

文章编号: 2095-2163(2022)11-0224-06

中图分类号: TP399

文献标志码: A

# 面向疫情防控的口罩佩戴检测系统

蔡雨晨, 马雨生, 凌晨, 刘巧红

(上海健康医学院 医疗器械学院, 上海 201318)

**摘要:** 为有效阻断疫情传播, 以及满足人口密集处疫情防控的需要, 本文设计了一个基于深度迁移学习的人脸口罩佩戴检测系统 MaskNet。首先, 以 VGG16 网络为原型设计了人脸口罩检测的深度神经网络模型。为解决样本不足问题, 利用 ImageNet 对 VGG16 网络进行预训练, 随后利用口罩数据集和迁移学习进行模型的微调优化。在应用中, 采用 OpenCV 和 Dlib 库实现视频中人脸区域的实时检测, 再利用训练好的 MaskNet 模型进行口罩实时检测。MaskNet 不仅可以实现图像中的人脸口罩检测, 也可以用在视频中出现的单个或多个人脸口罩佩戴的检测。实验中, 系统的识别准确率可达 99.2%, 可以应用在疫情防控的场合, 对未佩戴口罩人员进行提示, 达到阻止病毒扩散传播的目的。

**关键词:** 深度学习; 迁移学习; 卷积神经网络; 人脸口罩检测; 疫情防控

## Mask wearing detection system for epidemic prevention and control

CAI Yuchen, MA Yusheng, LING Chen, LIU Qiaohong

(School of Medical Instruments, Shanghai University of Medicine and Health Sciences, Shanghai 201318, China)

**【Abstract】** To effectively block the spread of the epidemic and meet the needs of epidemic prevention and control in densely populated areas, a mask wearing detection system based on deep transfer learning, named MaskNet is proposed. First, a deep neural network model for face mask detection is designed with the VGG16 network as a prototype. In order to solve the problem of insufficient samples, ImageNet is used to pre-train the VGG16 network. Then, the mask data set and transfer learning are used to fine-tune and optimize the model. In the application, OpenCV and Dlib libraries are applied to realize real-time detection of face regions in videos, and the trained MaskNet model is used for the real-time detection of masks. MaskNet can not only realize the detection of face masks in images, but also can be used to detect single or multiple face masks that appear in videos. In the experiment, the recognition accuracy of the system can reach 99.2%, which can be used in epidemic prevention and control situations to prompt people who are not wearing masks to achieve the purpose of preventing the spread of the virus.

**【Key words】** deep learning; transfer learning; convolutional neural network; face mask detection; epidemic prevention and control

## 0 引言

2019 年末, 新型冠状病毒<sup>[1]</sup>首次爆发。研究表明, 新冠病毒具有极强的传染性以及较长的潜伏期<sup>[2]</sup>, 其传播途径主要通过飞沫传播、接触传播、血液、粪口传播的方式<sup>[3]</sup>。现如今, 多个国家发现了新冠肺炎变异病株, 并且已有确诊病例, 新型冠状病毒的防疫越发严峻。

当前, 国内采取了强有力的防控措施, 有效控制了疫情的发展, 抗击新冠已取得巨大成功<sup>[4]</sup>。随着国内疫情防控形势的好转, 疫情防控工作开始常态化, 人们也开始外出、复工复学。此时, 疫情防控再一次面临考验, 佩戴口罩是有效抑制疫情反弹的重要措施之一。随着人工智能技术的快速发展, 通过

智能化的方式实现人脸是否佩戴口罩的检测和预警显得尤为重要。尤其是在如地铁、公交、高铁、学校、医院等人流密集的进入场所, 增加智能化的口罩识别系统能大大提供检测效率, 减少人工检测带来的漏查错查等现象。因此, 本文针对疫情防控的需要, 利用深度学习技术设计与开发了一个人脸口罩佩戴识别系统。

近年来, 学者们针对口罩识别也开展了一些研究。万子伦等人<sup>[5]</sup>提出了一种基于改进 Faster R-CNN 的口罩佩戴检测算法, 将传统的单一 RPN 网络模型改进使用多任务增强 RPN 模型以提高检测识别精度, 其最终平均识别精准率可达 90.18%。但在目标过多的情况下会出现卡顿现象。刘国明等人<sup>[6]</sup>利用 ResNet-34 深度神经网络, 经适当预处理,

**基金项目:** 国家自然科学基金(61801288); 国家级大学生创新创业训练计划项目(202110262069, S202110262114, S202110262115)。

**作者简介:** 蔡雨晨(2003-), 男, 本科生, 主要研究方向: 健康医疗大数据分析; 刘巧红(1979-), 女, 博士, 副教授, 主要研究方向: 医学图像处理、医学影像大数据分析。

**通讯作者:** 刘巧红 Email: hqllqh@163.com

**收稿日期:** 2022-03-25

调整学习率大小和批数据量大小,网络在验证集上准确率为 98.41%,在测试集上准确率为 97.25%。李国进等人<sup>[7]</sup>提出了一种基于 DCN-SERes-YOLOv3 的人脸佩戴口罩检测算法,算法的平均精度值高达 95.36%,比传统的 YOLOv3 算法提高了约 4.1 个百分点,且检测速度提高了 11.7 fps。Cao 等人<sup>[8]</sup>基于 YOLOv4 系列,使用一种改进的 Mosaic 数据增强的方法,增强在夜间对于人脸口罩检测的辨别能力。Sen 等人<sup>[9]</sup>提出一种使用 PyTorch 框架的深度学习方法,从视频流中检测出任意类型的口罩和不同形状的口罩。

总体来说,目前直接应用于人脸佩戴口罩的算法较少,且存在识别精准度有待提高、识别速度有待提升等问题。同时由于真实环境下人脸佩戴口罩的复杂性以及环境的不可控性,真实场景下的检测难度也大大增加。此外,由于数据量有限,现有的口罩数据集未能覆盖所有的真实场景,造成模型拟合不足,带来漏检问题。

为了进一步提高人脸佩戴口罩的识别准确度,应用在实时的视频监控检测中,本文基于卷积神经网络和迁移学习技术设计了一个人脸口罩自动检测系统,用于自动检测视频和图像中人脸佩戴口罩情况。本文研究的主要创新在于:

(1) 以 VGG16 网络为原型,结合自定义的全连接网络,提出了一个用于口罩检测的深度神经网络模型,命名为 MaskNet。

(2) 为解决样本量不足的问题,采用了模型迁移学习技术。首先利用 ImageNet 对 VGG16 网络进行预训练,随后利用口罩数据集进行模型的参数微调优化。

(3) 为了实现视频下的人脸佩戴口罩的实时检测,利用 OpenCV 和 Dlib 库打开摄像头并动态采集人脸,进一步捕获人脸所在区域后送入训练好的 MaskNet,实现口罩检测。提出模型的检测准确率可以达到 99.2%,可以应用在对实时要求较高的场合。

(4) 实验表明,本文提出的 MaskNet,不仅可以检测单张图像或实时视频中人脸佩戴口罩情况,还可以同时检测多人存在的口罩佩戴情况。

## 1 理论基础

### 1.1 VGG16

VGG 是牛津大学的 Visual Geometry 团队提出,可以说 VGG 是加深版的 AlexNet<sup>[10]</sup>。其中,VGG16<sup>[11]</sup>将 AlexNet 中的大尺寸卷积核全部更改成

小尺寸的  $3 \times 3$  卷积核,利用多个连续小卷积核的堆叠增大感受野。VGG 在保证感受野大小不受影响的同时,通过多个非线性层<sup>[12]</sup>可以增加网络深度来保证学习更复杂的模式,并且产生的参数更少。VGG16 是常被用于迁移学习的网络,由于标准的 VGG16 网络在整个 ImageNet 上已经进行了预训练,因此在完成新任务的时候,只需要通过微调网络就可以应用在新任务上。本文采用了 VGG16 网络的迁移学习,达到快速模型训练和提取特征的目的。VGG16 网络模型的结构<sup>[11]</sup>如图 1 所示。

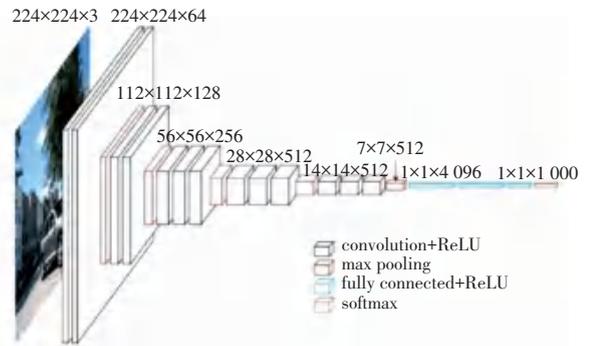


图 1 VGG16 模型结构示意图

Fig. 1 Schematic diagram of VGG16 model structure

### 1.2 迁移学习

迁移学习<sup>[13]</sup>是一种机器学习的方法,即把为任务 A 开发的模型作为初始点,重新使用在为任务 B 开发模型的过程中。迁移学习一词首次来源于教育心理学,Pratt 最早在机器学习领域引用迁移,在 1993 年研发了基于可区分性转移算法。后经过一系列发展,使得迁移学习在深度学习中越来越重要。迁移学习<sup>[14]</sup>是可以利用从其他任务中获取的知识来帮助进行当前任务试验的一系列方法,利用此方法,可以有效解决过拟合等问题,关键点就是找出新问题和原问题两者之间有哪些相似性<sup>[15]</sup>。在小样本数据的条件下,由于数据量不足,会导致训练 CNN 模型时发生过拟合,无法获得理想识别结果。迁移学习的应用范围非常广泛,比如用来解决标注数据稀缺问题、误差分享、进行机器人训练等等。目前,作为一个新兴的研究领域,迁移学习的研究主要还是集中在算法开发方面,基础理论研究还并不成熟,因此值得进一步的研究。

## 2 系统设计和开发

### 2.1 基于迁移学习和 VGG16 的人脸口罩检测模型

原始的 VGG16 网络模型包含了 16 个隐藏层,分别为 13 个卷积层和 3 个全连接层,接着将最后一

个输出张量输入到一个分类器网络中,由分类器处理向量得到分类结果。在处理二分类问题时,最后一层采用的激活函数为 *Softmax*。本文提出的 MaskNet 模型将利用去除全连接层后的 VGG16 网络结构完成对口罩图像数据集的特征提取工作。初始权重直接采用 VGG16 在 ImageNet 数据集上预训练好的权重,预训练的 VGG16 的结构和参数见表 1。

表 1 去除全连接层的 VGG16 模型结构

Tab. 1 VGG16 model structure removing full connection layer

层(类型)	输出层	参数
<i>input_1</i> (Input Layer)	[ (None, 150, 150, 3) ]	0
<i>block1_conv1</i> (Conv2D)	(None, 150, 150, 164)	1 792
<i>block1_conv2</i> (Conv2D)	(None, 150, 150, 64)	36 928
<i>block1_pool</i> (MaxPooling2D)	(None, 75, 75, 128)	0
<i>block2_conv1</i> (Conv2D)	(None, 75, 75, 128)	73 856
<i>block2_conv2</i> (Conv2D)	(None, 37, 37, 128)	147 584
<i>block2_pool</i> (MaxPooling2D)	(None, 37, 37, 128)	0
<i>block3_conv1</i> (Conv2D)	(None, 37, 37, 256)	295 168
<i>block3_conv2</i> (Conv2D)	(None, 37, 37, 256)	590 080
<i>block3_conv3</i> (Conv2D)	(None, 37, 37, 256)	590 080
<i>block3_pool</i> (MaxPooling2D)	(None, 18, 18, 256)	0
<i>block4_conv1</i> (Conv2D)	(None, 18, 18, 512)	1 180 160
<i>block4_conv2</i> (Conv2D)	(None, 18, 18, 512)	2 359 808
<i>block4_conv3</i> (Conv2D)	(None, 18, 18, 512)	2 359 808
<i>block4_pool</i> (MaxPooling2D)	(None, 9, 9, 512)	0
<i>block5_conv1</i> (Conv2D)	(None, 9, 9, 512)	2 359 808
<i>block5_conv2</i> (Conv2D)	(None, 9, 9, 512)	2 359 808
<i>block5_conv3</i> (Conv2D)	(None, 9, 9, 512)	2 359 808
<i>block5_pool</i> (MaxPooling2D)	(None, 4, 4, 512)	0
<i>Totalparams</i> : 14 714 688		
<i>Trainablepara, s</i> : 14 714 688		
<i>Non - trainableparams</i> : 0		

本文提出的 MaskNet 网络以 Sequential 顺序模型为基础,调用了去除全连接层的 VGG16 网络的权重参数进行迁移学习,只训练自定义的全连接层。预训练的 VGG16 的输出经过一个拉平层后,再通过 2 个全连接层来处理后续的分类识别。MaskNet 的结构和参数见表 2,模型示意图如图 2 所示。

表 2 MaskNet 模型结构

Tab. 2 MaskNet model structure

层(类型)	输出层	参数
VGG16 (Model)	(None, 4, 4, 512)	14 714 688
<i>flatten</i> (Flatten)	(None, 8 192)	0
<i>dense</i> (Dense)	(None, 256)	2 097 408
<i>dense_1</i> (Dense)	(None, 1)	257
<i>Totalparams</i> : 16 812 353		
<i>Trainableparams</i> : 16 812 353		
<i>Non - trainableparams</i> : 0		

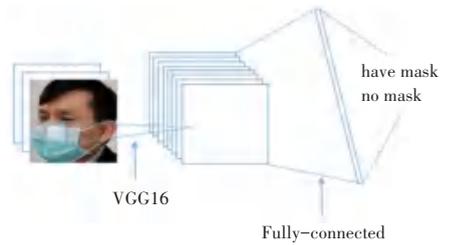


图 2 MaskNet 口罩检测模型示意图

Fig. 2 Schematic diagram of MaskNet mask detection model

### 2.2 人脸口罩佩戴检测系统

本文设计和开发的人脸口罩佩戴检测系统流程如图 3 所示。首先,将获得的数据集划分为正负样本,通过本文提出的 MaskNet 网络进行特征提取,训练分类器。然后,可以调用已训练完成的分类器对图像以及视频中的人脸口罩佩戴情况进行检测分类,最后输出分类结果。系统经过模型判断后,给出是否佩戴口罩的结果,分为 2 种情况。如果检测出没有佩戴口罩的,将会用红色的矩形框标记人脸区域并显示提示字样;如果是检测出佩戴口罩,将会用绿色的矩形框标记人脸区域并显示提示字样。

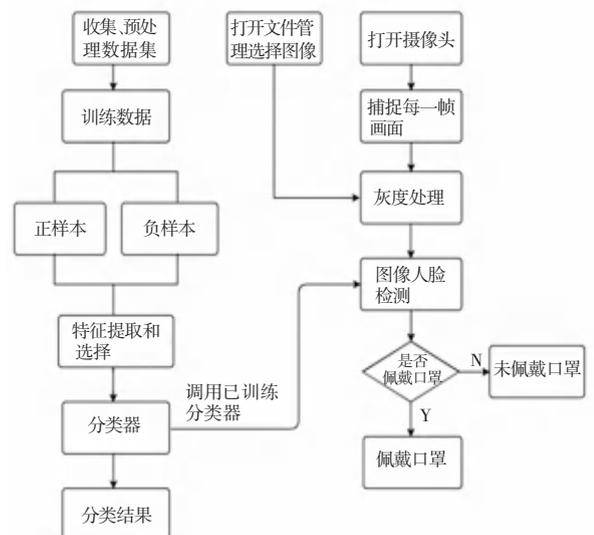


图 3 系统流程图

Fig. 3 Flow chart of the system

### 3 实验

#### 3.1 实验设置

##### 3.1.1 数据集

数据集来自 GitHub 官方网站上的公开数据集, 这是 2020 年 3 月初武汉大学发布的一类开源的口罩人脸识别数据集。收集到的图像数据总共为 2 055 张, 本次数据均采用的是 50 \* 50 \* 3 大小的彩色图片。部分图像数据如图 4 所示。



图 4 部分图像数据

Fig. 4 Partial images data

##### 3.1.2 数据预处理

将下载的数据集划分为训练集、验证集和测试集三部分, 划分后会在目录中生成 train、test、validation 三个文件夹, 同时这 3 个文件夹下面都会创建名为 have\_mask 与 no\_mask 文件夹。这里的 have\_mask 与 no\_mask 作为佩戴口罩和未佩戴口罩的标签。每个文件夹下包含相应佩戴口罩和未佩戴口罩的图像数据, 相应的数据量见表 3。

表 3 数据集的划分

Tab. 3 Division of data sets

类型	数量
<i>totaltrainingmaskimages</i>	600
<i>totaltrainingnomaskimages</i>	600
<i>totalvalidationmaskimages</i>	400
<i>totalvalidationnomaskimages</i>	400
<i>totaltestmaskimages</i>	8
<i>totaltestnomaskimages</i>	47

为了获取更多的训练数据, 本文对原始数据进行增强操作, 主要利用了 Keras 框架下的 ImageDataGenerator 类来实现数据增强的功能。具体过程分为 2 个部分。首先, 是图片生成器, 主要是生成各个批次的图片, 以供生成器的形式给模型训练; 其次, 就是要训练每一批次的图像, 并适当进行数据增强。本文主要采用的数据增强操作见表 4。

表 4 数据增强方法

Tab. 4 Data enhancement methods

数据增强方法	方法操作
翻转变换 (flip)	对图像进行上下左右的翻转
平移变化 (shift)	平移图像
尺度变化 (scale)	改变图像的尺度
旋转变换 (rotation)	对图像进行旋转操作
反射变化 (reflection)	对图像进行反射变化
噪声扰动 (noise)	在图像中增加噪声

#### 3.2 开发环境

本系统的开发环境基于 Pycharm 集成平台, 采用了 Tensorflow 框架的 Keras 模块来搭建网络结构, Python3.6 为开发语言。在进行视频流处理和实时人脸区域标记时, 采用了计算机视觉库, 即 OpenCV 和 Dlib 库。利用 OpenCV 库开启摄像头, 并捕捉每一帧实时画面, 再用 Dlib 库中的正向人脸检测器进行人脸检测, 提取人脸外部矩形框。接着利用 Opencv 库进行图像灰度处理, 获取检测到的人脸坐标, 并调用训练后的 MaskNet 网络对人脸口罩的情况进行预测。

#### 3.3 实验结果

##### 3.3.1 照片中的口罩检测

本文提出的 MaskNet 可以检测单张图像中的口罩佩戴情况, 验证结果将使用数值与中文表示, 其数值以 0.5 作分界线。如果数值结果大于 0.5、则为未佩戴口罩, 反之数值小于 0.5、则为戴口罩。图 5 给出了 2 张佩戴口罩和未佩戴口罩的图像的预测结果。

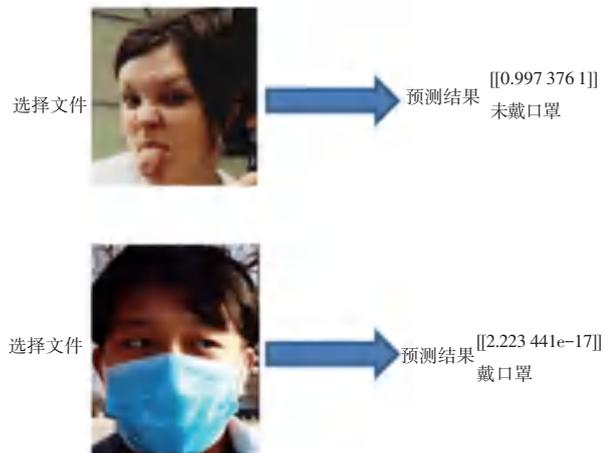


图 5 照片预测验证结果

Fig. 5 Verification results of photos prediction

##### 3.3.2 视频实时口罩检测

本文开发的人脸口罩佩戴检测系统主界面如图

6所示。点击主界面中的口罩识别按钮,进入视频口罩检测界面,同时调用电脑摄像头进行视频的实时采集。将摄像头捕捉到的画面中的人脸用候选框标记出来,并调用 MaskNet 网络进行口罩检测。被检测的人脸是否佩戴口罩的结果将直接呈现在视频监控画面中,使用不同颜色的候选框和提示文字来加以表示。未佩戴口罩将用红色候选框和 NoMask 提示,佩戴口罩将用绿色候选框和 Mask 提示,并显示准确率的数值。单人检测示例如图 7 所示。多人检测示例如图 8 所示。



图 6 系统界面

Fig. 6 System interface

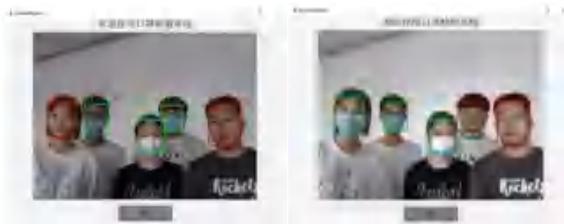


(a) 没戴口罩

(b) 戴口罩

图 7 单人验证结果

Fig. 7 Single-person verification results



(a) 结果 1

(b) 结果 2

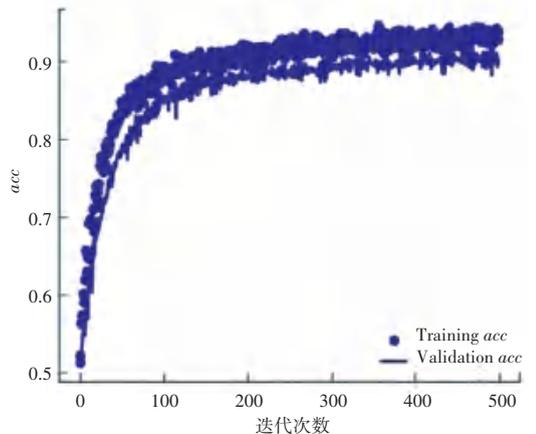
图 8 多人验证结果

Fig. 8 Multi-person verification results

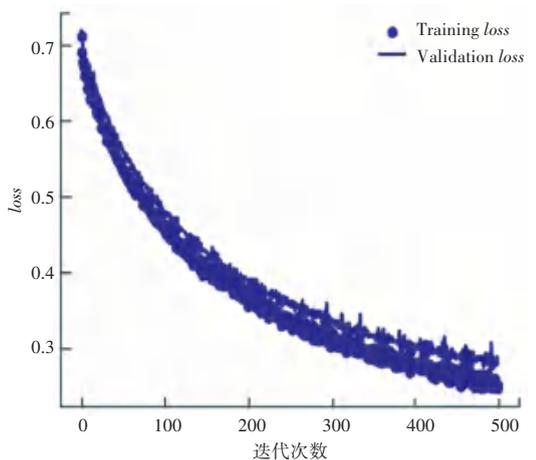
### 3.3.3 模型精度

图 9 是本文提出模型在训练时的精确度曲线和损失曲线。由图 9 中曲线的趋势可以看出,本文提出的模型总体上精确度呈现上升趋势,并在约 500 次训

练后趋近收敛,由此证明本文提出模型的有效性。



(a) 精确度曲线



(b) 损失曲线

图 9 精确度曲线与损失曲线

Fig. 9 Accuracy curve and loss curve

## 4 结束语

阻断新冠疫情主要是阻断病毒的传播,而阻断病毒传播最有效的方法就是佩戴口罩。因此本文基于迁移学习和卷积神经网络,展开对目标人脸图片或视频口罩佩戴的有效检测研究。研究中采用 VGG16 模型,利用 ImageNet 进行预训练,结合 OpenCV 和 Dlib 库实现单或多个人脸口罩佩戴检测。相较于其他的模型算法,MaskNet 的准确率达到 99.2%,同时可以检测多个目标。在全球预防新冠疫情的背景下,将其应用于公共场所,能准确做到人脸口罩自动检测并可及时发出预警提示,对疫情防控做出积极贡献,具有重大实用意义。

## 参考文献

- [1] 彭麟,王卫红. 新冠肺炎(COVID-19)的致死率分析与治疗对策[J]. 基因组学与应用生物学,2020,39(09):4405-4408.

(下转第 235 页)