深度学习模型，使其能够更匹配无法提供足够算力的场景。有学者提出设计专门的轻量化网络结构，如 ShuffleNet、MobileNet、ESPNet、EfficientNet 和 GhostNet[1-5] 等，这些网络结构相较于原本的深度学习网络结构都更加侧重于模型轻量化设计，从而减小深度学习网络对于计算资源的需求。其中文献［2］提出的 MobileNetV1 使用深度可分离卷积来构建轻量级网络；文献［6］提出的 MobileNetV2 是用了 inverted residual with linear bottleneck 单元，进一步提升网络性能；之后文献［7］在 MobileNetV3 又使用 AutoML 技术进行了更轻量级的网络构建。文献［5］使用其设计的 ghost 特征映射实现了很好的模型压缩效果。

轻量化网络结构设计是在某一训练任务开始前对模型进行调整，而更多情况下需要对已经训练好的模型进行剪枝处理。模型剪枝技术就是直接删除模型的部分权重，即剔除模型中不重要的权重，从而使模型的参数量减小。文献［8］发现神经网络的部分神经元处在完全未激活的状态，据此提出 ApoZ 指标来衡量神经元激活的百分比，删除未激活的神经元。文献［9］人提出每个通道后 BN 层的缩放因子即可以作为通道重要程度的重要判据，根据缩放因子选择需要剪枝的通道。文献［10］提出的基于层自适应幅度的剪枝（Layer-Adaptive Magnitude-based Pruning，LAMP）采用了一种评价性能感知全局通道修剪全局剪枝重要性评分，为每一层剪枝的比例提供参考。文献［11］提出的 PAGCP（Performance-Aware Global Channel Pruning）是一种全局通道剪枝方法，通过联合滤波器显著性来评价滤波器重要性为算法提供剪枝参考。

人体姿态通常蕴含了丰富的信息，导览机器人对人体姿态的分析有助于其更好地为人类提供服务。随着深度学习方法进入人体动作识别领域，研究人员提出了一系列高效的人体动作识别方法。文献［12］提出的 TSM（Temporal Shift Module）使用一个简单的模块来对输入的特征图进行时序上的移位，从而增强时序信息的传递，同时保持计算量和参数量不变。文献［13］提出的 SlowFast 方法中，研究人员基于三维卷积神经网络的方法，它使用两个并行的网络来分别处理视频中的慢速通道和快速通道，从而捕捉不同时间尺度的运动信息，同时使用一些跨通道的连接来增强信息的交互。文献［14］又提出了 Non-local 基于自注意力机制的方法，它使用一个非局部模块来计算视频中任意两个位置之间的相似度，从而捕捉长距离的时序依赖，同时可以嵌入到任何卷积神经网络中。

为了能够减小深度学习网络的算力需求，使其能够符合导览机器人大多数移动端设备的算力需求。本文提出了一种轻量化设计的 YOLOv5 目标检测网络，并将其与 DeepSORT 目标跟踪算法和 SlowFast 人体动作识别算法相结合，从而实现在一段连续的视频帧中准确识别并跟踪人类目标并对其进行动作分类，为导览机器人更好地分析人类行为信息提供帮助。

1 展厅环境数据集建立

1.1 数据采集

本文所设定的场景为一处室内展厅，主要识别目标是展柜和人类，展柜是展示机械工程相关知识和成果的设备，人类是展厅中的参观者或工作人员。数据集中应当包含有展柜的多角度图像信息以及丰富的人类目标图像信息。

对于展柜来说，展柜的主要结构是一块透明的玻璃柜板，而其识别的难点在于展柜的检测受到光线和拍摄角度的较大影响。为了确保后续模型能够充分学习目标特征，提升模型的泛化性能，本文特别注重在数据采集过程中保持多样性。

在数据采集时，使用了智能迎宾机器人搭载的深度相机对展柜图像进行拍摄。如图 1 所示，本文考虑了弱光条件、强光条件、角度变化导致的目标形状变化以及玻璃柜板反光等不利条件，同时还包括了部分含有人类目标的图像。通过这些不同条件的选择，最终成功采集到了展柜类原始图像共计 425 张。

这样的数据集设计不仅能够满足对展柜进行准确识别的需求，同时也为模型提供了更为全面和具有挑战性的训练样本，能够提高模型在实际场景中的鲁棒性和泛化性。

对于人类的信息来说，由于用户目标的姿态和穿着变化多端，仅仅依赖有限的数据可能无法实现对用户目

图 1　展柜图像部分展示

标的精确检测。因此，考虑从公开数据集中单独提取人类目标的图像和标注数据，将其与之前提到的展柜图像数据合并，形成一个更为全面的数据集，以用于后续的模型训练。

PASCAL VOC（Visual Object Classes）是一个用于目标检测和图像分割的重要计算机视觉数据集。最初由英国牛津大学的计算机视觉小组创建，PASCAL VOC 数据集旨在推动目标检测和图像分割任务的研究和发展。在计算机视觉领域，PASCAL VOC 数据集是最为著名且被广泛使用的数据集之一。

本文选择了 PASCAL VOC 2012 的最新版本。每个版本都包含有关 20 个不同类别的对象的图像及其相应的标注信息。在本文中，专门提取了以“person”为标签的部分数据，总计 4 139 条。为了保持两个目标数据量的平衡，避免由于某一目标数据量过大而导致模型学习的偏斜，最终选择采用其中的 1 727 条数据进行后续的训练工作。这样的数据选择和平衡策略旨在确保训练集具有足够的多样性和代表性，从而提高模型的性能和泛化能力。

1.2　数据集制作

对采集到的数据需要进行标注处理，以明确图像中物体的位置和类别。在 YOLO 中，采用边界框（bounding box）来标注物体的位置。标注的一般格式为：class，x_center，y_center，width，height，其中 class 表示物体的类别，x_center 和 y_center 是边界框的中心坐标，width 和 height 分别表示边界框的宽度和高度。

为了进行标注，本文使用了 labelme 工具，它能够方便地生成符合 YOLO 系列模型训练需求的“.txt”格式标注文件。标注好的数据集随后被划分为 80%的训练集和 20%的验证集，以确保模型在不同数据集上有充分的训练和测试。最终，数据集的分布如表 1 所示。

表 1　数据集分布

数据类型	训练集	验证集	小计	总计
展柜	344	81	425	2 152
人类	1 376	351	1 727	

2　YOLOv5 轻量化结构设计

2.1　YOLOv5 模型

YOLOv5 算法在 YOLOv4 的基础上经过一系列改进，具备轻量高速、高准确率和易移植的特点。模型整体由 Backbone：用于提取图像特征的卷积神经网络；Neck：用于融合和传递不同尺度的特征的网络层；Head：用于预测边界框和类别的网络层，三部分组成。YOLOv5n 是在模型基本结构的基础上减小通道宽度提出的一个轻量化分支，

在 Backbone 的结构设计中 YOLOv5 引入了基于 CSPNet（Cross-Stage Partial Network）的骨干网络思想，实现了对图像特征的高效提取。此外，为了实现更全面的特征融合，YOLOv5 的 Neck 部分采用了特征金字塔网络（Feature Pyramid Network，FPN）和结合路径聚合网络（Path Aggregation Network，PAN）结构。YOLOv5 使用了 YOLOv3 的 Head 网络结构，它由 3 个输出分支组成，每个分支负责预测不同尺度的目标。YOLOv5 的头部网络层使用了自适应的锚框，它根据每个特征层的大小和步长动态生成锚框，从而适应不同的数据集。YOLOv5n 模型结构如图 2 所示。

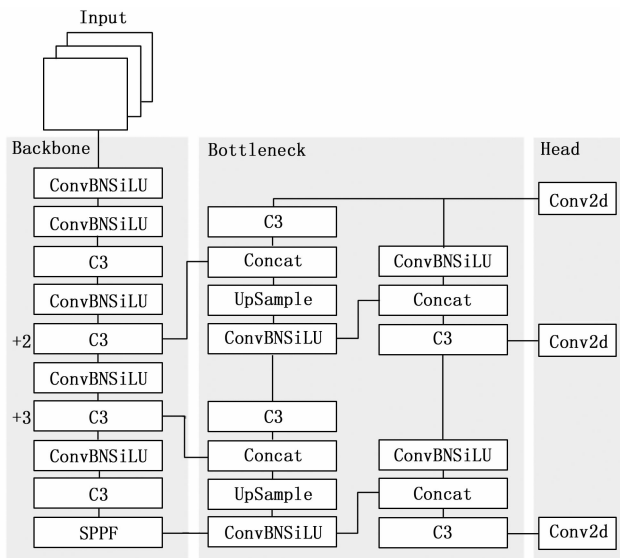

图 2　YOLOv5n 模型结构

2.2　Backbone 部分结构设计

GhostNet 基于深度可分离卷积（DSC，depthwise separable convolution）设计[15]，是一种轻量级神经网络架构，旨在实现模型压缩的同时保持较高的性能。GhostNet 的核心是 Ghost 模块，它引入了 ghost 特征映射，这是一种利用线性变换增加特征映射数量的低成本方式。Ghost 模块由两个部分组成：主要分支和 ghost 分支。主要分支是普通的卷积操作，而 ghost 分支是一

个基于主要分支输出的较小的卷积操作，它产生的特征图比主要分支的特征图要少。通过将主要分支的特征图与 ghost 分支的特征图相结合，GhostNet 能够在不损失准确率的情况下保持模型的表现力，同时降低了计算和内存开销。

如图 3 所示，在 GhostNet 的理论的基础上设计了 GhostConv 模块对输入的卷积分为 ghost 分支上的一次卷积和主要分支上的两次卷积，并将两条支路的特征提取结果进行拼接。基于 GhostConv 和深度可分离卷积设计残差结构，得到 GhostBottleneck 模块。最后，使用 GhostBottleneck 代替原 YOLOv5 结构中的 C3 模块中的 Bottleneck 得到 C3Ghost 模块[16]。

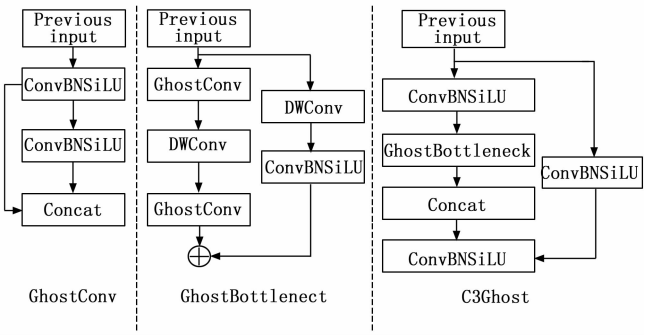

图 3 GhostConv、GhostBottleneck、C3Ghost 模型结构

2.3 Neck 部分结构设计

在 YOLOv5n 的 Neck 部分，采用了 GSConvns 结构来替代原本的 ConvBNSiLU，并引入了 VoV-GSCSP 来替代传统的 C3 结构。GSConvns 结构是一种轻量化的卷积层结构，它基于 GSConv，这是一种结合了标准卷积（SC）和深度可分离卷积（DSC）的卷积操作。GSConv 的原理是将 SC 的输出通道分为两部分，一部分直接作为 GSConv 的输出，另一部分经过 DSC 操作后与前一部分拼接，从而实现了通道信息的交互和融合。GSConv 的优点是既保留了 SC 的特征表达能力，又降低了 DSC 的计算和内存开销。GSConvns 结构进一步融合了标准卷积、深度可分离卷积和通道洗牌的特性，以实现在精度和速度之间的平衡。这种综合结构不仅考虑到模型性能的提升，还注重在计算效率和模型尺寸之间找到最佳平衡点。

在 GSConv 的基础上，进一步提出了一种使用一次性聚合的跨阶段局部网络 VoV-GSCSP，它可以实现特征的分割和融合，以及多尺度的特征提取。VoV-GSCSP 使用了 GSBottleneck 模块，它是一种基于 GSConv 的残差模块，可以保留 SC 的特征表达能力，同时降低 DSC 的计算和内存开销。GSConvns、GSBottleneck 和 VoV-GSCSP 结构如图 4 所示。

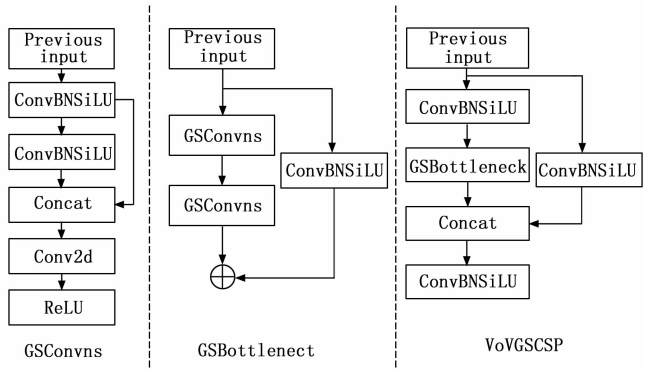

图 4 GSConvns、GSBottleneck、VoV-GSCSP 模型结构

2.4 轻量化结构 YOLOv5n

为了减小 Backbone 部分的参数量，本文使用基于 GhostNet 网络架构的 C3Ghost 模块代替原本 YOLOv5 中的 C3 模块。在 YOLOv5n 的 Neck 部分，采用了 GSConvns 结构来替代原本的 ConvBNSiLU，并引入了 VoV-GSCSP 来替代传统的 C3 结构。改进后的 YOLOv5n 模型结构如图 5 所示。

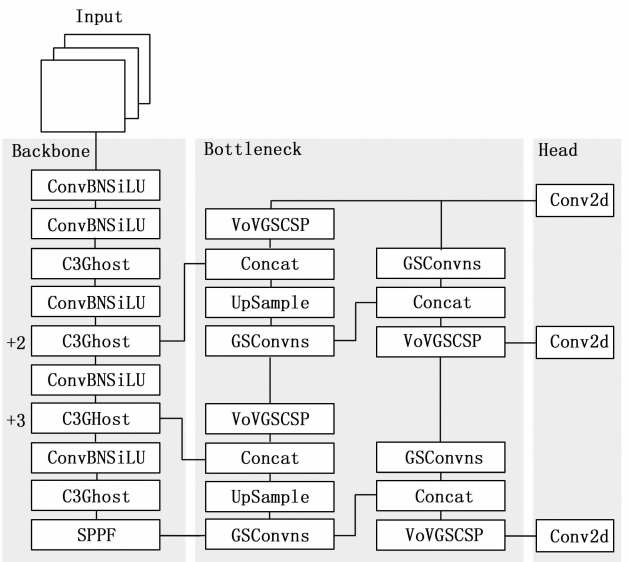

图 5 YOLOv5n 轻量化结构设计

3 基于剪枝技术的模型稀疏化

3.1 网络瘦身（Network Slimming）

Network Slimming（以下简称 Slim）是一种深度 CNN 的通道级剪枝方法，其核心思想是为每个通道分配一个缩放因子，该因子反映了该通道的重要性。在训练过程中，网络权重和缩放因子同时更新，并受到 L_1 正则化的约束，使得部分缩放因子趋于零。在剪枝阶段，根据缩放因子的大小，去除那些不重要的通道，并对剪枝后的网络进行微调，以恢复模型性能。Slim 的

训练目标由下式给出：

$$L=\sum_{(x,y)}l[f(x,W),y]+\lambda\sum_{\lambda\in\Gamma}g(\gamma) \tag{1}$$

其中：$(x,\ y)$ 表示训练输入和目标，W 表示可训练权重，第一个和项对应于 CNN 的正常训练损失，$g(\cdot)$ 是稀疏性引起的缩放因子惩罚，并且 λ 平衡了这两项。在本课题的实验中，选择 $g(s)=|s|$，即 L_1 正则化[17]。

通过对通道的剪除可以直接获得较窄的网络。在 Slim 算法的计算过程中，缩放因子与网络权重共同优化，并将其作为通道选择的标准，自动筛选网络中的不重要通道。该设计能够确保剪枝过程中安全地删除部分通道而不至于影响模型性能。Slim 算法流程如图 6 所示。

3.2　基于性能感知的全局通道修剪（PAGCP）

PAGCP 算法考虑了层内和层间滤波器的关系，将联合显著性纳入剪枝目标选择的分析过程中。PAGCP 不仅注重单个滤波器的性能，更加强调在整个深度模型中通道的全局影响。

PAGCP 采用了一种贪婪的剪枝策略，并在对每一层的剪枝过程中都引入上一步剪枝后模型的多任务损失状态向量用以估计当前层的剪枝率，以确保对当前层压缩实现任务敏感性可知和模型状态性能可控，从而确保剪枝后的模型性能。在对层内的滤波器进行筛选时，引入了联合显著性来评价滤波器的重要性，联合显著性计算的目标函数如下：

$$\min_{\theta_l}\sum_{i=1}^{K_l}S_l^k(\theta)\ \text{s.t.}\ g(\theta_l)\leqslant\alpha_l \tag{2}$$

将显著性准测 $S_l^{(k_i)}(\theta)$ 设计为 l_1 的形式，得到计算公式为：

$$S_{l_1}^{k_1,\cdots,k_N}(x)=\sum_{i=1}^{N}\|\omega_{l_ik_i}(\theta_{l_1k_1}\cdots,\theta_{l_Nk_N}\|_1 \tag{3}$$

其中：θ_l 为第 l 层的剪枝选择，K_l 是第 l 层的滤波器。

通过式（2）、（3）的计算输出当前层待删除滤波器的列表，并以此为参考对当前层进行压缩。压缩所有层后，再次通过评估其压缩贡献来对剪枝后的层重新排序，再次压缩排序后的前 P 层，然后重新训练剪枝后的模型并重复上述过程，直到满足 FLOPs 或参数的减少要求。PAGCP 算法流程如图 7 所示。

3.3　轻量化改进方法

本文提出的方法是在对 YOLOv5n 进行轻量化结构设计后，使用两种剪枝算法对其分别进行剪枝，再对剪枝后的模型进行重训练，对比改进前后的模型性能，从而选取最高效的轻量化方法作为下一步行人姿态检测算法的目标检测部分。

4　行人姿态检测算法

在分析环境中行人的姿态时除了能够准确地识别到目标对象外，还要确保在检测过程中能够持续跟踪目标不产生跟丢的情况。最后，再对一段连续帧中捕捉到的人类目标进行姿态检测及分类。

4.1　DeepSORT 目标跟踪算法

简单的在线实时跟踪（SORT，simple online and realtime tracking）[18]是一种经典的多目标检测算法。其基于目标检测进行跟踪（TBD，tracking-by-detecton），实现的效果是根据目标检测的结果实现对于多个目标的跟踪。SORT 算法使用卡尔曼滤波器（KF，kalman filter）[19]实现根据当前帧的目标检测结果预计后来帧的目标位置即跟踪器的功能，使用匈牙利算法实现检测结果和跟踪器预测结果的匹配。

在 SORT 算法中仅使用目标框和跟踪器的 IoU 来匹配相邻帧之间的目标。这种方式在遇到遮挡或是未检测到目标的情况时很容易丢失跟踪过程，因此在 DeepSORT 算法[20]中引入了深度学习方法对目标的外观特征信息进行提取和保存。之后每执行一步都要执行一次当前帧被检测物体外观特征与之前存储的外观特征的相似度计算，这个相似度将作为一个重要的判别依据。此外，DeepSORT 算法在 SORT 算法的基础上增加了级联匹配（Matching Cascade）同时考虑 IoU 和目标相似度以及新跟踪器确认。将卡尔曼滤波器得到的跟踪器分为确认态和不确认态。DeepSORT 的算法流程如图 8 所示。首先为第一帧得到的目标框创建

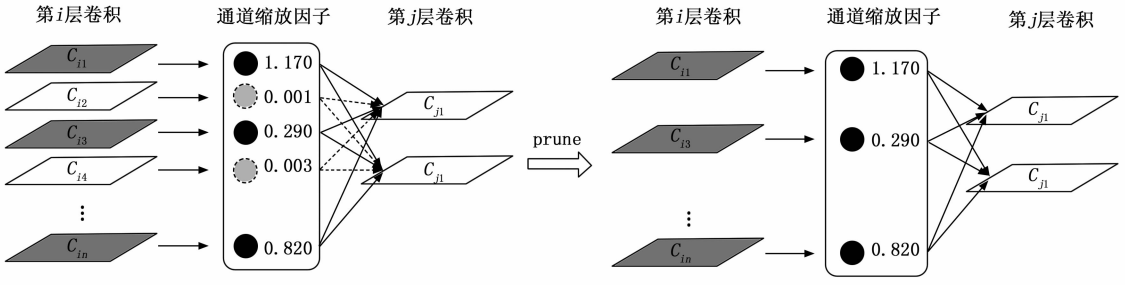

图 6　Slim 算法流程

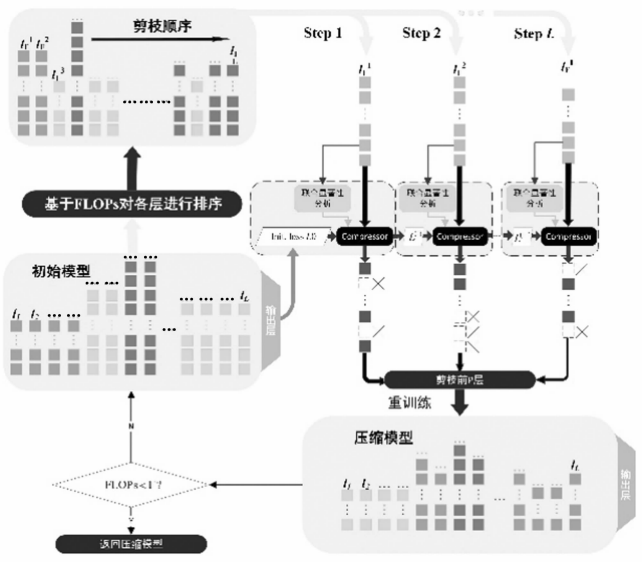

图 7 PAGCP 算法流程

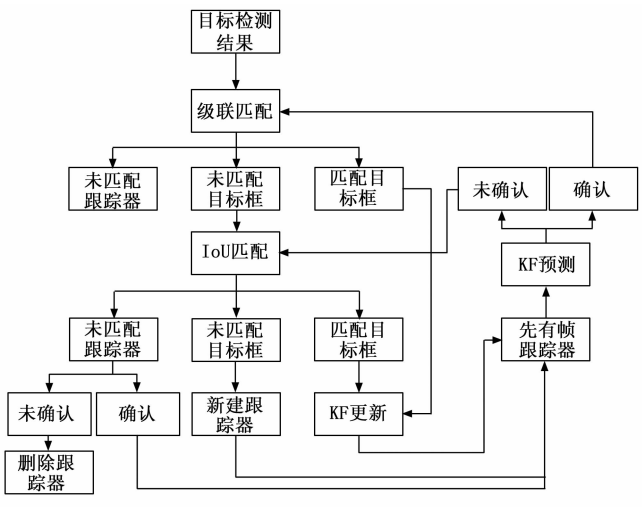

图 8 DeepSort 算法框图

对应的跟踪器，然后通过卡尔曼滤波器预测下一帧的跟踪器，需要注意的是这里的跟踪器一定是不确定态；在下一帧中将该帧的目标检测框和上一帧卡尔曼滤波器得到的跟踪框进行 IoU 匹配，再通过 IoU 匹配的结果计算其代价矩阵，将此处得到的代价矩阵输入到匈牙利算法中进行线性匹配从而得到最优的匹配结果，并将匹配成功的跟踪器置为确认态。下一步将确认态的跟踪器与目标检测框进行级联匹配，匹配的结果分为 3 种情况，对失配的检测器和跟踪器再次进行 IoU 匹配并计算代价矩阵，得到的结果输入到匈牙利算法中从而实现算法的迭代更新，不断产生新的跟踪器和跟踪轨迹直至视频结束。

4.2 SlowFast 姿态检测算法

SlowFast 是一种用于视频理解的深度学习架构，专门设计用于解决处理视频中的空间和时间信息的挑战。其主要动机是观察到视频包含不同时间尺度的信息。某些动作或事件可能发展缓慢，需要更长的时间上下文进行识别，而其他动作则发生迅速，需要更高的帧率进行准确的识别。传统的深度学习架构可能难以有效处理这两个时间方面。

SlowFast 通过引入两个路径来应对这一挑战：一个慢速路径处理低帧率视频片段，通过较大的通道数学习空间语义信息；另一个快速路径处理高帧率片段，以较少的通道数捕捉学习运动信息。需要注意的是，尽管快速路径处理更多的帧，但由于使用较少的通道数，其计算量并未显著增加，仅占整体计算量的 20%。通过这两个路径的巧妙设置，SlowFast 有效地平衡了时间信息，实现了更为准确的视频理解。这种架构设计使得 SlowFast 在处理视频时能够更好地捕捉不同时间尺度下的关键信息。如图 9 所示为 SlowFast 网络的关键组成部分。

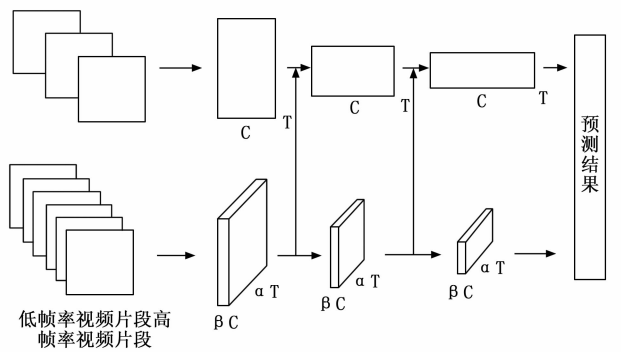

图 9 SlowFast 关键组成

分支路径包含慢速路径和快速路径两部分：(1) 慢速路径：慢速路径以较低的帧率运行，通常以一个较大的时间步长采集视频帧，默认为 1 秒采集 2 帧。对采集到的视频帧进行 3 D 卷积，卷积核形状为 $\{T\times S^2, C\}$ 其中 T、S 和 C 分别表示时序 (temporal)，空间 (spatial) 和频道 (Channel) 的数目。它包含更深和更宽的卷积层，这在计算上更昂贵，但可以捕获更复杂和抽象的特征。(2) 快速路径：快速路径以较高的帧率运行，以一个较小的时间步长采集视频帧，默认为 1 秒 15 帧。它专注于捕获短期的时间动态，负责检测快速变化和局部运动模式。快速路径通常比慢速路径更浅更窄，使用较小的通道数 ($\beta=1/8$) 以减少计算成本。

视频片段进行分支处理后需要对低帧率路径和高帧率路劲进行分支融合。分支融合是慢速和快速路径的输出通过融合机制合并，并产生最终的视频表示。融合过程使模型能够有效地利用长期上下文和短期细节。在融

合过程中单一数据样本在两条通路的形状是不同的，采用一个 5×12 的核进行 3 d 卷积，$2\beta C$ 输出频道，$stride = \alpha$，对 Fast 通道的结果进行数据变换然后融入到 Slow 通道。

4.3　行人检测算法

利用本文所提出的轻量化 YOLOv5n 进行目标检测，使用 DeepSORT 算法对检测到的目标进行跟踪将物体前后类别联系起来[21]，使得行为类别信息更加连贯，最后采用 SlowFast 算法对人类目标动作进行分类。

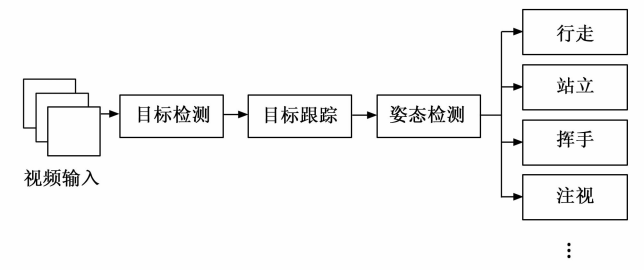

图 10　行人检测算法框图

5　实验验证

5.1　轻量化 YOLOv5 实验

模型的训练设备使用一台 Windows 主机，CPU 配置为 Intel Core i9-10980XE；GPU 采用 NvidiaGeforceRTX3090 * 2；RAM 为 256 G；深度学习模型框架基于 pytorch1.12.0＋torchvision0.13.0。以下为模型训练结果。

5.1.1　损失及精度

如图 11 所示，以官方提供的 YOLOv5n 作为 baseline，展示了轻量化结构设计的 YOLOv5 以及在经过模型剪枝后重训练的模型的训练损失。

训练损失主要包含 3 个部分：

1）box_loss：用于监督检测框的回归，预测框与标定框之间的误差（CIoU）。

2）obj_loss：用于监督网格中是否存在物体，计算网络的置信度。

3）cls_loss：用于监督类别分类，计算锚框与对应的标定分类是否正确。

5.1.2　模型改进效果

本文选取 FLOPs、Parameters 以及 $mAP50$ 作为评估模型改进效果的指标。

FLOPs 是浮点运算数的缩写，衡量一个模型或算法的计算量，通常 FLOPs 越小则说明模型的计算量越小。Parameters 表征了模型的参数量，其值越小，则说明模型所需的内存资源越小。$mAP50$ 表示 IoU 阈值为 0.5 时的平均精度，是常用的用于评估模型检测精度的

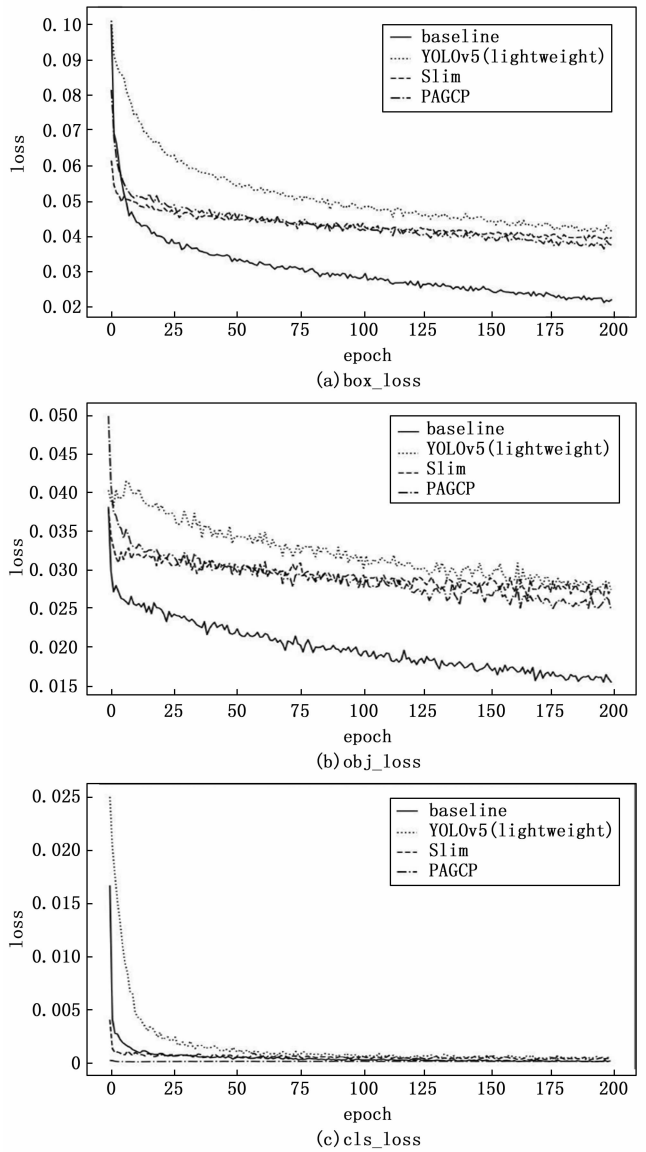

图 11　使用多种剪枝方法后的模型训练损失

指标。如图 12 所示，以 YOLOv5n 作为 baseline，对比使用不同方法改进后的模型性能。

5.1.3　结果分析

根据图 11 中提供的结果可以得出，使用不同方法对 YOLOv5n 进行改进后模型的训练损失均能维持在一个较低的水平。其中，只做轻量化结构改进，而不做剪枝改进的 YOLOv5n（lightweight）的损失是最高的。两种剪枝算法改进后的模型损失基本一致，且维持在一个可以接受的范围内。

根据图 12 提供的结果可以得出，轻量化改进后的 YOLOv5n 模型的 FLOPs 均有大幅下降，其中 YOLOv5n（lightweight）下降 40%，Slim 算法下降 71%，PAGCP 算法下降 55%。考虑 Parameters，YOLOv5n（lightweight）减少 35%，Slim 算法减少 68%，PAGCP 算法减少 56%。此外，各种轻量化改进方法的模型精度均有所

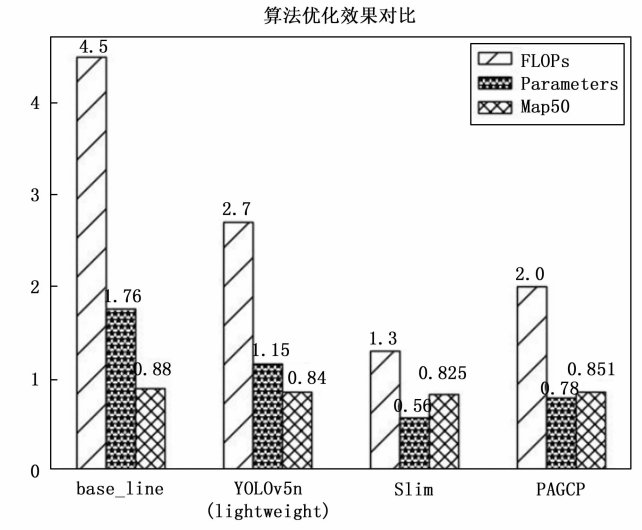

图 12 模型优化效果

下降，*mAP*50 的下降基本维持在 4%左右。

综合考虑上述分析，以模型压缩量为首要标准，选择 Slim 算法剪枝后的模型作为行人姿态检测算法的目标检测方法。

5.2 行人姿态检测实验

对采集到的视频输入到动作识别算法中进行检测，输出检测到的人类目标姿态的结果如图 13 所示。可以看出，在实验条件下，本文所提出的算法能够准确识别目标对象并能够分析出人类目标的姿态信息。

图 13 行人姿态检测实验结果

表 2 行人姿态检测实验结果

姿态分类	实验数据/组	检测成功/组	成功率/%
行走(walk)	100	95	95
站立(stand)	100	98	98
挥手(hand wave)	150	117	78
打电话(answer phone)	150	122	81
注视(watch)	150	124	83

6 结束语

针对搭载移动端低算力设备的迎宾机器人提供一种分析用户行为的解决方案。

通过使用轻量化模型结构和模型稀疏化算法降低了 YOLOv5n 模型的计算资源需求。在轻量化 YOLOv5n 的基础上使用 DeepSORT 算法实现目标跟踪，SlowFast 算法实现人体姿态检测，从而完成对于视频中连续的人类目标进行目标检测以及姿态检测。

根据实验结果，本文设计的轻量化 YOLOv5n 相较原模型压缩率在 70%以上。在此基础上进行的行人姿态检测算法能够对视频中的人类目标进行准确的姿态检测。

参考文献：

[1] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices [J]. arXiv, 2018: 6848 - 6856.

[2] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications [J]. arXiv, 2017: 566 - 578.

[3] MEHTA S, RASTEGARI M, CASPI A, et al. ESPNet: Efficient Spatial Pyramid of Dilated Convolutions for Semantic Segmentation [Z]. 2018: 552 - 568.

[4] TAN M, LE Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks [A]. Proceedings of the 36th International Conference on Machine Learning [C] //PMLR, 2019: 6105 - 6114.

[5] HAN K, WANG Y, TIAN Q, et al. GhostNet: More Features From Cheap Operations [Z]. 2020: 1580 - 1589.

[6] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks [Z]. 2018: 4510 - 4520.

[7] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3 [Z]. 2019: 1314 - 1324.

[8] HU H, PENG R, TAI Y W, et al. Network Trimming: A Data-Driven Neuron Pruning Approach towards Efficient

Deep Architectures [J]. arXiv, 2016.
[9] LIU Z, LI J, SHEN Z, et al. Learning Efficient Convolutional Networks Through Network Slimming [Z]. 2017: 2736 - 2744.
[10] LEE J, PARK S, MO S, et al. Layer-adaptive sparsity for the Magnitude-based Pruning [EB/OL]. arXiv. org. (2020-10-15). [2023-12-06]. https://arxiv.org/abs/2010.07611v2.
[11] YE H, ZHANG B, CHEN T, et al. Performance-aware approximation of global channel pruning for multitask CNNs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45 (8): 10267 - 10284.
[12] LIN J, GAN C, HAN S. TSM: Temporal shift module for efficient video understanding [Z]. 2019: 7083 - 7093.
[13] FEICHTENHOFER C, FAN H, MALIK J, et al. Slow-Fast Networks for Video Recognition [J]. arXiv, 2019: 134 - 146
[14] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks [Z]. 2018: 7794 - 7803.
[15] 王　军，冯孙铖，程　勇. 深度学习的轻量化神经网络结构研究综述 [J]. 计算机工程，2021，47 (8)：1 - 13.
[16] CAO M, FU H, ZHU J, et al. Lightweight tea bud recognition network integrating GhostNet and YOLOv5 [J]. Mathematical biosciences and engineering: MBE, 2022, 19 (12): 12897 - 12914.
[17] 刘馨柔. 基于 YOLOv3 的流水线动态目标检测算法研究 [D]. 长春：长春理工大学，2022.
[18] BEWLEY A, GE Z, Ott L, et al. Simple online and realtime tracking [A] //2016 IEEE International Conference on Image Processing (ICIP) [C], 2016: 3464 - 3468.
[19] KALMAN R E. A new approach to linear filtering and prediction problems [J]. Journal of Basic Engineering, 1960, 82 (1): 35 - 45.
[20] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric [A] // 2017 IEEE International Conference on Image Processing (ICIP) [C] // 2017: 3645 - 3649.
[21] 高　切，李登华，丁　勇. 基于 M-DBT 框架的岩质边坡落石跟踪算法研究 [J]. 水利水运工程学报，2024 (3)：166 - 176.

(上接第 44 页)

[5] 谷梦勖，万　衡，黄怡婷，等. 上海轨道交通不间断电源设备状态评估系统的设计和应用 [J]. 城市轨道交通研究，2022，25 (10)：187 - 191.
[6] 王　森，杨晓峰，李世翔，等. 城市轨道交通直流自耦变压器牵引供电系统故障保护研究 [J]. 电工技术学报，2022，37 (4)：976 - 989.
[7] 李广明. 地铁刚性接触网供电系统弓网状态在线检测装置 [J]. 城市轨道交通研究，2023，(S1)：152 - 157.
[8] 刘　超，张红星，高兴华，等. 厦门地铁 1 号线辅助供电系统相电流不平衡故障分析与改进 [J]. 城市轨道交通研究，2023，26 (1)：196 - 200.
[9] 谢国财，温　锐，陈　琛. 基于模糊神经网络的高压电力设备故障预测模型 [J]. 电网与清洁能源，2022，38 (9)：120 - 125.
[10] 李　奎，王超然，王　尧，等. 单相供用电系统故障漏电电流特征及保护方法 [J]. 天津工业大学学报，2023，42 (4)：63 - 70.
[11] 唐圣学，马晨阳，勾　泽. 基于时频特征融合与 GWO-ELM 的棒控电源早期故障状态辨识方法 [J]. 仪器仪表学报，2023，44 (1)：121 - 130.
[12] 聂同攀，曾继炎，程玉杰，等. 面向飞机电源系统故障诊断的知识图谱构建技术及应用 [J]. 航空学报，2022，43 (8)：46 - 62.
[13] 李　炜，韩寅龙，孙晓静. 基于特征优选与深度学习的车载电源微小故障诊断方法 [J]. 兵工学报，2022，43 (11)：2935 - 2944.
[14] 程红丽，史金鑫，李　勇. ISOP 系统电源模块故障诊断策略设计与实现 [J]. 电子器件，2022，45 (1)：154 - 159.
[15] 阳璞琼，关志全，王　磊. EAST 离子回旋共振加热系统电子管电源打火故障分析 [J]. 核电子学与探测技术，2022，42 (4)：672 - 676.
[16] 宫　琦，陈秉智，李永华，等. 地铁电源模块故障分析及预测方法 [J]. 东北大学学报 (自然科学版)，2022，43 (1)：8 - 16.
[17] 朱吉然，牟龙华，郭文明. 考虑并网运行微电网故障方向识别的逆变型分布式电源故障控制 [J]. 电工技术学报，2022，37 (3)：634 - 644.
[18] 杨丽琼，章隆兵，肖俊华，等. 高性能 CPU 电源 Droop 检测优化设计实现 [J]. 高技术通讯，2022，32 (9)：894 - 902.
[19] 王云艳，罗　帅，王子健. 基于改进型 BP 神经网络的光伏功率预测 [J]. 计算机仿真，2022，39 (11)：153 - 157.
[20] 托　娅，王　伟，毛华敏，等. 基于机器学习的电力设备故障红外智能诊断方法 [J]. 河南理工大学学报 (自然科学版)，2022，41 (5)：121 - 126.
[21] 梁　剑，黄志鸿，张可人. 基于多尺度引导滤波和决策融合的电力设备热故障诊断方法研究 [J]. 红外技术，2022，44 (12)：1344 - 1350.