

成洪豪, 王学贺. 部分遮挡人脸识别在门禁管理中的应用[J]. 智能计算机与应用, 2024, 14(9): 70-75. DOI: 10. 20169/j. issn. 2095-2163. 240910

部分遮挡人脸识别在门禁管理中的应用

成洪豪, 王学贺

(菏泽医学专科学校 计算机教研室, 山东 菏泽 274000)

摘要: 近年来, 由于新冠疫情等公共卫生事件的影响, 佩戴口罩成了日常生活的新常态, 这也给传统的人脸识别技术带来了巨大挑战。在高校实验室等特殊场景下, 由于实验环境的敏感性和安全要求, 有必要对实验人员进行身份识别和行为监控, 因此亟需一种能够在口罩遮挡情况下准确识别人脸的解决方案。本文设计了一种基于 MTCNN 和改进 FaceNet 的佩戴口罩人脸识别系统, 并将其应用于高校实验室场景。人脸识别结果将被传回客户端, 同时系统还会在 Kurento 服务器中存储一段时间的视频片段, 可以有效支持实验室安全管理和行为审计。

关键词: 人脸识别; 部分遮挡; 口罩识别; 多任务级联卷积网络; 身份识别

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 2095-2163(2024)09-0070-06

Application of partially occluded face recognition in access control management

CHENG Honghao, WANG Xuehe

(Division of Computer Science, Heze Medical College, Heze 274000, Shandong, China)

Abstract: In recent years, due to the impact of public health events such as the COVID-19 pandemic, wearing masks has become the new normal in daily life. This has posed significant challenges to traditional facial recognition technologies. In special scenarios such as university laboratories, where sensitivity and safety requirements demand the identification and monitoring of individuals who often need to wear masks, there is an urgent need for a solution that can accurately recognize faces even when they are partially obscured by masks. This paper proposes a mask-wearing face recognition system based on MTCNN (Multi-Task Cascaded Convolutional Networks) and FaceNet, which is applied in the context of university laboratories. The results of face recognition are transmitted back to the client, and the system also stores video clips for a certain period in the Kurento server, effectively supporting laboratory security management and behavioral auditing. The system has been piloted in multiple laboratories and has achieved good application results. It is expected to provide an efficient and accurate personnel management and security solution for special scenarios such as university laboratories.

Key words: facial recognition; partial occlusion; mask recognition; multi-task cascaded convolutional network; identity recognition

0 引言

人脸识别研究历来专注于局部特征和浅层特征学习^[1], 直至 2015 年 FaceNet^[2] 的问世, 将研究重点转向深度学习领域。目前, ArcFace^[3] 和 CosFace^[4] 等深度学习技术已经在 LFW^[5] 的数据集中实现了超过 99.5% 的精确度, 这表明深度学习在这个领域的进步是巨大的。尽管如此, 这些基于深度学习的解决方案仍然无法应对诸如光线变化、姿势变动及遮挡等问题带来的影响。

在人脸识别中, 面部遮挡一直是一个棘手问题,

过去的研究主要处理眼镜、面饰等局部遮挡^[6-7]。然而, 口罩的出现使处理遮挡问题更加复杂, 导致鼻子、嘴巴等关键结构信息丢失, 给关键点位置信息引入更多噪音。

目前, 关于遮挡人脸识别的研究主要集中在三方面:

(1) 利用注意力机制的技术被提及。如文献^[8]通过剪切和注意力来解决戴口罩人的眼部问题。但是, 此技术在无遮挡情况下精确度有所降低。

(2) 如文献^[9]提出了一种名为 MTARCFACE 的损失函数, 其融合了 ArcFace 与口罩人脸分类损

作者简介: 成洪豪(1994-), 男, 硕士, 助教, 主要研究方向: 软件工程, 强化学习; 王学贺(1976-), 男, 硕士, 教授, 主要研究方向: 情感计算, 文本挖掘研究。Email: hzyzwangxuehe@163.com

收稿日期: 2024-03-21

哈尔滨工业大学主办 ◆ 学术研究与应用

失函数,旨在适应遮挡情况下的面部识别需求。

(3)文献[10]中的 ConstraintTripletLoss 也被用作优化嵌入过程,特别注重对眼睛部位的关注。第三条途径均采用了人脸修复的方式,如文献[11]中提及了使用生成对抗网络 GAN 来预测被遮挡的面部区域特征,以修复面部信息,接着输入到识别网络中。

本文使用改进的 FaceNet 的人脸识别网络^[12],包括应用于 InceptionResNet-v1 网络中的轻量级混合卷积注意模块,以增强其特征提取的能力并提高识别的精确度;使用 SoftMaxLoss^[13]与 TripletLoss 相结合的方式构建了新的损失函数。该系统的创新之处在于:利用多任务级联卷积网络(Multi-Task Cascaded Convolutional Networks, MTCNN)^[14]进行口罩检测,实现了对口罩佩戴情况的动态判断;根据判断结果选择合适的 FaceNet 模型权重,有效提高了识别精度;将人脸识别与视频存储相结合,为实验室安全管理提供了数据支持。

1 基础工作

1.1 MTCNN 模型基本原理

(1)对于识别各种大小的人脸, MTCNN 采用了一种名为图像金字塔技术。该技术是通过将初始图片进行一系列放大或缩小操作,来创建出一组新的图像。这个过程由 P-Net(Proposal Network)、R-Net(Refine Network)和 O-Net(Output Network)三级连贯的卷积神经网络完成,这些网络依次处理着各个尺度的图像,以实现人类面部检测功能。其中:

P-Net 的主要功能在于生成人脸区域内的候选窗格和对应的边界框回归向量。该网络利用卷积运算来提取图像特征,并运用全连接层计算出候选窗格存在的几率及相应的边界框偏移量;非极大值抑制(NMS)被应用于消除重叠的候选窗口。

R-Net 负责对 P-Net 生成的候选窗口进行进一步筛选,通过卷积层提取特征,并使用全连接层预测候选窗口的置信度以及边界框的偏移量。与 P-Net 类似, NMS 被应用于消除误检区域(false-positive regions)。R-Net 相较于 P-Net 增加了一个全连接层,从而更好地抑制误检区域。

O-Net 是最后一个级联网络,负责对人脸区域进行更精确的检测,并输出关键点(landmarks)。O-Net 增加了一层卷积层以提取更详细的特征,除预测人脸的存在概率和边界框偏移量外, O-Net 还输出 5 个关键点(landmarks)的位置,其中包括左右

眼、鼻子和左右嘴角。

(2)损失函数: MTCNN 采用多个损失函数来优化网络,人脸检测任务中使用了交叉熵损失函数,而在边界框回归和关键点定位任务中则采用了 L2 损失函数(均方误差)。这些损失函数分别乘以权重并相加,形成总损失。

MTCNN 对输入图像的处理流程如图 1 所示。通过 3 个级联网络的协作, MTCNN 能够在不同尺度的图像上实现高效且准确的人脸检测和关键点定位。

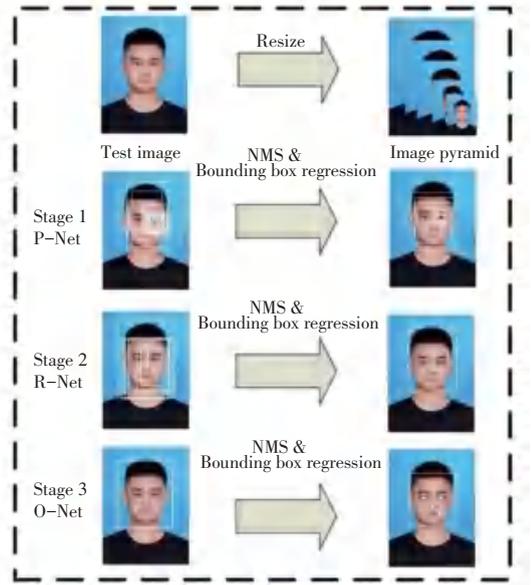


图 1 MTCNN 处理流程

Fig. 1 Processing flow of MTCNN

当整个图片的局部信息都进行处理之后,就能得到所有的局部人脸信息,或有或无,进行校正处理后可以得到最后的结果。

1.2 FaceNet 模型

FaceNet 是一种基于深度学习的人脸识别方法,其主要目的是把人的面部图像映射到一个高维的欧式空间(即 128 维度)。在欧式空间中,如果人脸图像相似,则其之间的距离会很小;反之,则较大。因此,通过计算人脸图像在这个空间中的距离,从而完成诸如身份辨认、确认及分组等工作。FaceNet 网络结构如图 2 所示。

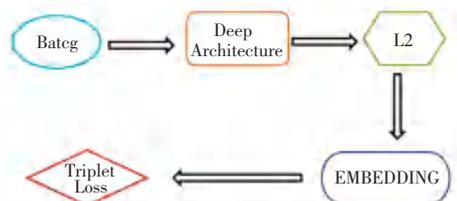


图 2 FaceNet 网络结构图

Fig. 2 Structure diagram of the FaceNet network

1.2.1 轻量型混合注意力模块

在识别被口罩遮挡的人脸时,因遮挡导致的部分信息无法识别,而未被遮挡的面部含有重要的识别信息。本文使用的改进 FaceNet 网络加入了一种轻量级的混合注意力机制^[15],将其应用至 InceptionResNet-v1

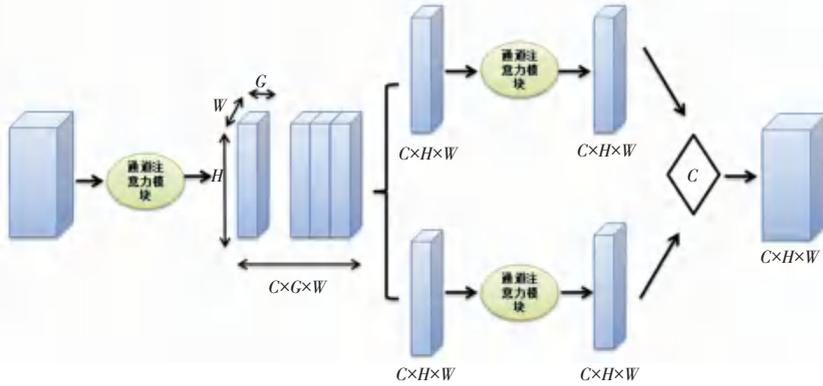


图3 轻量型混合注意力模块整体结构图

Fig. 3 Overall structure diagram of the lightweight hybrid attention module

使用 $F \in R^{C \times H \times W}$ 来表示输入的特征图。其中, C 是通道数, H 和 W 代表特征图的空间维度。通道注意力模块输出特征图的计算公式如下:

$$A_{out} = \text{Softmax}(\text{Transpose}(\text{AvgPool}^{2 \times 2}(\text{PWConv}^{1 \times 1}(F_{in})))) \quad (1)$$

$$\hat{F}_{out} = \text{Transpose}(\text{AvgPool}^{2 \times 2}(F_{in})) \quad (2)$$

$$F_{out} = F_{in} \oplus \text{Expand}(\text{PWConv}^{1 \times 1}(\hat{F}_{out} \odot A_{out})) \quad (3)$$

其中,用 PWConv 表示点卷积; \oplus 表示按位相加;用 \odot 表示矩阵乘法。

空间注意力模块沿通道采用平均池化操作和最大池化操作,计算空间注意力;然后将其连接成一个有效的特征图;最后通过卷积操作得到大小为 $R \times H \times W$ 的空间关注权重特征图。对于任意一组子空间中输入的特征图,其输出特征图的计算公式如下:

$$\hat{F}_{out} = \text{AvgPool}^{3 \times 3}(\text{DWConv}^{1 \times 1}(F_{in})) \oplus \text{MaxPool}^{3 \times 3}(\text{DWConv}^{1 \times 1}(F_{in})) \quad (4)$$

$$A_{out} = \text{Expand}(\text{Softmax}(\text{PWConv}^{1 \times 1}(\hat{F}_{out}))) \quad (5)$$

$$F_{out} = (F_{in} \otimes A_{out}) \oplus F_{in} \quad (6)$$

其中, \otimes 表示逐位相乘。

通过对初始影像进行低频与高频成分的拆解发现,低频部分主要反映了相对稳定的区域信息,而高频部分则揭示出更细致的图像元素。如果类别间的差别较大,采用多频特征学习的方法,可以让模型专注于非遮挡区段中人脸特征的辨识。

1.2.2 损失函数

在进行算法模型的训练过程中,关键要素是实

架构中的核心区域瓶颈层,以增强对人面部非遮挡部位特征的重视程度;依据待提取特征的重要性和评价标准,合理地调整特征图中空间和通道部分的权重,从而获得更具辨识度的人脸特征^[16]。轻量型混合注意力模块整体结构如图3所示。

际数据和输出值之间的误差距离,这一点在损失函数中有所体现。在 FaceNet 网络模型的损失函数中,采用了结合 SoftmaxLoss 和 TripletLoss 的加权融合方式。其中 SoftmaxLoss 主要关注类别的差异性,但对于分类内的分布表现并不出色;相比之下, TripletLoss 常被用以优化具有高相似性的样本^[17]。结合这两个损失函数来进行共同训练,能带来更好的模型收敛性能。

本文使用的联合损失函数:

$$Loss = \gamma L_t + \lambda L_s \quad (7)$$

式中: L_t 代表 TripletLoss, L_s 代表 SoftmaxLoss, γ 和 λ 分别代表 TripletLoss 和 SoftmaxLoss 这两个损失函数的权重。

1.3 Kurento 服务器

Kurento 作为一个功能丰富的 WebRTC 开源开发平台,其模块化的设计和易于集成的特点,使其成为开发现代视频类应用程序的理想选择。其核心部分包括高性能的 WebRTC 流媒介服务器。Kurento Media Server 构成了 Kurento 的主要组成元素,该服务器具有多种功能:如视频流转换、录像、混音等。由于 Kurento 的体系结构不需要依赖第三方插件,极大地降低了开发和维护的成本,同时也使得维护工作变得更加简便^[18]。

2 部分遮挡人脸识别系统总体设计

2.1 部分遮挡人脸识别系统运行时序

人脸识别系统运行主要步骤如下:

1) 视频获取

摄像头实时采集视频数据,并通过网络传输协议(RTMP)将视频流传输到 Kurento 服务器。

2) Kurento 媒体管道构建

Kurento 服务器接收到视频流后,创建新的媒体管道(MediaPipeline)。在管道中添加一个 PlayerEndPoint,用于接收视频流;添加一个 FaceOverlayFilter,用于在视频流中叠加人脸检测和识别的结果。

3) 人脸检测(MTCNN)

在 FaceOverlayFilter 中集成 MTCNN 模型,对视频流中的每一帧图像进行人脸检测。MTCNN 会输出每个检测到的人脸边界框坐标。

4) 口罩检测(MTCNN)

对于每个检测到的人脸,使用 MTCNN 的口罩检测功能判断该人脸是否佩戴了口罩,并输出该人脸是否佩戴口罩的标识。

5) 人脸识别(FaceNet)

对于每个检测到的人脸,将其 crop 输出作为单独的图像。根据该人脸是否佩戴口罩,选择对应的 FaceNet 模型权重。若佩戴口罩,使用专门针对口罩人脸训练的 FaceNet 模型权重;否则,使用通用的 FaceNet 模型权重。使用选定的 FaceNet 模型提取该人脸图像的特征向量,将提取的特征向量与已知的人脸特征向量库进行比对,找到最匹配的身份。

6) 结果叠加

将识别的结果(身份信息)通过 FaceOverlayFilter 叠加到原始视频流中。

7) 结果输出

在管道中添加一个 WebRtcEndPoint,用于将处理后的视频流传输回客户端应用。客户端应用接收并解码 WebRTC 视频流,展示带有识别结果的视频。

8) 事件处理

Kurento 服务器从人脸检测成功开始进行视频录制,当前设置为 10 min,方便实验室后续查看和审计,维护实验室的安全运行。当出现人脸检测成功但识别失败的情况,会对视频进行标注并发送警示给实验室管理人员。Kurento 服务器发送识别的事件到客户端应用,客户端应用根据接收到的事件数据,判断是否进行开门操作。

在上述过程中,关键环节是步骤 4 和步骤 5。通过 MTCNN 来判断人脸是否佩戴口罩,根据判断结果选择适当的 FaceNet 模型权重,从而更精确地

辨识出佩戴口罩的人脸。

2.2 人脸识别模块

本模块主要任务是构建人脸识别部分,负责对即将被检查的人员进行身份确认。首先,将摄像头捕捉到的影像立即传递给 Kurento 服务器,每张图片都会经过 MTCNN 处理以确定是否存在人脸,然后把这些已知有人的照片输入至人脸识别系统;最后,带有面部识别信息的视频将会反向播放回到屏幕上。人脸识别模块架构如图 4 所示,其主要实现步骤如下:

(1) 图像传入:接收经 MTCNN 检测到的人脸图像。

(2) 口罩检测:使用 MTCNN 对人脸图像进行口罩检测,如果检测图像为戴口罩图像,则传入由口罩图像集训练好的 FaceNet 模型,否则传入由正常图像集训练好的 FaceNet 模型。

(3) 特征提取:由 FaceNet 提取面部特征。

(4) 特征对比:将收集到的特征与已经储存在数据库里的人脸特征进行对比。本环节采用欧氏距离来评估两个特征间的相似度。如果匹配程度高于某个阈值,则认为这两个特征属于同一个人。

(5) 结论呈现:计算欧式距离(d),然后将其与设置好的阈值作对比。若欧式距离超过了该阈值,则认定为识别错误,给出“未知”的结果;若欧式距离低于阈值,则输出对应于此特征的身份标识信息。

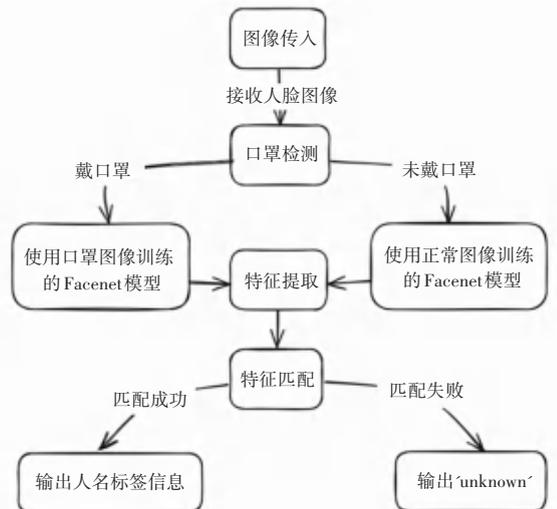


图 4 人脸识别模块的工作流程图

Fig. 4 Workflow diagram of the face recognition module

3 实验

3.1 实验环境

为了评估本文所设计的 MTCNN 与 FaceNet 构

建的部分遮挡人脸识别系统的表现,进行了相关的对比实验。硬件环境为:CPU为Intel Core i7-12700 K, GPU为NVIDIA GeForce RTX 3080,内存为32 GB DDR4 3200 MHz,内存,硬盘为1 TB NVMe M.2 SSD,摄像头为外置Logitech C920网络摄像头;软件环境:Windows 10操作系统,Pytorch框架,Python语言。

3.2 实验数据集

RMFRD^[19]中描述了3种独特类型的遮挡人脸数据集,其中包括口罩遮挡人脸检测数据集(MFDD)、真实口罩遮挡人脸识别数据集(RMFRD)和模拟蒙面人脸识别数据集(SMFRD)。MFDD数据集的部分样本来源于相关研究,其余则通过网络爬虫技术获取,该数据集共包含24 771张有遮挡的人脸图像。真实遮挡和模拟遮挡的样本如图5所示。

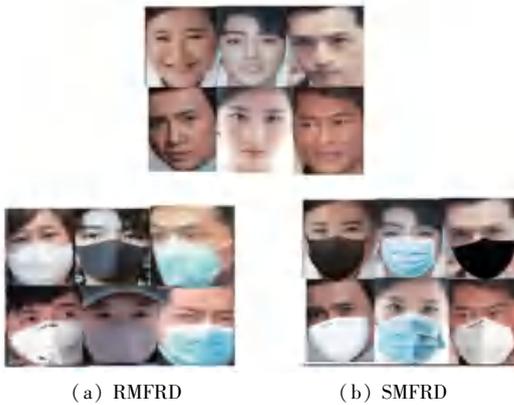


图5 真实遮挡和模拟遮挡样例

Fig. 5 Examples of real occlusion and simulated occlusion

3.3 结果分析

部分遮挡下的人脸识别能否真正应用到现实生活中主要体现在识别率和识别速度方面。为了检验本文采用的MTCNN+FaceNet模型在识别率和识别速度方面的性能,在相同数据集的情况下,与ArcFace、SphereFace、CosFace、FaceNet 4种在部分遮挡环境下表现优越的人脸识别模型进行了对比实验,实验结果如下:

1) 识别率比较

为了检验算法效果,按照8:2的比例,将数据集分为了训练集和测试集。与本文所提模型进行对比实验,4种模型的差异主要体现在使用的损失函数不同。SphereFace使用角度SoftmaxLoss作为损失函数,增强网络生成特征中同类别样本的相似性以及不同类别样本的区分度;CosFace使用了余弦距离;FaceNet使用了TripletLoss作为损失函数,提升

相似性,扩大差异性。数据对比结果见表1。

表1 识别率对比

Table 1 Comparison of recognition rates

方法	识别率/%			
	RMFRD (无遮挡)	SMFRD (无遮挡)	RMFRD (遮挡)	SMFRD (遮挡)
ArcFace	99.38	99.40	85.40	86.50
SphereFace	99.42	99.48	86.30	86.60
CosFace	99.60	99.70	86.70	87.20
FaceNet	99.63	99.65	87.68	88.20
本文算法			91.30	92.50

从表中数据可以看出,在无遮挡的情况下,现有人脸识别算法的准确度均在较高水平,但在有遮挡的情况下,所有的人脸识别算法准确率均有下降,而本文使用的改进FaceNet网络模型更专注于未遮挡部分的人脸特征,在准确率上超过其他网络。

2) 响应时间对比

在本系统中,首先使用MTCNN算法判断人是否佩戴口罩。根据口罩佩戴情况,采用不同权重的FaceNet模型进行人脸识别,系统的响应时间差异可参考表2。

表2 算法响应时间对比

Table 2 Comparison of algorithm response times

方法	识别速度/ms
ArcFace	14.7
SphereFace	16.8
CosFace	15.6
FaceNet	16.9
本文算法	25.6

本文所采用的模型较其他模型多了口罩检测环节,因此在响应时间方面略多于其他模型。根据结果分析,虽然响应时间有所增加,但仍然符合实际应用中的实时性需求。

3) 实验效果

实验中对4个人分别进行了100次测试,总共400次,实验结果见表3,识别效果如图6所示。

表3 真实环境测试

Table 3 Real environment test

测试人员	测试次数	正确次数	识别率/%
a	100	87	87
b	100	92	92
c	100	91	91
d	100	90	90



图6 运行测试图

Fig. 6 Operating test diagram

经过对上述测试结果的分析,本文设计的人脸辨识系统在一般环境下能够实现约90%的辨识准确率。从性能测试方面来看,平均响应时间约为0.53 s,所有响应时间均稳定在1 s以内。系统在用户体验良好的时间段内表现出优异性能,符合实时需求。

总体来看,MTCNN+FaceNet具有结构清晰、泛化能力强的特点,是一种通用且易于部署的解决方案。考虑到模型准确率、高校实验室的特殊环境、实时性需求、摄像设备以及部分遮挡数据集等因素,认为MTCNN+FaceNet方法更为适用。

4 结束语

本文提出了一种基于MTCNN和改进的FaceNet的部分遮挡人脸检测与识别系统,并将其应用于高校实验室环境中。该系统利用Kurento服务器作为后端,通过不同权重的FaceNet模型来识别戴口罩和不戴口罩的人员,从而提高了整体识别精度。该系统可实时检测进入实验室人员的身份,一旦发现无权限人员,将自动报警并存储一段时间的视频以备后查,有效加强了实验室的安全防范。

未来将进一步优化算法,增强对不同遮挡和角度的人脸识别能力。同时,计划将人脸识别系统与体温检测等其他安防模块集成,构建一个更加完善的实验室智能管理系统,为学校的师生创造一个更加安全可靠的科研环境。

参考文献

[1] WANG M, DENG W. Deep face recognition: A survey [J]. *Neurocomputing*, 2021, 429: 215-244.

[2] SCHROFF F, KALENICHENKO D, PHILBIN J. Facenet: A unified embedding for face recognition and clustering [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2015: 815-823.

[3] DENG J, GUO J, XUE N, et al. Arcface: Additive angular margin loss for deep face recognition [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 2019: 4690-4699.

[4] WANG H, WANG Y, ZHOU Z, et al. Cosface: Large margin cosine loss for deep face recognition [C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2018: 5265-5274.

[5] HUANG G B, MATTAR M, BERG T, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments [R]. University of Massachusetts Amherst Technical Report, 2008.

[6] JIA H, MARTINEZ A M. Support vector machines in face recognition with occlusions [C]// *Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009: 136-141.

[7] WU C Y, DING J J. Occluded face recognition using low-rank regression with generalized gradient direction [J]. *Pattern Recognition*, 2018, 80: 256-268.

[8] LI Y, GUO K, LU Y, et al. Cropping and attention based approach for masked face recognition [J]. *Applied Intelligence*, 2021, 51: 3012-3025.

[9] MONTERO D, NIETO M, LESKOVSKY P, et al. Boosting masked face recognition with multi-task arcface [C]// *Proceedings of 2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. IEEE, 2022: 184-189.

[10] NETO P C, BOUTROS F, PINTO J R, et al. My eyes are up here: Promoting focus on uncovered regions in masked face recognition [C]// *Proceedings of 2021 International Conference of the Biometrics Special Interest Group (BIOSIG)*. IEEE, 2021: 1-5.

[11] FARAHANIPAD F, REZAEI M, NASR M, et al. GAN-based face reconstruction for masked-face [C]// *Proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments*. IEEE, 2022: 583-587.

[12] 移洁,侯劲,石浩德. 基于轻量型网络的口罩遮挡人脸识别方法 [J]. *电子测量技术*, 2023, 46(6): 159-165.

[13] LI X, WANG W. Learning discriminative features via weights-biased softmax loss [J]. *Pattern Recognition*, 2020, 107: 107405.

[14] ZHANG K, ZHANG Z, LI Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks [J]. *IEEE Signal Processing Letters*, 2016, 23(10): 1499-1503.

[15] 张泽哈. 基于局部遮挡人脸识别的移动端人员管控系统设计与实现 [D]. 北京: 北京邮电大学, 2021.

[16] 冯宇平,管玉宇,杨旭睿,等. 融合注意力机制的实时行人检测算法 [J]. *电子测量技术*, 2021, 44(17): 123-130.

[17] 赵佳鑫,张雅丽,施新凯,等. 基于联合损失函数的多视角步态识别方法研究 [J]. *智能计算机与应用*, 2022, 12(2): 13-17.

[18] 刘杰,孙立民. 深度学习人脸识别技术在考勤系统的应用 [J]. *智能计算机与应用*, 2020, 10(2): 17-22.

[19] 张庆辉,张媛,张梦雅. 有遮挡人脸识别进展综述 [J]. *计算机应用研究*, 2023, 40(8): 2250-2257, 2273.