



(12) 发明专利申请

(10) 申请公布号 CN 112232199 A

(43) 申请公布日 2021.01.15

(21) 申请号 202011104863.5

G06N 3/08 (2006.01)

(22) 申请日 2020.10.15

(71) 申请人 燕山大学

地址 066004 河北省秦皇岛市海港区河北大街西段438号

(72) 发明人 张立国 张勇 金梅 孙胜春
郎梦园 张子豪 张少阔 刘博
李福昆

(74) 专利代理机构 北京孚睿湾知识产权代理事务
所(普通合伙) 11474

代理人 王冬杰

(51) Int. Cl.

G06K 9/00 (2006.01)

G06K 9/62 (2006.01)

G06N 3/04 (2006.01)

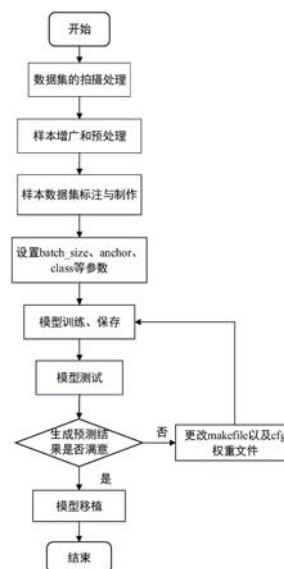
权利要求书1页 说明书4页 附图5页

(54) 发明名称

基于深度学习的佩戴口罩检测方法

(57) 摘要

本发明公开了一种基于深度学习的佩戴口罩检测方法,其包括:进行佩戴口罩检测模型的训练以及与语音提示模块的连接。佩戴口罩检测模型的训练过程包括:优化yolov3网络检测模型,将yolov3网络模型中的基础网络BN层和卷积层进行合并,并将yolo检测层由三层变为两层,通过k-means等方法进行参数优化,训练得到鲁棒性较高,检测速度较快的佩戴口罩检测模型,实现对于人们是否佩戴口罩的精确检测。同时,将佩戴口罩检测结果与语音模块相连接,当佩戴口罩检测结果不同时,语音模块发出对应的提示语音,实现语音提醒的实际功能。该方法能够实现对于是否佩戴口罩的快速,精准识别以及准确的语音提示功能,在日常生活工作方面有广泛的应用前景,具有较强的实用性。



1. 一种基于深度学习的佩戴口罩检测方法,其特征在于,其包括以下步骤:

S1、采集公共场合下人们佩戴口罩和未佩戴口罩的图像,按清晰度进行筛选;

S11、选择不同地点及环境,使用高清摄像机采集5000张佩戴口罩照片与未佩戴口罩照片作为样本,保证训练数据集的多样性;

S12、对样本数据集进行扩充和旋转处理,使训练出的模型具有高鲁棒性;

S2、对所述图像进行预处理,构建佩戴口罩检测数据集,使用标注工具对数据集图像生成对应的标注、标签信息和位置信息,将数据集按4:1比例随机分为训练数据集和测试集;

S21、使用标注工具对图像进行标注时,对图像中的人脸部分进行标注,当人脸部分有口罩时,标注mask,当人脸部分没有口罩时,标注为nomask;

S3、将所述训练集放入深度神经网络进行模型训练,提取戴口罩与未佩戴口罩的相关特征,通过聚类算法k-means产生预测框;

S4、删减yolov3网络最大尺度的yolo检测层,将卷积层和BN层合并,对所述训练集输入改进后的yolov3网络进行重复训练,获得优选的权重值与偏置值;

S41、将检测层的数量更改至与检测类别的数量相一致,以提高检测精度;

S42、对改进后的网络进行重复训练,获得适用本数据集的最优参数、权重值和偏置;

S5、将目标检测模块与语音模块相连接,根据不同的检测情况发出不同提示声音。

2. 根据权利要求1所述的基于深度学习的佩戴口罩检测方法,其特征在于:在训练深度学习网络过程中,对实际现场中出现的人员进行标注,通过更改原有的yolov3网络模型对应的参数与网络层数,能保证精度并加快目标检测速度,通过改变网络层数和参数,能减少检测时间,加强目标检测的实时性。

3. 根据权利要求2所述的基于深度学习的佩戴口罩检测方法,其特征在于:将卷积层 $x_1 = \omega * x + b$ 与BN层的函数公式 $x_2 = \gamma * \frac{x_1 - u}{\sqrt{\sigma^2 + \epsilon}} + \beta$ 合并为公式 $x_2 = \omega_1 * x + b_1$,其中, ω 为卷积权重、 b 为卷积偏置、 u 为BN层中均值、 σ 为方差、 γ 为缩放因子、 β 为偏移、 $\epsilon \neq 0$, x_1 为卷积层计算结果, x_2 为经过卷积计算和BN层后的结果, x 为进入卷积计算的计算参数, ω_1 为合并后新的卷积权重, b_1 为合并后新的卷积偏置。

4. 根据权利要求1所述的基于深度学习的佩戴口罩检测方法,其特征在于,对yolov3网络进行重复性训练时,将网络的输入尺寸调整为 416×416 大小,设置交并比IOU阈值为0.5、置信度阈值为0.5、冲量momentum为0.9,以提高模型的检测精度。

5. 根据权利要求4所述的基于深度学习的佩戴口罩检测方法,其特征在于:将目标检测模块与语音模块相连接,当检测到人佩戴好口罩时,则提示请进,当检测到人未佩戴好口罩时,则提示请佩戴好口罩。

基于深度学习的佩戴口罩检测方法

技术领域

[0001] 本发明属于机器学习领域,特别是涉及一种基于深度学习的佩戴口罩检测方法。

背景技术

[0002] 目前急需一种能够在公共区域内自动检测进出人员是否佩戴口罩的检测提醒的应用系统。

[0003] 目前传统的检测方法利用阈值分割和帧差法检测目标区域,再使用目标区域的色彩分布来进行判断,基于人工设计的传统图像处理算法应用场景单一、不稳定、当场景变化时,准确率严重变化。而现有的深度学习检测方法检测速度相对较慢,而且在获取速度时又损失了相应检测的精度且占用内存较大。

[0004] 目前在口罩佩戴检测方面需要一种检测速度快,鲁棒性好的检测系统,为此我们提出了基于深度学习的佩戴口罩检测方法。

发明内容

[0005] 针对现有的目标检测跟踪系统存在应用场景单一、检测识别速度慢、鲁棒性差、占用内存大等缺点,本发明提出了一种基于深度学习的目标检测方法,其包括以下步骤:

[0006] S1、采集公共场合下人们佩戴口罩和未佩戴口罩的图像,按清晰度进行筛选;

[0007] S11、选择不同地点及环境,使用高清摄像机采集5000张佩戴口罩照片与未佩戴口罩照片作为样本,保证训练数据集的多样性;

[0008] S12、对样本数据集进行扩充和旋转处理,使训练出的模型具有高鲁棒性;

[0009] S2、对所述图像进行预处理,构建佩戴口罩检测数据集,使用标注工具对数据集图像生成对应的标注、标签信息和位置信息,将数据集按4:1比例随机分为训练数据集和测试集;

[0010] S21、使用标注工具对图像进行标注时,对图像中的人脸部分进行标注,当人脸部分有口罩时,标注mask,当人脸部分没有口罩时,标注为nomask;

[0011] S3、将所述训练集放入深度神经网络进行模型训练,提取戴口罩与未佩戴口罩的相关特征,通过聚类算法k-means产生预测框;

[0012] S4、删减yolov3网络最大尺度的yolo检测层,将卷积层和BN层合并,对所述训练集输入改进后的yolov3网络进行重复训练,获得优选的权重值与偏置值;

[0013] S41、将检测层的数量更改至与检测类别的数量相一致,以提高检测精度;

[0014] S42、对改进后的网络进行重复训练,获得适用本数据集的最优参数、权重值和偏置;

[0015] S5、将目标检测模块与语音模块相连接,根据不同的检测情况发出不同提示声音。

[0016] 优选地,在训练深度学习网络过程中,对实际现场中出现的人员进行标注,通过更改原有的yolov3网络模型对应的参数与网络层数,能保证精度并加快目标检测速度,通

过改变网络层数和参数,能减少检测时间,加强目标检测的实时性。

[0017] 优选地,将卷积层 $x_1 = \omega * x + b$ 与BN层的函数公式 $x_2 = \gamma * \frac{x_1 - u}{\sqrt{\sigma^2 + \varepsilon}} + \beta$ 合并为公式 $x_2 = \omega_1 * x + b_1$,其中, ω 为卷积权重、 b 为卷积偏置、 u 为BN层中均值、 σ 为方差、 γ 为缩放因子、 β 为偏移、 $\varepsilon \neq 0$, x_1 为卷积层计算结果, x_2 为经过卷积计算和BN层后的结果, x 为进入卷积计算的计算参数, ω_1 为合并后新的卷积权重, b_1 为合并后新的卷积偏置。

[0018] 优选地,对yolov3网络进行重复性训练时,将网络的输入尺寸调整为 416×416 大小,设置交并比IOU阈值为0.5、置信度阈值为0.5、冲量momentum为0.9,以提高模型的检测精度。

[0019] 优选地,将目标检测模块与语音模块相连接,当检测到人佩戴好口罩时,则提示请进,当检测到人未佩戴好口罩时,则提示请佩戴好口罩。

[0020] 本发明与现有技术相比具有如下优点:

[0021] 1) 通过分析更改yolov3网络,加快检测速度,提高检测精度;

[0022] 2) 提出了基于深度学习的目标检测模块,结合语音播报模块实现对进入公共活动区域人员的佩戴口罩提醒;

[0023] 3) 具有检测速度快,检测目标精准,适宜多场景检测,鲁棒性较好等优点,能够满足目标检测系统的要求。

附图说明

[0024] 图1为本发明的模型优化流程图;

[0025] 图2a为本发明的原始网络结构图、图2b为残差组件改进图;

[0026] 图3为本发明的改进网络结构图;

[0027] 图4为本发明的整体流程图;以及

[0028] 图5为本发明的检测效果图。

具体实施方式

[0029] 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。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。

[0030] 需要注意的,除非另有说明,本申请使用的技术术语或者科学术语应当为本发明所属领域技术人员所理解的通常意义。

[0031] 一种基于深度学习的佩戴口罩检测方法,其包括如下步骤:

[0032] 深度模型的优化过程如图1,具体训练过程描述如下:

[0033] S1、首先在公路上、景区、超市、公园等各种生活区域拍摄4000-5000张人们佩戴口罩照片以及同等数量的未佩戴口罩照片,并将拍摄到的图片中模糊不清的照片清除出去;

[0034] S2、对S1中采集的图像进行预处理,并构建佩戴口罩检测的数据集:

[0035] 数据集图像使用Labelmming标注工具生成对应的标注、标签信息和位置信息,划定脸部区域,标签信息即样本中目标所属类别分为mask和no mask两类,标注信息即样本

中目标的位置信息。将数据集使用python文件按照比例4:1随机分为训练数据集和测试集,同时对数据进行扩充操作:图像进行 水平翻转、随机裁剪、更改图像对比度。

[0036] S3、将S2中处理制作后的数据集放入未改进的yolov3网络进行训练,由于yolov4在做bounding box预测的时候,用到了anchor boxes,anchors为最 有可能的object的width与height,可事先通过聚类得到,比如某一个像素单元,通过这个像素单元预测出一个object,围绕这个像素单元,可以预测出无数种object的形状,并不是随便预测的,要参考anchor box的大小,即从已标注的数据中通过聚类统计到的最有可能的object的形状,所以为了使模型参数 优化本发明使用K-means算法获得最优参数,为深度模型的准确性提供了保证,具体参数如图1所示。详细来说,先将网络的输入尺寸调整为416×416,并设置交并比IOU阈值为0.5,IOU为目标检测中使用的一个概念,是产生的候选框candidate bound与原标记框ground truth bound的交叠率,即它们的交集与并集的比值,最理想情况是完全重叠,即比值为1,,进行训练,获取最优 参数值。所有训练图片的一个批次batch,批次大小batch设置置信度阈值为0.5、冲量momentum设置为0.9ch_size的划分根据我们在.cfg文件中设置的 subdivisions参数。在我使用的.cfg文件中batch=64,subdivision=8,所以在训练输出中,训练迭代包含了8组,每组又包含了8张图片,跟设定的batch和subdivision的值一致。也就是说每轮迭代会从所有训练集里随机抽取batch=64 个样本参与训练,所有这些batch个样本又被均分为subdivision=8次送入网络 参与训练,以减轻内存占用的压力。Class参数为训练检测的目标数,由于本 目标检测的目的为是否佩戴口罩,将class参数设置为2,两种即戴口罩mask 和未佩戴口罩no mask。修改好后,修改编译文件makefile,生成可操作的执行文件进行模型训练。

[0037] S4、在优化模型参数后,在进行训练前对yolov3网络进行两方面的改进,合并BN层与屏蔽最小尺度的yolo层。

[0038] 关于合并BN层,本方法首次创新性的对yolov3网络进行BN层与卷积层的合并。详情原理如下:在深度学习中,Conv+BN+Relu已成为标配,在网络推理阶段,将Conv层和BN层进行合并,因为他们都是线性运算。合并后就 相当于少了BN层的计算,这样可以加快网络推理。在数学计算上讲,合并的 本质其实就是改变了卷积层的权重和偏置。

[0039] 卷积层的计算 $x_1 = \omega * x + b$

[0040] BN层的计算 $x_2 = \gamma * \frac{x_1 - u}{\sqrt{\sigma^2 + \epsilon}} + \beta$

[0041] 将BN层合并于Conv层后,Conv的计算方式如下:

[0042] $x_2 = \gamma * \frac{\omega * x + b - u}{\sqrt{\sigma^2 + \epsilon}} + \beta$

[0043] 即

$x_2 = \frac{\gamma * \omega}{\sqrt{\sigma^2 + \epsilon}} * x + \beta + \gamma * \frac{b - u}{\sqrt{\sigma^2 + \epsilon}}$ 、 $\omega_1 = \frac{\gamma * \omega}{\sqrt{\sigma^2 + \epsilon}}$ 、 $b_1 = \beta + \gamma * \frac{b - u}{\sqrt{\sigma^2 + \epsilon}}$ 、 $x_2 = \omega_1 * x + b_1$,其中, ω 为卷积权重、 b 为卷积偏置、 u 为BN层中均值、 σ 为方差、 γ 为缩放因子、 β 为偏移、 $\epsilon \neq 0$, x_1 为卷积层计算结果, x_2 为经过卷积 计算和BN层后的结果, x 为进入卷积计算的计算参数, ω_1 为合并后新的卷积 权重, b_1 为合并后新的卷积偏置。

[0044] 这样处理后合并了Conv层和BN层加速了模型的推理速度在保证精度度 后加快了

检测速度,原理如图2b所示。其中CBL为yolov3基本组件由Conv 层+BN层+Leaky_relu激活函数三者组成。在训练深度网络模型时,BN层能够加速网络收敛,并且能够控制过拟合,一般放在卷积层之后。BN层将数据 归一化后,能够有效解决梯度消失与梯度爆炸问题。虽然BN层在训练时起到了积极作用,然而,在网络前向推断时多了一些层的运算,影响了模型的性能,且占用了更多的内存或者显存空间。目前,很多先进的网络模型都使用了BN 技术,如ResNet、MobileNet、Xception和ShuffleNet等,因此,我们有必要 将BN层的参数合并到卷积层,来提升模型前向推断的速度。因此本方法创新 性的将yolov3.结构中两个基本组件中的卷积层与BN层进行了合并,从而达到加快检测速度的目的。

[0045] 删减yolo检测层结构的详情如图3所示,yolov3在经过多个卷积和上采 样之后最终得到的是三个卷积结果,每一个卷积结果的长和宽分别是 $(13 \times 13, 26 \times 26, 52 \times 52)$,深度信息是 $[4(\text{box信息}) + 1(\text{物体判别信息}) + 80(\text{classNum置信 度})] * 3$ (每个点上面计算几个anchor)yolo层的作用就是对这些信息进行解码操 作。第一个Yolo层是最大的特征图 $52 * 52$,mask=0,1,2,对应最小的anchor box,目标检测算法通常对输入图像中的大量区域进行采样,判断这些区域是否包含 感兴趣的目標,并调整这些区域的边缘,以便更准确地预测目标的真实边界框。不同的模型可能使用不同的区域采样方法。在这里,我们介绍一种这样的方法:它生成多个大小和纵横比不同的边框,同时以每个像素为中心。这些 边框称为锚框即anchor box。第二个Yolo层是中等的特征图 $38 * 38$,mask=3,4,5,对应中等的anchor box。第三个Yolo层是最小的特征图 $13 * 13$,mask=6,7,8,对应最大的anchor box。首先需要明白的是yolo层对3个anchor box信息不去重,如果三个anchor都满足条件的话都会把信息写入到输出。输出会有一个内存地址专门填写满足条件的检测框的数量信息,这里的满足条件 就是在物体判别信息概率大于门限同时在类别中选择置信度最大的一个物体 当作类别信息。本发明为对目标进行目标检测。通过删减最大尺度的yolo检测层即最大特征图 52×52 对应的yolo检测层,加快了模型检测速度。变化前 后的结构图对比如图2a到图3的变化。

[0046] S5、将通过S4已经训练好的目标检测模块与对应的语音模块相链接,具体检测系统运行行为:当检测来人佩戴口罩则发出“请进”的语音,当检测到来人 未佩戴口罩时则发出提示语音“请您佩戴好口罩”。整体流程图如图4所示。

[0047] 本发明利用了基于深度学习的目标检测技术,通过多元化的数据集的收集 与处理,本发明克服了检测背景单一这一缺点,可以在日常生活中的各个场景 实现对未佩戴口罩人员的检测提醒,通过使用最新的yolov3网络并对其网络 进行合并BN层和删除尺度最大的yolo检测层提升了检测速度,克服了检测速 度慢的缺点,同时通过多元数据集和对图像进行尺度变换的图像处理方法,加 强了系统的鲁棒性。克服了鲁棒性差的缺点。具体效果如图5所示。由图5 可以看出本方法可以准确的识别出人们是否佩戴口罩。正如图5所示所有出现 在画面中的人都检测到佩戴了口罩标记mask。

[0048] 以上所述各实施例仅用于说明本发明的技术方案,而非对其限制;尽管参 照前述实施例对本发明进行了详细的说明,本领域的普通技术人员应该理解:其依然能对前述实施例所记载的技术方案进行修改,或者对其中部分或全部技 术特征进行等同替换;而这些修改或替换,并不使相应技术方案的本質脱离本 发明各实施例技术方案的范围。

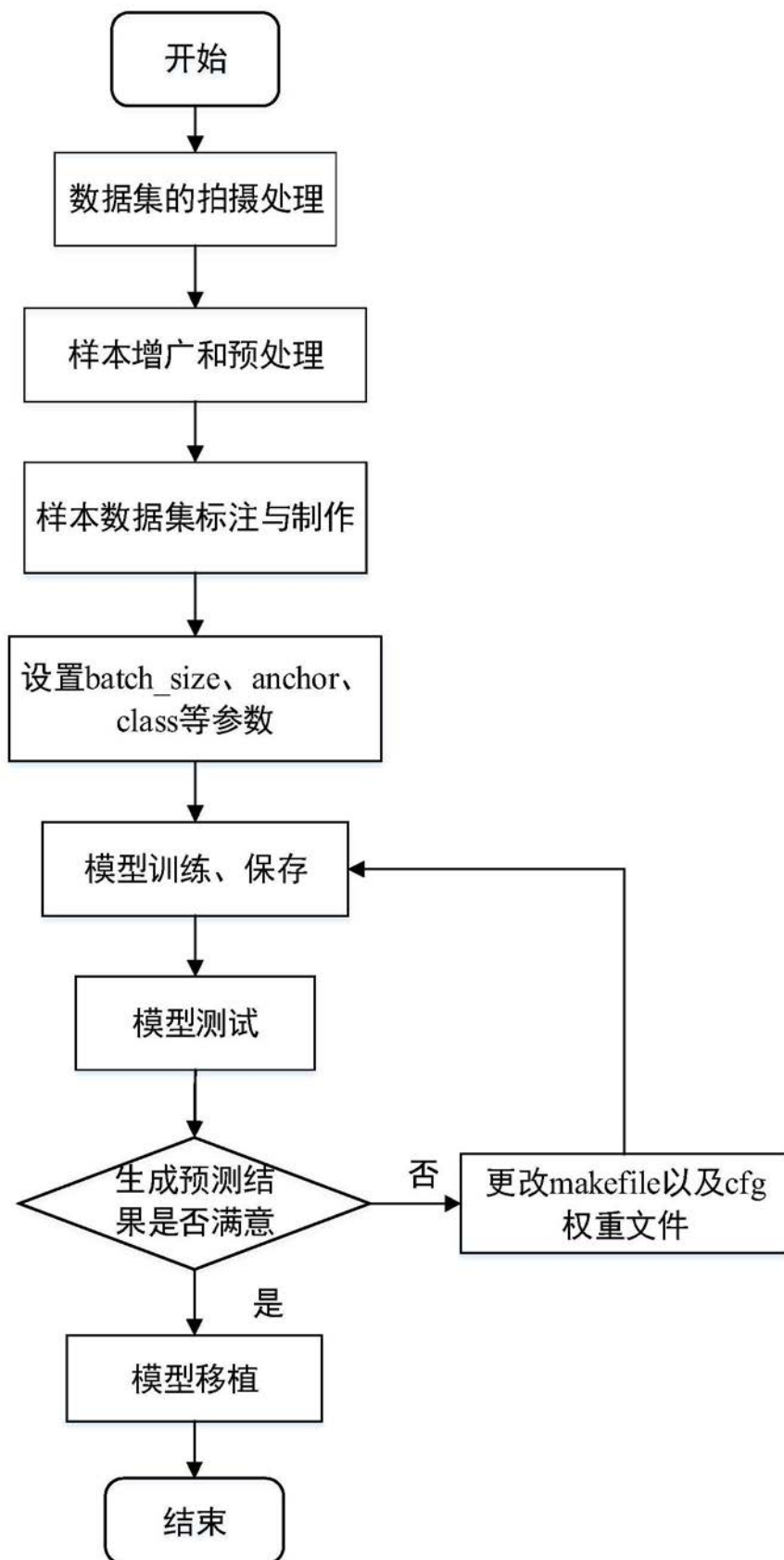


图1

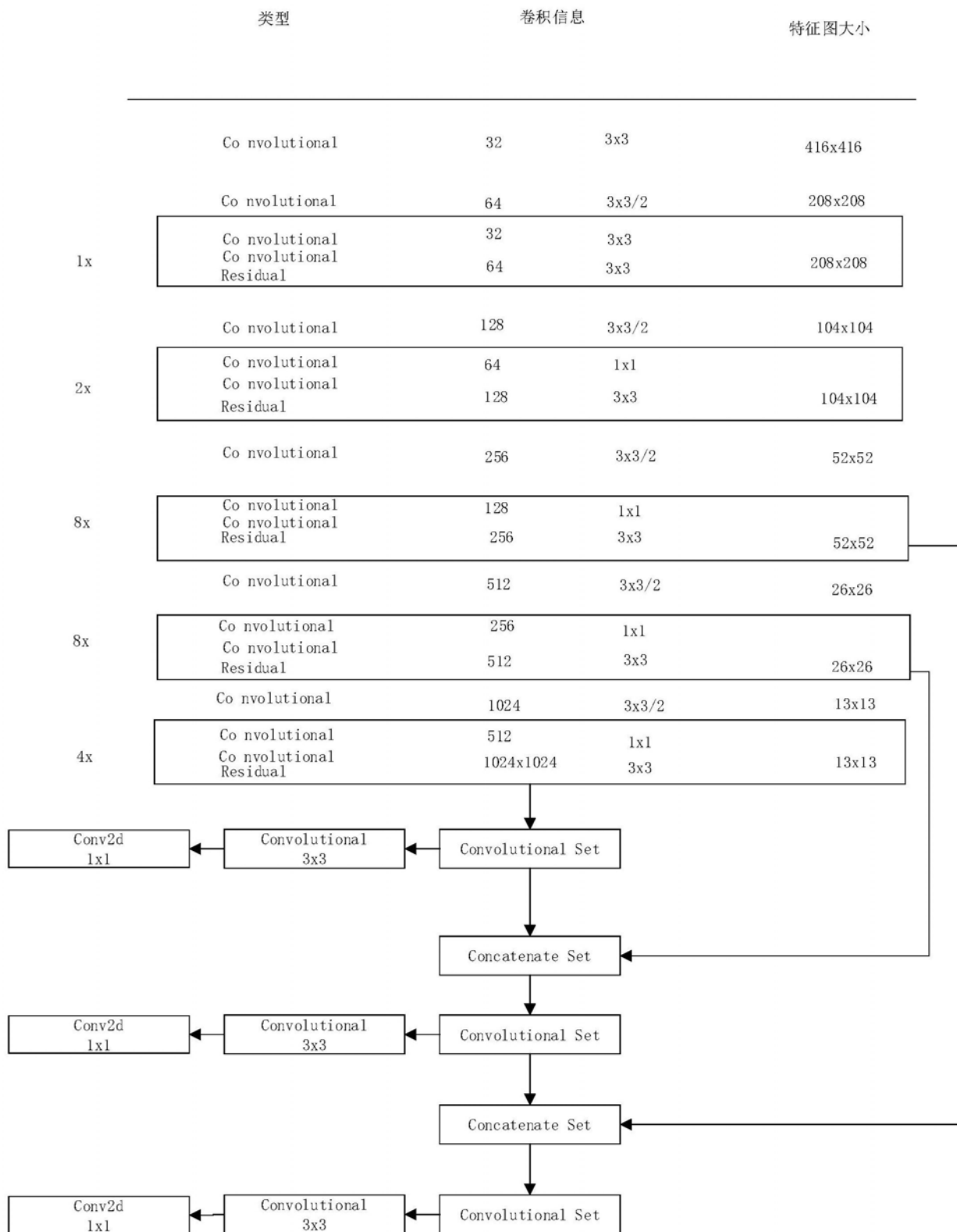


图2a

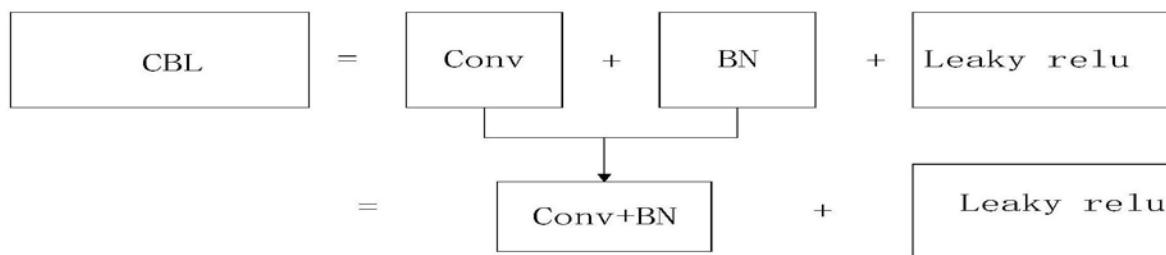


图2b

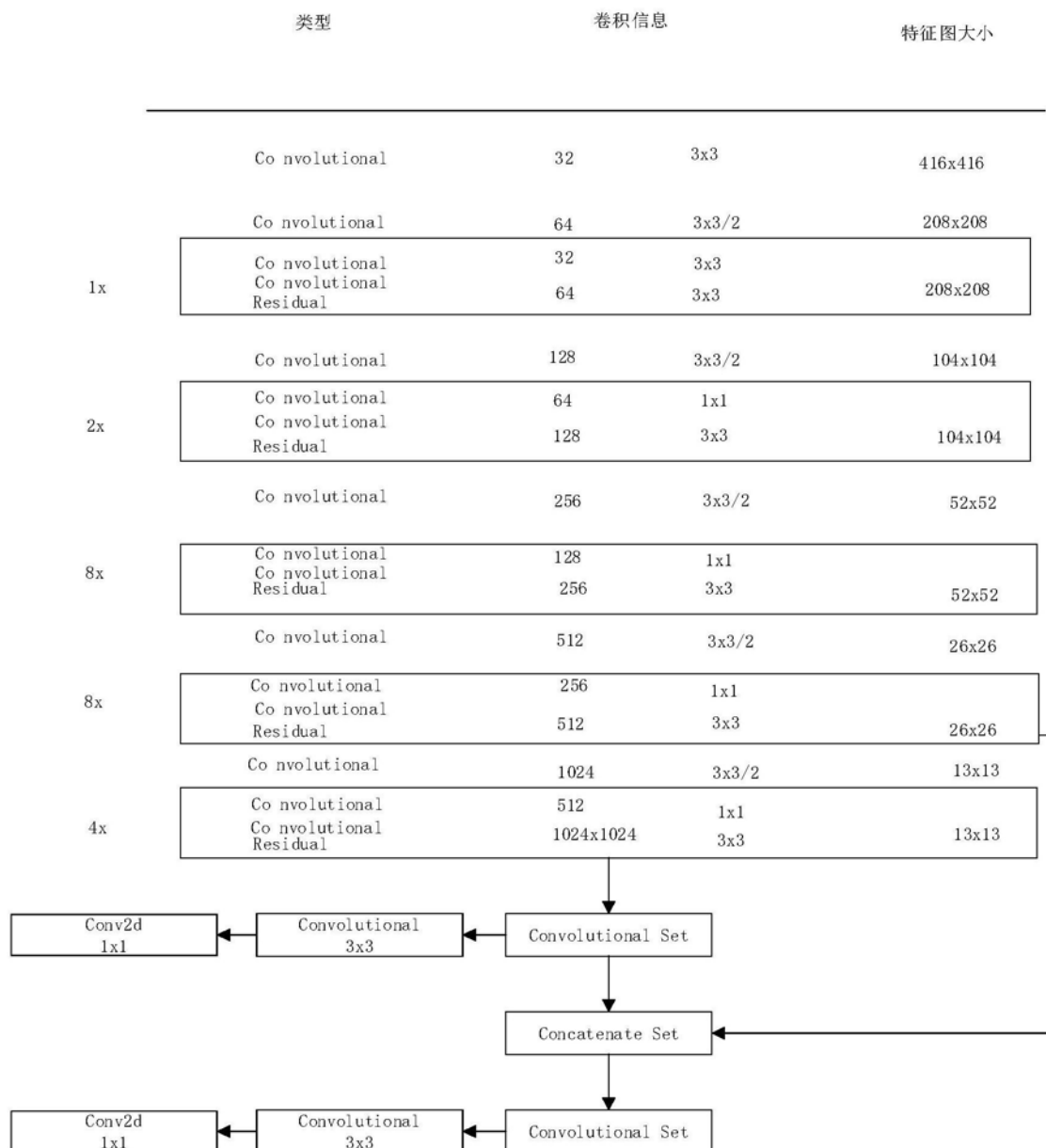


图3

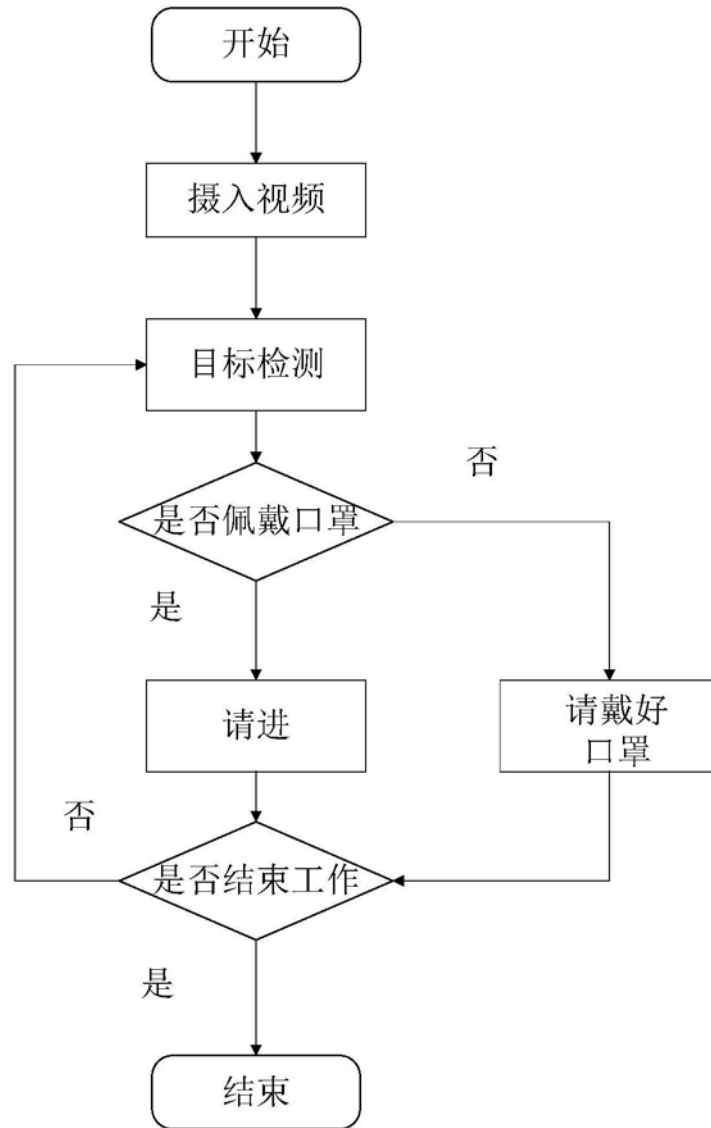


图4



图5