(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 113361397 A
(43)申请公布日 2021.09.07

(21)申请号 202110625104.1

(22)申请日 2021.06.04

(71)申请人 重庆邮电大学
地址 400065 重庆市南岸区南山街道崇文路2号

(72)发明人 张伟 虞继敏 周尚波 张鑫 秦毛伟 吴涛 王首刚

(74)专利代理机构 重庆辉腾律师事务所 50215
代理人 卢胜斌

(51)Int.Cl.
G06K 9/00(2006.01)
G06K 9/62(2006.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

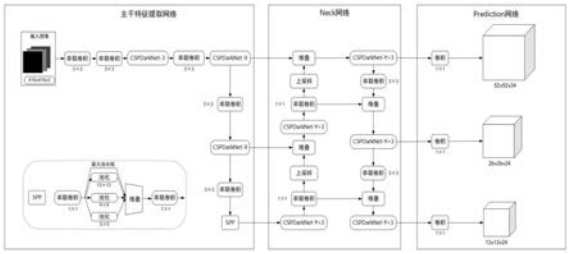
权利要求书2页 说明书13页 附图9页

(54)发明名称

一种基于深度学习的人脸口罩佩戴情况检测方法

(57)摘要

本发明属于深度学习领域,具体涉及一种基于深度学习的人脸口罩佩戴情况检测方法,该方法包括:实时获取待检测图像数据,将待检测图像输入到训练好的口罩检测网络模型中,得到检测结果;根据检测结果对待检测图像进行标记;所述口罩检测网络模型包括主干特征提取网络模、Neck网络模块以及Prediction网络;本发明在口罩检测模型中的主干特征提取网络中使用CSPDarkNet-X模块不但可以加强模型的特征提取能力,同时可以降低模型的参数量并且简化了模型的骨干网络的结构,使得模型的特征学习能力得到提升。



CN 113361397 A

1.一种基于深度学习的人脸口罩佩戴情况检测方法,其特征在于,包括:实时获取待检测图像数据,对待检测数据进行图像增强处理;将增强后的待检测图像输入到训练好的口罩检测网络模型中,得到检测结果;根据检测结果对待检测图像进行标记;所述口罩检测网络模型包括主干特征提取网络模、Neck网络模块以及Prediction网络;所述主干特征提取模块包括残差块和CSPDarkNet-X网络模块;所述Neck网络模块包括改进的PANet网络结构和FPN网络;采用CSPDarkNet-Y网络模块对PANet网络结构进行改进优化。

2.根据权利要求1所述的一种基于深度学习的人脸口罩佩戴情况检测方法,其特征在于,对口罩检测网络模型进行训练的过程包括:

S1:获取原始图像数据,对原始图像数据进行分类处理,即分为未佩戴口罩图像、未规范佩戴口罩图像以及规范佩戴口罩图;将分后的图像数据作为训练集;

S2:将训练集中的数据输入到主干特征提取模块中进行特征提取,得到全局特征;

S3:将主干特征网络模块提取的全局特征输入到Neck网络模块,得到多尺度局部特征;

S4:将全局特征和多尺度局部特征进行融合处理,得到融合特征图;

S5:将融合特征图输入到Prediction网络中,得到目标定位结果;

S6:根据目标定位结果计算模型的损失函数,并采用模型优化器不断调整模型的参数;当损失函数的值最小时,完成模型的训练。

3.根据权利要求2所述的一种基于深度学习的人脸口罩佩戴情况检测方法,其特征在于,主干特征提取模块包括三个CSPDarkNet-X网络模块,每个CSPDarkNet-X网络模块中包含X个残差块,X个残差块串联构成一个残差单元组;将三个CSPDarkNet-X网络模块串联,构成主干特征提取模块。

4.根据权利要求2所述的一种基于深度学习的人脸口罩佩戴情况检测方法,其特征在于,采用主干特征提取模块提取图像的全局特征的过程包括:

S21:对训练集中的图像进行卷积操作,得到浅层特征图;

S22:将浅层特征图输入到CSPDarkNet-X网络进行纹理特征提取并进行堆叠操作获取更丰富的特征层信息;

S23:将经过卷积操作和CSPDarkNet-X模块特征提取后的特征信息输入到SPPNet模块中进行池化操作,从而提高主干特征的感受野并分离出特征层的上下文信息;

S24:将经过主干特征提取模块处理后的特征信息分为三部分输入到Neck网络中进行特征融合,得到全局特征图。

5.根据权利要求2所述的一种基于深度学习的人脸口罩佩戴情况检测方法,其特征在于,采用Neck网络模块对全局特征进行处理的过程包括:

S31:将来自主干特征提取网络的全局特征信息进行堆叠和上采样操作,得到多尺度特征信息;

S32:将多尺度特征信息输入到CSPDarkNet-Y模块中进行多尺度特征信息的提取;

S33:将提取的多尺度特征信息进行3×3的卷积操作,即让特征层的高和宽变成原来的1/2;

S34:将上述步骤得到的多尺度特征信息进行特征融合,得到多尺度局部特征图。

6.根据权利要求2所述的一种基于深度学习的人脸口罩佩戴情况检测方法,其特征在于,采用Prediction网络对融合特征图进行处理的过程包括:

步骤1:将Neck网络融合后的特征图划分为三个有效特征层,将三个有效特征层输入到预测网络中;三个有效特征层依次为13×13×24、26×26×24、52×52×24的网格;

步骤2:对三个有效特征层的每一个网格设置三个先验框,当图像的中心落在先验框中时,则由该先验框对该图像进行检测;

步骤3:采用非极大值抑制算法不断对先验框的尺寸进行学习并调整,从而不断地趋近于预测框的位置;当先验框的尺寸达到极大值时筛选出最终的预测框,该预测框为模型的检测结果。

7.根据权利要求2所述的一种基于深度学习的人脸口罩佩戴情况检测方法,其特征在于,模型的损失函数表达式为:

$$Loss\text{-}function = CIOU_{loss} + Confidence_{loss} + Class_{loss}$$

$$Confidence_{loss} = -\sum_{i=0}^{S^2}\sum_{j=0}^{B} 1_{ij}^{obj}\left[\hat{C}_i^j \log\left(C_i^j\right) + \left(1-\hat{C}_i^j\right)\log\left(1-C_i^j\right)\right] -$$

$$\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B} 1_{ij}^{noobj}\left[\hat{C}_i^j \log\left(C_i^j\right) + \left(1-\hat{C}_i^j\right)\log\left(1-C_i^j\right)\right]$$

$$Class_{loss} = -\sum_{i=0}^{S^2} 1_{ij}^{obj} \sum_{c\subset class}\left[\hat{P}_{i,c}^j \log\left(P_{i,c}^j\right) + \left(1-\hat{P}_{i,c}^j\right)\log\left(1-P_{i,c}^j\right)\right]$$

其中,Loss表示损失函数,function表示函数,$CIOU_{loss}$表示目标的定位损失函数,$Confidence_{loss}$表示目标的置信度损失函数,$Class_{loss}$表示目标的分类损失函数,S表示网格大小,B表示每一个网格所对应的先验框个数,1_{ij}^{obj}表示特征层上第i个网格点中的第j个先验框,obj表示目标,$\hat{C}_i^j$表示是否负责物体的检测,C_i^j表示口罩检测模型的实际输出值,λ_{noobj}表示权重系数,1_{ij}^{noobj}表示特征层上第i个网格点中的第j个先验框内不包含物体,$\hat{P}_{i,c}^j$表示真实框中的物体属于类别c的概率,$P_{i,c}^j$表示模型预测第i个网格点中的第j个先验框是类别c的概率。

8.根据权利要求2所述的一种基于深度学习的人脸口罩佩戴情况检测方法,其特征在于,采用模型优化器不断调整模型的参数的过程包括:对模型参数初始化;采用Adam优化器对模型的参数进行学习并优化;在对参数进行学习优化过程中设置批大小为8,IOU为0.5,模型训练世代为50次,得到优化后的模型参数。

一种基于深度学习的人脸口罩佩戴情况检测方法

技术领域

[0001] 本发明属于深度学习领域，具体涉及一种基于深度学习的人脸口罩佩戴情况检测方法。

背景技术

[0002] 呼吸道传染疾病病毒、有毒有害气体、粉尘等对人体有害物质可以通过人体的呼吸循环系统进行体内，从而对人体造成损伤，严重时甚至危及到生命。而规范地佩戴口罩可有效地阻止病毒、粉尘、有毒有害气体进入肺部、神经等生命组织，并降低人员被感染、作为传播源、中毒损伤的风险。人脸规范佩戴口罩检测算法可应用与小区门禁、机场、校园、街道、医院等公共场所，该算法通过对人脸佩戴口罩的情况进行特征分析，从而确定人脸是否规范地佩戴口罩，以及是否佩戴口罩。

[0003] 目前，针对于人脸口罩识别的相关研究较少，其关键是：首先定位出人脸在图像中位置；然后识别出数据集所给出的人脸是否佩戴口罩以及是否正确佩戴口罩的类别。而现有的口罩数据集较少而且存在类别、环境、尺度信息不丰富等问题。复杂的环境可以通常为脸部遮挡、人脸尺度多变、光照不均衡、密集等，这些问题则是影响检测算法性能的主要原因，同时这类问题易导致模型的泛化能力差、信息冗余、精度低、实时性差等问题。

[0004] 目前，基于深度学习的目标检测算法存在的问题是：实时性较差、模型规模大、小目标检测效果差、鲁棒性差、适应性差、泛化能力弱等问题。同时，由于算法在实际的复杂环境中使用时，由于目标存在遮挡、多尺度等外在干扰，直接使用当前主流的目标检测算法进行人脸口罩检测仍然存在些许不足之处。其主要表现为：针对于多尺度的目标还存在着浅层特征提取不充分等问题；在实际的使用中，模型存在大量的冗余信息，使得模型在推理上具有更多的运行成本；模型存在训练成本高、计算消耗大、模型过于复杂以至于难以部署等问题。因此，本发明基于现有的目标检测算法，构建出一种实时性高、鲁棒性强、具有较高检测精度的目标检测算法，并将其应用在口罩检测任务上。

发明内容

[0005] 为解决以上现有技术存在的问题，本发明提出了一种基于深度学习的人脸口罩佩戴情况检测方法，该方法包括：实时获取待检测图像数据，对待检测数据进行图像增强处理；将增强后的待检测图像输入到训练好的口罩检测网络模型中，得到检测结果；根据检测结果对待检测图像进行标记；所述口罩检测网络模型包括主干特征提取网络模、Neck网络模块以及Prediction网络；所述主干特征提取模块包括残差块和CSPDarkNet-X网络模块；所述Neck网络模块包括改进的PANet网络结构和FPN网络；采用CSPDarkNet-Y网络模块对PANet网络结构进行改进优化。

[0006] 优选的，对口罩检测网络模型进行训练的过程包括：

[0007] S1：获取原始图像数据，对原始图像数据进行分类处理，即分为未佩戴口罩图像、未规范佩戴口罩图像以及规范佩戴口罩图；将分后的图像数据作为训练集；

[0008] S2:将训练集中的数据输入到主干特征提取模块中进行特征提取,得到全局特征;

[0009] S3:将主干特征网络模块提取的全局特征输入到Neck网络模块,得到多尺度局部特征;

[0010] S4:将全局特征和多尺度局部特征进行融合处理,得到融合特征图;

[0011] S5:将融合特征图输入到Prediction网络中,得到目标定位结果;

[0012] S6:根据目标定位结果计算模型的损失函数,不断调整模型的参数,当损失函数的值最小时,完成模型的训练。

[0013] 进一步的,主干特征提取模块包括三个CSPDarkNet-X网络模块,每个CSPDarkNet-X网络模块中包含X个残差块,X个残差块串联构成一个残差单元组;将三个CSPDarkNet-X网络模块串联,构成主干特征提取模块。

[0014] 优选的,采用主干特征提取模块提取图像的全局特征的过程包括:

[0015] S21:对训练集中的图像进行卷积操作,得到浅层特征图;

[0016] S22:将浅层特征图输入到CSPDarkNet-X网络进行纹理特征提取并进行堆叠操作获取更丰富的特征层信息;

[0017] S23:将经过卷积操作和CSPDarkNet-X模块特征提取后的特征信息输入到SPPNet模块中进行池化操作,从而提高主干特征的感受野并分离出特征层的上下文信息;

[0018] S24:将经过主干特征提取模块处理后的特征信息分为三部分输入到Neck网络中进行特征融合。

[0019] 优选的,采用Neck网络模块对全局特征进行处理的过程包括:

[0020] S31:将来自主干特征提取网络的全局特征信息进行堆叠和上采样操作,得到多尺度特征信息;

[0021] S32:将多尺度特征信息输入到CSPDarkNet-Y模块中进行多尺度特征信息的提取;

[0022] S33:将提取的多尺度特征信息进行3×3的卷积操作,即让特征层的高和宽变成原来的1/2;

[0023] S34:将上述步骤得到的多尺度特征信息进行特征融合,从而实现多尺度特征信息的融合并且加强了特征信息的流通,这有利于提高模型的检测精度。

[0024] 优选的,采用Prediction网络对融合特征图进行处理的过程包括:

[0025] 步骤1:将Neck网络融合后的特征图划分为三个有效特征层,将三个有效特征层输入到预测网络中;三个有效特征层依次为13×13×24、26×26×24、52×52×24的网格;

[0026] 步骤2:对三个有效特征层的每一个网格设置三个先验框,当图像的中心落在先验框中时,则由该先验框对该图像进行检测;

[0027] 步骤3:采用非极大值抑制算法不断对先验框的尺寸进行学习并调整,从而不断地趋近于预测框的位置;当先验框的尺寸达到极大值时筛选出最终的预测框,该预测框为模型的检测结果。

[0028] 优选的,模型的损失函数表达式为:

[0029] $\text{Loss-function}=\text{CIOU}_{\text{loss}}+\text{Confidence}_{\text{loss}}+\text{Class}_{\text{loss}}$

[0030] $$Confidence_{loss}=-\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{obj}\left[\hat{C}_i^j\log\left(C_i^j\right)+\left(1-\hat{C}_i^j\right)\log\left(1-C_i^j\right)\right]-$$

[0031] $$\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{noobj}\left[\hat{C}_i^j\log\left(C_i^j\right)+\left(1-\hat{C}_i^j\right)\log\left(1-C_i^j\right)\right]$$

[0032] $$Class_{loss}=-\sum_{i=0}^{S^2}1_{ij}^{obj}\sum_{c\subset class}\left[\hat{P}_{i,c}^j\log\left(P_{i,c}^j\right)+\left(1-\hat{P}_{i,c}^j\right)\log\left(1-P_{i,c}^j\right)\right]$$

[0033] 优选的,采用模型优化器不断调整模型的参数的过程包括:对模型参数初始化;采用Adam优化器对模型的参数进行学习并优化;在对参数进行学习优化过程中设置批大小为8,IOU为0.5,模型训练世代为50次,得到优化后的模型参数。

[0034] 本发明的有益效果:

[0035] (1):在口罩检测模型中的主干特征提取网络中使用CSPDarkNet-X模块不但可以加强模型的特征提取能力,同时可以降低模型的参数量并且简化了模型的骨干网络的结构,使得模型的特征学习能力得到提升。

[0036] (2):在口罩检测模型的颈部网络中使用CSPDarkNet-Y网络模块可以有效地简化模型结构,同时提高模型的特征提取能力以及加强特征融合的能力。

[0037] (3):在Neck中应用CSPDarkNet-Y模块来增加信息融合。在网络的低层特征中通常含有更多的目标位置信息,而高层特征中则含有更多的目标语义信息。在Neck网络模块中,FPN模块是模型在下采样操作结束后再进行上采样,并通过横向连接获取同级下采的信息的操作。

[0038] (4):人脸口罩检测算法用于实现复杂的函数表示,而非线性激活函数能够使得人脸口罩检测模型逼近于任意的复杂函数。一般而言,人脸佩戴口罩处于的复杂环境可以总结为:人脸处于遮挡、多尺度、光照、密集等环境,这些环境会导致人脸特征缺失、不足等问题,而这类问题易降低模型的检测精度,造成漏检和误检等情况出现。为此,本专利使用Hard-Swish作为激活函数,以此提高模型的非线性特征学习能力,从而加强模型对特征的提取,同时提高模型的检测精度。

附图说明

[0039] 图1为本发明的串联卷积操作模块结构图;

[0040] 图2为本发明的残差模块结构图;

[0041] 图3为本发明的CSPDarkNet-X网络模块的结构图;

[0042] 图4为本发明的CSPDarkNet-Y网络模块的结构图;

[0043] 图5为本发明的一种口罩检测网络模型的结构图;

[0044] 图6为本发明的一种13×13的有效特征层示意图;

[0045] 图7为本发明的真实框和预测框的关系图;

[0046] 图8为本发明的对口罩检测网络模型进行训练的一种实施例流程图;

[0047] 图9为本发明的对口罩检测网络模型训练的结果图;

[0048] 图10为本发明的对网络进行预测的流程图;

[0049] 图11为本发明的口罩检测网络模型的P-R曲线图;

[0050] 图12为本发明的口罩检测网络模型部署流程图。

具体实施方式

[0051] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0052] 一种基于深度学习的人脸口罩佩戴情况检测方法，如图5所示，该方法包括：实时获取待检测图像数据，对待检测数据进行图像增强处理；将增强后的待检测图像输入到训练好的口罩检测网络模型中，得到检测结果；根据检测结果对待检测图像进行标记；所述口罩检测网络模型包括主干特征提取网络模、Neck网络模块以及Prediction网络；所述主干特征提取模块包括残差块和CSPDarkNet-X网络模块；所述Neck网络模块包括改进的PANet网络结构和FPN网络；采用CSPDarkNet-Y网络模块对PANet网络结构进行改进优化。

[0053] 对待检测的数据进行图像增强处理包括：对待检测图像进行裁剪、翻转、旋转、缩放、扭曲等几何变换，同时对图像进行像素扰动、添加噪声、光照调节、对比度调节、马赛克数据增强等操作。

[0054] 对口罩检测网络模型进行训练的过程包括：

[0055] S1：获取原始图像数据，对原始图像数据进行分类处理，即分为未佩戴口罩图像、未规范佩戴口罩图像以及规范佩戴口罩图；将分后的图像数据作为训练集；

[0056] S2：将训练集中的数据输入到主干特征提取模块中进行特征提取，得到全局特征；

[0057] S3：将主干特征网络模块提取的全局特征输入到Neck网络模块，得到多尺度局部特征；

[0058] S4：将全局特征和多尺度局部特征进行融合处理，得到融合特征图；

[0059] S5：将融合特征图输入到Prediction网络中，得到目标定位结果；

[0060] S6：根据目标定位结果计算模型的损失函数，并采用模型优化器不断调整模型的参数；当损失函数的值最小时，完成模型的训练。

[0061] 由于获取的原始数据集中的数据较少，且存在内容不丰富、质量差、背景单一等问题，不能直接适用于复杂环境下的人脸口罩检测任务。为此，本专利制作出共计10000张图像的数据集，从中选取7000张用于训练，1000用于验证，2000张用于测试。整个数据集中人脸未佩戴口罩图像3400张、人脸规范佩戴口罩3300张、人脸未规范佩戴口罩3300张。每张图片中的每一个人脸对应一个标签，每个标签对应一个序号，本文将检测任务分为三个类别：序号0对应face，表示人脸未佩戴口罩；序号1对应face_mask，表示人脸规范佩戴口罩；序号2对应WMI，表示人脸未规范佩戴口罩。

[0062] 由于模型在训练过程中会遇到数据信息分布不均、模型泛化能力差等问题，而批处理模块则是一种有效解决数据特征离散问题的有效方法。因此，口罩检测网络模型的输入是一组多维数据，输出是一组经过批处理后的多维数据。为了保证特征信息在网络层中的分布，引入γ和β作为模型中的学习参数，使得网络在训练的过程中能够恢复网络所学习到的特征分布，从而实现特征信息在每一次的归一化处理。批处理模块处理数据的公式为：

[0063] $B=\{x_1,x_2,x_3,.....x_m\}$

[0064] $$\mu_B \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i$$

[0065] $$\sigma_B^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m}(x_i-\mu_B)^2$$

[0066] $$\hat{x}_i \leftarrow \frac{x_i-\mu_B}{\sqrt{\sigma_B^2+\varepsilon}}$$

[0067] $$y_i \leftarrow \gamma \hat{x}_i+\beta$$

[0068] 其中，B表示批处理的输入，x_m表示图像数据，μ_B表示期望，m表示输入图像的维度，σ_B^2表示方差，$\hat{x}_i$表示图像数据的标准化，ε表示标准化参数，y_i表示原特征信息，γ 表示尺度调整参数，β表示变换调整参数。

[0069] 主干特征提取模块中设置有残差模块和CSPDarkNet-X网络模块，残差模块中设置有串联卷积模块。

[0070] 串联卷积模块的结构如图1所示，包括卷积层、批处理模块以及Hard-Swish激活函数，其中，“卷积”表示普通的卷积操作，在本专利中卷积核的大小有两种尺寸：1×1和3×3，1×1的卷积操作对特征层的通道进行整合，3×3的卷积操作加强特征提取并对特征层的通道数进行扩充。“批处理”表示批量归一化操作，可有效地改善模型的梯度并提高模型的训练速度。

[0071] 残差模块用于减少模型参数量的同时加强模型特征提取的能力，该模块的结构如图2所示。在残差模块中首先进行1x1的卷积，再进行3x3的卷积，最后将模块的两个输出进行加权操作来增加特征层的特征信息，并且不改变特征层的维度。

[0072] 在保证主干特征提取网络模块进行特征提取的情况下，能进一步降低模型对算力的要求，从而构建出CSPDarkNet-X网络模块来优化模型结构，其结构如图3所示。在CSPDarkNet-X中首先将输入到残差块的特征层分成两条分支：一条分支作为残差边进行卷积操作。另一条分支作为主干部分，首先进行1x1的卷积操作，待进入残差块后先进行1x1的卷积进行通道调整，然后进行3x3的卷积加强特征提取，最后将两条分支进行堆叠，从而将通道进行合并获取更多的特征层信息。在本专利的主干特征提取网络中共使用三个CSPDarkNet-X模块，X表示残差结构内部残差加权操作的次数，最后堆叠完成后再利用一个1x1的小卷积对通道进行整合。

[0073] 在CSPDarkNet-X模块中，每一个串联卷积操作中皆是采用Hard-Swish激活函数，堆叠操作是将流通到堆叠模块的特征层信息进行堆叠，从而使得输出的特征层的维度增加，而不改变每一个通道下的信息量。相比于ReLU函数，Leaky-ReLU函数则是将所有的负值赋值为具有非零的斜率，这保证模型在快速收敛的同时，不出现梯度消失的现象。因此，在确保不出现梯度爆炸、消失等问题的情况下，借助Hard-Swish函数的优点来降低模型对运行时间的要求。同时，提高模型的检测精度。

[0074] Hard-Swish激活函数的表达式为：

[0075] $$Hard-Swish = x\frac{\mathrm{ReLU6}(x+3)}{6}$$

[0076] 其中，x表示特征信息，ReLU6表示激活函数。

[0077] $$Leaky-\mathrm{Re}LU=\begin{cases}x_i, x_i \geq 0 \\ \frac{x_i}{\alpha_i}, x_i \leq 0\end{cases}$$

[0078] 其中，Leaky表示泄露，α_i表示激活参数，x_i表示特征信息。

[0079] 一种Neck网络模块的结构，如图4所示，卷积神经网络需要输入图像具有固定的尺寸，在以往的卷积神经网络中则通过裁剪、扭曲等操作来获取固定的输入，但易导致目标缺失、目标形变等问题。为了消除上述问题，SPPNet(Spatial Pyramid Pooling network)网络模块可以移除网络对固定输入尺寸的要求，从而获取多尺度局部特征。为了进一步将多尺度局部特征信息与全局特征信息进行融合，本专利在PANet网络结构中加入CSPDarkNet-Y模块，从而加强FPN网络与PANet网络对特征的提取，加快特征信息之间的流通，同时提高模型的精度。

[0080] 口罩检测算法的网络模型结构如图5所示，在主干特征提取网络中共使用了三个CSPDarkNet-X模块，每一个CSPDarkNet-X模块中有X个残差单元。本专利在考虑硬件设备的计算成本时，将残差模块串联成X个残差单元的组合，这样的操作可将两个3×3的卷积操作替换为1×1+3×3+1×1卷积模块的组合，第一个1×1的卷积层可将通道数压缩为原来的一半，同时减少参数量。3×3的卷积层可加强特征提取并还原通道数。最后一个1×1的卷积操作对3×3的卷积层的输出又做了还原，这样交替进行的卷积操作有助于特征提取，保证了精度的同时又降低了计算量。

[0081] Neck网络主要由SPPNet网络和PANet网络构成，采用SPPNet模块中的不同大小的最大池化核来进行池化操作，池化核的尺寸有四种:5×5,9×9,13×13,最大池化核。这样的操作在堆叠模块中进行堆叠，从而有效地增加主干特征的接受范围，显著地分离出最重要的上下文特征。而在Neck网络模块中的“上采样”操作的主要作用是让输入进来的特征层的高和宽变成原来的两倍，从而让上采样得到的特征图和主干特征提取网络中的特征图进行堆叠，从而实现特征融合。而模型推理的计算代价过高主要是由于网络在优化过程中梯度信息反复出现造成的，因此本文从网络模型的设计角度出发，在PANet网络中引入CSPDarkNet-Y模块，将来自于主干特征提取网络的基础特征层划分成两个部分，然后通过跨阶段操作减少重复梯度信息的使用。同样地，CSPDarkNet-Y模块中使用1×1+3×3+1×1卷积模块的组合，从而实现减少计算量的同时保证精度。

[0082] Prediction网络会利用网络提取到的特征进行预测。本专利将Prediction网络划为13×13×24、26×26×24、52×52×24三个有效特征层，分别对应大目标、中目标、小目标。24＝3×(4+3+1)其中4表示预测框的四个位置参数，1即是用来判断先验框内是否包含物体，3表示本文口罩检测的类别。

[0083] 本发明所使用的模型在特征的利用部分提取了三个有效特征层，最后由Prediction网络进行预测。对于13×13×24有效特征层而言，相当于将输入的图片划分成13×13的网格，每一个网格会负责这个网格所对应的区域的物体检测。当某一个物体的中心落在这个区域时，则需要利用这个网格去负责这个物体的检测。每一个网格会预先设定三个先验框，网络的预测结果会对三个先验框的位置参数进行调整，从而获得最终的预测的结果。同样地，26×26×24、52×52×24的有效特征层的预测过程与13×13×24的特征层一样。

[0084] 如图6所示，用特征层被划分为13×13的网格尺寸来说明目标定位与预测过程。原始输入图像为尺寸为416×416×3的三通道彩色图像。输入图像经过网络的处理后得到13×13的有效特征层。该特征层被划为13×13的网格，每一个网格对应三个先验框。三个先验框用绿色的框表示，它们的中心点为(c_x,c_y)，宽和高分别为p_w,p_h。而最终的预测框是中心点为(t_x,t_y)，宽和高分别为b_w,b_h的蓝色框。因此，当人脸的中心落在橘色框内时，则由这个网格负责这个人脸的检测。而网络的预测结果会调整三个先验框的位置，然后经过置信度大小的排序和非极大值抑制算法筛选出最终的预测框，从而得到网络的检测结果。

[0085] 在预测过程中，(t_x,t_y,t_w,t_h)为网络需要学习四个参数，即参数学习的表达式为：

[0086] $b_x=\sigma(t_x)+c_x$

[0087] $b_y=\sigma(t_y)+c_y$

[0088] $$b_w = p_w e^{t_w}$$

[0089] $$b_h = p_h e^{t_h}$$

[0090] 其中，(b_x,b_y)表示预测框的中心点x轴和y轴方向的坐标，σ表示逻辑约束，(t_x,t_y)表示表示预测框的中心点，(c_x,c_y)表示包含预测框的网格的左上角坐标，，b_w表示预测框的宽度，p_w表示先验框的宽度，e^{t_w}表示对预测框的宽度预测偏差取e底数，t_w表示预测框的宽度预测偏差，b_h表示预测框的高度，p_h表示先验框的高度，e^{t_h}表示对预测框的高度预测偏差取e底数，t_h表示预测框的高度预测偏差。

[0091] 网络在训练的过程中不断地去学习(t_x,t_y,t_w,t_h)四个参数，从而不断地调整先验框的位置去接近预测框的位置，进而得到最终的预测结果。$\sigma(t_x)$，$\sigma(t_y)$分别表示对t_x,t_y进行Sigmoid函数的约束，从而使得预测框的中心落在网格内。置信度得分反映了模型预测某个物体是某个类别的准确度，则置信度的公式为：

[0092] $$Confidence = P_r\left(Class_i \mid Object\right)\times P_r\left(Object\right)\times IoU_{pred}^{truth}$$

[0093] 其中，Confidence表示置信度，$P_r(Class_i|Object)$表示已知是物体的情况下，是哪一类物体的概率，$Class_i$表示第i个类别，Object表示目标，$P_r(Object)$表示是否包含物体的概率，如果包含物体，$P_r(Object)=1$，否则等于0；IoU_{pred}^{truth}表示预测框与真实框之间的重叠率，重叠部分如图7所示；truth表示真实框，pred表示预测框。

[0094] $$IOU=\frac{A\cap B}{A\cup B}$$

[0095] 其中，A表示预测框，B表示真实框，∩表示交运算，∪表示并运算。

[0096] 将二元交叉熵作为模型的置信度和分类的损失函数，使得网络可以利用一个边界框就可以实现多个类别的预测。为充分地考虑边框的重合度、中心距离和宽高比的尺度信息，引入CIOU作为目标的定位损失函数，使得预测框的回归过程变得更加地稳定。模型总的目标损失函数为：

[0097] $Loss\text{-}function=CIOU_{loss}+Confidence_{loss}+Class_{loss}$

[0098] 其中，CIOU充分地考虑了目标与预测框之间的距离、重合度、尺度以及惩罚项等因子，这使得模型对目标的多尺度问题有了更好的处理方法，Loss表示损失函数，function表示函数，$CIOU_{loss}$表示定位损失函数，$Confidence_{loss}$置信度损失函数表示，$Class_{loss}$分类损失函数表示。

[0099] CIOU的表达式为：

[0100] $$CIOU = IOU - \frac{\rho^2(b,b^{gt})}{c^2} - \alpha v$$

[0101] 其中，IOU表示交并比，$\rho^2(b,b^{gt})$表示预测框的中心点与真实框的中心点之间的欧式距离，b表示预测框，b^{gt}表示真实框，gt表示真实框，c表示同时包含预测框和真实框的最小闭包区域的对角线距离，α表示权重系数，v表示预测框与真实框的长宽比的距离。

[0102] α和v的表达式分别为：

[0103] $$\alpha = \frac{v}{1 - IOU + v}$$

[0104] $$v = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2$$

[0105] 其中，w^{gt}表示真实框的宽度，h^{gt}表示真实框的高度，w表示预测框的宽度，h表示预测框高度。

[0106] 目标的CIOU定位损失函数为：

[0107] $$CIOU_{loss} = 1 - IOU + \frac{\rho^2(b,b^{gt})}{c^2} + \alpha v$$

[0108] Confidence_loss反映了模型对于某个网格点内物体预测的准确度的损失。因此，Confidence_loss由两部分二元交叉熵组成，即：

[0109] $$Confidence_{loss} = -\sum_{i=0}^{S^2}\sum_{j=0}^{B} 1_{ij}^{obj}\left[\hat{C}_i^j \log\left(C_i^j\right) + \left(1-\hat{C}_i^j\right)\log\left(1-C_i^j\right)\right] -$$

[0110] $$\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B} 1_{ij}^{noobj}\left[\hat{C}_i^j \log\left(C_i^j\right) + \left(1-\hat{C}_i^j\right)\log\left(1-C_i^j\right)\right]$$

[0111] 其中，S表示网格大小，B表示每一个网格所对应的先验框个数，1_{ij}^{obj}表示特征层上第i个网格点中的第j个先验框，如果这个先验框内存在物体，即负责这个物体的检测，那么$1_{ij}^{obj}=1$，否则为0；obj表示目标，$\hat{C}_i^j$表示是否负责物体的检测，若是则$\hat{C}_i^j=1$，否则为0；C_i^j表示口罩检测模型的实际输出值，λ_{noobj}表示权重系数，1_{ij}^{noobj}表示特征层上第i个网格点中的第j个先验框内不包含物体，此时$1_{ij}^{noobj}=1$，否则$1_{ij}^{noobj}=0$；noobj表示先验框内不存在物体。式中，特征层被划分为S×S个网格点，每一个网格点对应B个先验框。由于不包含物体的网格点占多数，因此需要添加权重系数λ_{noobj}对不包含物体的置信度损失函数进行约束，从而平衡正负样本。

[0112] 同理，当第i个网格点中的第j个先验框负责某一种物体检测时，那么就由这个先验框去负责计算这个物体的分类损失函数，分类损失函数的表达式为：

[0113] $$Class_{loss} = -\sum_{i=0}^{S^2} 1_{ij}^{obj}\sum_{c\subset class}\left[\hat{P}_{i,c}^j \log\left(P_{i,c}^j\right) + \left(1-\hat{P}_{i,c}^j\right)\log\left(1-P_{i,c}^j\right)\right]$$

[0114] 其中，class类别个数为3，$\hat{P}_{i,c}^{j}$表示真实框中的物体属于类别c的概率，如果此时物体的类别是c，则$\hat{P}_{i,c}^{j}=1$，反之为0。$P_{i,c}^{j}$表示模型预测第i个网格点中的第j个先验框是类别c的概率。

[0115] 模型优化器是一种自适应学习率的方法，本专利在动量一阶矩估计的基础上加入了二阶矩估计，并利用梯度的一阶矩估计和二阶矩估计动态调整每个参数的学习率。这样每一次迭代学习率都有个确定范围，使得参数比较平稳。而传统的梯度下降算法保持单一的学习率更新所有的权重，学习率在训练过程中并不会改变。为此本专利通过计算梯度的一阶矩估计和二阶矩估计为不同的参数设计独立的自适应性学习率。而模型训练的过程实质为不断地进行前向传播和反向传播来更新网络权重和偏置的过程，因此本专利的优化器对模型的权重进行更新，从而使得模型训练过程不断地进行和收敛，模型优化器算法如下所示：

[0116] $v_{dw}\leftarrow\beta_1 v_{dw}+(1-\beta_1)\,dw$

[0117] $v_{db}\leftarrow\beta_1 v_{db}+(1-\beta_1)\,db$

[0118] $S_{dw}\leftarrow\beta_2 S_{dw}+(1-\beta_2)\,dw\odot dw$

[0119] $S_{db}\leftarrow\beta_2 S_{db}+(1-\beta_2)\,db\odot db$

[0120] $$v_{dw}^{t}\leftarrow\frac{v_{dw}}{1-\beta_1^{t}}$$

[0121] $$v_{db}^{t}\leftarrow\frac{v_{db}}{1-\beta_1^{t}}$$

[0122] $$S_{dw}^{t}\leftarrow\frac{S_{dw}}{1-\beta_2^{t}}$$

[0123] $$S_{db}^{t}\leftarrow\frac{S_{db}}{1-\beta_2^{t}}$$

[0124] $$w\leftarrow w-\eta\frac{v_{dw}^{t}\times dw}{\sqrt{S_{dw}^{t}+\varepsilon}}$$

[0125] $$b\leftarrow b-\eta\frac{v_{db}^{t}\times db}{\sqrt{S_{db}^{t}+\varepsilon}}$$

[0126] 其中，w，b分别表示模型的权重和偏置，η为模型的学习率，本专利采用余弦退火衰减算法对学习率进行衰减，β_1，β_2等于β_1^{t}，β_2^{t}，分别为0.9，0.999，主要用于控制指数衰减，t表示更新的次数。$\varepsilon=10^{-8}$，防止出现除零现象。v_{dw}，v_{db}为有偏一阶矩估计变量，分别初始化为0。S_{dw}，S_{db}为有偏二阶矩估计变量，同样分别初始化为0，v_{dw}^{t}，v_{db}^{t}，S_{dw}^{t}，S_{db}^{t}分别是v_{dw}，v_{db}，S_{dw}，S_{db}的纠正。式中$dw=\frac{\partial L}{\partial w}$，$db=\frac{\partial L}{\partial b}$，其中L表示模型的目标函数Loss-function。

[0127] 在神经网络中，需要合理的设置学习率来使模型能够正确地进行训练，当学习率过小时易导致模型训练成本增加，若学习率过大时则易导致模型在学习的过程中呈现发散状态而不能正常学习特征。因此，需要设置学习率在训练的过程中不断地减小，从而实现学

习率衰减过程，而余弦退火衰减学习率是一种很有效的方法，本专利利用余弦退火衰减进行学习率的调整，公式为：

[0128] $$\eta_t = \eta_{\min}^i + \frac{1}{2}(\eta_{\max}^i - \eta_{\min}^i)(1+\cos(\frac{T_{cur}}{T_i}\pi))$$

[0129] 其中，i表示第几次epoch，每一个epoch表示学习中所有训练数据均被使用过一次时的更新次数。$\eta_{\max}^i$和$\eta_{\min}^i$分别表示学习率的最大值和最小值，定义了学习率的范围。T_{cur}则表示当前执行了多少个epoch，但是T_{cur}是在每个batch运行之后就会更新，而此时一个epoch还没有执行完，所以T_{cur}的值可以为小数，T_i表示第i次训练中总的epoch数。

[0130] 针对本专利口罩检测数据集，为获得精准的预测结果，需要设置合适的先验框尺寸。通过k-means聚类算法得到先验框的尺寸如表1所示：

[0131]

特征层	感受野	先验框尺寸
13x13	大目标	（212x242） （243x218） （240x249）
26x26	中目标	（143x154） （210x221） （207x289）
52x52	小目标	（15x32） （53x89） （78x118）

[0132] 表1先验框的尺寸

[0133] 在本专利所使用的模型开始训练之前，需初始化设置模型的超参数，从而模型在训练的过程中不断地对参数进行优化，使得模型在训练的过程中加快网络的收敛，又防止模型出现过拟合等现象，本专利所有实验均在epoch为100，Batch_size为8，输入图像的尺寸为(416x416x3)的情况下进行，参数的调整变化如表2所示：

[0134]

超参数	初始化	优化后
学习率 L	0.02000	0.00340
权重衰减系数 W	0.00100	0.00038
动量系数 M	0.94000	0.84700
分类损失系数 C	0.49000	0.24600
有无物体的损失系数 O	1.00000	0.30100
余弦退火衰减系数 Cosin	0.18000	0.11800
色调度系数 hsv_h	0.01500	0.01350
饱和度系数 hsv_s	0.69000	0.66800
明亮度系数 hsv_v	0.38000	0.45900
缩放参数 scale	0.49000	0.87800
剪切系数 shear	0.00000	0.61100
马赛克数据增强 Mosaic	1.00000	1.00000
图像混合系数 Mixup	0.00000	0.23900
上下翻转系数 Flipud	0.47000	0.00853

[0135] 表2模型的超参数

[0136] 模型的空间复杂度可由总参数量和每层输出的特征图尺寸来反映，表达式为：

[0137] $$Spcace\_Complexity=\sum_{l=1}^{D}K_l^2\cdot C_{l-1}\cdot C_l+\sum_{l=1}^{D}M^2\cdot C_l$$

[0138] 其中，可以看出空间复杂度与卷积核的尺寸K、通道数C、层数D、输出特征图尺寸M有关。

[0139] M的计算公式为：

[0140] $$M=\frac{X+2\cdot Padding-K}{Stride}+1$$

[0141] 其中X表示输入特征图的尺寸、Padding为填充、Stride为步幅。

[0142] 而模型的算力可以由时间复杂度FLOPs来分析，也即是浮点运算次数，其公式为：

[0143] $$Time\_Complexity=\sum_{l=1}^{D}M_l^2\cdot K_l^2\cdot C_{l-1}\cdot C_l$$

[0144] 本专利从模型的各个网络部分对参数量的分布进行分析，从而验证本专利所使用的主干特征提取网络、PANet的有效性。

[0145] 当模型训练完成后，将训练好的权重用于模型的测试，并从多个方面对模型进行评估。针对本专利的人脸口罩数据集，其测试结果可以归为三个类别：TP(真阳性)表示测试样本中的类别与检测结果相同；FP(假阳性)表示错误地将其他类别检测为该类别；FN(假阴性)表示将真实的样本检测为与之相反的结果、未检测出的类别。对于模型判断出来的所有正例而言，其个数为(TP+FP)，因此将真正例(TP)所占的比例称为查准率或者是精确率，其表征了被模型检测为正例的样本中，实际为真正例的样本在正例中占的比例，其表达式为：

[0146] $$\mathrm{Pr}ecision=\frac{TP}{TP+FP}$$

[0147] 对于测试集中的所有正例而言，其个数为(TP+FN)。因此，用召回率衡量模型将测试集中的真正例检测出来的能力，其表达式为：

[0148] $$\mathrm{Re}call=\frac{TP}{TP+FN}$$

[0149] 为表征模型的准确率，本专利引入AP(Average Precision)、mAP(Mean Average Precision)指标对模型进行精度评估，评估公式为：

[0150] $$AP=\int_0^1 P(R)dR$$

[0151] $$mAP=\frac{\sum_{i=1}^{N}AP_i}{N}$$

[0152] 其中，P、R、N分别表示精确度、召回率、所有类别中目标的总数。由于精确率与召回率之间存在矛盾的情况，使用综合评价指标F-Measure用于评估检测模型的好坏，其公式为：

[0153] $$F_\alpha=\frac{(\alpha^2+1)\times P\times R}{\alpha^2(P+R)}$$

[0154] 其中，F_α表示综合评价指标，α表示综合评价指标系数，P表示精确度，R表示召回率。

[0155] 一种模型训练的具体实施方式，如图8所示，该过程包括：

[0156] (1)：收集人脸佩戴口罩的图像来制作口罩检测数据集，数据集中的图像大小可以是任意尺寸的图像，且每一张图像有一个独立的编号，并将数据集划分为训练集、验证集、测试集。

[0157] (2)：使用LabelImg工具对数据集中的每一张图像进行目标标注，并确定标签名称，从而获得与每一张图像编号一样的xml文件。

[0158] (3)：初始化模型的超参数，将其用于模型的训练过程，加快网络的训练和防止模型的过拟合，并提高模型的准确度，同时加强模型的鲁棒性和适应性。

[0159] (4)：将模型的训练时代epoch设置为50。每一个epoch表示使用训练集中的全部样本训练一次模型，epoch越大模型越精确，训练的时间越长。当每一个训练时代结束后，启用验证集进行交叉验证，从而进行步骤(3)以此更新模型的超参数。同时，将batch_size设置为8，表示模型一次训练所选取的样本数，输入图像的大小设置为(416x416x3)。

[0160] (5)：在模型的训练过程中使用adam优化器来优化模型的损失函数，从而生成模型best.pt文件。

[0161] 训练结果如图9所示，图中Box表示训练集上预测框的CIOU_loss值，Objectness表示训练集上目标的Confidence_loss值，Classification表示训练上的目标分类的Class_loss值，val则表示测试集上的测试结果。Precision表示准确度P，Recall表示召回率R，mAP@0.5表示IOU为0.5条件下的所有类别的平均精确度，mAP@0.5:0.95表示IOU以0.05为递增单位，从IOU为0.5递增到IOU为0.95的所有类别的平均精确度的平均值。

[0162] 当模型进行训练与验证后即可将模型进行性能的测试。测试流程如图10所示，将训练过程中生成的best.pt文件作为测试过程中的测试模型，模型的测试步骤如下：

[0163] (1)：将best.pt文件作为训练好的模型进行测试。然后将测试集中的每一张图像作为模型的输入，其输入可以是任意大小的图片。

[0164] (2)：将目标的置信度阈值设置为0.5，同时设置非极大值抑制(NMS)算法的交并比阈值IOU＝0.5来筛选出置信度得分最高的框。

[0165] (3)：将输入到网络的图像经过图像自适应缩放算法的处理，然后生成固定尺寸的图像作为预测的输入数据。

[0166] (4)：图像数据经过主干特征提取网络来提取更多的细节信息，从而得到三个有效特征层。然后有效特征层在Neck网络模块中进行更为复杂的特征融合和流通，最后在Prediction网络被划分为S×S的网格。

[0167] (5)：每一个网格点对应三个先验框，当某一个物体的中心落在这个网格点以内时，则由这个网格点负责这个物体的预测。预测的结果会不断地迭代更新模型的损失函数，从而对每一个框的参数进行调整，使得预测框不断地趋近于xml文件中每一个目标的位置信息。最后由非极大值抑制算法筛选出得分最高的预测框作为最终的预测结果。

[0168] 在测试集上对模型进行测试，同时以每一个类别的P-R曲线来评估模型对每种类别的检测能力，如图11所示。图中，横轴为召回率Recall，纵轴为精确度Precision。在IOU为0.5的情况下，每一种类别对应着一条P-R曲线，每一条P-R曲线与横轴，纵轴围城的面积即

是mAP。从图中可以观察到，本文的模型对于三种类别的目标都具有较好的检测能力。

[0169] 本发明的模型依托计算机和GPU进行训练和部署，在实际使用场合中该模型的实施步骤如图12所示。监控设备负责获取人群流动的实时视频信息，该数据通过摄像头USB传输到计算机，实现获取视频与视频解码为每一帧的同步处理。每一帧对应每一张图像，该模型预测一张图像的平均时间为0.022s，这代表该模型可以实时性地处理视频数据，并快速地预测出数据中的检测结果。该模型会将检测为未佩戴口罩以及未规范地佩戴口罩的结果返回到计算机设备，由计算机设备的扬声器设备进行语音提示，同时监控设备依旧在跟踪这两类检测结果。

[0170] 综上所述，本专利的模型在数据集上的测试效果达到了当前所有模型的最优。同时，本专利所所使用的数据集充分囊括了现实环境中的各种复杂情况。因此，这使得本专利的模型可直接部署于实际使用环境中，有效地提高了模型的适用性。

[0171] 以上所举实施例，对本发明的目的、技术方案和优点进行了进一步的详细说明，所应理解的是，以上所举实施例仅为本发明的优选实施方式而已，并不用以限制本发明，凡在本发明的精神和原则之内对本发明所作的任何修改、等同替换、改进等，均应包含在本发明的保护范围之内。

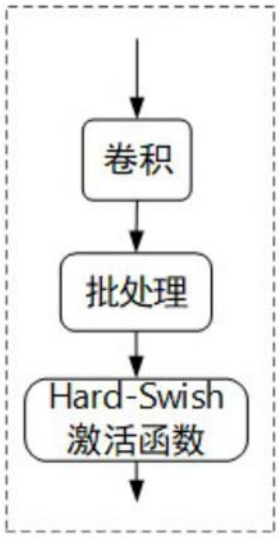


图1

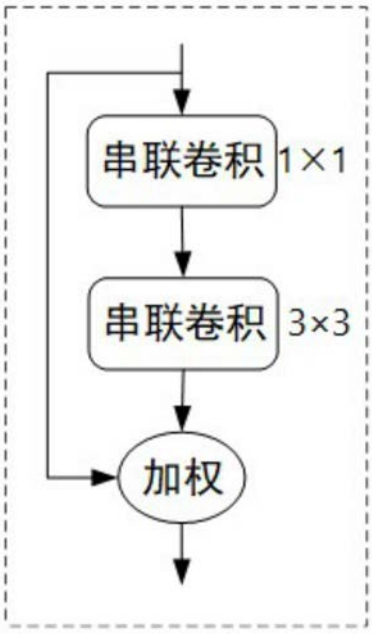


图2

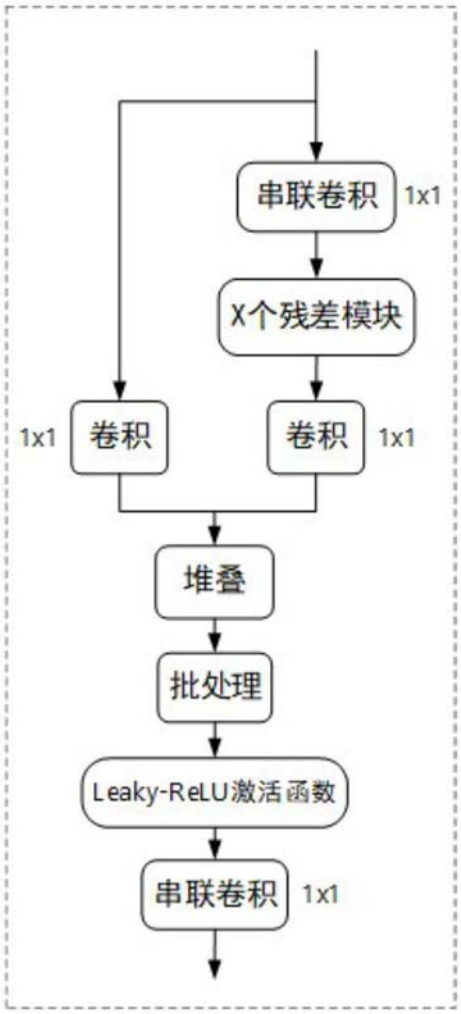


图3

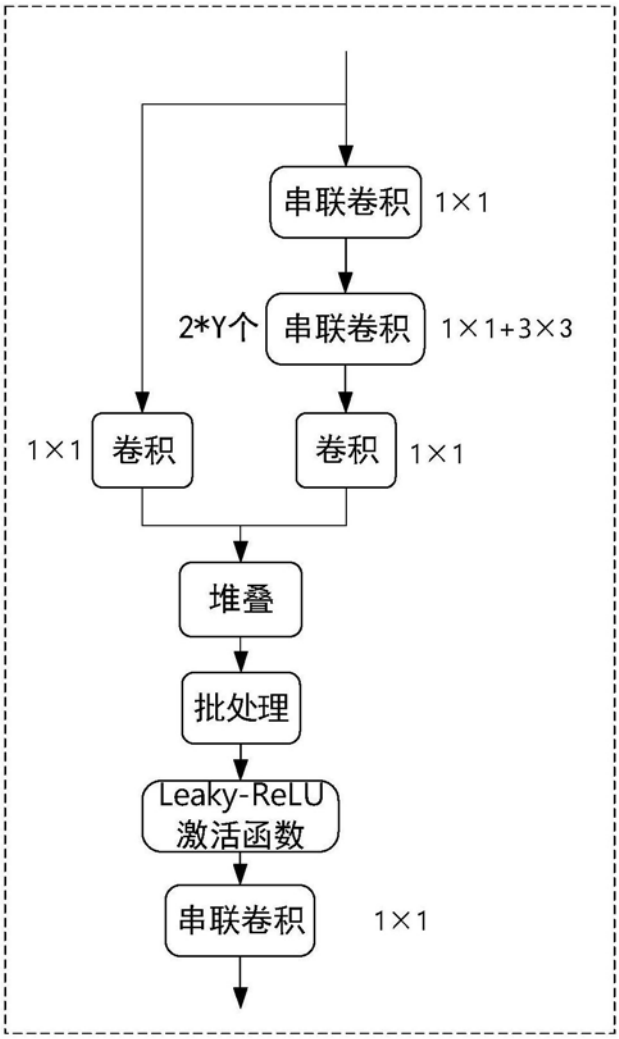

串联卷积 1×1
2*Y个 串联卷积 1×1+3×3
1×1 卷积
卷积 1×1
堆叠
批处理
Leaky-ReLU 激活函数
串联卷积 1×1

图4

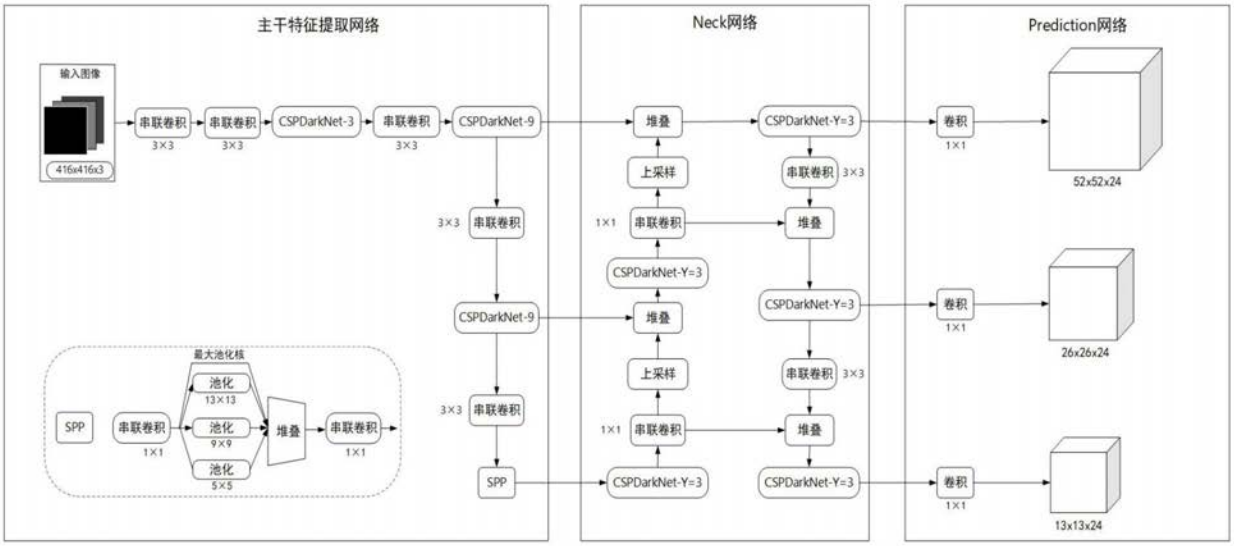

主干特征提取网络
Neck网络
Prediction网络
输入图像
416x416x3
串联卷积
CSPDarkNet-3
CSPDarkNet-9
堆叠
CSPDarkNet-Y=3
上采样
最大池化核
池化
SPP
卷积
52x52x24
26x26x24
13x13x24

图5

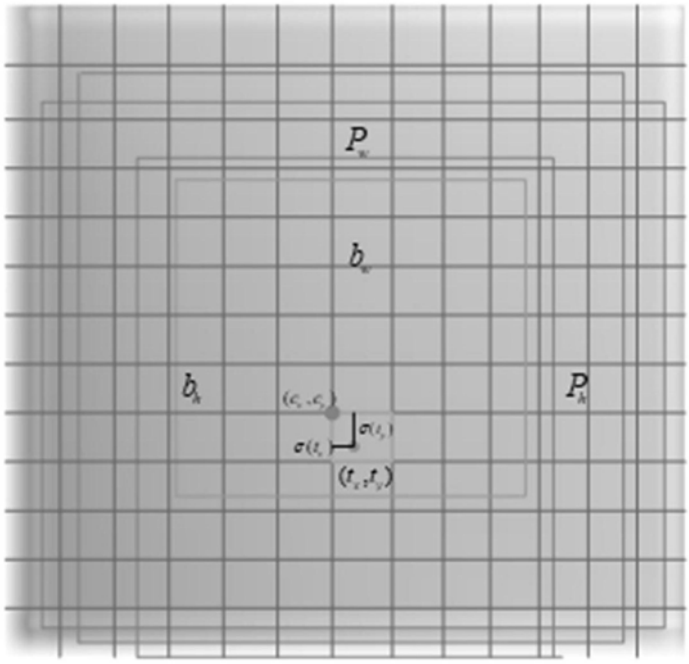


图6

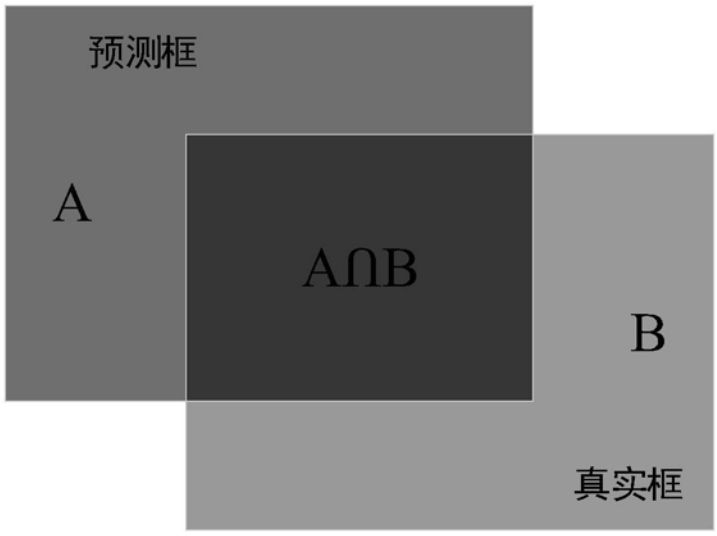


图7


制作口罩检测数据集，并将数据集划分为训练集、验证集、测试集。
使用LabelImg工具对数据集中的每一张图像进行目标标注，并确定标签名称，从而获得与标注图像名称一样得xml文件。
初始化模型的超参数，使其不断地在网络训练过程中进行优化，进而加快网络的训练和防止模型出现过拟合现象，同时加强模型的鲁棒性并提高模型的准确度和适应性。
将模型的训练时代epoch设置为50，每一个训练时代结束后进行交叉验证，从而更新模型的超参数。同时，将batch_size设置为8，输入图像的大小设置为（416x416x3）。
使用adam优化器对模型进行训练，从而生成模型best.pt文件。

图8

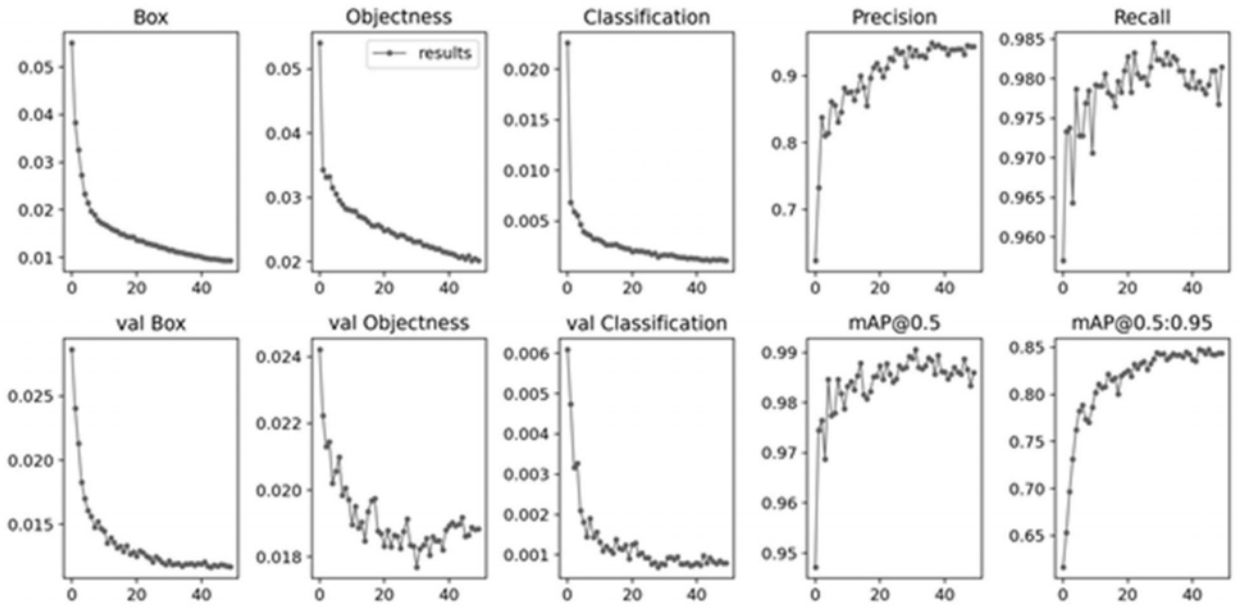

Box
Objectness
results
Classification
Precision
Recall
val Box
val Objectness
val Classification
mAP@0.5
mAP@0.5:0.95

图9

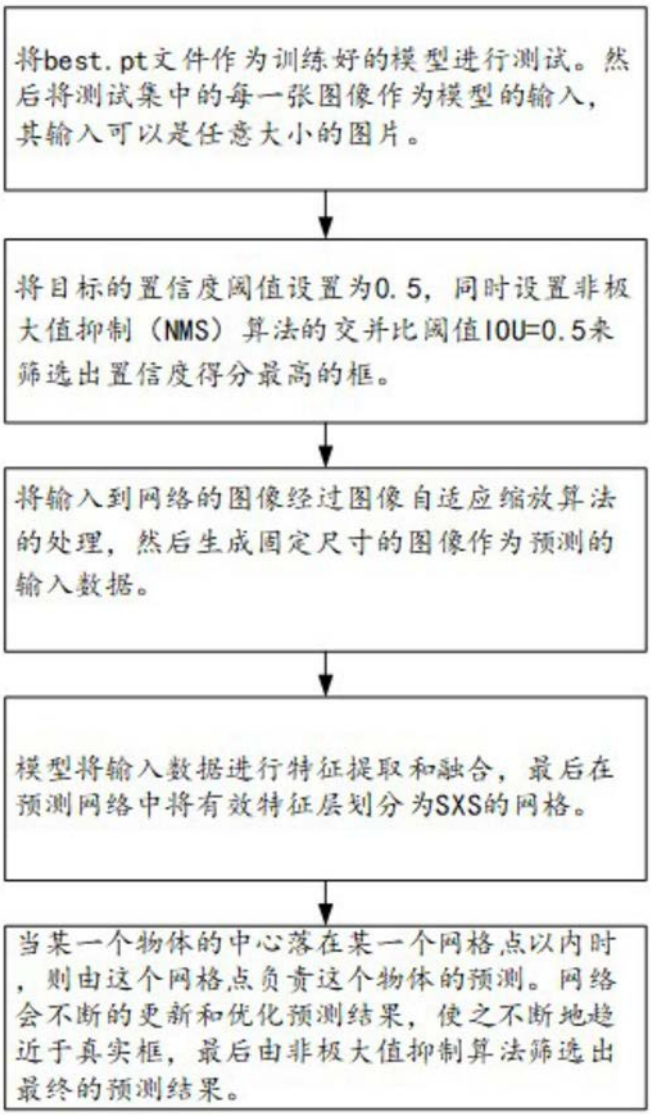

将best.pt文件作为训练好的模型进行测试。然后将测试集中的每一张图像作为模型的输入，其输入可以是任意大小的图片。
将目标的置信度阈值设置为0.5，同时设置非极大值抑制（NMS）算法的交并比阈值IOU=0.5来筛选出置信度得分最高的框。
将输入到网络的图像经过图像自适应缩放算法的处理，然后生成固定尺寸的图像作为预测的输入数据。
模型将输入数据进行特征提取和融合，最后在预测网络中将有效特征层划分为SXS的网格。
当某一个物体的中心落在某一个网格点以内时，则由这个网格点负责这个物体的预测。网络会不断的更新和优化预测结果，使之不断地趋近于真实框，最后由非极大值抑制算法筛选出最终的预测结果。

图10

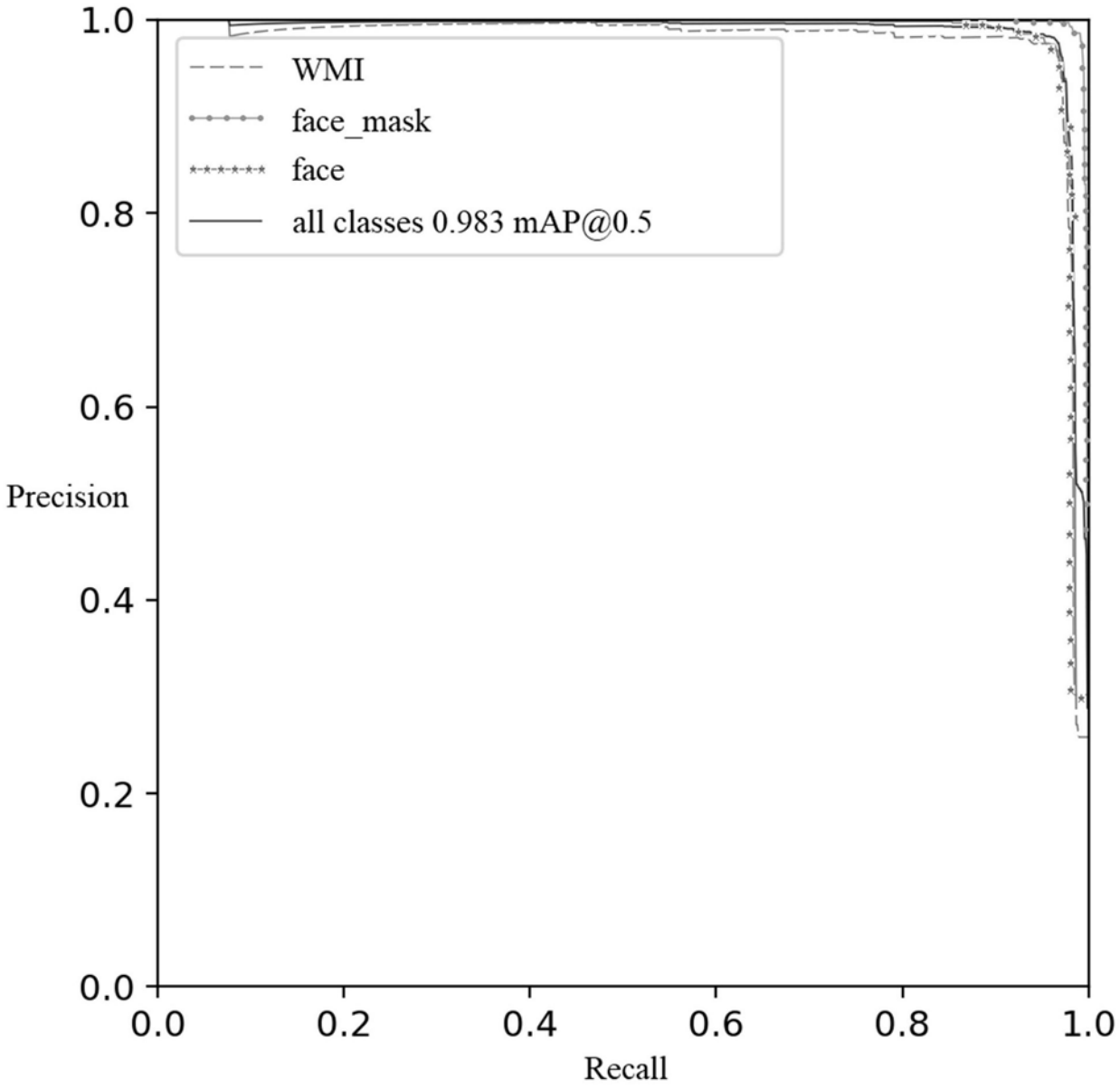

WMI
face_mask
face
all classes 0.983 mAP@0.5
Precision
Recall
0.0
0.2
0.4
0.6
0.8
1.0

图11

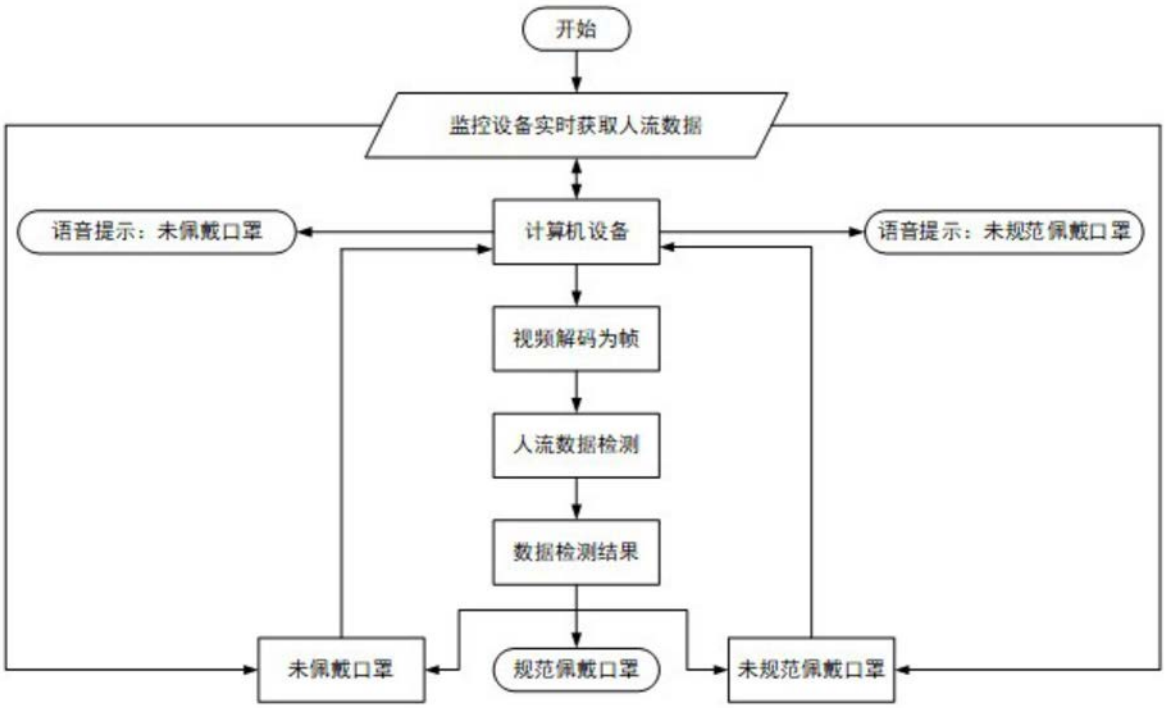

开始
监控设备实时获取人流数据
语音提示：未佩戴口罩
计算机设备
语音提示：未规范佩戴口罩
视频解码为帧
人流数据检测
数据检测结果
未佩戴口罩
规范佩戴口罩
未规范佩戴口罩

图12