



(12) 发明专利申请

(10) 申请公布号 CN 112488034 A

(43) 申请公布日 2021. 03. 12

(21) 申请号 202011462599.2

(22) 申请日 2020.12.14

(71) 申请人 上海交通大学

地址 200240 上海市闵行区东川路800号

(72) 发明人 沈耀 刘韵

(74) 专利代理机构 上海交达专利事务所 31201

代理人 王毓理 王锡麟

(51) Int.Cl.

G06K 9/00 (2006.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

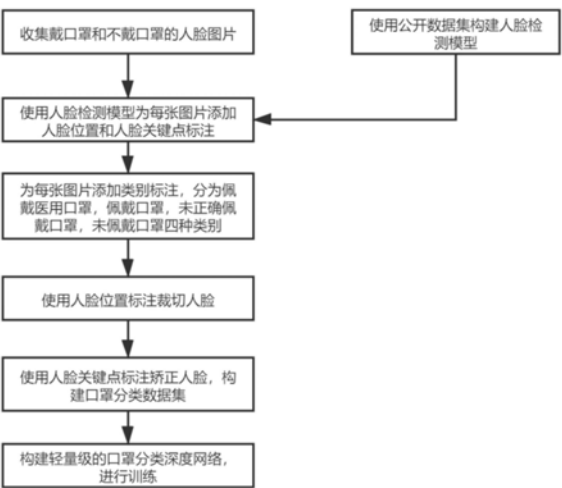
权利要求书2页 说明书5页 附图6页

(54) 发明名称

基于轻量化人脸口罩检测模型的视频处理方法

(57) 摘要

一种基于轻量化人脸口罩检测模型的视频处理方法,使用人脸数据集构建得到的人脸检测模型向样本图像中添加人脸位置标注和人脸关键点标注,并进一步在样本图像上对应的人脸位置添加人工标注后经裁剪处理,得到训练样本校正后用于对轻量级口罩分类深度模型进行训练,再将训练后得到的模型与人脸检测模型使用区域图像抽取模块进行级联得到视频流检测分类模块,并将视频流中每一帧连续输入到人脸检测模型,模型输出当前帧所含所有人脸的位置、人脸关键点和佩戴口罩情况的信息。本发明能够有效地节约模型存储空间和计算量,方便部署在低算力设备中。而且模型训练仅需要人为标注部分数据,易于结合多种已有的人脸检测模块,降低开发成本。



1. 一种基于轻量化人脸口罩检测模型的视频处理方法, 其特征在于, 使用人脸数据集构建得到的人脸检测模型向样本图像中添加人脸位置标注和人脸关键点标注, 并进一步在样本图像上对应的人脸位置添加人工标注后经裁剪处理, 得到训练样本校正后用于对轻量级口罩分类深度模型进行训练, 再将训练后得到的模型与人脸检测模型使用区域图像抽取模块进行级联得到视频流检测分类模块, 并将视频流中每一帧连续输入到人脸检测模型, 模型输出当前帧所含所有人脸的位置、人脸关键点和佩戴口罩情况的信息;

所述的样本图像中的每张图片中至少含有1张完整的人脸图像, 且同时包括带口罩和不带口罩的人脸照片;

所述的人工标注包括: 正确佩戴医用口罩、正确佩戴口罩、未正确佩戴口罩、未佩戴口罩。

2. 根据权利要求1所述的基于轻量化人脸口罩检测模型的视频处理方法, 其特征是, 所述的裁剪是指: 基于人脸位置标注对样本图像进行裁切; 并对裁剪后的人脸图片中的侧脸图片进行校正处理, 具体为: 根据人脸关键点对裁剪后的人脸图片中的侧脸图片进行校正, 得到校正后的正脸图片后进一步添加人工标注。

3. 根据权利要求1或2所述的基于轻量化人脸口罩检测模型的视频处理方法, 其特征是, 所述的人脸位置标注是指: 为样本图像中的每一张的人脸边框在图像中的左下角坐标和右上角坐标;

所述的人脸关键点标注是指: 为样本图像中的每一张图像的每一个人脸的左瞳孔中心坐标、右瞳孔中心坐标、鼻尖坐标、左嘴角坐标和右嘴角坐标; 标注中的每一个坐标为一个二维点, 原点为图片左上角, 单位为像素, 第一个维度代表从上到下的像素坐标, 第二个维度代表从左到右的像素坐标。

4. 根据权利要求2所述的基于轻量化人脸口罩检测模型的视频处理方法, 其特征是, 所述的裁切, 根据人脸边框先进行衬垫操作, 对于上下两边, 衬垫长度为人脸边框的高度乘以0.25后四舍五入取整, 对于左右两边衬垫长度为人脸边框的宽度乘以0.2后四舍五入取整, 衬垫操作后如果有边框落在图像外, 则以欧氏距离最近的像素进行填充。

5. 根据权利要求2所述的基于轻量化人脸口罩检测模型的视频处理方法, 其特征是, 所述的校正, 参考标准正脸的人脸关键点, 列出从当前关键点二维坐标至参考标准正脸关键点二维坐标的仿射变换方程, 从仿射变换方程解出原图到参考正脸图仿射变换的线性变换矩阵, 再对原图中的每一个点乘以变换矩阵得到校正后的人脸, 得到校正后的人脸, 之后以正脸为中心进行裁剪, 使用双线性插值的方法得到正脸图片并进一步通过人脸类别标注, 得到校正后的人脸图像数据集。

6. 根据权利要求2所述的基于轻量化人脸口罩检测模型的视频处理方法, 其特征是, 所述的轻量级口罩分类深度模型采用深度可分离卷积模块作为骨干网络的分类深度模型; 其训练, 优选采用交叉熵作为损失函数, 通过自适应矩估计 (Adam) 优化器进行梯度下降求得参数优化值的过程。

7. 一种实现上述任一权利要求所述方法的系统, 其特征在于, 包括: 人脸检测模块、人脸校正模块和口罩分类模块, 其中: 人脸检测模块对输入的视频流中每一帧的RGB三通道图像信息进行前向传播运算, 得到人脸位置信息和人脸关键点信息以及人脸位置置信度后, 再进行非极大值抑制运算得到人脸位置和人脸关键点信息, 人脸校正模块根据人脸位置和

人脸关键点信息进行图像裁切和仿射变换矩阵求解,得到仿射变换矩阵并根据矩阵对裁切出的图像进行仿射变换,得到正脸RGB三通道图像,口罩分类模块根据正脸RGB三通道图像经过双线性插值后进行前向传播计算,得到四种佩戴口罩情况的置信度。

8. 根据权利要求7所述的系统,其特征是,所述的人脸检测模块包括:视频帧抽取单元、模型计算单元、非极大值抑制单元以及输出整合单元,其中:视频帧抽取单元与视频流输入相连并抽取有效输入帧作为RGB三通道图片输入模型计算单元和人脸矫正模块输入单元,模型计算单元与非极大值抑制单元相连并输出人脸位置信息和人脸关键点信息以及人脸位置置信度,非极大值抑制单元与输出整合单元相连并筛选人脸位置信息,输出整合单元与人脸矫正模块相连并传输整合后的信息。

9. 根据权利要求7所述的系统,其特征是,所述的人脸校正模块包括:输入单元、校正单元以及输出单元,其中:输入单元与检测模块输入单元和输出单元相连,按照人脸位置和关键点裁切出人脸图像,输入校正单元,校正单元与输出单元相连并输出校正后的人脸图像,输出单元传输校正后的人脸图像信息。

10. 根据权利要求7所述的系统,其特征是,所述的口罩分类模块包括:输入单元、模型计算单元、以及输出单元,其中:输入单元与人脸校正模块的输出单元相连并传输校正后的人脸图像信息,模型计算单元与输出单元相连并传输四种不同类别的置信度信息,输出单元与模型计算单元相连并通过Softmax函数计算产生每张人脸对应类别的置信度信息。

基于轻量化人脸口罩检测模型的视频处理方法

技术领域

[0001] 本发明涉及的是一种图像处理领域的技术,具体是一种基于轻量化人脸口罩检测模型的视频处理方法。

背景技术

[0002] 当发生突发公共卫生事件时,人员在公共区域佩戴口罩是防止疾病传播的重要手段,当前主要有两种方案解决口罩检测问题。

[0003] 第一种方案是派遣防疫人员进行现场人工检查,这种方案的主要缺点是效率较低,增加人员聚集的风险。同时人力成本较高,需要多名防疫人员轮班作业才能保证时刻的监测。

[0004] 第二种方案是采用计算机视觉中目标检测的方法,训练或者微调(fine-tune)目标检测模型,得到口罩检测模型。这种方案的主要缺点是需要全新的训练和测试数据集。新数据集中需要同时标注人脸位置、是否佩戴口罩的信息。若要追求更高的准确度,还需要人脸关键点信息。然而相对于人脸是否佩戴口罩的标注,不同的人工标注者标注的人脸位置信息和人脸关键点信息很难保持一致,从而导致目标检测模型的损失在训练或者微调时难以收敛。同时,大规模的标注需要大量的人力成本,而且难以与现有的人脸检测模型融合。

发明内容

[0005] 本发明针对现有技术存在的上述不足,提出一种基于轻量化人脸口罩检测模型的视频处理方法,能够有效地节约模型存储空间和计算量,方便部署在低算力设备中。而且模型训练仅需要人为标注部分数据,易于结合多种已有的人脸检测模块,降低开发成本。

[0006] 所述的部分位置标注,包括:人脸位置标注和人脸关键点标注,

[0007] 本发明是通过以下技术方案实现的:

[0008] 本发明涉及一种基于轻量化人脸口罩检测模型的视频处理方法,使用人脸数据集构建得到的人脸检测模型向样本图像中添加人脸位置标注和人脸关键点标注,并进一步在样本图像上对应的人脸位置添加人工标注后经裁剪处理,得到训练样本校正后用于对轻量级口罩分类深度模型进行训练,再将训练后得到的模型与人脸检测模型使用区域图像抽取模块进行级联得到视频流检测分类模块,并将视频流中每一帧连续输入到人脸检测模型,模型输出当前帧所含所有人脸的位置、人脸关键点和佩戴口罩情况的信息。

[0009] 所述的样本图像优选为RGB三通道JPEG格式的图片,每张图片中至少含有1张完整的人脸图像,每个人脸图像不小于 30×30 像素。所有图片中的总人脸数目不少于10000个,其中佩戴医用口罩和佩戴口罩的人脸比例不少于20%。

[0010] 所述的样本图像优选同时包括带口罩和不带口罩的人脸照片。

[0011] 所述的人工标注包括:正确佩戴医用口罩、正确佩戴口罩、未正确佩戴口罩、未佩戴口罩。

[0012] 所述的裁剪是指:基于人脸位置标注对样本图像进行裁切。

[0013] 优选地,对裁剪后的人脸图片中的侧脸图片进行校正处理,具体为:根据人脸关键点裁剪后的人脸图片中的侧脸图片进行校正,得到校正后的正脸图片后进一步添加人工标注。

[0014] 所述的人脸数据集优选为公开的数据集为具有人脸关键点标注的WIDER FACE数据集。任意一种检测模型在使用WIDER FACE数据集训练精调后均可作为机器标注的标准使用。

[0015] 所述的人脸位置标注是指:为样本图像中的每一张的人脸边框在图像中的左下角坐标和右上角坐标。

[0016] 所述的人脸关键点标注是指:为样本图像中的每一张图像的每一个人脸的左瞳孔中心坐标、右瞳孔中心坐标、鼻尖坐标、左嘴角坐标和右嘴角坐标;标注中的每一个坐标为一个二维点,原点为图片左上角,单位为像素,第一个维度代表从上到下的像素坐标,第二个维度代表从左到右的像素坐标。

[0017] 所述的人工标注,优选在标注前首先人工筛选舍弃非人脸的机器标注和大小为 30×30 像素以下的机器标注,然后向每一个符合要求的机器标注中添加四种分类,对应为0对应未佩戴口罩;1对应未正确佩戴口罩;2对应正确佩戴口罩;3对应正确佩戴医用口罩。佩戴医用口罩是指正确佩戴满足YY/T0969-2013或YY0469-2011或GB19083-2010的口罩。正确佩戴是指佩戴口罩的方法满足中华人民共和国卫生部医院隔离技术规范附录A的佩戴要求。使用毛巾、纱巾、围巾等织物或者使用手等非口罩物品覆盖口鼻分类为未佩戴口罩。

[0018] 所述的裁切,优选为根据人脸边框先进行衬垫(padding)操作,对于上下两边,衬垫长度为人脸边框的高度乘以0.25后四舍五入取整,对于左右两边衬垫长度为人脸边框的宽度乘以0.2后四舍五入取整。衬垫操作后如果有边框落在图像外,则以欧氏距离最近的像素进行填充。

[0019] 所述的校正,优选参考标准正脸的人脸关键点,列出从当前关键点二维坐标至参考标准正脸关键点二维坐标的仿射变换方程,从仿射变换方程解出原图到参考正脸图仿射变换的线性变换矩阵,再对原图中的每一个点乘以变换矩阵得到校正后的人脸,得到校正后的人脸。之后以正脸为中心进行裁剪,使用双线性插值的方法得到大小为 128×128 像素的正脸图片,复制步骤4中的人脸类别标注,得到校正后的人脸图像数据集。

[0020] 所述的轻量级口罩分类深度模型采用深度可分离卷积模块作为骨干网络的分类深度模型。

[0021] 所述的训练,优选采用交叉熵作为损失函数,通过自适应矩估计(Adam)优化器进行梯度下降求得参数优化值的过程。

[0022] 本发明涉及一种实现上述方法的系统,包括:人脸检测模块、人脸校正模块和口罩分类模块,其中:人脸检测模块对输入的视频流中每一帧的RGB三通道图像信息进行前向传播运算,得到人脸位置信息和人脸关键点信息以及人脸位置置信度后,再进行非极大值抑制运算得到人脸位置和人脸关键点信息,人脸校正模块根据人脸位置和人脸关键点信息进行图像裁切和仿射变换矩阵求解,得到仿射变换矩阵并根据矩阵对裁切出的图像进行仿射变换,得到正脸RGB三通道图像,口罩分类模块根据正脸RGB三通道图像经过双线性插值后进行前向传播计算,得到四种佩戴口罩情况的置信度。

技术效果

[0023] 本发明整体解决了现有技术无法在视频流中实时进行人脸口罩佩戴情况的检测；与现有技术相比，本发明通过机器视觉对全新数据集进行部分位置标注的同时为每一个人脸位置人工标注四种不同的佩戴口罩的类型以保证标注准确性和一致性的同时有效降低标注成本，并采用卷积神经网络作为特征提取器，将提取到的特征进行分类，并以包含深度可分离卷积模块(depwise separable convblock, CONV_DW)的口罩分类网络，采用双步式目标检测器的设计思想，利用现有的经WIDER FACE公开数据集训练的人脸检测网络作为特征提取器，得到人脸位置信息的同时产生是否正确佩戴口罩以及佩戴口罩的类型精确分类结果。

附图说明

[0024] 图1为本发明实施流程图；

[0025] 图2为基本卷积模块(Conv Block)示意图；

[0026] 图3为深度可分离卷积模块(Depwise Separable ConvBlock, Conv_DW Block)示意图；

[0027] 图4为口罩分类网络结构示意图；

[0028] 图5为实施例系统示意图；

[0029] 图6为实施例口罩检测效果展示示意图。

具体实施方式

[0030] 如图1所示，为本实施例涉及一种基于轻量化人脸口罩检测模型的视频处理方法，具体包括以下步骤：

[0031] 步骤1、收集包含15412人脸的RGB三通道JPEG格式的图片，每张图片中至少含有1张完整的人脸图像，每个人脸图像不小于 30×30 像素。其中佩戴医用口罩和佩戴口罩的人脸有8935个。

[0032] 步骤2、使用轻量化后的单步式检测器SSD检测模型作为基础模型，只保留前三个检测头，使用WIDER FACE数据集训练精调，作为机器标注的标准模型使用。

[0033] 步骤3、使用轻量化后的SSD标注人脸位置。人脸位置为步骤1中的每一张图像的人脸边框在图像中的左下角坐标和右上角坐标；人脸关键点为步骤1中的每一张图像的每一个人脸的左瞳孔中心坐标、右瞳孔中心坐标、鼻尖坐标、左嘴角坐标和右嘴角坐标。标注中的每一个坐标为一个二维点，原点为图片左上角，单位为像素，第一个维度代表从上到下的像素坐标，第二个维度代表从左到右的像素坐标。

[0034] 步骤4、为步骤三中每一个检测出的人脸进行四种不同的人工标注。在标注前会要求标注人员首先人工筛选舍弃非人脸的机器标注和大小为 30×30 像素以下的机器标注，再，向每一个符合要求的机器标注中添加四种分类，使用0, 1, 2, 3四种数字表示。0对应未佩戴口罩；1对应未正确佩戴口罩；2对应正确佩戴口罩；3对应正确佩戴医用口罩。佩戴医用口罩是指正确佩戴满足YY/T0969-2013或YY0469-2011或GB19083-2010的口罩。正确佩戴是指佩戴口罩的方法满足中华人民共和国卫生部医院隔离技术规范附录A的佩戴要求。使用毛巾、纱巾、围巾等纺织物或者使用手等非口罩物品覆盖口鼻分类为未佩戴口罩。

[0035] 步骤5、根据步骤4中的人脸边框先进行衬垫(padding)操作，对于上下两边，衬垫

长度为人脸边框的高度乘以0.25后四舍五入取整,对于左右两边衬垫长度为人脸边框的宽度乘以0.2后四舍五入取整。衬垫操作后如果有边框落在图像外,则以欧氏距离最近的像素进行填充。

[0036] 步骤6、人脸图像校正。使用步骤3中生成的人脸关键点,参考标准正脸的人脸关键点,解出从原图到正脸图仿射变换的线性变换矩阵,再对原图中的每一个点乘以变换矩阵得到校正后的人脸,得到校正后的人脸。之后以正脸为中心进行裁剪,使用双线性插值的方法得到大小为 128×128 像素的正脸图片,复制步骤4中的人脸类别标注,得到校正后的人脸图像数据集。

[0037] 步骤7、使用图1中展示的基本卷积模块和图2中展示的深度可分离卷积模块构建如图3展示的轻量级口罩分类深度模型。使用步骤6中的数据集利用自适应矩估计优化器优化参数,得到口罩分类模型。

[0038] 最后,将口罩分类模型封装为一个输入校正后的人脸图片,输出口罩分类结果的模块。

[0039] 如图5所示,为本实施例涉及的一种实现上述方法的系统,包括:轻量化后的单步式检测器SSD检测模型封装而成的人脸检测模块、线性变换矩阵操作封装而成的人脸校正模块、经过口罩覆盖后的FDDB数据集训练得到的Resnet模型封装而成的人脸识别模块和口罩分类模块,其中:人脸检测模块接受输入的视频流中每一帧的RGB三通道图像信息进行前向传播运算,得到人脸位置信息和人脸关键点信息以及人脸位置置信度,经非极大值抑制运算后输出至人脸校正模块,人脸校正模块根据人脸位置和人脸关键点信息进行图像裁切和仿射变换矩阵求解,得到仿射变换矩阵并根据矩阵对裁切出的图像进行仿射变换,得到正脸RGB三通道图像,口罩分类模块根据正脸RGB三通道图像经过双线性插值后进行前向传播计算,得到四种佩戴口罩情况的置信度。

[0040] 所述的人脸检测模块包括:视频帧抽取单元、模型计算单元、非极大值抑制单元以及输出整合单元,其中:视频帧抽取单元与视频流输入相连并抽取有效输入帧作为RGB三通道图片输入模型计算单元和人脸校正模块输入单元,模型计算单元与非极大值抑制单元相连并输出人脸位置信息和人脸关键点信息以及人脸位置置信度,非极大值抑制单元与输出整合单元相连并筛选人脸位置信息,输出整合单元与人脸校正模块相连并传输整合后的信息。

[0041] 所述的人脸校正模块包括:输入单元、校正单元以及输出单元,其中:输入单元与检测模块输入单元和输出单元相连,按照人脸位置和关键点裁切出人脸图像,输入校正单元,校正单元与输出单元相连并输出校正后的人脸图像,输出单元与人脸识别模块的输入单元和口罩分类的输入单元相连并传输校正后的人脸图像信息。

[0042] 所述的人脸识别模块包括:输入单元、模型计算单元、以及比对单元,其中:输入单元与人脸校正模块的输出单元相连并传输校正后的人脸图像信息,模型计算单元与比对单元相连并传输512维正则化后的浮点数人脸特征信息,比对单元与模型计算单元相连并传输人脸身份信息。

[0043] 所述的口罩分类模块包括:输入单元、模型计算单元、以及输出单元,其中:输入单元与人脸校正模块的输出单元相连并传输校正后的人脸图像信息,模型计算单元与输出单元相连并传输四种不同类别的置信度信息,输出单元与模型计算单元相连并通过Softmax

函数计算产生每张人脸对应类别的置信度信息。

[0044] 所述的系统可部署在服务器上作为API被用户远程调用,也可部署在如树莓派的嵌入式设备中作为本地服务提供无网络下的服务。

[0045] 经过具体实际实验,在Intel i7-7700 CPU作为算力单元,采用python3.7语言环境,以非极大值抑制阈值0.4,非极大值抑制最佳值取5000,筛选最佳值取750,人脸检测模块置信度阈值0.6,人脸检测模块模型计算单元输入长边为640像素,宽边等比例缩放,人脸识别和口罩分离模块模型计算单元输入大小为 128×128 像素参数运行上述系统1000次。如图6所示,其中人脸由检测框标出,五个数字分别代表人脸检测的置信度,和正确佩戴口罩、正确佩戴医用口罩、未正确佩戴口罩、未佩戴口罩的置信度,其余点为人脸关键点位置标识;口罩分离模块模型计算单元平均每张人脸耗时6ms,单人脸视频流从视频流输入到口罩分类结果产生的平均耗时为153ms/帧。在WIDER FACE验证集中,人脸检测EasyAP超过0.7,在医用口罩分类测试集Medical Mask上,分类准确度超过70%。

[0046] 与现有技术相比,本发明提供多种分类结果,能够详细分析人员是否佩戴口罩、佩戴是否正确以及佩戴的口罩类型;分类速度较快,在实验中单张人脸的分类速度小于7ms。

[0047] 上述具体实施可由本领域技术人员在不背离本发明原理和宗旨的前提下以不同的方式对其进行局部调整,本发明的保护范围以权利要求书为准且不由上述具体实施所限,在其范围内的各个实现方案均受本发明之约束。

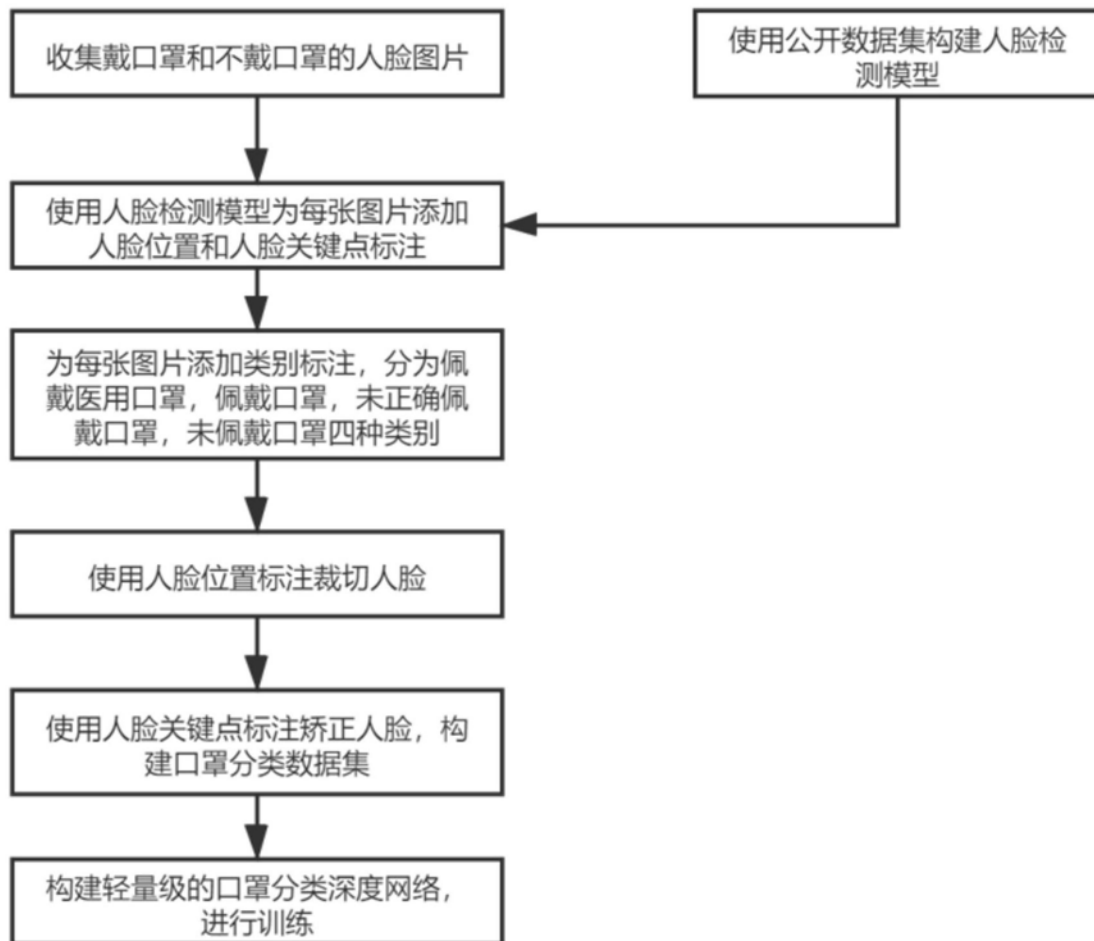


图1

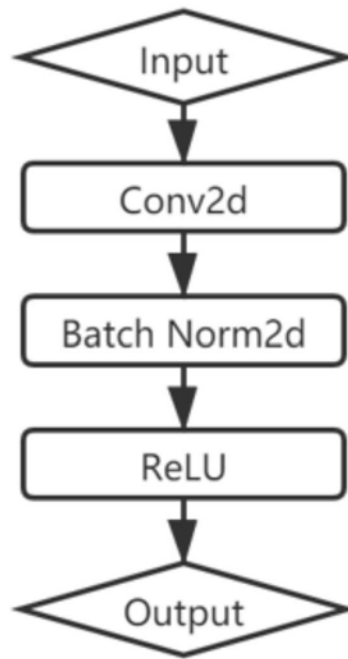


图2

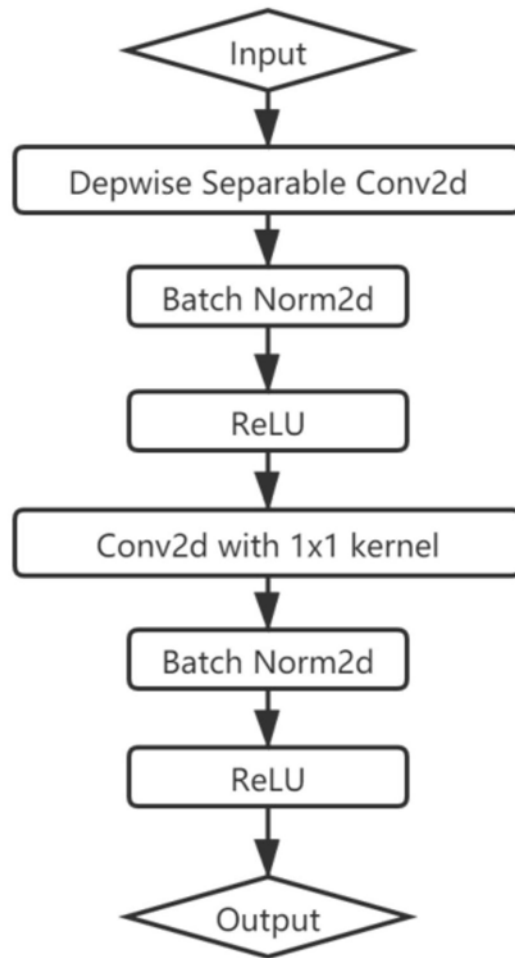


图3

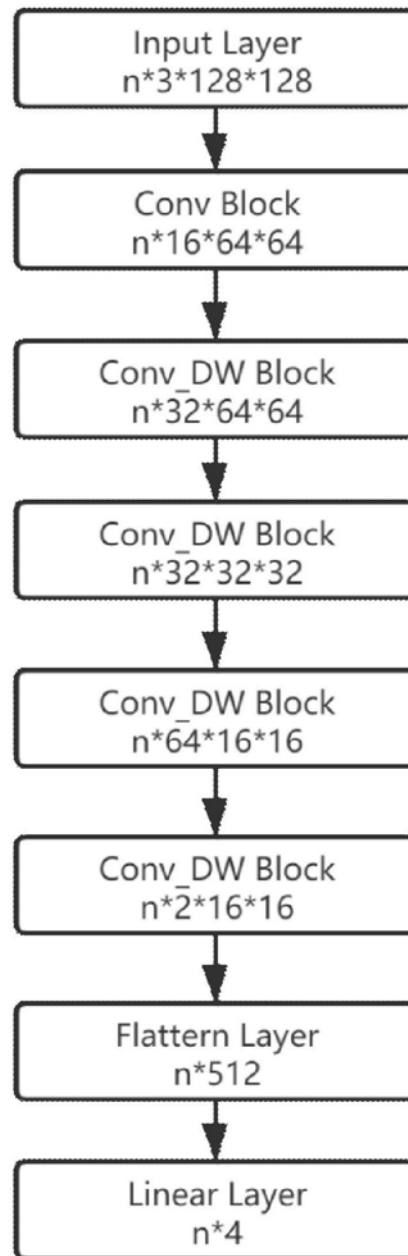


图4

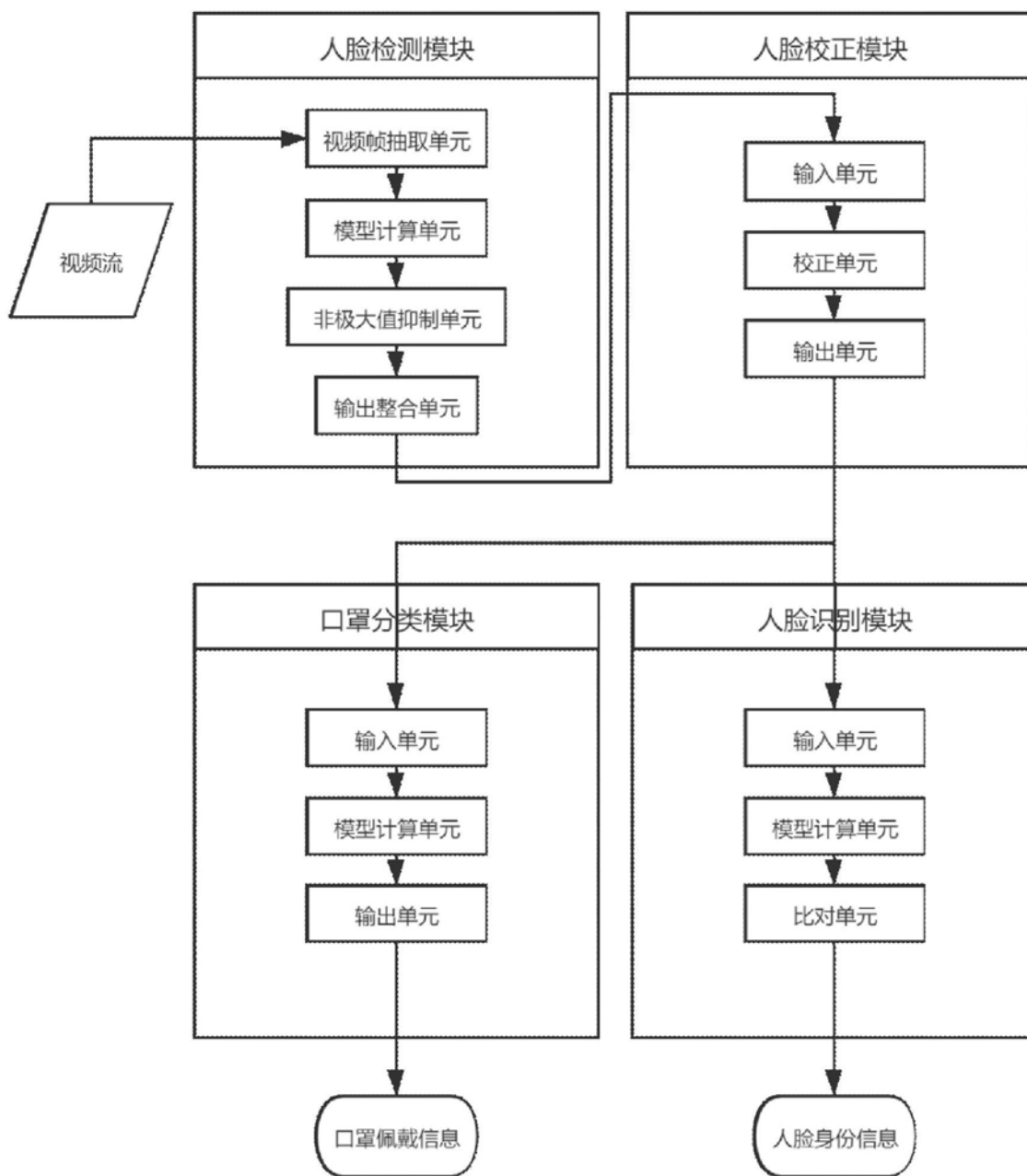


图5



图6