基于改进YOLO-v4的室内人脸快速检测方法

巢 渊,刘文汇,唐寒冰,马成霞,王雅倩
江苏理工学院 机械工程学院,江苏 常州 213000

摘 要:针对室内安防工程应用中检测人脸角度不同、光照变化、部分遮挡、模糊等复杂工况,提出一种基于改进YOLO-v4的室内人脸快速检测方法。基于深度可分离残差网络结构改进YOLO-v4主干网络,提升模型检测效率;在构建特征金字塔过程中引入注意力机制,自适应调整通道特征与空间特征权重,提升模型特征提取能力。实验结果表明,该方法对室内人脸图像的检测精度与速度分别为92.53%与35 frame/s,相比原YOLO-v4算法及其他主流人脸检测算法,具有更好的检测精度与效率,因此可应用于移动机器人的室内人脸实时检测。

关键词:深度学习;特征融合;人脸检测;注意力机制
文献标志码:A **中图分类号:**TP391.4 doi:10.3778/j.issn.1002-8331.2201-0309

Fast Indoor Face Detection Method Based on Improved YOLO-v4

CHAO Yuan, LIU Wenhui, TANG Hanbing, MA Chengxia, WANG Yaqian
School of Mechanical Engineering, Jiangsu University of Technology, Changzhou, Jiangsu 213000, China

Abstract: A fast indoor face detection method based on improved YOLO-v4 is proposed in this paper, under the complex working conditions of human faces with different angles, varying illumination, partial occlusion, and blurring in indoor security engineering applications. The YOLO-v4 backbone network is improved based on the deep separable residual network structure to increase the detection efficiency of the model. The attention mechanism is introduced during the process of constructing the feature pyramid, which can adaptively adjust the weights of the channel features and spatial features, to improve the feature extraction capability of the model. The experimental results show that the accuracy and speed of the proposed method are 92.53% and 35 frame/s, respectively, for indoor face images, which has relatively better detection precision and efficiency, compared with the original YOLO-v4 algorithm, and other mainstream face detection algorithms. The proposed method therefore can be applied to indoor face detection of mobile robots in real time.

Key words: deep learning; feature fusion; face detection; attention mechanism

随着科技的不断进步和市场应用需求的不断推动,人脸检测技术与日常生活的关联日益密切。目前,视觉人脸检测技术及其应用在全世界范围内得到了快速发展与普及,如:移动支付、交通安全、视频监控、门禁识别与情绪识别等。当前人脸检测算法的研究主要围绕基于特征的传统人脸检测算法和基于深度学习的人脸检测算法两方面展开。传统人脸检测算法通过图像处理提取特征,通过分类器进行分类[1-3]。但该类算法依赖人为设计特征,因此在精度与效率上都有一定局限性。基于深度学习的人脸检测算法通过神经网络对大量数据的学习与分析找到人脸与非人脸之间的关系以完成人脸检测。深度学习人脸检测算法可分为两大类:两步检测算法与单步检测算法。两步人脸检测算法中最常用的包括R-CNN(regions with convolutional neural network features)、Fast-R-CNN以及Faster-R-CNN[4-6],总体来说,该类算法可实现较为准确的目标定位与识别,但识别过程中候选区域的反复选择会影响图像总体的检测效率。

单步人脸检测算法中最具代表性的包括MTCNN(mutil-task convolutional neural network)[7]、RetinaFace[8]和YOLO(you only look once)系列算法[9-11]等。MTCNN采用三个独立网络模块(P-Net、R-Net和O-Net)级联的方式检测与定位人脸关键点,速度较快,但其精度较低,常用于人脸主动检测与定位场合[12-13]。RetinaFace也是目前主流的人脸检测算法之一,精度方面表现较优秀,

基金项目:国家自然科学基金(51905235);江苏省自然科学基金(BK20191037);江苏省中以产业技术研究院开放课题(JSIITRI202101);江苏省研究生实践创新计划项目(SJCX20_1045);江苏理工学院研究生实践创新计划项目(XSJCX20_32)。

作者简介:巢渊(1988—),男,博士,讲师,研究方向为机器视觉检测与测量、机电一体化装备智能控制技术等,E-mail:chaoyuan@jsut.edu.cn。

收稿日期:2022-01-19 **修回日期:**2022-04-07 **文章编号:**1002-8331(2022)14-0105-09

但模型运算量较大[14]。YOLO系列算法省略了通过滑窗选择候选区域的步骤,直接将整幅图像输入网络中,通过深度神经网络进行一次前向传播,使用非极大值抑制后直接输出识别结果。得益于出色的目标检测效率,该算法被广泛应用于缺陷检测、仪表检测与鸟类检测等场景[15-17]。近年来YOLO系列算法正逐渐成为人脸检测领域的研究与应用热点,如文献[18]采用YOLO-v2算法与ResNet算法完成监控视频中的人脸检测与识别,算法在真实场景中检测速度为21 frame/s,能够满足实际检测需求,其不足之处在于被检人脸倾斜角度较大或模糊时易出现漏检情况。文献[19]通过将多尺度回归思想应用于YOLO模型,实现人脸年龄估计,有效提高了算法对小尺寸目标的提取能力。对于密集小尺寸人脸的检测,文献[20]通过改进目标框聚类算法与对不同层级特征图进行细粒度特征融合,提出一种改进YOLO的人脸检测方法,提升算法对小尺度人脸的检测精度。安防场景中的人脸检测多为非主动检测,部分检测对象会存在遮挡问题,如口罩、帽子和眼镜等。文献[21]针对口罩人脸检测容易出现误检、漏检等问题,通过增加特征层与多尺度融合的方式,提出一种基于改进YOLO-v3的人群口罩佩戴检测算法,增强口罩遮挡下人脸检测的精度。文献[22]提出一种融合环境特征与改进的安全帽佩戴检测方法。针对复杂多变的环境因素导致检测准确率降低等情况,利用数据增强以及对抗训练等方法对YOLO-v4进行了改进,提升了算法的准确率,使其在真实环境下有较稳定的表现。

综上所述,YOLO系列算法在人脸检测方面已取得较大的进展,但在室内安防应用场景下,人脸检测的精度易受被检人脸角度[23]、光照变化[24]、部分遮挡[25]、模糊[26]等方面因素的影响,且大部分主流算法网络存在结构复杂、运算量大等问题,难以适应嵌入式设备实时检测需求。因此,如何平衡室内安防场景下人脸检测的准确率与效率,仍是一个值得研究的问题。针对上述问题,本文提出一种基于YOLO-v4的室内人脸快速检测方法。建立包含人脸不同角度、光照变化、部分遮挡、模糊等工况的数据集;基于深度可分离残差网络改进主干网络;在特征金字塔中引入注意力机制,自适应调整不同像素通道特征与空间特征权重。最后,通过实验对比与分析,验证本文方法的人脸检测精度与效率。

1　总体框架

YOLO-v4算法因其可在一次扫描中直接通过网络输出目标的位置与类别信息,相比其他类型算法具有更快的检测速度,且室内安防对人脸检测算法的效率要求较高,因此本文基于YOLO-v4[11]基础框架,研究提出一种适用于室内安防工程应用中的人脸快速检测算法。其思路描述如下,流程如图1所示。

(1)移动安防机器人实时采集、存储室内监控视频,并将序列图像输入本文人脸检测模型。

(2)将输入图像平均分成多个感兴趣区域,每个区域对应多个预测框。

(3)将图像输入主干网络进行特征提取,通过颈部网络进行特征融合,获取有效人脸特征。

(4)输出每个预测框的类别、位置以及置信度(*Confidence*)值,通过头部网络进行非极大值抑制运算,去除重合度(intersection over union,*IOU*)较大及置信度较低的预测框,输出检测结果。置信度与*IOU*计算公式如下:

$$Confidence = Pr(Face)IOU_{\text{pred}}^{\text{truth}} \tag{1}$$

$$IOU_{\text{pred}}^{\text{truth}} = \frac{Detection \cap GroundTruth}{Detection \cup GroundTruth} \tag{2}$$

其中,$Pr(Face)$表示是否有人脸的中心落在某栅格中,有则取1,反之取0;*IOU*表示预测框和标记框的重合度大小;*Detection*为预测人脸框;*GroundTruth*为样本中已标记的人脸框。

YOLO-v4模型主要由主干网络、颈部网络与头部网络等构成,网络结构如图2所示。其中主干网络部分主要完成特征提取,原YOLO-v4主干网络为基于残差

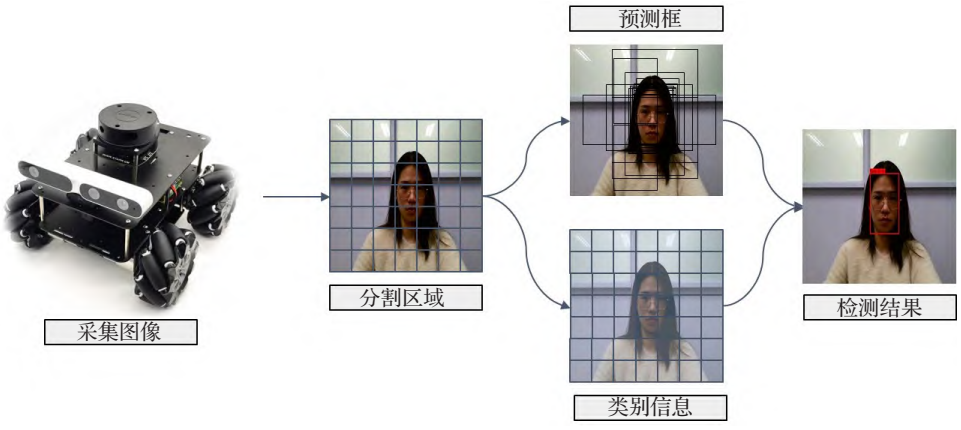


图1　人脸检测流程

Fig.1　Process of face detection

神经网络改进的Darknet-53,由1个普通的卷积层加23个不同维度的残差神经网络模块组成,存在结构复杂、运算量大的问题。本文改进的主干网络应用残差网络结构可有效减少模型的运算量。原YOLO-v4颈部网络部分通过对三个大小不一的特征图进行张量拼接,构建特征金字塔,使网络同时包含高层抽象特征与底层位置信息特征。本文改进的颈部网络在构建特征金字塔时引入通道与空间注意力模块,通过轻量级注意力模型调整不同像素权重,优化特征金字塔,提升模型特征提取能力。头部网络主要进行最终检测值的回归与预测。下文将对本文方法的改进部分进行详细描述。

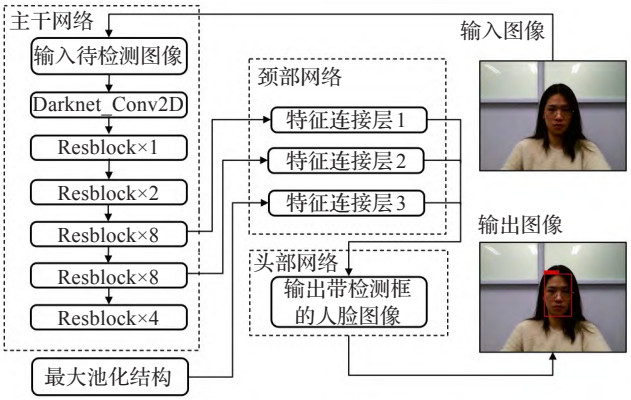


图2　基于YOLO-v4的人脸检测算法结构

Fig.2　Algorithm structure of face detection based on YOLO-v4

2　算法描述

本文网络具体结构如图3所示。待检测图像首先经过统一缩放变成416×416×3大小,进行感兴趣区域分割,再输入主干网络进行特征提取。进行特征融合的颈部网络主要由融合层(concat)、普通卷积块结构(CBL)、卷积层(COV)以及深度可分离卷积(DW)组合而成。头部网络的检测器对颈部网络得到的3个有效特征层结果进行堆叠,得出每个预测框最终置信度,通过非极大值抑制算法输出包含有预测框与置信度数据的图像。

2.1　数据集制作

本文主要研究室内安防场景下的非主动人脸检测问题,因此针对人脸角度不同、光照变化、部分遮挡、模糊等常见的五种不同工况,对网络公开的CelebFaces[27]、WIDER FACE[28]等人脸图像数据进行筛选、分类及标注,建立人脸数据集。数据集中的部分图片如图4所示,图(a)、(c)分别为不同角度人脸与部分遮挡人脸,此类图像同时存在人脸变形,且易丢失部分人脸结构特征;从图(b)中可看出受光照变化影响的人脸图像,其纹理特征丢失较多;图(d)为模糊人脸图像,通常表现为人脸面积占比较小或处于运动状态,其结构与纹理特征较弱,因此该类人脸检测依赖更深度的特征。

(a)角度不同　(b)光照变化　(c)部分遮挡　(d)模糊

图4　人脸数据集图片

Fig.4　Face images in facial dataset

由图4可以看出,本文数据集中不同图像中人脸面积变化相对较大,为得到适合本文研究对象的人脸预测框,首先应用K-means++算法[29]对数据集中人脸宽度及高度进行聚类分析,可得出9个预测框,尺寸分别为12×

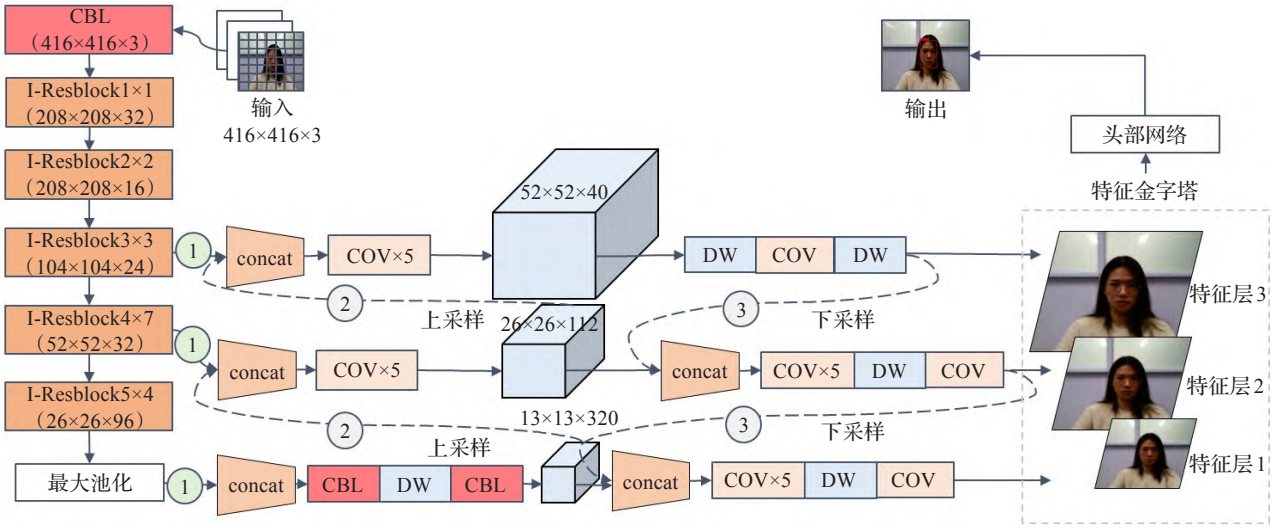


图3　本文网络结构

Fig.3　Proposed network structure

16、19×36、40×28、36×75、76×55、72×146、142×110、192×243和459×401。

针对图4(a)、(c)两类图像中的部分人脸缺失、整体结构不完整的问题,引入马赛克数据增强算法[11]进行数据集的扩充。马赛克数据增强算法的主要思想为:在数据集中随机挑选4幅图片,对其进行剪裁、缩放以及重新拼接,组成1幅新的图片,可以增强网络对人脸结构不完整时的鲁棒性。制作完成的人脸数据集主要包含5类图像:清晰人脸、角度不同人脸、光照变化人脸、部分遮挡人脸以及模糊人脸。其中,清晰人脸2 136幅,角度不同人脸2 560幅,光照变化人脸2 792幅,部分遮挡人脸2 944幅,模糊人脸3 092幅。其中有些图像中包含不止一类人脸,因此数据集包含总数为13 524幅的人脸图像,其中重复样本数量为3 524张。将其划随机分为训练集8 524张、测试集2 500张、验证集2 500张进行训练。

2.2　主干网络改进

2.2.1　激活函数

原YOLO-v4主干网络由1个单次卷积模块和一系列残差网络结构组成。卷积模块又由1个卷积层(COV),1个归一化层(BN)组成。其中,卷积层采用了Mish激活函数作为激活函数[11]。为保障室内人脸快速检测的实时性与准确率,本文将如式(3)所示的卷积块中运算量较大的Mish激活函数改进为ReLU6,如式(4)所示:

$$Mish(x)=x\times\tanh(\ln(1+e^{x})) \tag{3}$$

$$ReLU6=\min(6,\max(0,x)) \tag{4}$$

由式(1)和式(2)可以看出,ReLU6激活函数为线性,而Mish激活函数为非线性,理论上Mish函数比ReLU6函数的梯度下降效果更好,但同时也会增加运算的复杂程度,导致检测速度相对变慢。本文通过YOLO-v4搭建轻量级人脸检测网络,检测类别较少,使用ReLU6函数可使模型收敛的同时加快模型收敛速度。

2.2.2　深度可分离残差网络

原YOLO-v4主干网络中采用了残差网络结构[17]其具体过程为:输入数据首先经过步长为2×2的基础卷积层,改变维度;接着分为两部分,一部分作为主干部分(Resblock)在循环中进行迭代,得出权重与输入数据的运算关系,另一部分建立独立的残差边,将输入数据进行少量处理后直接输出;最后对两部分输出数据进行跨层相加,将求和结果作为本层的输出,过程如图5所示。采用该结构是为了通过分开梯度流,使梯度流在不同路径上传播,使网络学习到更多梯度流的相关性差异;同时通过减少循环堆叠计算量,以降低算力消耗,提升运算速度和网络的学习能力。

为进一步提升网络运算速度,满足人脸在线检测任务,本文将原YOLO-v4主干网络中的残差网络结构改进为深度可分离残差网络结构,具体结构如图6所示。此结构延续了YOLO-v4分开梯度流的思想,将一部分

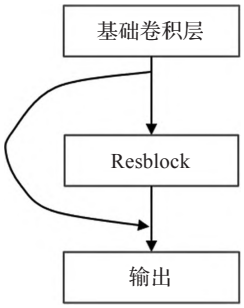


图5　残差块结构

Fig.5　Residual structure

输入数据继续循环迭代,另一部分数据跳接到最后。再将原来循环迭代的普通卷积残差块替换为深度可分离残差块,深度可分离残差块具体过程为:首先通过1×1的普通卷积块(CBL)扩张通道以便特征提取,然后引入深度可分离卷积(DW)进一步减少模型运算量,最后使用1×1的普通卷积块(CBL)进行降维,以提升后续网络的计算效率。

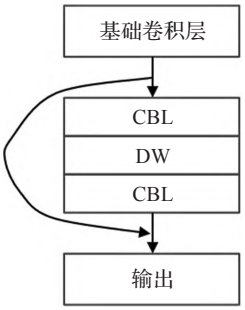


图6　深度可分离残差块结构

Fig.6　Depthwise separable residual structure

残差网络结构中,假设输入图片通道数为 C_{in},输出通道为 C_{out},普通卷积层大小则为 $N\times N\times C_{in}$。通道数为 C_{in} 的普通卷积层对输入量进行卷积计算,则普通卷积一次运算过程中,某卷积点参数量计算公式为:

$$M=N\times N\times C_{out}\times C_{in} \tag{5}$$

本文深度可分离残差网络结构中,深度可分离卷积一次运算具体过程为:假设输入图片通道数为 C_{in},输出通道为 C_{out},首先通过 $N\times N$ 深度卷积按 C_{in} 个不同通道分别对输入图片进行按位相乘的计算,得到第一步结果,此处图片宽高变化,但通道数不变;再使用 $1\times1\times C_{in}$ 普通卷积核对第一步结果进行卷积运算,此时结果为 $1\times1\times X$,数据维度也可按照需求调整为 $1\times1\times Y\times Z$。则深度可分离卷积的某卷积点参数量计算公式为:

$$M=C_{in}\times N\times N+C_{out}\times C_{in}\times1\times1 \tag{6}$$

综上所述,本文主干网络中采用的深度可分离残差网络模块输出通道数可控,且理论上可有效减少主干网络中模型体积与参数量。本文主干网络位于图3最左侧,从上到下共18个模块,由1个普通卷积块结构(CBL)和17个步长不一的深度可分离残差网络结构(I-Resblock)组成。这种设计可降低本文模型总体的运算量,从而提

高自然环境下人脸检测的速度,有利于实时人脸的快速检测。

2.3　注意力模块

在现实人脸检测场景中,一幅图片上不同位置的像素重要性可能不同,不同通道的像素重要性也可能不同。因此,引入注意力机制,可用特定网络调整不同像素对检测结果的影响力,从而分离出更显著的特征。为进一步融合主干网络提取到的3个尺度特征图,分离出更显著的特征,在颈部网络引入注意力机制,其具体思路为:首先学习特征图中不同位置或不同通道的重要性权值,然后将学习到的重要性权值与原特征图中值相乘,输出新特征图。当前应用广泛的注意力机制包括SENET(squeeze-and-excitation net)[30]与CBAM(convolutional block attention module)[31]等,其中SENET为通道注意力机制模块,CBAM则为结合了通道与空间的注意力机制模块。通道注意力模块保持通道维度不变,只压缩空间维度,因此该模块对人脸类别信息敏感;空间注意力模块空间维度不变,压缩通道维度,因此该模块对人脸位置信息敏感。综合考虑室内人脸检测的复杂工况,本文引入CBAM模块,可分离出更显著的特征,提高人脸检测的准确率。CBAM由1个通道注意力模块与1个空间注意力模块串联而成,其结构如图7所示。

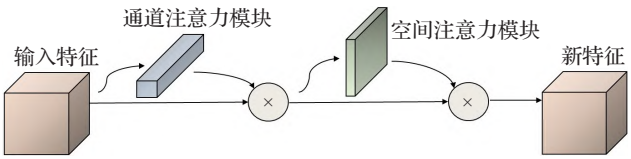


图7　CBAM注意力模块

Fig.7　Convolutional block attention module

本文先对主干网络引出的3个不同尺度特征层间进行上采样和下采样,然后搭建特征金字塔。具体过程如下:对于特征层1,由13×13×320特征图与26×26×112浅层特征图下采样特征融合而成;对于特征层2,由13×13×320深层次特征图特征与26×26×112特征图融合,再与52×52×40浅层特征图下采样融合;对于特征层3,由52×52×40特征图与26×26×112深层特征图上采样特征融合而成。接着引入注意力机制,通过轻量级注意力模型调整图像特征的权重,最终提取出有效的特征,即特征层1、特征层2以及特征层3。最后,将3个特征层送入头部网络,得出每个预测框的最终置信度,使用非极大值抑制算法去除*IOU*较大与置信度较低的预测框,输出带预测框与置信度的图像。

在图3网络中标记的①、②、③三类不同位置,进行7组CBAM注意力模块插入的实验对比,不同插入位置及对应网络名见表1,其中位置①为主干网络输出端,位置②为上采样,位置③为下采样。下文将对7组不同位置注意力模块对模型训练及检测的影响进行对比分析。

表1　不同位置插入注意力模块

Table 1　CBAM insertion at different positions

网络	CBAM插入位置	CBAM数量
YOLO-v4	无	0
改进YOLO-v4	无	0
改进YOLO-v4-Attention1	①	1
改进YOLO-v4-Attention2	②	1
改进YOLO-v4-Attention3	③	1
改进YOLO-v4-Attention12	①、②	2
改进YOLO-v4-Attention13	①、③	2
改进YOLO-v4-Attention23	②、③	2
改进YOLO-v4-Attention123	①、②、③	3

3　实验结果与分析

本文在2.1节建立的包含有13 524幅图片的数据集基础上,进行模型的训练。训练过程中采用迁移学习[32]的思想以提高训练效率。具体训练过程如下:

(1)使用大型数据集上的预训练权重对网络前152层进行冻结训练,以调整非主干网络的权重。冻结训练采用一次训练10个样本的小批量随机梯度下降法,以避免网络运算量过大。

(2)采用学习率动态衰减方式,设置初始学习率为0.001,连续10个迭代后模型性能不提升,则学习率减少一半继续训练,以防止模型过拟合。

(3)解冻网络所有的层进行解冻训练,解冻训练参数量较大,一次选取4个样本,其余参数与冻结训练相同。

本文实验平台图像处理单元采用GeForce RTX 1060显卡,显存为8 GB。操作系统为Windows 10,深度学习框架为Tensorflow 1.15,GPU加速工具为CUDA 8.0。搭建移动安防机器人,并采集室内监控视频,进行真实场景下的人脸检测实验验证。

为较为正确地评价算法性能的优劣,本文引入的主要评价指标包括召回率(*Recall*)、准确率(*Precision*)和平均准确率(average precision,*AP*)、浮点运算数(floating point operations,*FLOPs*)等。其中召回率指被正确检测出人脸占验证集中所有人脸的比例;准确率指被正确检测出人脸占检测出人脸的比例。召回率、准确率、平均准确率表达式分别如式(7)、(8)、(9)所示:

$$Recall=\frac{TP}{TP+FN}\times 100\% \tag{7}$$

$$Precision=\frac{TP}{TP+FP}\times 100\% \tag{8}$$

$$AP=\sum_{k=1}^{n}J(Precision,Recall) \tag{9}$$

其中,*TP*(true positives)表示检测到人脸,且实际图片中也存在人脸的样本个数;*FP*(false positives)表示检测到人脸,但实际图像中不存在人脸的样本个数;*FN*(false negatives)表示未检测到人脸,但实际图像中存在人脸的样本个数。$J(Precision,Recall)$为平均精度函

数,计算方式为准确率 *Precision* 与召回率 *Recall* 构成的 P-R 曲线下方面积。

浮点运算数的计算公式如式(10)所示:

$$FLOPs=(2C_{in}K^2+1)HWC_{out} \tag{10}$$

其中,C_{in} 为输入通道数,K 为卷积核大小,H、W 分别为输出特征图的尺寸,C_{out} 为输出通道数。

3.1 定量分析

在如2.1节所描述的验证集上进行本文方法与原YOLO-v4模型的定量测试对比,以验证本文网络结构改进及引入注意力机制的有效性。不同网络及不同种类人脸的准确率与平均准确率 AP 数据分别如表2与表3所示。

表2　人脸检测准确率

Table 2　Precision of face detection　%

网络	清晰人脸	角度不同	光照变化	部分遮挡	模糊
YOLO-v4	89.20	87.95	84.73	83.80	76.74
改进YOLO-v4	94.48	89.03	86.09	88.42	81.20
改进YOLO-v4-attention1	**97.38**	**91.36**	**89.66**	**90.34**	**84.77**
改进YOLO-v4-attention2	**98.87**	**96.05**	**95.09**	**92.30**	**89.72**
改进YOLO-v4-attention3	86.93	82.21	79.51	80.81	77.32
改进YOLO-v4-attention12	89.78	86.86	86.60	85.05	82.28
改进YOLO-v4-attention13	85.02	84.36	80.82	83.66	79.58
改进YOLO-v4-attention23	77.20	71.23	71.38	72.32	71.64
改进YOLO-v4-attention123	70.36	67.51	68.23	67.16	67.01

表3　人脸检测平均准确率 *AP*

Table 3　*AP* of face detection　%

网络	清晰人脸	角度不同	光照变化	部分遮挡	模糊
YOLO-v4	82.34	64.16	66.71	58.39	51.13
改进YOLO-v4	93.60	86.11	84.99	75.91	81.49
改进YOLO-v4-attention1	**94.11**	**88.80**	**87.81**	**78.78**	**69.12**
改进YOLO-v4-attention2	88.81	73.05	78.69	65.12	67.79
改进YOLO-v4-attention3	68.89	53.12	61.30	55.10	49.94
改进YOLO-v4-attention12	71.73	55.10	60.07	55.18	46.40
改进YOLO-v4-attention13	61.23	42.33	39.70	44.61	37.45
改进YOLO-v4-attention23	55.95	46.94	46.80	49.14	46.77
改进YOLO-v4-attention123	56.57	42.13	46.47	40.83	38.46

由表2可以看出,改进后的YOLO-v4网络相比未改进网络,在各类人脸的检测准确率上都有所提升。其中,清晰人脸与遮挡人脸两个类别的准确率提升最为明显。引入改进YOLO-v4-Attention1与改进YOLO-v4-Attention2注意力机制后,各类人脸检测准确率都有进一步提升。而改进YOLO-v4-attention3插入的注意力模块位于下采样中,提取到的特征相对更少,因此引入YOLO-v4-attention3后模型的准确率较低。进一步地,在依次进行不同注意力模块的叠加后,模型的准确率也开始降低。

由表3可以看出,依次加入不同注意力模块后,改进模型的平均准确率开始降低。其中,改进YOLO-v4-Attention12相对于改进YOLO-v4-Attention1各类别 AP 均降低20%以上,而改进YOLO-v4-Attention123相对于其他模型的 AP 值则更低。虽然YOLO-v4-Attention2具有最高的人脸检测准确率,但其召回率较低,因此从 AP 的角度分析,YOLO-v4-Attention1的总体性能更佳。总体来看,改进后的网络能更好地识别各类人脸,尤其是清晰人脸与遮挡人脸,且在图3位置①主干网络输出端插入1组CBAM注意力模块,可使模型的综合性能进一步提升。

引入损失函数(loss function)判断当前迭代轮次内的模型训练状态,计算公式为:

$$Loss=l_{object}+l_{box}+l_{class} \tag{11}$$

其中,l_{object}、l_{box} 与 l_{class} 分别为置信度损失、回归框损失与类别损失。图8为不同改进网络的训练、验证损失曲线图,训练损失(*loss*)曲线、验证损失(*val*-*loss*)曲线分别为训练集上的损失曲线与验证集上的损失曲线。各网络的训练、验证损失曲线最终波动都较小,说明网络稳定性都较好。其中改进YOLO-v4在第180迭代轮次时损失曲线不再下降,模型收敛完成。改进YOLO-v4-Attention3在第145迭代轮次时损失曲线不再下降,具有最快的收敛速度。而YOLO-v4-Attention1与YOLO-v4-Attention2的收敛速度仅次于YOLO-v4-Attention3,且在叠加不同注意力模块后,模型收敛速度逐渐降低。因此,引入注意力机制,可有效提升模型整体准确率与收敛速度,且引入1组注意力模块的YOLO-v4-Attention1、YOLO-v4-Attention2与YOLO-v4-Attention3具有更快的收敛速度。

表4为5种网络在验证集上的准确率、召回率与不同人脸平均准确率均值 mAP(mean average precision)。改进YOLO-v4-Attention1具有总体最高的 mAP 值,为86.82%,说明该算法的准确率与召回率能够达到较好的平衡。而改进YOLO-v4-Attention12及改进YOLO-v4-Attention123,由于关注了更多特征,取得了较高召回率,但同样存在干扰特征导致准确率较低,模型综合性

表4　不同网络的检测结果对比

Table 4　Comparison of detection results of different networks　%

网络	Precision	Recall	mAP
YOLO-v4	86.44	43.90	65.73
改进YOLO-v4	91.54	37.96	77.65
改进YOLO-v4-Attention1	92.53	57.77	**86.82**
改进YOLO-v4-Attention2	93.87	42.65	73.81
改进YOLO-v4-Attention3	52.92	61.55	56.94
改进YOLO-v4-Attention12	63.59	74.24	58.29
改进YOLO-v4-Attention13	58.02	46.67	51.23
改进YOLO-v4-Attention23	44.64	57.98	49.77
改进YOLO-v4-Attention123	40.40	93.75	45.08

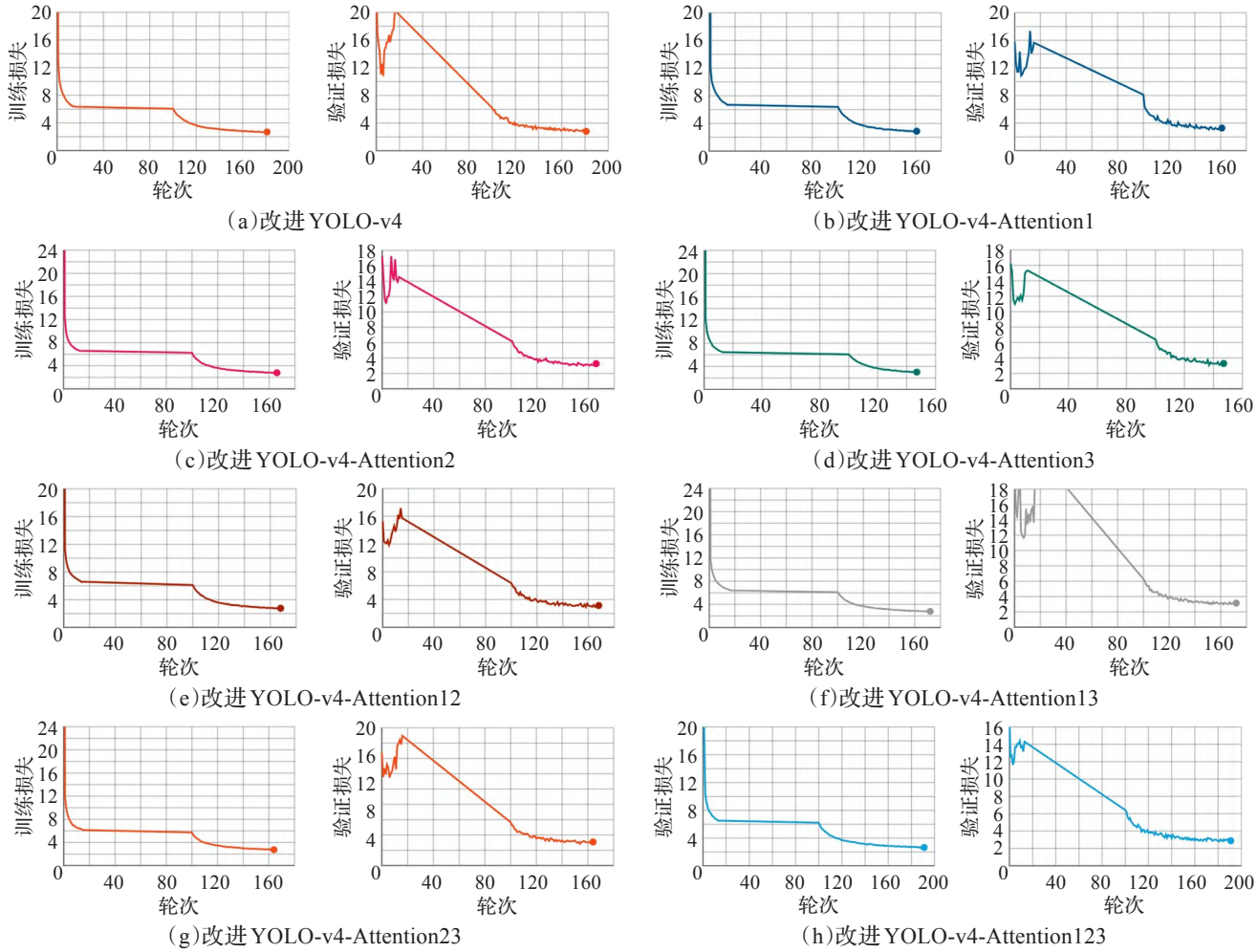


图8　Loss曲线

Fig.8　Curves of Loss

能欠佳。因此,本文采用仅在主干网络输出端插入注意力模块(图3位置①)的改进YOLO-v4,即改进YOLO-v4-Attention1,为最终的人脸检测算法。

为进一步体现本文算法在检测精度及检测效率方面的优越性,选取Faster-R-CNN[4]、MTCNN[7]、Retina-face[8]以及YOLO-v4[11]等主流人脸检测算法在验证集上进行对比测试,结果如表5所示。本文算法在验证集上的准确率为92.53%,相较于Faster-R-CNN提高了15.58个百分点,相较于MTCNN提高了10.96个百分点,相较于Retina-face提高了0.86个百分点,相较于原YOLO-v4提高了6.09个百分点。本文算法的召回率为57.77%,略高于Retina-face,且明显优于原YOLO-v4。

表5　人脸检测算法性能对比

Table 5　Performance comparison of different face detection methods

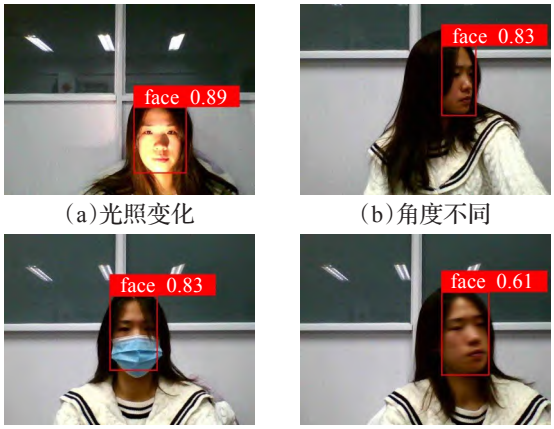
网络	*Precision*/%	*Recall*/%	*FLOPs*/10^9	参数量/MB	单次运行时间/ms
Faster-R-CNN[4]	76.95	40.83	0.270	95	253
MTCNN[7]	81.57	40.46	0.001	2	157
Retina-face[8]	91.67	56.42	0.048	105	85
YOLO-v4[11]	86.44	43.90	0.120	246	41
改进YOLO-v4	91.54	37.96	0.022	40	30
本文算法	**92.53**	**57.77**	**0.022**	41	**28**

从表5中浮点运算数FLOPs指标可以看出,本文的改进算法显著降低了YOLO-v4的计算量,且引入注意力机制不会使模型计算量有明显的提升。表5中参数量数据反应了完整模型参数的存储大小。本文算法的FLOPs与参数量虽大于MTCNN,但其检测准确率与检测速度明显优于MTCNN。故本文方法综合能力优于其他算法,可更好地应用于室内安防场景下实时人脸检测与识别。

3.2　定性分析

为进一步验证本文算法在实际室内安防场景下的检测效果,搭建如图1所示移动安防机器人进行真实室内场景下的人脸检测实验验证。移动安防机器人下位机采用树莓派4B主控,上位机采用GeForce RTX 1060显卡,摄像头采用乐视LeTMC-520。机器人下位机通过摄像头完成室内监控视频的采集与存储。将监控视频序列图像由TCP/IP协议传输到上位机,输入人脸检测模型,进行人脸的实时检测验证。验证图像数据即来源于机器人采集的室内监控视频中的序列图像。本文

算法与原YOLO-v4的部分检测效果分别如图9、图10所示。从测试结果图中可以看出,本文方法在室内安防场景下对光照变化、角度不同、部分遮挡、模糊人脸检测的置信度皆高于原YOLO-v4,尤其对于如图9(c)所示的部分遮挡人脸的图像样本,置信度提升较高。因此本文方法对实际室内人脸的实时检测整体效果要优于原YOLO-v4。

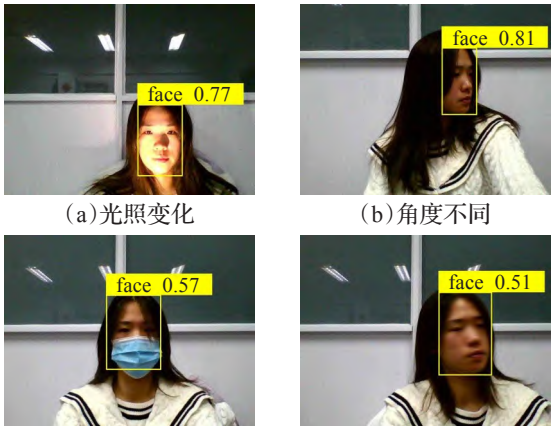


(a)光照变化　(b)角度不同

(c)部分遮挡　(d)模糊

图9　本文算法检测效果

Fig.9　Detection results of proposed method



(a)光照变化　(b)角度不同

(c)部分遮挡　(d)模糊

图10　原YOLO-v4算法检测效果

Fig.10　Detection results of original YOLO-v4

表6为不同人脸检测算法在实际室内安防场景中的每秒检测帧数(frames per second,FPS)(最低值)对比。由表6可知,本文方法在实际室内场景中的人脸检测FPS为35左右,而原YOLO-v4为30左右。因此,本文算法具有更高的实时性,可满足室内安防等场景下的人脸快速检测需求。

表6　人脸检测算法帧数对比

Table 6　FPS comparison of different face detection methods

网络	每秒帧数FPS
Faster-R-CNN[4]	5
MTCNN[7]	7
Retina-face[8]	10
YOLO-v4[11]	30
本文算法	**35**

4　结束语

本文提出了一种基于改进YOLO-v4的室内人脸快速检测方法。针对室内安防工程应用中检测人脸角度不同、光照变化、部分遮挡、模糊等复杂工况,制作人脸数据集;在特征提取阶段引入深度可分离残差网络结构改进主干网络,应用可分离残差网络结构提升模型检测效率;在构建特征金字塔阶段引入注意力机制,通过轻量级注意力模型优化特征金字塔,提高模型特征提取能力。本文方法与原YOLO-v4、Faster-R-CNN、MTCNN以及Retina-face等主流人脸检测算法的对比实验表明,本文方法精度均值为92.53%,检测速度达到35 frame/s,具有更高的检测与检测效率,可应用于室内移动机器人实现人脸实时检测。

参考文献:

[1] VIOLA P,JONES M.Rapid object detection using aboosted cascade of simple features[C]//Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,2001:511-518.

[2] IENHART R,MAYDT J.An extended set of Haar-like features for rapid object detection[C]//Proceedings of International Conference on Image Processing.New York:IEEE,2002:900-903.

[3] 胡丽乔,仇润鹤.一种自适应加权HOG特征的人脸识别算法[J].计算机工程与应用,2017,53(3):164-168.
HU L Q,QIU R H.Face recognition based on adaptively weighted HOG[J].Computer Engineering and Applications,2017,53(3):164-168.

[4] JIANG H,LEARNED-MILLER E.Face detection with the Faster R-CNN[C]//IEEE International Conference on Automatic Face & Gesture Recognition,2017:650-657.

[5] 李泽琛,李恒超,胡文帅,等.多尺度注意力学习的Faster R-CNN口罩人脸检测模型[J].西南交通大学学报,2021,56(5):1002-1010.
LI Z C,LI H C,HU W S,et al.Masked face detection model based on multi-scale attention-driven Faster RCNN[J].Journal of Southwest Jiaotong University,2021,56(5):1002-1010.

[6] WU W,YIN Y,WANG X,et al.Face detection with different scales based on faster R-CNN[J].IEEE Transactions on Cybernetics,2019,49(11):4017-4028.

[7] ZHANG K,ZHANG Z,LI Z,et al.Joint face detection and alignment using multi-task cascaded convolutional networks[J].arXiv:1604.02878,2016.

[8] DENG J,GUO J,ZHOU Y,et al.RetinaFace:single-stage dense face localisation in the wild[J].arXiv:1905.00641,2019.

[9] REDMON J,FARHADI A.YOLO9000 better,faster,stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:7263-7271.

[10] REDMON J,FARHADI A.YOLOv3:an incremental improvement[J].arXiv:1804.02767,2018.

[11] BOCHKOVSKIY A,WANG C Y,LIAO H Y M.YOLOv4:optimal speed and accuracy of object detection[J].arXiv:2004.10934,2020.

[12] 薛均晓,程君进,张其斌,等.改进轻量级卷积神经网络的复杂场景口罩佩戴检测方法[J].计算机辅助设计与图形学学报,2021,33(7):1045-1054.
XUE J X,CHENG J J,ZHANG Q B,et al.Improved efficient convolutional neural networks for complex scene mask-wearing detection[J].Journal of Computer-Aided Design & Computer Graphics,2021,33(7):1045-1054.

[13] 李亚,张雨楠,彭程,等.基于多任务学习的人脸属性识别方法[J].计算机工程,2020,46(3):229-236.
LI Y,ZHANG Y N,PENG C,et al.Face attributes recognition method based on multi-task learning[J].Computer Engineering,2020,46(3):229-236.

[14] NGUYEN Q H,TRUONG H V.Real-time human ear detection based on the joint of Yolo and RetinaFace[J].Complexity,2021:1-11.

[15] 李闻,李小春,闫昊雷.基于改进YOLO v3的PCB缺陷检测[J].电光与控制,2022,29(4):106-111.
LI W,LI X C,YAN H L.PCB defect detection based on improved YOLO v3[J].Electronics Optics & Control,2022,29(4):106-111.

[16] 华泽玺,施会斌,罗彦,等.基于轻量级YOLO-v4模型的变电站数字仪表检测识别[J/OL].西南交通大学学报:1-11[2021-11-28].http://kns.cnki.net/kcms/detail/51.1277.U.20211027.1050.003.html.
HUA Z X,SHI H B,LUO Y,et al.Detection and recognition of digital instruments based on lightweight YOLO-v4 model at substations[J/OL].Journal of Southwest Jiaotong University:1-11[2021-11-28].http://kns.cnki.net/kcms/detail/51.1277.U.20211027.1050.003.html.

[17] 邹聪,梁永全.基于YOLO V3算法的输电线路鸟类检测[J].计算机应用与软件,2021,38(10):164-167.
ZOU C,LIANG Y Q.Bird detection of transmission line based on YOLO V3 algorithm[J].Computer Applications and Software,2021,38(10):164-167.

[18] 朱超平,杨艺.基于YOLO2和ResNet算法的监控视频中的人脸检测与识别[J].重庆理工大学学报(自然科学),2018,32(8):170-175.
ZHU C P,YANG Y.Face detection and recognition in monitoring video based on YOLO2 and ResNet algorithm[J].Journal of Chongqing University of Technology (Natural Science),2018,32(8):170-175.

[19] 房国志,孙康瞳.多尺度YOLO人脸年龄估计方法研究[J].计算机工程与应用,2019,55(21):135-141.
FANG G Z,SUN K T.Research on face age estimation based on multi-scale YOLO model[J].Computer Engineering and Applications,2019,55(21):135-141.

[20] 邓珍荣,白善今,马富欣,等.改进YOLO的密集小尺度人脸检测方法[J].计算机工程与设计,2020,41(3):874-879.
DENG Z R,BAI S J,MA F X,et al.Improved YOLO dense small-scale face detection method[J].Computer Engineering and Design,2020,41(3):874-879.

[21] 张路达,邓超.多尺度融合的YOLOv3人群口罩佩戴检测方法[J].计算机工程与应用,2021,57(16):283-290.
ZHANG L D,DENG C.Multi-scale fusion of YOLOv3 crowd mask wearing detection method[J].Computer Engineering and Applications,2021,57(16):283-290.

[22] 葛青青,张智杰,袁珑,等.融合环境特征与改进YOLOv4的安全帽佩戴检测[J].中国图象图形学报,2021,26(12):2904-2917.
GE Q Q,ZHANG Z J,YUAN L,et al.Safety helmet wearing detection method of fusing environmental features and improved YOLOv4[J].Journal of Image and Graphics,2021,26(12):2904-2917.

[23] YANG F,LU H,YANG M H.Robust superpixel tracking[J].IEEE Transactions on Image Processing,2014,23(4):1639-1651.

[24] YU Q,DINH T B,MEDIONI G.Online tracking and reacquisition using co-trained generative and discriminative trackers[C]//European Conference on Computer Vision,2008:678-691.

[25] WU Y,CHENG J,WANG J,et al.Real-time probabilistic covariance tracking with efficient model update[J].IEEE Transactions on Image Processing,2012,21(5):2824-2837.

[26] JEPSON A D,FLEET D J,EL-MARAGHI T F.Robust online appearance models for visual tracking[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2003,25(10):1296-1311.

[27] LIU Z,LUO P,WANG X,et al.Deep learning face attributes in the wild[C]//IEEE International Conference on Computer Vision(ICCV),2015:3730-3738.

[28] YANG S,LUO P,LOY C,et al.WIDER FACE:a face detection benchmark[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR),2016:5525-5533.

[29] ARTHUR D,VASSILVITSKII S.K-Means++:the advantages of careful seeding[C]//Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms,2007:7-9.

[30] JIE H,LI S,GANG S,et al.Squeeze-and-excitation networks[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,42(8):2011-2023.

[31] WOO S,PARK J,LEE J Y,et al.CBAM:convolutional block attention module[C]//European Conference on Computer Vision.Cham:Springer,2018:3-19.

[32] PAN S J,QIANG Y.A survey on transfer learning[J].IEEE Transactions on Knowledge and Data Engineering,2010,22(10):1345-1359.