文章编号:1671-4598(2024)09-0066-07 DOI:10.16526/j.cnki.11-4762/tp.2024.09.010 中图分类号:TP242.6 文献标识码:A

基于改进 YOLOv5 的室内楼梯检测方法研究

韩飞燕[1],赵 伟[2],吴子英[2]
(1. 西安航空职业技术学院 航空制造工程学院,西安 710048;
2. 西安理工大学 机械与精密仪器工程学院,西安 710048)

摘要:移动机器人视觉 SLAM 的楼梯建图过程需要对楼梯特征进行检测识别,传统的边缘检测、直线提取等楼梯检测技术往往视角较为理想、背景较为简单,无法实现栏杆遮挡、复杂背景下的楼梯特征提取;为了解决以上问题,提出了一种可用于移动机器人的改进 YOLOv5 的楼梯目标检测方法,在输入端引入 FenceMask 数据增强策略,增加对遮挡楼梯的训练样本数量;通道注意力模块 CAM 与空间注意力模块 SAM 采用并行连接的方式组成注意力模块 CBAM,加强在复杂环境下对楼梯的特征提取能力;在预测端将 NMS 与 WBF 结合,将 NMS 筛选之后置信度较高且位置相邻的边框进行融合为新的边框,在满足精度要求的情况下改善了 Faster-RCNN 与 SSD 检测算法存在的单段多阶楼梯检测速度问题;仿真表明改进的 YOLOv5s 可以在模型大小 18.4 MB 的情况下达到 82.9%的平均精度,改进的 YOLOv5m 在增大模型为 45.5 MB 的情况下平均精度提高为 86.5%,均可有效识别栏杆遮挡、复杂背景以及单段长阶梯。

关键词:机器视觉;楼梯目标检测;YOLOv5;CBAM;边框融合

Study on Indoor Staircase Detection Method Based on Improved YOLOv5

HAN Feiyan[1], ZHAO Wei[2], WU Ziying[2]
(1. School of Aeronautical Manufacture Engineering, Xi'an Aeronautical Polytechnic Institute, Xi'an 710048, China;
2. School of Mechanical and Instrumental Engineering, Xi'an University of Technology, Xi'an 710048, China)

Abstract: During the process of building stairs by using mobile robot visual simultaneous localization and mapping (SLAM), it is necessary to detect and recognize the features of stairs. Traditional stair detection technologies, such as edge detection and line extraction, often have the characteristic of ideal visual angle and simple background, but they can not extract the stair features under the railing occlusion and complex backgrounds. In order to solve the above problems, a stair target detection method based on improved YOLOv5 used for mobile robots is proposed. The FenceMask data enhancement strategy at the input end is introduced to increase the number of training samples for occluded stairs. The channel attention module (CAM) and spatial attention module (SAM) are connected in parallel to form the convolution block attention module (CBAM), enhancing the ability to extract stair features in complex backgrounds. At the prediction end, the non-maximum (NMS) and weighted boxes fusion (WBF) are combined, and the high confidence and close position bounding boxes filtered by the NMS are fused into new bounding boxes, improving the detection speed of single segment and multi-step stairs in the Faster-RCNN and single short multi-box detector (SSD) detection algorithms while meeting accuracy requirements. Simulation results show that the improved YOLOv5 reaches an average accuracy 82.9% with a model size of 18.4 MB, and improves the average accuracy of 86.5% with a model size of 45.5 MB. The improved YOLOv5 can effectively identify the conditions of railing occlusion, complex backgrounds and single segment long stairs.

Keywords: machine vision; stair target detection; YOLOv5; CBAM; border fusion

0 引言

传统的机器视觉楼梯目标检测方法主要采用边缘检测[1]与直线提取[2]等方法。边缘检测技术是将采集到的楼梯图像进行校正与降噪处理,对检测的边缘轮廓进行分析定位从而实现楼梯识别;直线提取算法主要借助了楼梯台阶规律平行分布的特性,通过提取图像的水平直线,筛除倾角过大以及长度不符的直线,对符合特性的直线进行分析从而估计楼梯在图像中的位置。传统的楼梯检测方法存在 3 个问题:(1)在有栏杆、扶手等遮挡的环境中,无法提取到有效的边缘信息;(2)传统方法检测背景往往过于理想,对拍摄角度要求较高;(3)检测模型泛用性较差,无法在复杂的动态环境中进行高准确率的检测识别。

深度学习的目标检测方法是目前机器视觉领域的重要研究方向,2012 年,文献[3]在 ImageNet 大赛中采用卷积神经网络(CNN, convolutional neural network)开启了深度学习的研究热潮;2014 年,文献[4]提出了先对图像划分候选区域,再通过 CNN 对候选区域进行特征提取,最后通过 SVM 分类的 RCNN 给出了目标检测算法;2015 年,

收稿日期:2024-05-08; **修回日期**:2024-06-17。
作者简介:韩飞燕(1989-)女,硕士,讲师。
吴子英(1975-),男,博士,副教授。
引用格式:韩飞燕,赵 伟,吴子英.基于改进 YOLOv5 的室内楼梯检测方法研究[J].计算机测量与控制,2024,32(9):66-72,79.

文献［5］改进 RCNN 提出了 Fast RCNN 算法，将边框回归引入到 CNN 特征提取中并加入了 ROI 池化层，极大幅度地缩短了训练时间与检测耗时；同年，文献［6］提出的 Faster RCNN 采用端到端的目标检测框架，进一步提高了检测精度的同时缩短了检测时间。RCNN、Fast RCNN、Faster RCNN 是二阶段目标检测算法的代表，二阶段目标检测算法的特点为精度较高但实时性较差，一阶段目标检测算法的出现解决了目标检测实时性的问题。一阶段目标检测算法的代表为 SSD 算法与 YOLO 系列，2015 年，文献［7］推出的 YOLOv1 采用 Darknet-19 网络结构，可以实现较高的实时检测；2016 年，文献［8］提出的 SSD 算法兼顾了目标检测检测精度与实时性的双重需求，一阶段目标检测算法将目标检测与分类结合为一个回归问题，检测精度逊于二阶段目标检测算法，但是明显提高了检测速度，具有较高的实时性。

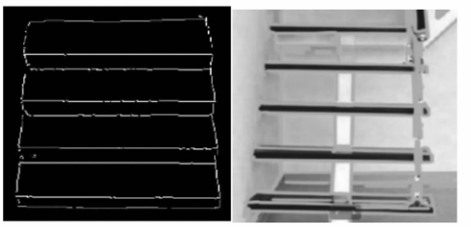
YOLOv5 相较于之前的 SSD、YOLOv4 算法，YOLOv5 算法拥有更高检测速度与轻量级的检测框架，这使得 YOLOv5 在移动机器人与嵌入式设备上的部署具有一定优势。近些年来，我国学者针对 YOLO 系列算法进行了多方位的改进，如文献［9］提出了一种基于注意力机制与跨尺度特征融合的 YOLOv5 检测算法，可以实现复杂背景下目标的高精度检测；文献［10］采用视觉信息与空间分布关系相结合的方案，将 YOLOv3 算法与 BGRU-L 模型相融合，可以有效地检测呈现一定规律分布的目标，但是增加了额外的计算量与检测时间；文献［11］通过引入 Mixup 数据增强策略与 WBF 边框融合策略，实现了多种重叠目标的检测，但是 WBF 边框融合策略相较于 NMS 非极大值抑制策略实时性较差；文献［12］改进了 SPP 网络结结构，并使用 BiFPN 来融合多尺度特征，用 ELU 作为激活函数，实现了低运算开销下的高精度检测。

本文提出的改进改进 YOLOv5 的楼梯检测方法，降低栏杆、墙壁、扶手以及周围环境带来的视觉遮挡；提高了室内楼道复杂环境下算法网络特征聚焦速度；一定程度上解决了不同拍摄视角下楼梯特征存在的差异性特征。

1　识别对象分析

楼梯是高层建筑常见的通道结构，其结构分布具有较强的规律性。传统的边缘检测与直线提取技术可以通过对比图像特征与楼梯的结构分布实现检测。如图 1 所示，传统的楼梯检测对楼梯图像的拍摄视角较为严格，且图像背景往往较为单一理想化，若存在栏杆、扶手等视觉遮挡以及其他背景干扰，则此类方法无法有效提取楼梯特征。

在常规的室内建筑结构中，楼梯的台阶尺寸按照“国标 GB 50352-2019”来规划建造，单段楼梯的台阶数量一般为 4～12 阶。将单段楼梯视作一个单独检测目标，可以满足大部分常规情况下的楼梯识别。在面对少数 18 阶以上的单段长台阶时，由于其图像外形特征与常规楼梯存在较大差距，容易会产生“多检”与错检的情况。若将任意台阶数量的单段楼梯视作一个整体目标，由于长、短楼梯造成



(a)边缘检测　　(b)直线提取

图 1　传统的楼梯检测方法

的样本差异，最终训练所得的模型整体检测效果较差。

为了解决以上问题，本文将连续的 4～12 台阶数单段楼梯视作一个识别目标，在预测框后处理中采用融合 NMS 与 WBF 的边框融合策略，将靠近程度较近的预测框进行融合，解决长台阶的多检问题。在不同的拍摄角度下，楼梯的图像特征存在较大的差异，楼梯正面与侧面拍摄的楼梯图像特征匹配度较低。考虑到相机在移动机器人的空间位置，本文的楼梯检测需要对拍摄视角进行统一，要求数据集中的楼梯图片符合移动机器人常规的拍摄视角。如图 2 所示，移动机器人需要在拍摄视线与水平面夹角［－60°，60°］以及与竖直平面夹角为［60°，150°］的情况下完成对楼梯的目标检测识别。

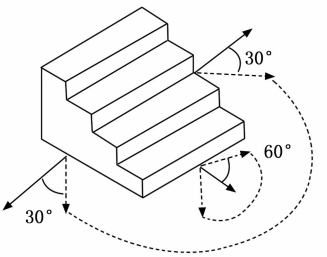


图 2　楼梯检测视角

2　改进的 YOLOv5 目标检测算法

2.1　改进 YOLOv5 算法框架

针对复杂室内环境中的楼梯检测问题，本文以 YOLOv5 算法为基本框架[9]，将独立连续的楼梯视作一个单独的识别对象进行训练与检测。在输入端引入 FenceMask 数据增强策略，模拟栏杆对楼梯的遮挡效果；引入注意力模块 CBAM，提高复杂背景下的楼梯检测效果；将 NMS 筛选策略与 WBF 边框融合策略结合，提高了面对连续长楼梯的检测效果。改进的 YOLOv5 框架如图 4 所示，改进的 YOLOv5 网络结构由“Input”“Backbone”、“Neck”、“Prediction”4 部分组成。

Input：为了解决栏杆、扶手等环境因素对楼梯的遮挡造成识别率低的问题，在输入端引入 FenceMask 数据增强策略来替换 Masoic 数据增强策略，丰富样本的同时强化复杂遮挡环境下算法对楼梯的识别检测效果；同时将不同尺寸的输入图像进行自适应缩放填充，统一调整为 640×640。

Backbone：BackBone 部分用于提取输入图像的多尺度特征信息。包括切片层 Focus、4 个 CBL 卷积层、3 个 CSP-

Net 模块、SSP 空间金字塔池化模块。Focus 层通过横纵间隔取像素点的方式对像素点切分从而扩大通道数，原始的 RGB 坐标系下的输入图像 640×640×3 切分为 320×320×12。之后多通道的特征图进行一系列的卷积操作提取特征，提取特征经过 CBAM 模块进行特征空间与通道上的聚焦，最后通过 SSP 金字塔池化结构将特征图进行合并。

Neck：Neck 部分用于将 Backbone 部分提取的特征进行融合，采用 FPN＋PAN 的结构，首先通过 FPN 自顶向下进行传递语义融合，再通过 PAN 结构自下向定的进行传递空间定位特征。交并比示意图如图 3 所示。

$$IOU_{i,j} = \frac{|\ i \cap j\ |}{|\ i \cup j\ |} \tag{1}$$

$$L_{GIOU} = 1 - IOU_{i,j} + \frac{|\ k/(i \cup j)\ |}{|\ k\ |} \tag{2}$$

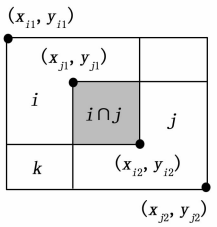


图 3 交并比示意图

Prediction：预测阶段的损失函数采用式（2）的 GIOU _ loss，GIOU _ loss 弥补了式（1）所示的 IOU _ loss 在面对预测框与标注框不重叠或是两框 IOU 相同的情况无法正确处理的问题。为了解决面对长阶梯的多检问题，在预测框的后处理中，先通过 NMS 筛选出置信度较高且重叠区域较小的预测框，再通过判断其邻近比 NS 的方式，将靠近程度较近的预测框进行 WBF 融合，生成最终的预测输出框。

2.2 FenceMask 数据增强策略

由于楼梯扶手、栏杆以及墙壁的存在，摄像头对于楼梯的拍摄往往容易受到遮挡。为了提高复杂环境中栏杆、扶手遮挡下楼梯的识别分辨能力，防止模型出现过拟合现象，本算法在训练之前采用离线 FenceMask 数据增强的方式，来扩大数据集规模。

FenceMask[13] 数据增强方式是在 GridMask[14] 数据增强方式的基础上改进得来的，如图 5 所示，FenceMask 与 GridMask 数据增强策略的基本思想都是生成一定数量的掩盖区域，模拟复杂环境下的遮挡效果。不同之处在于 GridMask 数据增强方式生成的掩盖区域为均匀分布的多个小正方形，通过预先设定单个小正方形的边长以及小正方之间的距离来确定掩盖区域的分布；FenceMask 数据增强方式改进了 GridMask 数据增强方式的小正方形遮挡区域，改为具有一定稀疏性的连续块状遮挡物分布。两种数据增强方式如图 5 所示。

通过对比图 5（c）与图 5（d）可以看出，FenceMask 数据增强方式在进行有效数据增强的同时总体遮挡范围更小，不容易造成图像特征的丢失，而且遮挡区域更接近楼梯栏杆的遮挡效果。

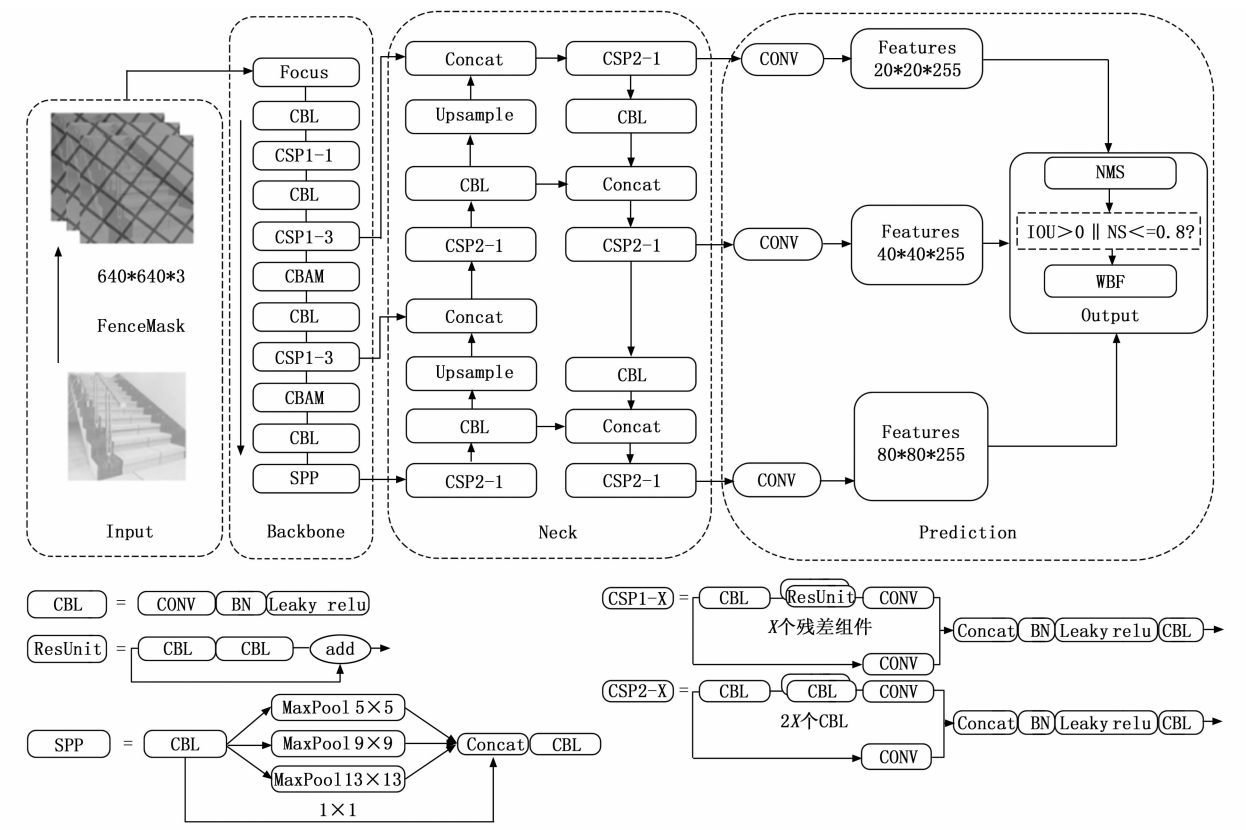
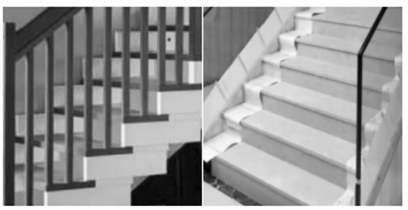


图 4 改进 YOLOv5 网络框架



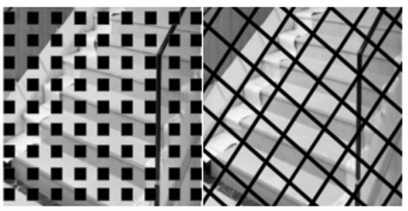


图 5　数据增强示意图

2.3　CBAM 注意力模块

室内楼道环境通常较为复杂，为了使得目标检测算法可以快速地聚焦楼梯特征，本算法引入轻量级的注意力模块 CBAM。CBAM 模块由通道注意力模块 CAM 与空间注意力模块 SAM 组成，可以实现通道与空间上的双重聚焦[15]。

通道注意力 CAM 模块用于提取并保留特征较为明显的通道。如图 6 所示，输入的特征图为 F [$H\times W\times C$]（长为 H、宽为 W、通道数为 C），经过并行的最大池化（MaxPool）与平均池化（AvgPool）处理得到两个特征图 F [$1\times 1\times C$]；将 F [$1\times 1\times C$] 送入两层的共享神经网络 MLP，MLP 网络第一层由 C/r 个神经元组成，其中 r 为减少率，第一层的作用是将 C 个通道的特征提取为 C/r 个，第二层为 C 个神经元，作用是将第一层的 C/r 特特征进行特征重组为 C 个特征图；将 MLP 输出的两个特征矩阵进行张量相加，通过激活函数 sigmoid 生成式（3）通道注意力特征 M_C。

$$M_C(F)=\sigma\{\mathrm{MLP}[\mathrm{AvgPool}(F)]+\mathrm{MLP}[\mathrm{MaxPool}(F)]\}=\sigma\{W_1[W_0(F^C_{\mathrm{Avg}})]+W_1[W_0(F^C_{\mathrm{Max}})]\}\in R^{C\times 1\times 1} \quad (3)$$

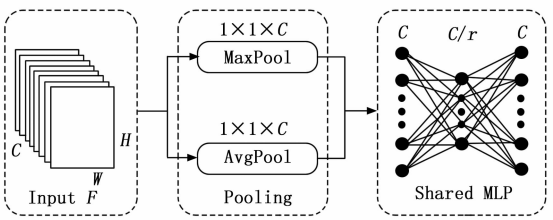


图 6　CAM 模块示意图[15]

空间注意力模块 SAM 的作用是提高对聚焦目标空间上的定位效果。如图 7 所示，其流程为将通道注意力特征图 F'作为输入，进行全局最大池化与全局平均池化，将池化后的特征进行通道拼接；拼接后的特征送入到一个 7×7 的卷积核进行卷积降维为单通道的特征 F' [$1\times 1\times C$]，最后通过激活函数输出式（4）空间注意力特征 M_S：

$$M_S(F)=\sigma(f^{7\times 7}([\mathrm{AvgPool}(F);\mathrm{MaxPool}(F)]))=$$

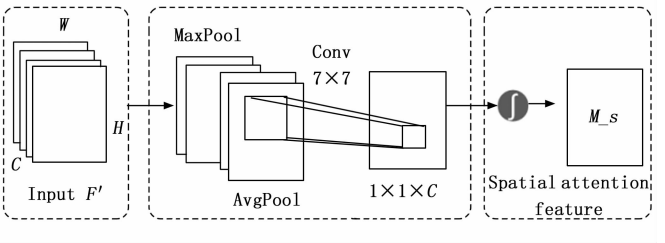


图 7　SAM 模块示意图[15]

$$\sigma(f^{7\times 7}([F^S_{\mathrm{Avg}};F^S_{\mathrm{Max}}])) \quad (4)$$

通道注意力模块 CAM 与空间注意力模块 SAM 可以通过串行与并行的方式组合连接，前者实验表明并行连接的效果更加理想。如图 8 所示，通道注意力模块 CAM 与空间注意力模块 SAM 采用并行连接的方式构成 CBAM 模块，CBAM 模块可以灵活的掺差在残差块与卷积块之间。

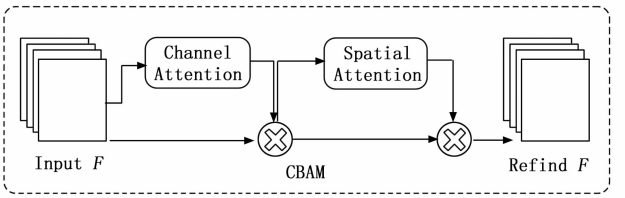


图 8　CBAM 模块示意图[15]

2.4　融合 NMS 与 WBF 的预测框生成

目标检测算法在对目标图片进行识别时，会生成一定数量的预测框。每个预测框都有其对应的置信度 C_i 与空间区域。YOLOv5 目标检测算法采用非极大值抑制 NMS[16]（NMS，non-maximum）的方法来对预测候选框进行筛选，其筛选策略是按照置信度对候选框进行排序，通过式（1）依次计算每个候选框与最大置信度候选框的交比并 IOU，若交比并 IOU 大于预设数值（YOLOv5 为 0.5），则认为该候选框与最大置信度候选框过重叠且置信度不足，删除该候选框；若交比并 IOU 小于预设数值，则将其保留，直到所有候选框都与最大置信度候选框比较完成。对单段多阶楼梯的检测中，会生成多个小交并比或者无交并比的候选框。为了满足对楼梯检测过程中目标与数量的对应匹配，需要对单段多阶楼梯的多个预测框进行融合。

加权边框融合 WBF[17]（Weighted Boxes Fusion）是一种有效的预测框融合方法，可以将多个预测框根据其置信度融合成新的预测框。WBF 的处理流程为将生成的预测框归一化处理并按置信度 C_i 排序；创建两个容器用于存放归一化的预测框与融合新生成的预测框；通过式（1）计算交并比 IOU 寻找匹配的两个预测框 Box_i、Box_j；将匹配的预测框按照边框融合式（6）和式（7）计算融合为新的预测框 $Boxf$（X_1，Y_1、X_2，Y_2 为边框对角点坐标）。WBF 加权边框融合可以通过调整交并匹配阈值来实现多阶楼梯的边框融合，但是 WBF 相较于 NMS 增大了计算量，降低了目标检测的实时性。

$$C=max(C_i),i=0,1,2\cdots \quad (5)$$

$$X_{1,2}=\frac{\sum_{i=0}^{N}C_i\cdot X_{1i,2i}}{\sum_{i=0}^{N}C_i} \tag{6}$$

$$Y_{1,2}=\frac{\sum_{i=0}^{N}C_i\cdot Y_{1i,2i}}{\sum_{i=0}^{N}C_i} \tag{7}$$

为了满足移动机器人目标检测过程中实时性与目标数量准确性的双重需求，本文融合了 NMS 与 WBF 预测框处理方法，将 NMS 筛选之后空间位置邻近的边框进行融合。为了描述两预测框 Box_i、Box_j 空间位置的邻近程度，本文提出了邻近比 $NX_{i,j}$ 的概念，邻近比 $NX_{i,j}$ 表达式如式（10）所示，其表示为当前两矩形边框 Box_i、Box_j 几何中心连线长度 $\delta_{cur\_i,j}$（式（8）所示）与 Box_i、Box_j 呈对角分布时几何中心连线长度 $\delta_{fix\_i,j}$（式（9）所示）的比值。如图 9 和图 10 所示，$\delta_{fix\_i,j}$ 是两矩形框的固有属性，只与 Box_i、Box_j 的几何大小有关，与矩形框位置无关；$\delta_{cur\_i,j}$ 则是描述两矩形框 Box_i、Box_j 在检测图片中空间分布的距离。

其处理流程如下所示：

1）输入测试图片 F，目标检测算法模型预测阶段生成 N 个候选定位框 Box_i（$i=0$，1，2…$N-1$），候选框 Box_j 置信度为 C_i，创建存储候选框的容器 V_1、V_2、V_3；

2）按照置信度 C_i 对 Box_i 进行升序排序，生成新的候选框序列 Box_i 并存储在容器 V_1 中；

3）依次比较 Box_i（$i=0$，1…$N-2$）与 Box_{N-1} 的交并比 $IOU_{1,N-1}$，若 $IOU_{1,N-1}$ 大于预设值（本文为 0.2），则将 Box_1 从容器 V_1 中删除；若 $IOU_{1,N-1}$ 小于等于预设值，将 Box_i 添加到容器 V_2 中并从 V_1 中删除；遍历比较完所有的 Box_i（$i=0$，1…$N-2$）之后，将 Box_{N-1} 添加到 V_2 中；

4）将 V_1 中的剩余候选框重复 3）的计算筛选过程，直到 V_1 遍历完中所有的 Box_i，此时 V_2 中存储的是 NMS 处理后的 n 个候选框；

5）将 V_2 中的候选框 Box_j（$i=0$，1…$n-1$）按照置信度 C_i 升序排列，遍历 Box_j（$i=0$，1…$n-2$）与 Box_{n-1} 进行条件判断；

6）若 Box_j（$i=0$，1…$n-2$）与 Box_{n-1} 的交并比 $IOU_{j,n-1}$ 大于 0 或邻近比 $NS_{i,j}$ 小于等于预设值（本文为 0.8），则将 Box_j 标记为邻近候选框 Box'_j，遍历完所有的 Box_j 之后将所有标记的 Box'_j 进行加权边框融合，将融合的结果存储到 V_3 中并将 Box'_j 从 V_2 中删除；

7）对 V_2 重复 6）过程，直到处理完毕所有的候选框，此时 V_3 中为最终的预测框。

$$\delta_{fix\_i,j}=\frac{1}{2}\left(\sqrt{(x_{i1}-x_{i2})^2+(y_{i1}-y_{i2})^2}+\sqrt{(x_{j1}-x_{j2})^2+(y_{j1}-y_{j2})^2}\right) \tag{8}$$

$$\delta_{cur\_i,j}=\frac{1}{2}\sqrt{(x_{i1}+x_{i2}-x_{j1}-x_{j2})^2+(y_{i1}+y_{i2}-y_{j1}-y_{j2})^2} \tag{9}$$

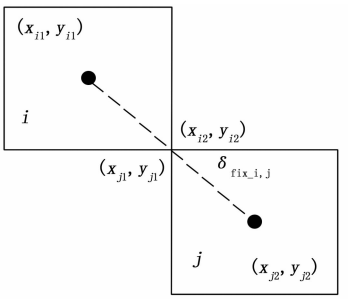


图 9 邻近比示意图 1

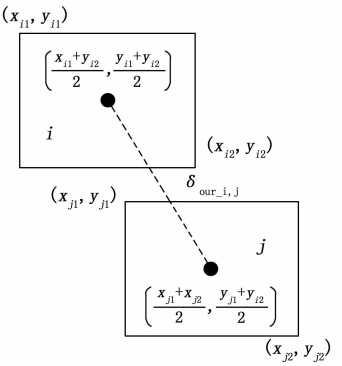


图 10 邻近比示意图 2

$$NS_{i,j}=\frac{\delta_{cur\_i,j}}{\delta_{fix\_i,j}} \tag{10}$$

以测试集中第 95 张图片为例，融合 NMS 与 WBF 的边框融合策略如图 11 所示，首先从将待检测图片输入到识别网络，识别网络经过一系列计算初步定为出 6 个候选框，通过 NMS 筛选出 3 个预测框，最后将这 3 个预测框进行加权融合，解决了单段长阶梯的“多检”问题。

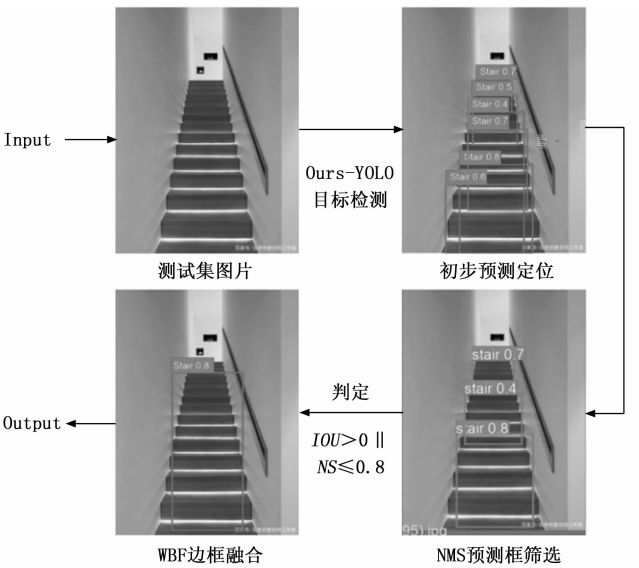


图 11 融合 NMS 与 WBF 的边框融合策略

3 仿真与结果分析

3.1 数据集准备

本文算法在自定义数据集上进行模型训练，自定义数

据集由：“ImageNet2012”No. 421 楼梯数据集、Pycharm 网络搜索下载以及现场拍摄的方式获得，总计获得 2 800 张左右的包含楼梯元素的图片素材。再将总计 2 800 多张图片进行筛选，将不符合室内环境要求以及移动机器人常规拍摄视角的图片删除，保留 1 534 张符合本文要求的图片作为训练模型的自定义数据集合。

如图 12 所示，数据集中包含室内楼道环境下的：常规楼梯、栏杆遮挡、墙壁遮挡、多阶楼梯、折转楼梯、不同材质（颜色）、不同视角楼梯等多种楼梯元素图片。将数据集图片统一调整为 640×640，并用 LabelImg 工具进行标注。标注要求为将独立连续楼梯视作一个单独检测目标，存在转向或平台等的多段楼梯视作多个检测目标则。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。

图 12　楼梯种类数据集

3.2　仿真配置

所用深度学习框架柱为 PyTorch、Python3. 8。硬件配置为：Intel i5 12400F、NVIDA GTX 2060、32GB 运行内存。训练轮数为 100 轮，batch 为 8，初始学习率为 0. 01。

3.3　评价指标

为了评价本算法对楼道环境下的楼梯检测效果，选取模型的平均精度 $mAP@0.5$、召回率 *Recall*、准确率 *Precision*、检测时间 Time 四项作为评价指标。其计算式如下：

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{11}$$

$$Recall = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{12}$$

$$Precision = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{13}$$

其中：AP 为 Precision-Recall 曲线的曲线积分，由于本算法为单目标检测模型，所以检测目标种类 N 为 1，即 $mAP=AP$；N_{TP}、N_{FF}、N_{FP}表示正确检测数、错误检测数以及漏检数目。

3.4　仿真结果分析

在测试数据集中选取了 5 张包含各类楼梯环境的测试图片，对检测实时性较好的二阶段目标检测算法 Faster-RCNN、一阶段目标检测算法 SSD、YOLOv5s、YOLOV5m、YOLOv5l 以及本文提出的改进的 Ours-YOLOv5s、Ours-YOLOv5m 进行测试，测试结果如图 13 和表 1 所示。

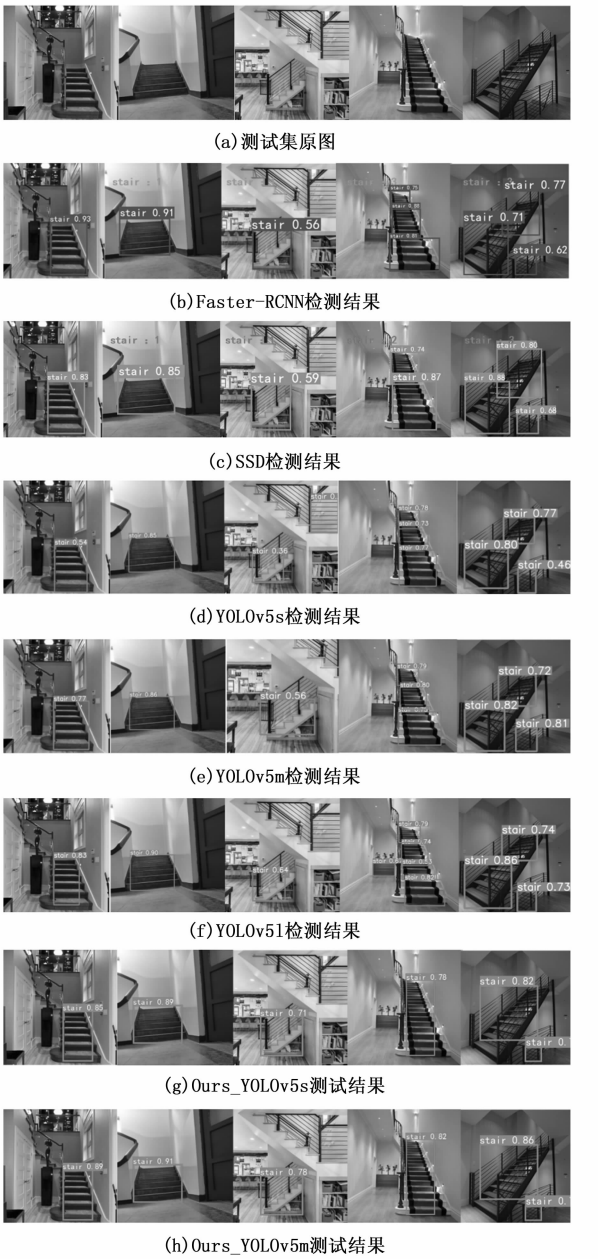


图 13　不同算法检测结果

表 1　目标检测算法测试结果

目标检测算法	*mAP*/%	*Recall*/%	Average Time/s	Modelsize/MB
Faster-RCNN	91.0	92.2	0.1120	159.412
SSD	87.2	84.6	0.0580	100.273
YOLOv5s	73.6	82.9	0.0133	14.073
YOLOv5m	79.1	73.1	0.0129	41.210
YOLOv5l	81.9	72.5	0.0125	90.641
Ours_YOLOv5s	82.9	84.4	0.0147	18.373
Ours_YOLOv5m	86.5	86.3	0.0151	45.511

结合图 13 及表 1 的检测结果可以看出：Faster-RCNN 的平均精度最高，达到 91%，但同时 Faster-RCNN 单张图片的平均推理时间最长且模型尺寸最大，单张图片推理时间达到 Ours _ YOLOv5s 单张推理耗时的 7.6 倍，模型尺寸达到 Ours _ YOLOv5s 的 8.7 倍，不满足移动机器人目标检测实时性与轻量级的检测要求；SSD 目标检测算法作为高精度一阶段目标检测算法的代表，其平均精度为 87.2%，单张推理时间为 Ours _ YOLOv5s 的 3.9 倍，模型尺寸为 Ours _ YOLOv5s 的 5.5 倍，虽然相比于 Faster-RCNN，SSD 大幅度加速了推理时间并降低了模型尺寸，但整体实时性与模型框架仍然较大，且 Faster-RCNN 与 SSD 均无法解决单段长楼梯的“多检”问题。YOLO 系列 3 种算法 YOLOv5s、YOLOv5m、YOLOv5l 相较于平均精度最高的 Faster-RCNN，平均精度分别下降了 17.4%、11.9%、9.1%，但推理时间仅为 Faster-RCNN 的 11%左右；面对复杂背景时，YOLOv5s 检测置信度较低且存在错检，随着网络结构加深加宽，检测置信度有了较为明显的提高；在面对带有栏杆遮挡的楼梯时候，3 种算法检测效果均较低；在面对单段长阶梯时，3 种算法均会产生多个检测框造成“多检”；在四号场景的识别中，YOLOv5l 出现错检；可以看出 YOLO 系列算法满足了实时性与轻量级的检测需求，但是整体检测精度较低，YOLOv5l 的模型尺寸为 YOLOv5m 的 2.2 倍，但平均精度小幅度提高了 2.8%，推理时间提高了 3%。Ours _ YOLOv5s 相比于 YOLOv5s，平均精度提高了 12.6%，模型大小增加了 30.6%，平均检测时间增加了 10.5%；Ours _ YOLOv5m 相比于 YOLOv5m，平均精度提高了 9.4%，模型大小增加了 10.4%，平均检测时间增加了 17.1%；Ours _ YOLOv5s 与 Ours _ YOLOv5m 在面对栏杆遮挡以及长台阶时均可以做出正确的预测结果。

综上所述，在移动机器人或是其他计算条件有限的嵌入式设备中，Ours _ YOLOv5s 可以做到在 18 MB 左右的轻量模型下实现 82.9%的较高精度楼梯目标识别，Ours _ YOLOv5m 比 Ours _ YOLOv5s 提升了 4.3%的平均精度，模型大小比 Ours _ YOLOv5s 增加了 27.138 MB。

4 结束语

针对室内环境中单段连续长楼梯存在的检测问题，改进了的 YOLOv5 识别算法通过在输入端引入 FenceMask 数据增强策略来替换 Masoic 数据增强策略，改善了栏杆、扶手对楼梯遮挡带来的识别率低的问题。通道注意力模块 CAM 与空间注意力模块 SAM 采用并行连接的方式组成注意力模块 CBAM，在骨干网络 Backbone 端引入双通道的 CBAM 注意力模块，实现通道与空间上对复杂背景下楼梯特征的双重聚焦，提高了楼梯特征的聚焦速度。在预测端提出了结合 NMS 与 WBF 的边框融合策略，将 NMS 筛选之后置信度较高且位置相邻的边框进行融合为新的边框。

在仿真模拟计算中，本文算法与 Faster RCNN 以及 SSD 目标检测算法进行了对比研究，测试结果表明，Faster-RCNN 的平均精度最高达到 91%，但单张图片推理时间达到本文算法单张推理耗时的 7.6 倍，模型尺寸达到本文算法的 8.7 倍，不满足移动机器人目标检测实时性与轻量级的检测要求。SSD 目标检测算法作为高精度平均精度为 87.2%，单张推理时间为本文算法的 3.9 倍，模型尺寸为本文算法的 5.5 倍。与上述两种算法相比，本文算法的平均精度为 82.9%，模型大小仅为 18 MB，单张平均推测时间为 0.014 7 s，可以满足了移动机器人在视觉 SLAM 楼梯建图过程中高精度、实时性以及轻量级的目标检测需求。

参考文献：

[1] 高 瑞. 基于图像特征的楼梯检测算法研究 [D]. 西安：西安科技大学，2017.

[2] 倪志鹏，李晓明. 复杂环境中楼梯检测问题研究 [J]. 电子测量技术，2019，42 (23)：158－163.

[3] KRIZHEVSKY A，SUTSKEVER I，HINTON G E. Imagenet classification with deep convolutional neural networks [C] // Advances in Neural Information Processing Systems，2012：1097－1105.

[4] GIRSHICK R，DONAHUE J，DARRELL T，et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] //Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (S9781479951178). Washington D. C.，USA，IEEE Press，2014：580－587.

[5] GIRSHICKR. FAST R-CNN [C] // IEEE International Conference on Computer Vision (ICCV)，Santiago，2015：1440－1448.

[6] REN S，HE K，GIRSHICK R，et al. Faster R-CNN：Towards real-time object detection with region proposal networks [C] // International Conference on Neural Information Processing Systems，MIT Press，2015：91－99.

[7] REDMON J，DIVVALA S，GIRSHICK R，et al. You only look once：Unified，real-Time object detection [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR)，Las Vegas，NV，2016：779－788.

[8] WEILIU，DRAGOMIR ANGUELOV，et al. SSD：single shot multiboxdetector [C] //Proceedings of European Conference on Computer Vision (S 978-3-319-46448-0). Berlin，Germany：Springer，2016：21－37.

[9] 郝 帅，杨 磊，马 旭，等. 基于注意力机制与跨尺度特征融合的 YOLOv5 输电线路故障检测 [J/OL]. 中国电机工程学，2022：1－12. http://kns.cnki.net/kcms/detail/11.2107.tm.20220126.1718.008.html.

[10] 张 翔，唐小林，黄岩军. 道路结构特征下的车道线智能检测 [J]. 中国图象图形学报，2021，26 (1)：123－134.

[11] 董乙杉，李兆鑫，郭靖圆，等. 一种改进 YOLOv5 的 X 光违禁品检测模型 [J/OL]. 激光与光电子学进展，2022：1－17. http://kns.cnki.net/kcms/detail/31.1690.TN.20220217.1141.008.html

(下转第 79 页)