



(12) 发明专利申请

(10) 申请公布号 CN 112464701 A

(43) 申请公布日 2021.03.09

(21) 申请号 202010870314.2

(22) 申请日 2020.08.26

(71) 申请人 北京交通大学

地址 100044 北京市海淀区西直门外上园村3号

(72) 发明人 秦勇 李永玲 曹志威 谢征宇 魏运

(74) 专利代理机构 北京市商泰律师事务所 11255

代理人 黄晓军

(51) Int. Cl.

G06K 9/00 (2006.01)

G06K 9/62 (2006.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

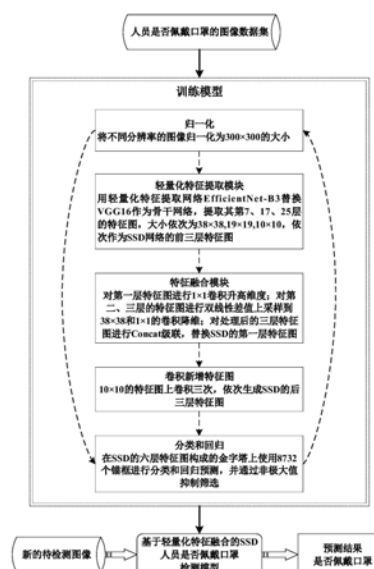
权利要求书3页 说明书9页 附图6页

(54) 发明名称

基于轻量化特征融合SSD的人员是否佩戴口罩检测方法

(57) 摘要

本发明提供了一种基于轻量化特征融合SSD的人员是否佩戴口罩检测方法。该方法包括：收集、整理各个场景人员佩戴口罩与不佩戴口罩的图像，构建口罩检测图像数据集；构建以EfficientNet-B3为骨干网络的特征提取模块，提取出三层特征图；对三层特征图通过Concat级联和归一化处理，得到轻量化特征融合SSD算法检测器的前三层特征图；对第三层特征图连续做卷积运算，得到轻量化特征融合SSD算法检测器的后三层特征图；对轻量化特征融合SSD算法检测器进行训练；对轻量化特征融合SSD算法检测器输入待检测图像，实时输出是否佩戴口罩结果。本发明的方法提高了小人脸的检测准确率，运行内存小、检测速度快、模型小，可以实时检测监控图像中众多小人脸是否佩戴口罩。



1. 一种基于轻量化特征融合SSD的人员是否佩戴口罩检测方法,其特征在于,包括:

收集、整理各个场景人员佩戴口罩与不佩戴口罩的图像,使用半监督图像自动标注构建自动标注佩戴口罩检测数据集,将所述自动标注佩戴口罩检测数据集和现有人脸基准数据集进行整合,得到口罩检测图像数据集;

在EfficientNet-B0网络的基础上,使用复合模型缩放方法扩大网络的深度、通道数、分辨率,得到EfficientNet-B3特征提取网络结构,提取出EfficientNet-B3特征提取网络结构的3个特征层;

基于所述EfficientNet-B3特征提取网络结构的3个特征层通过Concat级联和归一化处理,得到轻量化特征融合SSD算法检测器的第一层特征图、第二层特征图和第三层特征图,对所述第三层特征图连续做卷积运算,得到轻量化特征融合SSD算法检测器的第四层特征图、第五层特征图和第六层特征图;

基于所述轻量化特征融合SSD算法检测器的第一层特征图、第二层特征图、第三层特征图、第四层特征图、第五层特征图和第六层特征图,利用所述口罩检测图像数据集对所述轻量化特征融合SSD算法检测器进行训练,得到训练好的轻量化特征融合SSD算法检测器;

将所述训练好的轻量化特征融合SSD算法检测器安装在轻量化设备上,在轻量化设备上输入待检测的图像或者视频,轻量化设备利用轻量化特征融合SSD算法检测器对待检测的图像或者视频进行检查,实时输出人员是否佩戴口罩的检测结果。

2. 根据权利要求1所述的方法,其特征在于,所述的收集、整理各个场景人员佩戴口罩与不佩戴口罩的图像,使用半监督图像自动标注构建自动标注佩戴口罩检测数据集,将所述自动标注佩戴口罩检测数据集和现有人脸基准数据集进行整合,得到口罩检测图像数据集,包括:

收集整理各个人群密集场景的人员佩戴口罩与不佩戴口罩的图像,使用半监督学习图像自动标注方法对图像的种类进行标注,生成.xml文件,其中佩戴口罩的人脸类别标签为mask,不佩戴口罩的人脸类别标签为face,得到自动标注佩戴口罩检测数据集;

在公开的人脸检测数据集Wider Face中加入公开的遮挡人脸的公开数据集MAFA,将扩充后的人脸检测数据集Wider Face中的人脸的标签统一修改为face,得到公开佩戴口罩检测数据集;

将所述自动标注佩戴口罩检测数据集和所述公开佩戴口罩检测数据集进行合并,得到口罩检测图像数据集。

3. 根据权利要求1所述的方法,其特征在于,所述的在EfficientNet-B0网络的基础上,使用复合模型缩放方法扩大网络的深度、通道数、分辨率,得到EfficientNet-B3特征提取网络结构,提取出EfficientNet-B3特征提取网络结构的3个特征层,包括:

在轻量化网络EfficientNet-B0网络结构的基础上使用复合维度缩放方法,放大网络的深度、宽度、分辨率得到Efficient-B3特征提取网络,

所述复合维度缩放方法的具体公式如下:

深度: $d = a^p$

通道数: $w = b^p$

分辨率: $r = c^p$

约束条件:

$$a \cdot b^2 \cdot c^2 \approx 2$$

$$a \geq 1, b \geq 1, c \geq 1$$

a, b, c 为常数, φ 为复合系数, 首先固定 $\varphi=1$, 通过网络搜索求得 $a=1.2, b=1.1, c=1.15, \varphi=3$, 按照约束条件, $d=1.2, w=1.4, r=1.3$, 也就是深度、通道数在EfficientNet-B0的基础上扩大1.2, 1.4倍, 分辨率由原来的224变为300;

步骤2.3: 使用EfficientNet-B3替换VGG16作为新的骨干网络模型, 用来提取输入图像的特征;

步骤2.4: EfficientNet-B3包含26个移动倒置瓶颈模块MBConv, 提取第8、18、26个MBConv块处理之后的 38×38 、 19×19 、 10×10 的3个特征图。

4. 根据权利要求3所述的方法, 其特征在于, 所述的基于所述EfficientNet-B3特征提取网络结构的3个特征层通过Concat级联和归一化处理, 得到轻量化特征融合SSD算法检测器的第一层特征图、第二层特征图和第三层特征图, 包括:

步骤3.1: 对特征图尺寸 \times 特征图尺寸 \times 通道数为 $38 \times 38 \times 48$ 的特征图进行 1×1 的卷积, 升高特征图的维度到 $38 \times 38 \times 128$;

步骤3.2: 对特征图尺寸 \times 特征图尺寸 \times 通道数为 $19 \times 19 \times 136$ 特征图, 首先进行 1×1 的卷积降低维度为 $19 \times 19 \times 128$, 然后进行双线性插值上采样到 $38 \times 38 \times 128$;

步骤3.3: 对特征图尺寸 \times 特征图尺寸 \times 通道数为 $10 \times 10 \times 384$ 特征图, 首先进行 1×1 的卷积降低维度为 $10 \times 10 \times 128$, 然后进行双线性插值上采样到 $38 \times 38 \times 128$;

步骤3.4: 对步骤3.1、3.2、3.3操作后的3个 $38 \times 38 \times 128$ 的特征图进行Concat级联和L2归一化得到融合后的 $38 \times 38 \times 384$ 的特征图, 将 $38 \times 38 \times 384$ 的特征图作为轻量化特征融合SSD算法检测器的第一层特征图;

步骤3.5: 对步骤3.2提出的 $19 \times 19 \times 136$ 特征图进行 1×1 的卷积, 升高维度为 $19 \times 19 \times 384$ 的特征图, 将 $19 \times 19 \times 384$ 的特征图作为轻量化特征融合SSD算法检测器的第二层特征图; 通过EfficientNet-B3中的26个MBConv得到 $10 \times 10 \times 384$ 的特征图, 将 $10 \times 10 \times 384$ 的特征图作为轻量化特征融合SSD算法检测器的第三层特征图。

5. 根据权利要求4所述的方法, 其特征在于, 所述的对所述第三层特征图连续做卷积运算, 得到轻量化特征融合SSD算法检测器的第四层特征图、第五层特征图和第六层特征图, 包括:

对所述轻量化特征融合SSD算法检测器的第三层 $10 \times 10 \times 384$ 的特征图, 首先进行 1×1 的卷积降低维度得到 $10 \times 10 \times 128$ 的特征图, 然后进行 3×3 的卷积, 步长为2, 填充padding为1, 得到轻量化特征融合SSD的第四层特征图 $5 \times 5 \times 256$;

对所述第四层特征图 $5 \times 5 \times 256$, 首先进行 1×1 的卷积降低维度得到 $5 \times 5 \times 128$ 的特征图, 然后进行 3×3 的卷积, 步长为1, 得到轻量化特征融合SSD的第五层特征图 $3 \times 3 \times 256$;

对所述第五层特征图 $3 \times 3 \times 256$, 首先进行 1×1 的卷积降低维度得到 $3 \times 3 \times 128$ 的特征图, 然后进行 3×3 的卷积, 步长为1, 得到轻量化特征融合SSD的第六层特征图 $3 \times 3 \times 256$ 。

6. 根据权利要求5所述的方法, 其特征在于, 所述的基于所述轻量化特征融合SSD算法检测器的第一层特征图、第二层特征图、第三层特征图、第四层特征图、第五层特征图和第六层特征图, 利用所述口罩检测图像数据集对所述轻量化特征融合SSD算法检测器进行训练, 得到训练好的轻量化特征融合SSD算法检测器, 包括:

将所述轻量化特征融合SSD算法检测器的第一层特征图、第二层特征图、第三层特征图、第四层特征图、第五层特征图和第六层特征图构成特征图金字塔；

利用所述口罩检测图像数据集对所述轻量化特征融合SSD算法检测器进行训练，训练过程中使用轻量化特征融合SSD算法检测器的特征图金字塔中的8732个锚框对口罩检测图像数据集中的标签为face和mask图像进行预测，并通过非极大值抑制筛选，抑制邻域里分数低于0.5的锚框，得到图像中的人物是否佩戴口罩的判断结果，将判断结果与标注图像的真实值进行比较，确定判断结果是否正确，当所述轻量化特征融合SSD算法检测器的判断结果的正确率达到设定的判断阈值，则得到训练好的轻量化特征融合SSD算法检测器。

基于轻量化特征融合SSD的人员是否佩戴口罩检测方法

技术领域

[0001] 本发明涉及人脸目标检测技术领域,尤其涉及一种基于轻量化特征融合 SSD的人员是否佩戴口罩检测方法。

背景技术

[0002] 新型冠状病毒自爆发以来,在人群中迅速传播,给人们的日常生活及社会的经济都造成了严重的影响。相关研究人员认为新型冠状病毒的传播主要是由于患者排出的病毒被其他人吸入。研究测试表明人员佩戴外科口罩或 N95口罩可以减少病毒的传播。所以,许多国家的疾病控制和预防中心建议或要求人们在公共场所佩戴口罩,以遏制新型冠状病毒肺炎的传播。通过深度学习算法对各个人群密集的公共场所如地铁、公交、医院进行是否佩戴口罩的自动检测是亟待解决的问题。

[0003] SSD(single shot multibox detector,单步多尺度目标检测器)鲁棒性好,是工业界常用的一步法(one stage)目标检测器。人员是否佩戴口罩检测属于人脸检测领域的问题,然而SSD不经网络结构的改进应用于口罩检测领域效果并不好。首先,SSD通过不同深度的卷积层检测不同尺度的目标,用于检测小目标的浅层特征图由于缺乏语义信息导致小人脸的检测效果差,而实际应用场景中待检测的小人脸很多。其次,SSD虽然是分类和回归一步检测的算法,但是模型的参数量多,运行内存大,无法在各个应用场景现有的普通CPU设备上做到实时目标检测。

[0004] 目前现有技术中针对口罩检测的研究并不多。现有技术中的一种基于SSD的口罩检测模型为:通过K-means聚类的方法确定数据集中标注人脸框的长宽比分布,修改原有SSD算法锚框的比例。该口罩检测模型的缺点为:对小目标人脸的检测效果差。

[0005] 现有技术中的另一种基于SSD的口罩检测方法包括:在SSD网络的基础上增加了特征金字塔与Focal loss损失函数,该算法采用ResNet作为特征提取的骨干网络,增加了卷积块CBAM注意力机制来调整感受野的大小使其关注特定的检测区域。该口罩检测方法的缺点为:检测速度很慢、模型参数量大,无法实现人员佩戴口罩实时检测以及将训练好的模型安装在现有的CPU设备上。

发明内容

[0006] 本发明的实施例提供了一种基于轻量化特征融合SSD的人员是否佩戴口罩检测方法,以克服现有技术的问题。

[0007] 为了实现上述目的,本发明采取了如下技术方案。

[0008] 一种基于轻量化特征融合SSD的人员是否佩戴口罩检测方法,包括:

[0009] 收集、整理各个场景人员佩戴口罩与不佩戴口罩的图像,使用半监督图像自动标注构建自动标注佩戴口罩检测数据集,将所述自动标注佩戴口罩检测数据集和现有人脸基准数据集进行整合,得到口罩检测图像数据集;

[0010] 在EfficientNet-B0网络的基础上,使用复合模型缩放方法扩大网络的深度、通

道数、分辨率,得到EfficientNet-B3特征提取网络结构,提取出 EfficientNet-B3特征提取网络结构的3个特征层;

[0011] 基于所述EfficientNet-B3特征提取网络结构的3个特征层通过Concat级联和归一化处理,得到轻量化特征融合SSD算法检测器的第一层特征图、第二层特征图和第三层特征图,对所述第三层特征图连续做卷积运算,得到轻量化特征融合SSD算法检测器的第四层特征图、第五层特征图和第六层特征图;

[0012] 基于所述轻量化特征融合SSD算法检测器的第一层特征图、第二层特征图、第三层特征图、第四层特征图、第五层特征图和第六层特征图,利用所述口罩检测图像数据集对所述轻量化特征融合SSD算法检测器进行训练,得到训练好的轻量化特征融合SSD算法检测器;

[0013] 将所述训练好的轻量化特征融合SSD算法检测器安装在轻量化设备上,在轻量化设备上输入待检测的图像或者视频,轻量化设备利用轻量化特征融合SSD算法检测器对待检测的图像或者视频进行检查,实时输出人员是否佩戴口罩的检测结果。

[0014] 优选地,所述的收集、整理各个场景人员佩戴口罩与不佩戴口罩的图像,使用半监督图像自动标注构建自动标注佩戴口罩检测数据集,将所述自动标注佩戴口罩检测数据集和现有人脸基准数据集进行整合,得到口罩检测图像数据集,包括:

[0015] 收集整理各个人群密集场景的人员佩戴口罩与不佩戴口罩的图像,使用半监督学习图像自动标注方法对图像的种类进行标注,生成.xml文件,其中佩戴口罩的人脸类别标签为mask,不佩戴口罩的人脸类别标签为face,得到自动标注佩戴口罩检测数据集;

[0016] 在公开的人脸检测数据集Wider Face中加入公开的遮挡人脸的公开数据集MAFA,将扩充后的人脸检测数据集Wider Face中的人脸的标签统一修改为face,得到公开佩戴口罩检测数据集;

[0017] 将所述自动标注佩戴口罩检测数据集和所述公开佩戴口罩检测数据集进行合并,得到口罩检测图像数据集。

[0018] 优选地,所述的在EfficientNet-B0网络的基础上,使用复合模型缩放方法扩大网络的深度、通道数、分辨率,得到EfficientNet-B3特征提取网络结构,提取出EfficientNet-B3特征提取网络结构的3个特征层,包括:

[0019] 在轻量化网络EfficientNet-B0网络结构的基础上使用复合维度缩放方法,放大网络的深度、宽度、分辨率得到Efficient-B3特征提取网络,

[0020] 所述复合维度缩放方法的具体公式如下:

[0021] 深度: $d = a^{\varphi}$

[0022] 通道数: $w = b^{\varphi}$

[0023] 分辨率: $r = c^{\varphi}$

[0024] 约束条件:

[0025] $a \cdot b^2 \cdot c^2 \approx 2$

[0026] $a \geq 1, b \geq 1, c \geq 1$

[0027] a, b, c 为常数, φ 为复合系数,首先固定 $\varphi=1$,通过网络搜索求得 $a=1.2, b=1.1, c=1.15, \varphi=3$,按照约束条件, $d=1.2, w=1.4, r=1.3$,也就是深度、通道数在

EfficientNet-B0的基础上扩大1.2,1.4倍,分辨率由原来的224变为 300;

[0028] 步骤2.3:使用EfficientNet-B3替换VGG16作为新的骨干网络模型,用来提取输入图像的特征;

[0029] 步骤2.4:EfficientNet-B3包含26个移动倒置瓶颈模块MBConv,提取第 8、18、26个MBConv块处理之后的 38×38 、 19×19 、 10×10 的3个特征图。

[0030] 优选地,所述的基于所述EfficientNet-B3特征提取网络结构的3个特征层通过Concat级联和归一化处理,得到轻量化特征融合SSD算法检测器的第一层特征图、第二层特征图和第三层特征图,包括:

[0031] 步骤3.1:对特征图尺寸 \times 特征图尺寸 \times 通道数为 $38 \times 38 \times 48$ 的特征图进行 1×1 的卷积,升高特征图的维度到 $38 \times 38 \times 128$;

[0032] 步骤3.2:对特征图尺寸 \times 特征图尺寸 \times 通道数为 $19 \times 19 \times 136$ 特征图,首先进行 1×1 的卷积降低维度为 $19 \times 19 \times 128$,然后进行双线性插值上采样到 $38 \times 38 \times 128$;

[0033] 步骤3.3:对特征图尺寸 \times 特征图尺寸 \times 通道数为 $10 \times 10 \times 384$ 特征图,首先进行 1×1 的卷积降低维度为 $10 \times 10 \times 128$,然后进行双线性插值上采样到 $38 \times 38 \times 128$;

[0034] 步骤3.4:对步骤3.1、3.2、3.3操作后的3个 $38 \times 38 \times 128$ 的特征图进行Concat级联和L2归一化得到融合后的 $38 \times 38 \times 384$ 的特征图,将 $38 \times 38 \times 384$ 的特征图作为轻量化特征融合SSD算法检测器的第一层特征图;

[0035] 步骤3.5:对步骤3.2提出的 $19 \times 19 \times 136$ 特征图进行 1×1 的卷积,升高维度为 $19 \times 19 \times 384$ 的特征图,将 $19 \times 19 \times 384$ 的特征图作为轻量化特征融合SSD算法检测器的第二层特征图;通过EfficientNet-B3中的26个MBConv得到 $10 \times 10 \times 384$ 的特征图,将 $10 \times 10 \times 384$ 的特征图作为轻量化特征融合SSD算法检测器的第三层特征图。

[0036] 优选地,所述的对所述第三层特征图连续做卷积运算,得到轻量化特征融合SSD算法检测器的第四层特征图、第五层特征图和第六层特征图,包括:

[0037] 对所述轻量化特征融合SSD算法检测器的第三层 $10 \times 10 \times 384$ 的特征图,首先进行 1×1 的卷积降低维度得到 $10 \times 10 \times 128$ 的特征图,然后进行 3×3 的卷积,步长为2,填充padding为1,得到轻量化特征融合SSD的第四层特征图 $5 \times 5 \times 256$;

[0038] 对所述第四层特征图 $5 \times 5 \times 256$,首先进行 1×1 的卷积降低维度得到 $5 \times 5 \times 128$ 的特征图,然后进行 3×3 的卷积,步长为1,得到轻量化特征融合SSD的第五层特征图 $3 \times 3 \times 256$;

[0039] 对所述第五层特征图 $3 \times 3 \times 256$,首先进行 1×1 的卷积降低维度得到 $3 \times 3 \times 128$ 的特征图,然后进行 3×3 的卷积,步长为1,得到轻量化特征融合SSD的第六层特征图 $3 \times 3 \times 256$ 。

[0040] 优选地,所述的基于所述轻量化特征融合SSD算法检测器的第一层特征图、第二层特征图、第三层特征图、第四层特征图、第五层特征图和第六层特征图,利用所述口罩检测图像数据集对所述轻量化特征融合SSD算法检测器进行训练,得到训练好的轻量化特征融合SSD算法检测器,包括:

[0041] 将所述轻量化特征融合SSD算法检测器的第一层特征图、第二层特征图、第三层特征图、第四层特征图、第五层特征图和第六层特征图构成特征图金字塔;

[0042] 利用所述口罩检测图像数据集对所述轻量化特征融合SSD算法检测器进行训练,

训练过程中使用轻量化特征融合SSD算法检测器的特征图金字塔中的 8732个锚框对口罩检测图像数据集中的标签为face和mask图像进行预测,并通过非极大值抑制筛选,抑制邻域里分数低于0.5的锚框,得到图像中的人物是否佩戴口罩的判断结果,将判断结果与标注图像的真实值进行比较,确定判断结果是否正确,当所述轻量化特征融合SSD算法检测器的判断结果的正确率达到设定的判断阈值,则得到训练好的轻量化特征融合SSD算法检测器。

[0043] 由上述本发明的实施例提供的技术方案可以看出,本发明实施例的方法针对SSD用于检测小人脸的浅层特征图缺乏语义信息导致小人脸的检测效果差的问题,提出了浅层特征图与包含高级语义信息的高层特征图融合的轻量级特征融合模块,提高小人脸的检测准确率。并且运行内存小、检测速度快、模型小,可以实时检测监控图像中众多小人脸是否佩戴口罩。

[0044] 本发明附加的方面和优点将在下面的描述中部分给出,这些将从下面的描述中变得明显,或通过本发明的实践了解到。

附图说明

[0045] 为了更清楚地说明本发明实施例的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0046] 图1为本发明实施例提供的一种基于轻量化特征融合SSD的人员是否佩戴口罩检测方法的处理流程图。

[0047] 图2为本发明实施例提供的一种人员是否佩戴口罩的图像数据集标注结果示意图。

[0048] 图3为本发明实施例提供的一种的EfficientSSD特征提取网络。

[0049] 图4为本发明实施例提供的一种的MBConv移动倒置瓶颈模块结构图

[0050] 图5为本发明实施例提供的一种的人员佩戴口罩检测的EfficientSSD网络结构图。

[0051] 图6为本发明实施例提供的一种部分场所不同尺度人脸的检测效果图。

具体实施方式

[0052] 下面详细描述本发明的实施方式,所述实施方式的示例在附图中示出,其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施方式是示例性的,仅用于解释本发明,而不能解释为对本发明的限制。

[0053] 本技术领域技术人员可以理解,除非特意声明,这里使用的单数形式“一”、“一个”、“所述”和“该”也可包括复数形式。应该进一步理解的是,本发明的说明书中使用的措辞“包括”是指存在所述特征、整数、步骤、操作、元件和/或组件,但是并不排除存在或添加一个或多个其他特征、整数、步骤、操作、元件、组件和/或它们的组。应该理解,当我们称元件被“连接”或“耦接”到另一元件时,它可以直接连接或耦接到其他元件,或者也可以存

在中间元件。此外,这里使用的“连接”或“耦接”可以包括无线连接或耦接。这里使用的措辞“和/或”包括一个或更多个相关联的列出项的任一单元和全部组合。

[0054] 本技术领域技术人员可以理解,除非另外定义,这里使用的所有术语(包括技术术语和科学术语)具有与本发明所属领域中的普通技术人员的一般理解相同的意义。还应该理解的是,诸如通用字典中定义的那些术语应该被理解为具有与现有技术的上下文中的意义一致的意义,并且除非像这里一样定义,不会用理想化或过于正式的含义来解释。

[0055] 为便于对本发明实施例的理解,下面将结合附图以几个具体实施例为例做进一步的解释说明,且各个实施例并不构成对本发明实施例的限定。

[0056] 本发明实施例提供了一种对小目标人脸检测准确率高,同时模型参数量小、检测速度快、运行内存小的人员是否佩戴口罩的检测方法。该方法可以安装在地铁、公交、商场、医院等人流量大的场景的现有CPU设备上,实时检测监控图像里的人员是否佩戴口罩,并输出没有佩戴口罩的结果并及时通知工作人员。

[0057] 图1为本发明实施例提供的一种基于轻量化特征融合SSD的人员是否佩戴口罩检测方法的处理流程图,包括以下具体步骤:

[0058] 步骤1:收集、整理各个场景人员佩戴口罩与不佩戴口罩的图像,使用半监督图像自动标注构建自动标注佩戴口罩检测数据集;

[0059] 步骤2:构建轻量化特征提取网络模块,使用复合模型缩放方法,在EfficientNet-B0的基础上扩大深度、通道数、分辨率,得到EfficientNet-B3特征提取网络结构,用其替换VGG16作为轻量化的骨干网络,从中提取出3个特征层作为轻量化特征融合SSD算法的前三层特征图,特征图的大小分别为 38×38 , 19×19 , 10×10 ;

[0060] 步骤3:构建特征融合模块,双线性插值上采样步骤2所述的 19×19 和 10×10 特征图为 38×38 ,融合上采样后的两层特征图与原来的 38×38 的特征图,得到即包含全局信息又包含局部信息的适用于检测小人脸的轻量化特征融合SSD算法检测器的第一层特征图;

[0061] 步骤4:卷积新增特征图,在步骤2得到的第三层特征图上连续做3次卷积,得到轻量化特征融合SSD算法的后三层特征图;

[0062] 步骤5:训练步骤2、3、4改进后的算法,得到基于轻量化特征融合的轻量化特征融合SSD算法是否佩戴口罩检测器;

[0063] 步骤6:将所述基于轻量化特征融合的轻量化特征融合SSD算法是否佩戴口罩检测器的算法安装在轻量化设备上,在轻量化设备上输入待检测的图像或者视频,轻量化设备利用所述基于轻量化特征融合的轻量化特征融合SSD算法是否佩戴口罩检测器对待检测的图像或者视频进行检查,实时输出是否佩戴口罩的检测结果,如果没有戴口罩及时通知工作人员。

[0064] 上述步骤1具体包括如下步骤:

[0065] 步骤1.1:收集整理地铁、公交、医院、学校和商场等各个人群密集场景的人员佩戴口罩与不佩戴口罩的图像,使用半监督学习图像自动标注方法对图像的种类进行标注,生成.xml文件,其中佩戴口罩的人脸类别标签为“mask”,不佩戴口罩的人脸类别标签为“face”,得到自动标注佩戴口罩检测数据集1。图2为本发明实施例提供的一种人员是否佩戴口罩的图像数据集标注结果示意图。

[0066] 步骤1.2:更改目前世界上最大规模的人脸检测数据集Wider Face以及MAFA遮挡

人脸VOC数据集的标注标签为“face”。

[0067] 基于轻量化特征融合SSD检测器的训练需要大量的数据样本,为了使检测器学习到更多的佩戴口罩的人脸与不佩戴口罩的人脸的特征,使用目前世界上应用最广泛的人脸检测基准数据集WIDER FACE,该WIDER FACE数据集包含39万余张尺度、姿势和遮挡方面具有高度可变性的人脸。并且,在该WIDER FACE数据集中加入遮挡人脸的公开数据集MAFA,来进一步提高改进的SSD是否佩戴口罩检测器,使得数据集具有遮挡人脸的检测能力,减少误报,该WIDER FACE数据集包含3.5万余张不同方向和遮挡程度的人脸。以上两个公开数据集都是PASCAL VOC格式,本发明编程将其人脸的标签统一修改为“face”,得到扩充的公开佩戴口罩检测数据集2。

[0068] 步骤1.3:人工使用LabelImg检查步骤1.1中的自动标注佩戴口罩检测数据集1,少部分不合适的进行修改,合并检查后的自动标注佩戴口罩检测数据集1与步骤1.2得到的公开佩戴口罩检测数据集2,整合为最终的口罩检测图像数据集。

[0069] 图3为本发明实施例提供的一种步骤2中的EfficientSSD的特征提取网络结构图,上述步骤2具体包括如下步骤:

[0070] 步骤2.1:将步骤1.3得到的口罩检测图像数据集中的不同分辨率的图像进行预处理,归一化的图像的分辨率统一为 300×300 ,加速后续卷积神经网络的处理。

[0071] 步骤2.2:修改轻量化网络EfficientNet-B0的网络结构为Efficient-B3。在轻量化网络EfficientNet-B0网络结构的基础上使用复合维度缩放方法,放大网络的深度、宽度、分辨率得到本发明的Efficient-B3特征提取网络。

[0072] 具体的,EfficientNet提出网络缩放要平衡深度、通道数(宽度)、分辨率这三个维度来提高准确率、降低参数量、提高速度。所以本发明使用复合维度缩放方法,具体公式如下:

[0073] 深度: $d = a^{\varphi}$

[0074] 通道数: $w = b^{\varphi}$

[0075] 分辨率: $r = c^{\varphi}$

[0076] 约束条件:

[0077] $a \cdot b^2 \cdot c^2 \approx 2$

[0078] $a \geq 1, b \geq 1, c \geq 1$

[0079] a, b, c 为常数, φ 为复合系数,首先固定 $\varphi=1$,通过网络搜索求得 $a=1.2, b=1.1, c=1.15$ 。本发明中 $\varphi=3$,按照约束条件, $d=1.2, w=1.4, r=1.3$,也就是深度、通道数在EfficientNet-B0的基础上扩大1.2,1.4倍,分辨率由原来的224变为300,符合SSD对分辨率要求。

[0080] 步骤2.3:使用EfficientNet-B3替换VGG16作为新的骨干网络模型,用来提取输入图像的特征;

[0081] 步骤2.4:EfficientNet-B3包含26个移动倒置瓶颈模块(MBConv)。本发明提取第8、18、26个MBConv块处理之后的 $38 \times 38, 19 \times 19, 10 \times 10$ 的3个特征图。

[0082] 图4为步骤2.4的MBConv模块结构图,具体的首先使用 1×1 的卷积将特征层升高6倍维度,然后通过深度可分离卷积与SE注意力机制模块得到权重,与深度可分离卷积得到

的特征图作“乘”运算,接着降低6倍维度与最开始输入的特征图作“加”运算。需要强调的是:

[0083] EfficientNet-B3开始的两层MBconv1模块不需要升维。

[0084] 深度可分离卷积中的步长stride=2时,步骤2.2中的特征图长宽缩减一半。

[0085] 图5为步骤3中的人员佩戴口罩检测的EfficientSSD网络结构图,上述步骤3具体包括如下步骤:

[0086] 步骤3.1:对步骤2提取出的 $38 \times 38 \times 48$ (特征图尺寸 \times 特征图尺寸 \times 通道数)特征图,进行 1×1 的卷积升高维度到 $38 \times 38 \times 128$ 。

[0087] 步骤3.2:对步骤2提取出的 $19 \times 19 \times 136$ 特征图,首先进行 1×1 的卷积降低维度为 $19 \times 19 \times 128$,然后进行双线性插值上采样到 $38 \times 38 \times 128$ 。具体的,双线性差值上采样是常用的扩大图像分辨率的方法。主要分为2步,第一步在水平方向差值,第二步在第一步的基础上接着在垂直方向上插值。设 (x, y) 为原图的坐标点,插值 i 的具体计算公式如下:

[0088] 水平方向插值:相邻上方两点的坐标为 (x_0, y_0) , (x_1, y_0) , 像素值为 $i_{0top}=f(x_0, y_0)$, $i_{1top}=f(x_1, y_0)$, 可知 $\frac{i_{top}-i_{0top}}{x-x_0}=\frac{i_{1top}-i_{0top}}{x_1-x_0}$, 所以可得差值 $i_{top}=\frac{x_1-x}{x_1-x_0}i_{0top}+\frac{x-x_0}{x_1-x_0}i_{1top}$ 。

同理相邻下方的插值为 i_{bottom} 。

[0089] 垂直方向插值:总的插值 $i=i_{top}+i_{bottom}$ 。

[0090] 步骤3.3:对步骤2提取的 $10 \times 10 \times 384$ 特征图,首先进行 1×1 的卷积降低维度为 $10 \times 10 \times 128$,然后进行双线性插值上采样到 $38 \times 38 \times 128$ 。

[0091] 步骤3.4:对步骤3.1、3.2、3.3操作后的三层特征图进行Concat级联和L2归一化得到融合后的 $38 \times 38 \times 384$ 特征图,作为轻量化特征融合SSD算法检测器的第一层特征图。

[0092] 步骤3.5:对步骤2提出的 $19 \times 19 \times 136$ 特征图进行 1×1 的卷积升高维度为 $19 \times 19 \times 384$,作为轻量化特征融合SSD算法检测器的第二层特征图; $10 \times 10 \times 384$ 的特征图通过26个MBConv得到,具有很高的全局特征,直接作为轻量化特征融合SSD算法检测器的第三层特征图。

[0093] 步骤4:卷积新增特征图,对步骤2所述的第三层特征图连续进行3次卷积,得到轻量化特征融合SSD算法的轻量化特征融合SSD的后三层特征图;

[0094] 图5为步骤4中的人员佩戴口罩检测的EfficientSSD网络结构图,上述步骤4具体包括如下步骤:

[0095] 步骤4.1:对步骤3.5所述的轻量化特征融合SSD算法检测器的第三层 $10 \times 10 \times 384$ 的特征图,首先进行 1×1 的卷积降低维度得到 $10 \times 10 \times 128$ 的特征图,然后进行 3×3 的卷积,步长为2,填充(padding)为1,得到轻量化特征融合SSD算法检测器的第四层特征图 $5 \times 5 \times 256$ 。

[0096] 步骤4.2:对步骤4.1所述的第4层特征图,首先进行 1×1 的卷积降低维度得到 $5 \times 5 \times 128$ 的特征图,然后进行 3×3 的卷积,步长为1,得到轻量化特征融合SSD算法检测器的第五层特征图 $3 \times 3 \times 256$ 。

[0097] 步骤4.3:对步骤4.2所述的第五层特征图,首先进行 1×1 的卷积降低维度得到 $3 \times 3 \times 128$ 的特征图,然后进行 3×3 的卷积,步长为1,得到轻量化特征融合SSD算法检测器的第六层特征图 $3 \times 3 \times 256$ 。

[0098] 步骤4.4:训练步骤2、3、4改进后的算法,对上述步骤所述六层特征图构成轻量化特征融合SSD算法检测器的特征图金字塔,该特征图金字塔包括8732个锚框。其中第一层特征图检测5776 ($38 \times 38 \times 4$) 个锚框,第二层特征图检测2166 ($19 \times 19 \times 6$) 个锚框,第三层特征图检测600 ($10 \times 10 \times 6$) 个锚框,第四层特征图检测150 ($5 \times 5 \times 6$) 个锚框,第五层特征图检测36 ($3 \times 3 \times 4$) 个锚框,第六层特征图检测4 ($1 \times 1 \times 4$) 个锚框。其中6个锚框的比例为1:1、1:2、2:1、1:3、3:1。4个锚框的比例为6个锚框的比例,去除1:3、3:1。

[0099] 步骤5:然后,利用所述口罩检测图像数据集对上述轻量化特征融合SSD算法检测器进行训练,训练过程中使用轻量化特征融合SSD算法检测器的特征图金字塔中的8732个锚框对口罩检测图像数据集中的标签为“face”和“mask”图像的预测,并通过非极大值抑制筛选,抑制邻域里分数低于0.5的锚框,得到图像中的人物是否佩戴口罩的判断结果。将判断结果与标注图像的真实值进行比较,确定判断结果是否正确。当上述轻量化特征融合SSD算法检测器的判断结果的正确率达到设定的判断阈值,则得到训练好的轻量化特征融合SSD算法检测器。

[0100] 步骤6:将训练好的轻量化特征融合SSD算法检测器安装在轻量化设备(如CPU)上,在轻量化设备上输入待检测的图像或者视频,轻量化设备利用轻量化特征融合SSD算法检测器对待检测的图像或者视频进行检查,实时输出人员是否佩戴口罩的检测结果,如果没有戴口罩及时通知工作人员。

[0101] 图6为步骤6的部分场所不同尺度人脸的检测效果图。

[0102] 综上所述,本发明实施例的方法针对SSD用于检测小人脸的浅层特征图缺乏语义信息导致小人脸的检测效果差的问题,提出了浅层特征图与包含高级语义信息的高层特征图融合的轻量级特征融合模块,提高小人脸的检测准确率。并且运行内存小、检测速度快、模型小,可以实时检测监控图像中众多小人脸是否佩戴口罩。

[0103] 本发明实施例结合工业界对人员是否佩戴检测模型的需要,收集整理不同场景的人员佩戴口罩的图像,并利用半监督学习图像自动标注,自动标注构建口罩检测数据集。并且合并Wider Face与MAFA公共数据集构建包含3万余张图像的口罩佩戴检测PASCAL VOC数据集。

[0104] 本发明采用轻量化的网络EfficientNet-B3替代SSD的VGG16作为骨干网络,提取佩戴口罩与不佩戴口罩的人脸的特征,提高了SSD算法的检测速度,降低了运行时的内存占用量和模型的参数量,可以安装在现有的普通CPU设备上使用。

[0105] 本发明实施例的轻量化特征融合SSD算法检测器与其它算法的检测效果对比如下表1所示(速度与内存占用测试于在普通CPU设备Intel Core i5 CPU@3.2HZ)

[0106] 表1

检测算法	SSD (VGG16)	SSD (EfficientNet)	轻量化特征融合SSD
准确率	91.3%	90.5%	95.7%
速度(帧/秒)	11	24	21
内存占用(MiB)	1927	1537	1576
模型大小(MB)	182	88.5	90.8

[0108] 本领域普通技术人员可以理解:附图只是一个实施例的示意图,附图中模块或流程并不一定是实施本发明所必须的。

[0109] 通过以上的实施方式的描述可知,本领域的技术人员可以清楚地了解到 本发明可借助软件加必需的通用硬件平台的方式来实现。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品 的形式体现出来,该计算机软件产品可以存储在存储介质中,如ROM/RAM、磁 碟、光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本发明各个实施例或者实施例的某些部分所 述的方法。

[0110] 本说明书中的各个实施例均采用递进的方式描述,各个实施例之间相同 相似的部分互相参见即可,每个实施例重点说明的都是与其他实施例的不同 之处。尤其,对于装置或系统实施例而言,由于其基本相似于方法实施例,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。以上所描 述的装置及系统实施例仅仅是示意性的,其中所述作为分离部件说明的单元 可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可 以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案 的目的。本领域普通技术人员在不付出创造性劳动的情况下,即可以理解并 实施。

[0111] 以上所述,仅为本发明较佳的具体实施方式,但本发明的保护范围并不 局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可 轻易想到的变化或替换,都应涵盖在本发明的保护范围之内。因此,本发明 的保护范围应该以权利要求的保护范围为准。

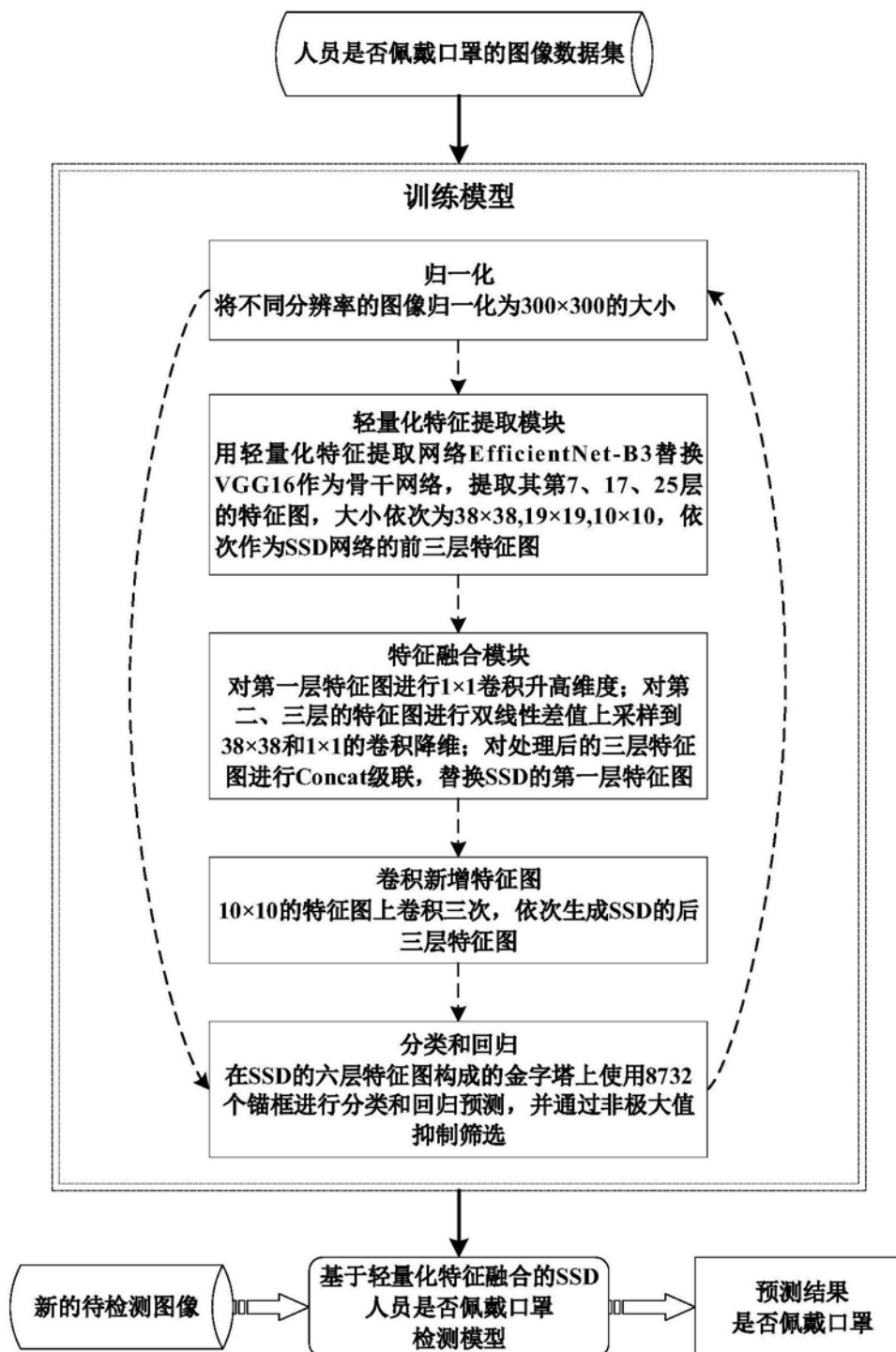


图1



图3

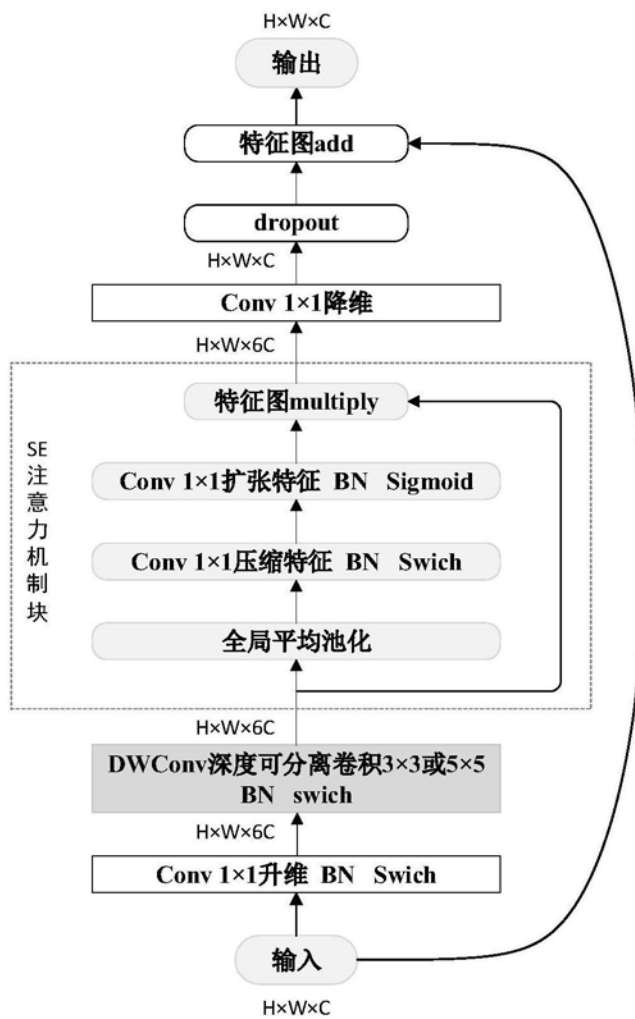


图4

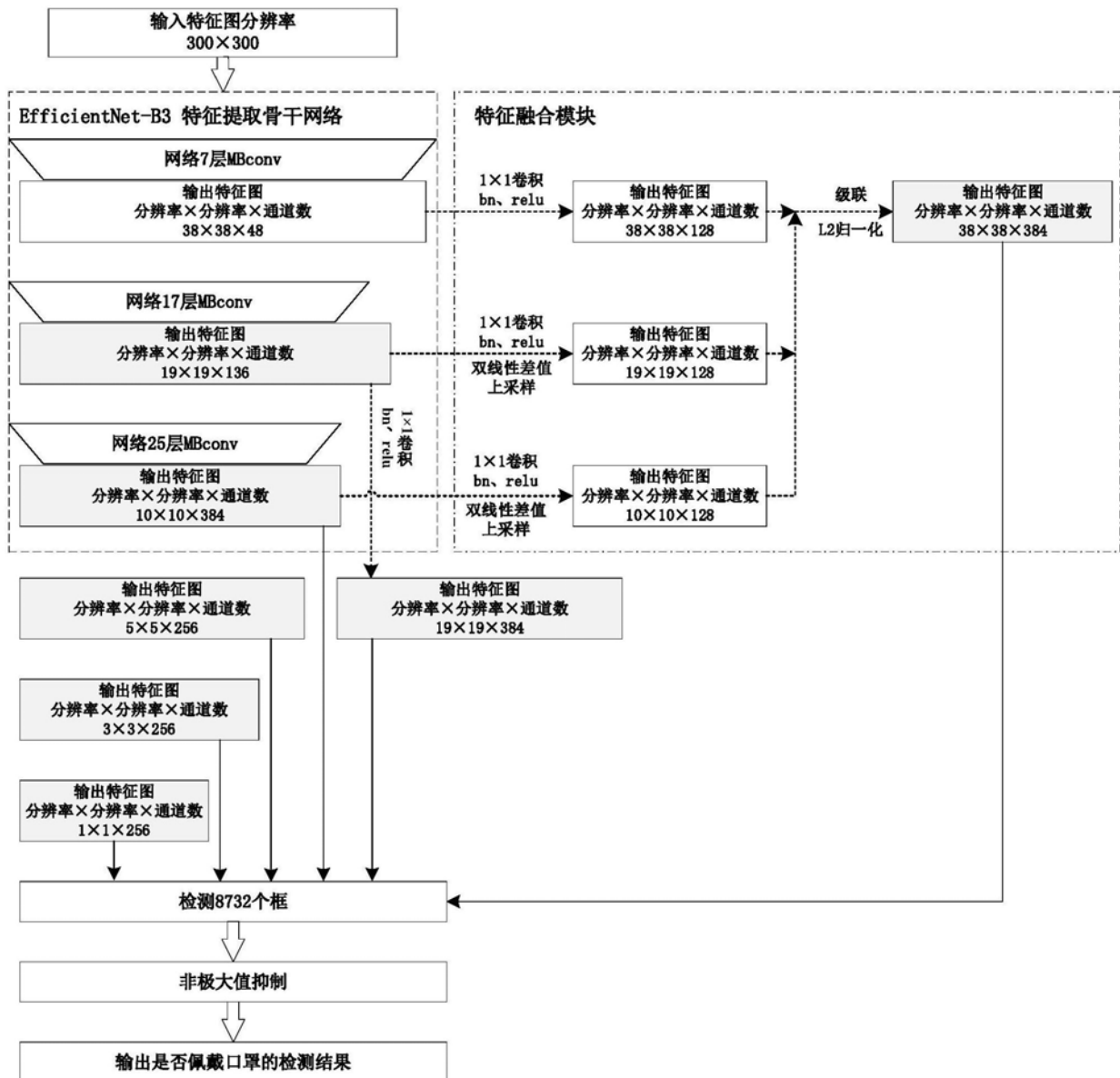


图5



图6