

文章编号: 1671-6833(2021)06-0085-08

## 基于改进 SSD 和 Jetson Nano 的口罩佩戴检测门禁系统

毛晓波<sup>1</sup>, 徐向阳<sup>1</sup>, 李楠<sup>1</sup>, 魏刘倩<sup>1</sup>, 刘玉玺<sup>1</sup>, 董梦超<sup>1</sup>, 焦淼鑫<sup>2</sup>

(1. 郑州大学 电气工程学院, 河南 郑州 450001; 2. 郑州大学 机械与动力工程学院, 河南 郑州 450001)

**摘要:** 为了减少疫情期间人们未佩戴口罩造成的交叉感染概率, 设计一款基于改进的 SSD 和 Jetson Nano 的口罩佩戴检测门禁系统, 以快速检测进出口行人是否佩戴口罩, 控制闸机的开合。首先, 从 MAFA 和 WIDER FACE 这 2 个数据集中抽取适合用于该系统的训练图片, 其中 6 000 张作为训练集, 2 000 张作为测试集; 其次, 利用随机色相、饱和度等像素级变换和随机扩展、随机裁剪等几何级变换, 对数据集中的小目标进行数据增强, 使数据集更加多样, 增强该检测网络的泛化能力; 再次, 将原始 SSD 的 VGG 特征提取网络替换为 MobileNet-V3, 利用其深度可分离卷积的速度优势, 以及计算量较小的 H-Swish 激活函数、轻量化的注意力机制等优化策略, 加速检测、提高精度; 最后, 将该检测网络移植到计算能力有限的人工智能边缘计算设备 Jetson Nano 上, 加装高清显示器, 并设计可折叠的平行四边形挡板, 选择合适的外围设备, 构成了一个具有防疫价值的快速检测公共场所进出口行人是否佩戴口罩的多功能门禁系统。在该嵌入式设备上的测试结果表明: 以 MobilNet-V3 为特征提取网络的目标检测算法 SSD, 取得了 78% 的 MAP, FPS 为 12, 与以 VGG 为特征提取网络的原始 SSD 算法 (FPS 为 2) 相比, 检测速度是原始 SSD 算法的 6 倍。该系统在保证实时性的同时也兼顾了检测精度, 达到了精度和速度的平衡。

**关键词:** 口罩佩戴检测; 门禁系统; 目标检测 SSD; Jetson Nano; MobileNet-V3

**中图分类号:** TP23

**文献标志码:** A

**doi:** 10.13705/j.issn.1671-6833.2021.06.002

### 0 引言

佩戴口罩是人们保护身体健康的简单而有效的途径, 而人脸佩戴口罩的智能识别可以高效监督人们是否佩戴口罩, 是抑制疾病传播和降低损失的重要技术手段<sup>[1]</sup>。针对该问题, 有以下 3 类检测方法。

第 1 类是基于深度学习的方法。特点是将图像送入计算量很大的卷积神经网络, 再加上各种后处理, 能够检测到复杂环境中的人脸是否佩戴口罩, 检测精度高但速度慢。Tang 等<sup>[2]</sup>提出了 PyramidBox 检测算法, 该算法是一种上下文辅助的单阶段检测算法, 设计了一种新的锚点 (anchors), 将高低维特征融合在一起。Tian 等<sup>[3]</sup>先检测人脸, 再使用 Attention 机制关注口罩区域, 以判定是否佩戴。有学者提出了使用  $k$ -means 聚类算法生成若干个 Anchors Boxes, 再结合 YoLo-V3 网络检测人脸是否佩戴口罩<sup>[4-6]</sup>。管军霖等<sup>[7]</sup>提出了基于 CSP-Darknet53 的 YoLo-V4 检测

框架检测人脸是否佩戴口罩。陈国特等<sup>[8]</sup>提出了先检测人头信息, 再用 Restnet34 和 Softmax 进行分类。但两级检测网络降低了检测速度。

第 2 类是基于传统的图像检测方法进行检测。该方法对输入图像要求严格, 不能掺杂多余的背景区域, 也不能检测 1 张图像中的多个人脸, 并且在面对门禁通道中排队的情况时, 无法提取队伍队首人脸图像 (因为门禁系统开门与否由排队队伍队首的人是否佩戴口罩决定), 检测速度快, 但检测精度低。金钰丰等<sup>[9]</sup>提出了将人脸关键点和口罩纹理等特征相结合的方法检测行人是否佩戴口罩。杨子扬等<sup>[10]</sup>提出了将 RGB 色彩空间转换到  $YCbCr$  色彩空间, 并结合二值化等传统的图像处理方式, 判断上半部分人脸和下半部分人脸像素平均值的异同, 从而判定行人是否佩戴口罩。此检测方法遇到的挑战是人脸和口罩颜色接近时很难判定。

第 3 类方法主要是检测模式的创新, 特点是现场图像采集工作线下进行, 图像处理工作线上

收稿日期: 2021-04-05; 修订日期: 2021-07-24

基金项目: 国家重点研发计划项目 (2020YFC2006100); 中央本级重大增减支项目 (2060302)

作者简介: 毛晓波 (1965—), 男, 河南开封人, 郑州大学教授, 博士, 主要从事智能仪器仪表、中医药智能化研究, E-mail: mxh@zzu.edu.cn。

进行。李超等<sup>[11]</sup>提出树莓派和云平台联合进行工作的模式,即利用嵌入式设备采集图像,然后将图像上传至云服务器进行检测并返回检测结果。该类方法需要稳定快速的网络、高性能的 GPU 服务器集群,成本很高,且结构复杂、延时大,降低了系统的鲁棒性。

本文构建了一个大型数据集,并对其进行数据增强,且将 SSD<sup>[12]</sup> (single shot multibox detector) 目标检测框架中的特征提取网络 VGG<sup>[13]</sup> 替换为轻量化的特征提取网络 MobileNet-V3<sup>[14]</sup>,使之适用于计算能力有限的边缘计算系统。本文设计了平行四边形连杆折叠结构的挡板,便于安装和部署,兼具广告信息播放功能,最终构成一个完整高效的门禁系统。

## 1 神经网络模型设计及训练

### 1.1 数据标注及预处理

由于口罩数据集少且佩戴场景多元化<sup>[15]</sup>,因此本文选择开源、权威的 MAFA 和 WIDER FACE 这 2 个数据库的部分图片,每 1 张均包含戴口罩的和未佩戴口罩的人脸,并且规定未戴口罩的人脸数据标签为 face,佩戴口罩的人脸数据标签为 face\_mask,大约 8 000 张图片数据,其中训练数据 6 000 余张,测试数据约 2 000 张。训练数据标注示例如图 1 所示。



图 1 数据标注示例

Figure 1 Example of data annotation

由于 SSD 是密集采样算法,所以数据增强是 SSD 目标检测算法中的重要环节,使用数据增强的训练模型明显优于未使用数据增强的模型<sup>[12]</sup>。数据增强包括随机扩展、随机颜色空间模型变换 (HSV 和 RGB 模型) 等内容。其中,随机扩展可以显著提高 SSD 算法对小目标的检测能力,如图 2 所示。图 2 (a) 为原始标注图像,图 2 (b) 和 2 (c) 为随机扩展的 2 种随机效果示意图。其他变换可以有效应对尺寸变化、光线变化等问题,具有

很好的鲁棒性和泛化性能。魏宏彬等<sup>[16]</sup>用均值聚类算法对标注框进行聚类,得到合适的先验框宽高比例。



图 2 随机扩展

Figure 2 Random expansion

### 1.2 原始特征提取器

原始 SSD 使用的是在 ImageNet 上预训练好的 VGG<sup>[13]</sup> 模型,并且替换掉了 VGG 模型中的全连接层,改用全卷积层,最后又添加了一些新的卷积层,继续提取图像特征。以 SSD300\_VGG (300 表示输入图像的尺寸为 300×300×3) 为例,其选用的 6 个特征图的具体信息如表 1 所示,每一层特征图上的先验框尺寸为

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m - 1} (k - 1) \quad (1)$$

式中:  $s_k$  为第  $k$  个特征图上先验框的最大尺寸占原始图像边长的比例;  $s_{\min}$ 、 $s_{\max}$  均为超参数,取值分别为 0.2、0.9;  $m$  表示选取的检测特征图的数量减去 1;  $k$  为特征图的序号。

从表 1 可以得出,6 个特征图上先验框总数为 8 732 个。每个先验框都具有一个类别得分和相应的坐标,SSD 目标检测算法的任务就是找出包含这 2 种目标的先验框,并给出正确的类别。原始 SSD 目标检测框架如图 3 所示。

表 1 SSD300\_VGG 特征图具体信息

Table 1 Specific information of SSD300\_VGG feature map

特征图名称	图像尺寸	先验框宽高比例	先验框数量
conv4_3	38×38×16	(1, 1, 0.5, 2)	5 776
conv7	19×19×24	(1, 1, 0.5, 2, 0.3, 3)	2 166
conv8_2	10×10×24	(1, 1, 0.5, 2, 0.3, 3)	600
conv9_2	5×5×24	(1, 1, 0.5, 2, 0.3, 3)	150
conv10_2	3×3×16	(1, 1, 0.5, 2)	36
conv11_2	1×1×16	(1, 1, 0.5, 2)	4

### 1.3 改进的特征提取器

由于 VGG 卷积层多、运算速度慢,因此,本文将其替换为适用于移动端运行的轻量化特征提取网络 MobileNet-V3<sup>[14]</sup> (输入图像的尺寸为 320×320×3),相比于原始 SSD 采用的 VGG 网络,模型

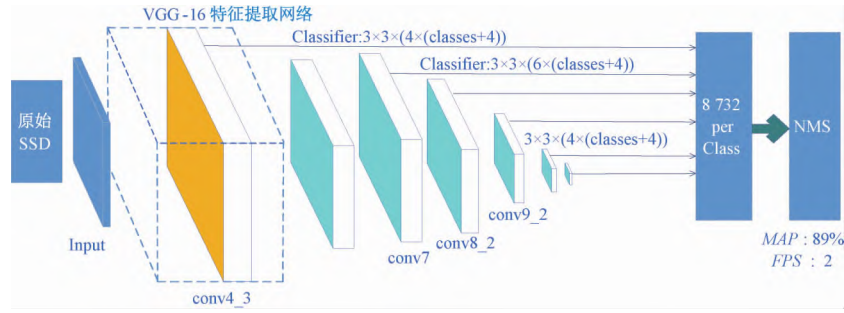


图 3 原始 SSD 目标检测框架

Figure 3 Original SSD object detection framework

大小从 201 MB 减至 29.9 MB, 约为原始模型大小的 1/7。MobileNet-V3( Large) 模型有 19 层, 第 2~15 层为集成模块( bneck 模块), 按顺序包含 2D 卷积层、BN 层、H-Swish 激活函数层、SE( squeeze and excite) 层<sup>[17]</sup>、2D 卷积层、BN 层, 如图 4 所示。

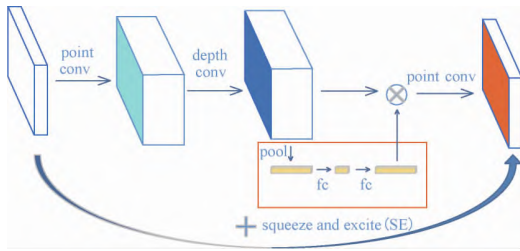


图 4 MobileNet-V3 特征提取网络的 bneck 模块

Figure 4 The bneck module of MobileNet-V3 feature extraction network

但是将其嵌入到 SSD 目标检测框架中, 就需要移除一些只适用于分类任务的层。这里移除 MobileNet-V3( Large) 中的第 17 层及以后的层, 并在其后添加 4 层 bneck, 共 20 层, 为 SSD 提供不同尺度的检测特征图, 称为 MobileNet-V3-SSD, 如图 5 所示。尽管其具有更多的卷积层, 但每层都经过了精心设计, 加入了深度可分离卷积、倒置的残差结构、轻量级的注意力机制、利用 H-Swish 代替计算量巨大的 Swish 激活函数等。深度可分离卷积将标准的卷积核进行分解, 减少了计算量, 提

高了训练推理速度。例如, 输入特征图的通道数为  $M$ , 卷积核数量为  $N$ , 卷积核尺寸为  $D_F \times D_F$ , 输出特征图尺寸为  $D_O \times D_O$ , 则使用普通卷积的计算量  $G$  为

$$G = D_F \times D_F \times M \times N \times D_O \times D_O。 \quad (2)$$

使用深度可分离卷积的计算量  $H$  为

$$H = M \times (D_F \times D_F \times 1) \times D_O \times D_O + N \times (1 \times 1 \times M) \times D_O \times D_O。 \quad (3)$$

$G$  和  $H$  的比值为

$$\frac{G}{H} = \frac{ND_F^2}{N + D_F^2}。 \quad (4)$$

在深度学习中,  $N$  一般为 32、64、128、256、512、1 024 等,  $D_F^2$  一般为 1、4、9, 因此, 式(4)的值通常远大于 1, 即: 普通卷积的计算量比深度可分离卷积的计算量高得多, 因而深度可分离卷积的运算速度比普通卷积快得多。

此网络结构中选择的具体特征图的信息如表 2 所示。其中, 先验框的总数为 3 234, 远少于表 1 中的总数 8 732, 这就为训练和推理速度的提高提供了理论基础。

这 6 个特征图上的先验框的最小、最大尺寸按照式(1)计算。将计算得到的  $s_k$  与输入图片的尺寸相乘, 即可得到该层特征图上的先验框的最大尺寸。在特征图 bneck12 中先验框的最小尺寸

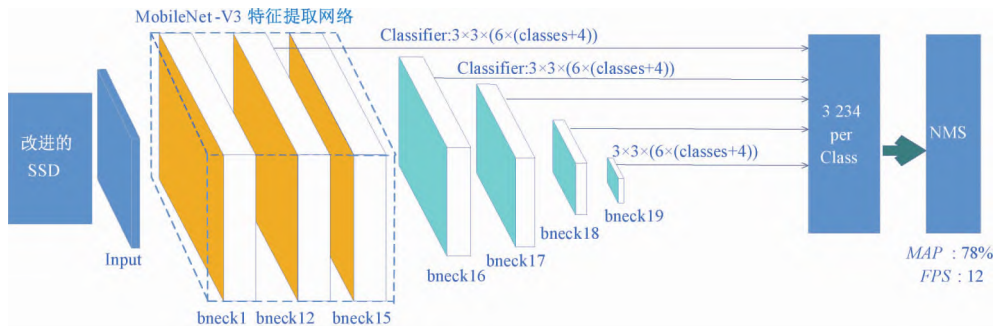


图 5 MobileNet-V3-SSD 目标检测框架

Figure 5 MobileNet-V3-SSD object detection framework

表 2 SSD320\_MobileNet\_V3 特征图具体信息

Table 2 Specific information of SSD320\_MobileNet\_V3

feature map			
特征图名称	图像尺寸	先验框宽高比例	先验框数量
bneck12	20×20×112	(1,1,0.5,2,0.3,3)	2 400
bneck15	10×10×960	(1,1,0.5,2,0.3,3)	600
bneck16	5×5×512	(1,1,0.5,2,0.3,3)	150
bneck17	3×3×256	(1,1,0.5,2,0.3,3)	54
bneck18	2×2×256	(1,1,0.5,2,0.3,3)	24
bneck19	1×1×64	(1,1,0.5,2,0.3,3)	6

选择 32, 最大尺寸由式(1) 计算可得 64, 特征图 bneck15~bneck19 中先验框的最小/最大尺寸分别为 64/109、109/154、154/198、198/243、243/332。

#### 1.4 训练编码

值得注意的是, SSD 算法并不是对目标框的位置进行直接输出, 而是对其进行编码(offset):

$$\begin{cases} l_{cx} = \frac{b_{cx} - d_{cx}}{d_w \times v_w}; \\ l_{cy} = \frac{b_{cy} - d_{cy}}{d_h \times v_h}; \\ l_w = \lg(b_w/d_w); \\ l_h = \lg(b_h/d_h). \end{cases} \quad (5)$$

式中:  $l_{cx}$ 、 $l_{cy}$  分别表示编码后的预测边界框位置信息的中心坐标的横、纵坐标;  $b_{cx}$ 、 $b_{cy}$  分别表示预测边界框的中心坐标的横、纵坐标;  $d_{cx}$ 、 $d_{cy}$  分别表示先验框的中心坐标的横、纵坐标;  $d_w$ 、 $d_h$  分别表示先验框的宽、高;  $v_w$ 、 $v_h$  为超参数, 其值均为 0.2;  $l_w$ 、 $l_h$  表示编码后的预测边界框位置信息的宽、高;  $b_w$ 、 $b_h$  表示预测边界框的宽、高。

#### 1.5 损失函数

为了使网络能够进行端到端的训练, 也需要将标注好的 Ground Truth 按照式(5) 进行编码。

SSD 的损失函数是一种联合损失函数, 即将分类损失和定位损失求和, 同时进行反向传播, 优化相关参数。由于交叉熵损失函数能够准确表达 2 个概率分布之间的关系, 因此这里采用交叉熵损失函数求得分类损失, 交叉熵损失函数为

$$L_{\text{conf}}(x, c) = - \sum_{i \in \text{Pos}} x_{ij}^p \lg \hat{c}_i^p - \sum_{i \in \text{Neg}} \lg \hat{c}_i^0. \quad (6)$$

式中:  $L_{\text{conf}}(x, c)$  为交叉熵损失函数;  $x$  为编码后的预测边界框位置信息;  $c$  为编码后的标注框位置信息;  $\text{Pos}$  为正样本集合;  $j$  为序号;  $p$  为类别;  $x_{ij}^p$  为一个因子, 取值范围为集合  $\{0, 1\}$ , 表示类别  $p$  的

第  $j$  个标注框与第  $i$  个先验框的匹配结果, 若匹配, 则取值为 1, 不匹配取值为 0;  $i$  为正样本的序号;  $\hat{c}_i^p$  为第  $i$  个先验框的类别被预测为  $p$  的置信度;  $N$  为正样本集合总数量;  $\text{Neg}$  为负样本集合;  $\hat{c}_i^0$  为第  $i$  个先验框的类别被预测为背景的置信度, 且  $\hat{c}_i^p = \frac{\exp c_i^p}{\sum \exp c_i^p}$ 。

由于正负样本数量相差很大, 若让所有的负样本损失全部参与反向传播、更新梯度是不合理的, 因为这会淹没为数不多的正样本损失。因此, SSD 采用了难负样本挖掘技术, 即: 将负样本的损失按照大小降序排列, 只取前面一部分损失最大的负样本。负样本损失的具体数量则根据正样本个数来确定, 一般选择正样本个数的 3 倍为宜。其中, 正负样本的划分规则: 若先验框与 Ground Truth 的交并比  $IoU \geq 0.5$ , 则将此先验框划分为正样本, 反之划分为负样本。 $IoU$  为

$$IoU = \frac{A \cap B}{A \cup B}. \quad (7)$$

式中:  $A$  为标注框;  $B$  为先验框。

对于定位损失, 采用  $\text{Smooth}_{L1}$  函数, 将编码后的矩阵与网络输出矩阵的相应位置作差, 将差值  $x$  作为  $\text{Smooth}_{L1}$  的自变量, 然后输出结果作为定位损失。

$$\text{smooth}_{L1}(x) = \begin{cases} 0.5x^2, & |x| < 1; \\ |x| - 0.5, & \text{其他。} \end{cases} \quad (8)$$

不同于分类损失的是, 定位损失中不含有负样本定位损失, 因为负样本是背景, 定位没有意义。

#### 1.6 网络训练

训练采用的深度学习框架为 Pytorch(1.6)、torchvision(0.7.0), 其他的依赖包为 yacs、tqdm、opencv-python、vizer、tensorboardX、six 等, 在 Ubuntu 18.04 LTS、电脑配置为 Intel(R) Xeon(R) Gold 5218 CPU @ 2.30 GHz, GeForce RTX 2080ti GPU 下, 设置 batch\_size 为 32, 初始学习率为 0.01, 在第 10 000 次迭代和第 20 000 次迭代时减小学习率, 每次迭代学习率缩小为前一次学习率的 1/10, 采用 SGD 优化器, Gamma 值为 0.1 进行网络训练, 总训练次数为 25 000 次。

## 2 结果分析

目标检测模型性能的主要评价指标包括: 平均精度均值 MAP 和 FPS(每秒处理的图片数量)。



MAP 随迭代次数而变化的曲线如图 6 所示。MAP 值等于每个类别的 *Precision-Recall* 曲线所围成的面积:

$$\begin{cases} Precision = \frac{TP}{TP + FP}; \\ Recall = \frac{TP}{TP + FN}. \end{cases} \quad (9)$$

式中: *TP* 指预测为正样本实际为正样本的数量; *FN* 指预测为负样本实际为正样本的数量; *FP* 指预测为正样本实际为负样本的数量<sup>[18]</sup>。

模型改进前后,不同检测网络的算法性能结果如表 3 所示,推理速度测试环境为 Jetson Nano 开发板。在表 3 中,MAP 指综合 MAP,即戴口罩人脸和不戴口罩人脸 2 种类别 MAP 的平均值; face\_MAP、face\_mask\_MAP 分别指不戴口罩的人脸类别的 MAP、戴口罩的人脸类别的 MAP; 训练耗时指从头开始训练本数据集至 MAP 收敛时所用的时间; FPS 指 Jetson Nano 控制器每秒处理  $1 \times 320 \times 320 \times 3$  图像的数量。

表 3 显示,以 VGG16 为特征提取网络的 SSD 检测框架的综合 MAP 为 89.40%,略高于以 MobileNet-V3 为特征提取网络的 SSD 检测框架

(78.49%),这可能是轻量化的特征提取网络在一定程度上丢失了部分信息所造成的。但是本文的主要目的是将其运行在边缘计算设备上,实时性更重要,因此算法更注重速度的提高。前者训练至 MAP 收敛时的耗时为 3.88 h,是后者的 2.27 倍;前者的 FPS 为 2( Jetson Nano 平台),是后者的 1/6,即实际效果上,改进后的网络检测速度是前者的 6 倍。主要原因是①轻量型 Mobilenet-V3 减少了参数量和计算量;②深度可分离卷积优化了普通卷积的计算形式,加速效果为 SSD+VGG16 的 1~9 倍,参见式(4)。

图 6 为在不同的特征提取网络(VGG16 和 Mobilenet-V3)下,face 和 face\_mask 的 MAP 以及综合 MAP 随着迭代次数的变化曲线。可以看出,MobileNet-V3 的 MAP(改进后的变化曲线)收敛速度更快,过程更加平坦,这可能是由于参数量和计算量较少,从而使网络更加容易训练,单步迭代耗时少。

表 4 展示了不同特征提取网络下,显存占用量、GPU 占用率等指标的具体数值。改进后的 SSD(以 MobileNet-V3 为特征提取网络)指标值均为该类最低,有利于减少模型存储空间、减弱对

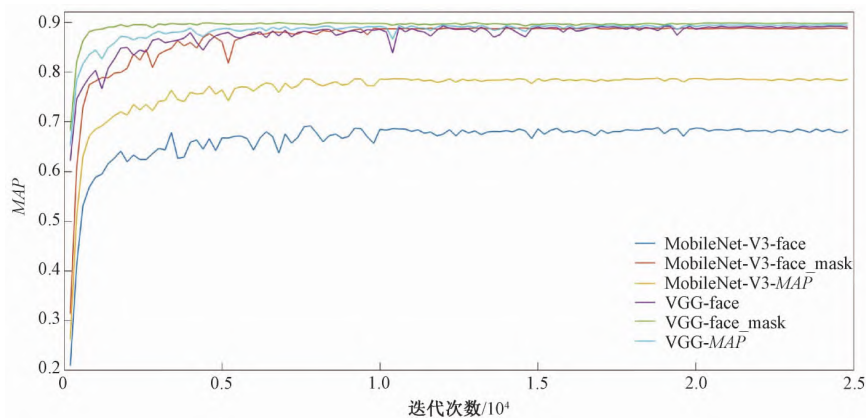


图 6 MAP 随着迭代次数而变化的曲线

Figure 6 Curve of MAP changing with the number of iterations

表 3 不同检测网络的算法性能

Table 3 Algorithm performance of different detection networks

特征提取器	MAP	face_MAP	face_mask_MAP	训练耗时/h	FPS	推理耗时/ms
SSD+VGG16	89.40%	89.14%	89.66%	3.88	2	475
SSD+MobileNet-V3	78.49%	68.31%	88.67%	1.71	12	86

表 4 不同检测网络的性能指标

Table 4 Performance indicators of different detection networks

特征提取器	输入图像 大小	显存 占用量/MB	GPU 占用率/%	参数量	GFLOPs	MAdd/G	内存 占用量/MB
SSD+VGG16	300×300×3	1 155	70	23 879 570	30.540 0	60.99	206.35
YOLO-V3	416×416×3	1 659	70	61 626 049	155.400 0	310.30	236.32
SSD+MobileNet-V3	320×320×3	1 025	20	3 626 508	0.483 7	0.95	108.60

硬件资源的依赖以及对 GPU 的损耗。以 VGG16 为特征提取网络的 SSD 检测框架和 YoLo-V3 检测框架的计算量分别是以 MobileNet-V3 为特征提取网络的 SSD 检测框架的 63 和 321 倍。因此, YoLo-V3 和原始 SSD 均无法胜任嵌入式平台应用的需求。

### 3 模型移植及机械部件设计

#### 3.1 中央控制器选择

MobileNet-V3-SSD 网络在输入  $1 \times 320 \times 320 \times 3$  时, 计算量以总浮点数运算操作次数 *GFLOPs* 表示, 为 0.483 7(FP32)。选择 Jetson Nano 作为中央处理器, 半精度(FP16)浮点数的计算速度以每秒浮点数运算操作次数 *TFLOPs* 表示, 为 0.5。NVIDIA 官网的 Jetpack 4.4.1 系统作为基础系统环境。

#### 3.2 外围电路

##### 3.2.1 光电开关输入信号

选择漫反射三线 PNP 常开型光电开关, 额定电压为 5 V, 额定电流为 30 mA。仅当检测到人脸, 中央控制器才会调用 GPU 进行推理检测, 以减少不必要的 GPU 和 CPU 损耗及电量消耗, 延长元器件使用寿命。

##### 3.2.2 USB 摄像头图像信号

当控制器收到光电开关高电平信号后, 立即调用 USB 摄像头(480 P, 60°, 焦距为 3.6 mm)捕捉图像信息, 进行推理检测。持续检测大约 20 帧图像, 如果超过 10 帧均检测到该行人佩戴了口罩, 即认为“佩戴口罩”, 打开挡板; 否则, 语音提示“您好, 您未佩戴口罩, 请佩戴口罩后通行”。为了在客流量大的时候依然保持鲁棒性, 将检测到的人脸框按照面积大小排序, 取面积最大的作为排队队伍的第 1 个人的脸, 是否放行依据此人是否佩戴口罩。

##### 3.2.3 动力输出

挡板的动力由伺服电机提供, 转动范围为  $0^\circ \sim 180^\circ$ , 扭矩为  $60 \text{ kg} \cdot \text{cm}$ 。依据摄像头中的人是否佩戴了口罩, 通过 python 编程提供的脉宽调制 PWM 信号, 经 I2C—GPIO 口和 PCA9685 模块输出控制伺服电机旋转。

#### 3.3 折叠挡板设计

常见的挡板类型如三辊闸、摆闸和翼闸等, 均不能自动调整挡板长度。采用平行四边形连杆折叠设计, 当伺服电机驱动挡板上升时, 由于重力的

影响, 挡板展开, 形成一张网状的门, 如图 7(a) 所示; 当伺服电机驱动挡板下降时, 由于外壳壁的阻力影响, 折叠机构收缩在外壳壁内, 如图 7(b) 所示。

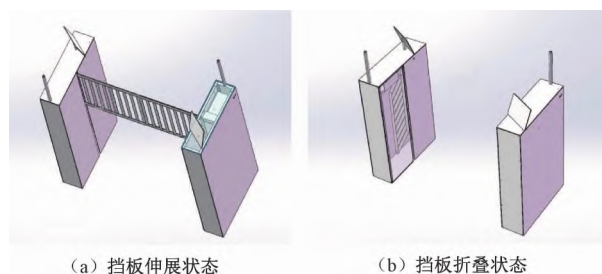


图 7 挡板的工作状态

Figure 7 Working states of baffles

### 4 结论

本文设计了一种新型门禁系统, 以 MobileNet-V3 为特征提取网络改进 SSD 检测算法检测公共场所出入口行人是否佩戴口罩, 平行四边形连杆折叠挡板节约占地空间, 留有以太网接口方便修改配置、更新神经网络模型。将原始 VGG 特征提取网络替换成轻量化 MobileNet-V3, *MAP* 虽有所下降, 但 *FPS* 却从 2 提高到了 12, 检测速度是原始 SSD 算法的 6 倍, 在保证精度的同时提高了速度。如何在 *MAP* 保持不变的情况下提高检测速度依然是一个值得研究的问题。下一步研究会尝试将门禁系统接入互联网, 利用云计算在云端进行监控, 实现集中统一管理和维护。

### 参考文献:

- [1] 张修宝, 林子原, 田万鑫, 等. 全天候自然场景下的人脸佩戴口罩识别技术 [J]. 中国科学: 信息科学, 2020, 50(7): 1110–1120.
- [2] TANG X, DU D K, HE Z Q, et al. PyramidBox: a context-assisted single shot face detector [J]. Computer vision-ECCV, 2018, 11213: 812–828.
- [3] TIAN W X, WANG Z X, SHEN H F, et al. Learning better features for face detection with feature fusion and segmentation supervision [EB/OL]. (2019-06-25) [2020-03-12]. <http://h-s.arxiv.org.zzulib.vpn358.com/abs/1811.08557>.
- [4] 南阳, 樊治国, 黄惠南, 等. 基于 YOLOV3 算法的二次检测口罩人脸识别方法: CN111414887B [P]. 2021-01-29.
- [5] 严安, 杨晓云, 周治尹. 基于深度学习的厨师帽和口

- 罩佩戴的检测方法: CN111062429A [P]. 2020-04-24.
- [6] MACQUEEN J. Some methods for classification and analysis of multivariate observations [J]. Berkeley symposium on mathematical statistics and probability, 1967, 1( 14): 281-297.
- [7] 管军霖, 智鑫. 基于 YOLOv4 卷积神经网络的口罩佩戴检测方法 [J]. 现代信息科技, 2020, 4( 11): 9-12.
- [8] 陈国特, 陈清梁, 程潇, 等. 用于疫情防控的直梯乘梯人员佩戴口罩监测方法: CN111382719A [P]. 2020-07-07.
- [9] 金钰丰, 闫鹏飞, 解艺博, 等. 一种口罩佩戴情况检测方法、装置、电子设备及存储介质: CN111428559A [P]. 2020-07-17.
- [10] 杨子扬, 江文涛, 苏松剑, 等. 一种人员佩戴口罩情况的检测方法以及装置: CN109101923B [P]. 2020-11-27.
- [11] 李超, 黄伟锋, 朱立学, 等. 基于 Raspberry Pi 和云平台的智能门禁系统研制与试验 [J]. 仲恺农业工程学院学报, 2020, 33( 1): 54-60.
- [12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [J]. Lecture notes in computer science, 2016, 9905: 21-37.
- [13] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. ( 2015-04-10) [2021-01-22]. <https://arxiv.org/abs/1409.1556>.
- [14] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3 [C] // 2019 IEEE/CVF International Conference on Computer Vision ( ICCV ). Piscataway: IEEE, 2019: 1314-1324.
- [15] 邹佰翰, 秦亚亭, 苑晓兵, 等. 基于轻量级 CNN 的口罩人脸检测方法现状研究 [J]. 软件, 2020, 41( 8): 186-188.
- [16] 魏宏彬, 张端金, 杜广明, 等. 基于改进型 YOLO v3 的蔬菜识别算法 [J]. 郑州大学学报( 工学版), 2020, 41( 2): 7-12, 31.
- [17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.
- [18] 罗荣辉, 袁航, 钟发海, 等. 基于卷积神经网络的道路拥堵识别研究 [J]. 郑州大学学报( 工学版), 2019, 40( 2): 18-22.

## A Mask Wear Detection Access Control System Based on Improved SSD and Jetson Nano

MAO Xiaobo<sup>1</sup>, XU Xiangyang<sup>1</sup>, LI Nan<sup>1</sup>, WEI Liujian<sup>1</sup>, LIU Yuxi<sup>1</sup>, DONG Mengchao<sup>1</sup>, JIAO Miaoxin<sup>2</sup>

( 1.School of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China; 2.School of Mechanical and Power Engineering, Zhengzhou University, Zhengzhou 450001, China)

**Abstract:** In order to reduce the chance of cross infection caused by people not wearing masks during the epidemic, a mask wearing detection access control system based on improved SSD and Jetson Nano is designed to quickly detect whether pedestrians at the entrance and exit wear masks and control the opening and closing of the gate. Firstly, 6 000 training pictures suitable for the system are extracted from the two data sets of MAFA and WIDER FACE, which are used as the training set and 2 000 as the test set; Secondly, the pixel level transformations such as random hue and saturation and geometric level transformations such as random expansion and random clipping are used to enhance the small targets in the data set, so as to add more samples to the data set and enhance the generalization ability of the detection network; Thirdly, the VGG feature extraction network of the original SSD is replaced by MobileNet-V3, which makes use of its speed advantage of depth-wise separable convolution, as well as the optimization strategies such as H-Swish activation function with less computation and lightweight attention mechanism ( squeeze and excite) to accelerate the detection and improve the accuracy. Finally, the detection network is transplanted to Jetson Nano, an artificial intelligence edge computing device with limited computing power, equipped with high-definition display, design a foldable parallelogram baffle, and select appropriate peripheral equipment to form a multi-functional access

control system with epidemic prevention value to quickly detect whether pedestrians at the entrance and exit of public places wear masks. The test results on the embedded device are as follows: the target detection algorithm SSD with MobileNet-V3 as the feature extraction network obtains 78% *MAP* and *FPS* is 12. Compared with the original SSD algorithm with VGG as the feature extraction network ( *FPS* is 2 ) , the detection speed is increased five times. Facts have proved that the system not only ensures the real-time performance, but also takes into account the detection accuracy, so achieves the balance of accuracy and speed.

**Keywords:** mask wear detection; access control system; object detection algorithm SSD; Jetson Nano; MobileNet-V3

---

( 上接第 84 页)

## Curing and Adhesive Characteristics of Monocomponent Polyurethane Binders

ZHONG Ke<sup>1,2,3</sup>, WANG Xue<sup>1</sup>, ZHANG Meng<sup>2,3</sup>, SUN Mingzhi<sup>2,3</sup>, SUN Shengkai<sup>1</sup>, GONG Yan<sup>2,3</sup>

( 1.College of Civil Engineering, Chongqing Jiaotong University, Chongqing 400041, China; 2.Research Institute of Highway Ministry of Transport, Beijing 100088, China; 3.Key Laboratory of Transport Industry of Road Structure and Material, Beijing 100088, China)

**Abstract:** Monocomponent polyurethane is a reactive adhesive, and its mechanical properties are better than those of conventional asphalt binders, but the reaction process is easily affected by environmental factors such as temperature, air and water, etc. During the paving and rolling construction of polyurethane mixes, it is necessary to strictly control the construction time nodes of each process in combination with the site temperature and humidity working conditions. In order to popularize the application of single-component polyurethane in road, curing and adhesive characteristics of monocomponent polyurethane were studied in this paper. Curing and adhesion characteristics of two single-component polyurethane ( PU ) binders were studied from FTIR, pull-out and shear tests. With the increase of natural curing conditions, the results showed that the isocyanate index of PU-I and PU-II decreased, while the urea index had the opposite trend. Both the isocyanate index and urea index of PU were stable after 4 days natural curing, and the variation coefficients of isocyanate index and urea index of PU-I and PU-II were 10% and 2%, 2% and 2%, respectively in 4-7 days, which belong to small variations, indicating that PU had basically reached the fully cured state. With the increase of curing time, the pull-out strength and shear strength of polyurethane bonding system generally increased, and the pull-out and shear strength of stone-polyurethane-stone interface were better than those stone-polyurethane-rubber, rubber-polyurethane-rubber bonding system. The pull-out strength of rubber-polyurethane-rubber interface was the worst, and the shear strength of stone-polyurethane-rubber interface was the worst.

**Keywords:** monocomponent polyurethane; curing; adhesion characteristics; isocyanate index; urea index