基于YOLOv5的口罩佩戴检测方法

**技术领域**

本发明应用场景涉及口罩佩戴检测、监控系统、安全防控领域，理论方面涉及计算机视觉和模式识别领域，尤其涉及一种基于YOLOv5的口罩佩戴检测方法，属于计算机视觉视频监控侦察信息化建设方向。

**背景技术**

新冠肺炎疫情暴发后，因新冠病毒传染性极强，人类健康受到巨大威胁，为了阻断新冠疫情的蔓延，规范佩戴口罩成为一项有效的防控手段。但是仅靠人工方式对流动人员进行口罩佩戴检测存在诸多问题，如效率较低，且会耗费大量人力物力，同时近距离接触待检测人员更是存在较大的安全风险。因此，构建口罩自动检测系统检测活动人员在各类车站、大型商场等公共场合的口罩佩戴情况，对疫情防控具有重要的现实意义。

口罩具有小目标、多样性等属性，再加上场景的多样性以及目标间交互的复杂性，使得多目标跟踪的研究成为难点。除此之外，口罩佩戴检测还存在以下难点：在昏暗条件下，由于光照强度不大，可见度低，难以对人脸进行精确定位，口罩佩戴检测任务的难度也更为困难。

已公布的多数目标检测方法已在口罩佩戴检测任务上取得了不错的成绩，但对可见度不高、光照强度不强的昏暗条件下，其检测精度仍有待提高。现有技术通常采用多阶段方法对脸部遮挡物进行检测，相对于一阶段方法，需要多级网络，检测速度慢，实时性不好；或者输入需要产生图像金字塔，增加了网络输入的数据量，计算速度慢；同时对人脸尺寸大小的范围进行限定，影响召回率。

**发明内容**

发明技术解决问题：为了克服在可见度不高、光线昏暗的场景下难以精确检测到目标，且难以对特殊目标人群（未佩戴口罩）进行精确检测并实时反馈检测信息。本发明提供了一种基于YOLOv5的人脸口罩检测方法，利用图像增强算法对图片进行预处理，将通道注意力和空间注意力结合，充分挖掘人脸口罩等关键特征点，综合考虑了实际检测中光线的明暗程度对检测精度的影响。通过注意力机制能更加准确的提取人脸口罩关键点特征，使得原始YOLOv5网络获得更强得特征表达能力，从而提高检测准确率。

为了克服上述现有技术中的缺陷，本发明提供了一种基于YOLOv5的人脸口罩检测方法，包括：

将待测人脸口罩数据输集入至目标YOLOv5算法模型，由1个Focus模块及4个Conv模块实现32倍下采样。

Focus模块将输入数据进行切片操作，在一张图片中每隔一个像素取一个值，类似于邻近下采样，切分为4份数据，每份数据都是相当于2倍下采样得到的。

将采样后的数据在通道维度上进行拼接，进行卷积操作。

卷积操作后的图像信息进入C3模块，C3模块参照CSPNet(Cross Stage Partial Network)结构[11]将一个阶段中基础层的特征图分成两部分，拆分和合并策略被跨阶段使用，将梯度的变化从头到尾地集成到特征图中。

从C3模块输出的信息传入至Head模块。Head部分里，通过将高层的特征信息通过上采样的方式与低层特征信息进行传递融合，传达强语义特征，实现自顶向下的信息流动。

再通过步长为2的卷积进行处理，将底层特征与高层特征进行拼接操作，使底层的特征信息容易传到上层去，从而实现了PANet操作。

进一步地，在所述将待测数据输入至目标YOLOV5算法模型之前，还包括：，通过网络爬取与自行拍摄结合制作数据集，共采集9000张图片，对采集的图片信息进行分类，规范数据集格式，手工标注。

进一步地，对所述数据集分类分为两种类别，分别是bad和good，bad表示人员未佩戴或未按规范佩戴口罩，good表示正确佩戴口罩。

进一步地，使用图像增强技术，对标签为good的样本图片进行图像平移、翻转、旋转、缩放，分离三个颜色通道并添加随机噪声来有效缓解类间不平衡的问题。

进一步地，所述规范数据集格式，手工标注包括：数据集采用YOLO格式，图片标注使用LabelImg，标注后的文件以.txt作为后缀，文件名称和图片名称一致。

进一步地，在原始YOLOv5的网络基础上，引入卷积注意力模块CBAM(Convolutional Block Attention Module Network)。CBAM包含两个子模块，分别是通道注意力模块CAM(Channel Attention Module)和空间注意力模块SAM(Spatial Attention Module)。引入注意力机制的YOLOv5模型总计367层，7150056个parameters。

进一步地，在网络结构中采用采用CIOU Loss作为目标框回归的损失函数。

进一步地，使用二元交叉熵损失函数来计算分类损失和目标置信度损失。

本发明提供了一种基于YOLOv5的人脸口罩检测方法，还包括：

对数据进行训练阶段时，重复佩戴口罩检测网络训练过程，不断修正佩戴口罩检测网络的参数，直至口罩检测网络学会找出图像中的人脸位置并能够正确的判断检测出来的人脸是否佩戴口罩，保存训练得到的参数。

在网络模型训练阶段，迭代批量大小设置为32，总迭代次数为600次。初始学习率设置为0.001，采用小批量梯度下降法，并使用Adam优化器计算每个参数的自适应学习率。

相对于现有技术，本发明地有益效果在于：

1. 本发明通过引入注意力机制，注意力机制作用在特征图之上，通过获取特征图中的可用注意力信息，能够达到更好的任务效果。
2. 本发明采用CIOU Loss作为目标框回归的损失函数，由于CIOU Loss从重叠面积、中心点距离和长宽比三个角度进行衡量，故预测框回归的效果更佳。
3. 本发明在YOLOv5网络中添加了CBAM模块，由于CBAM模型在通道注意力模块中加入了全局最大池化操作，它能在一定程度上弥补全局平均池化所丢失的信息。其次，生成的二维空间注意力图使用卷积核大小为7的卷积层进行编码，较大的卷积核对于保留重要大的卷积核对于保留重要的空间区域有良好的帮助。使YOLOv5网络不仅能更为准确的对目标进行分类识别，而且能更为精准的定位目标所在的位置。

综上所述，在图像增强和注意力机制的加持下，经过改进的YOLOv5模型实现了口罩佩戴的高效检测，对是否正确佩戴口罩作出了正确的判断。在可见度不高、光照强度不大的昏暗条件下，引入注意力机制能更加准确的提取人脸口罩关键点特征，提高了检测的准确率和口罩检测效率。本发明对YOLOv5网络的损失函数进行相应的改进，具有较强的鲁棒性和可扩展性，基本能够达到视频图像实时性的要求。

**附图说明**

为了更清楚地说明本发明的技术方案，下面将对实施方式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方式，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。

图1为YOLOv5网络结构图；

图2为口罩检测机制实现的全流程图；

图3为CBAM网络结构图；

图4为CAM结构图；

图5为SAM结构图；

图6为数据集部分图片；

图7为数据集类别分布图；

图8为图像增强后数据集可视化分析图；

图9为口罩检测效果实际场景对比实例图；

**具体实施方式**

为了使本发明的目的和优点更加清楚明白，下面结合实施例对本发明作进一步描述；应当理解，此处所描述的具体实施例仅仅用于解释本发明，并不用于限定本发明。

下面参照附图来描述本发明的优选实施方式。本领域技术人员应当理解的是，这些实施方式仅仅用于解释本发明的技术原理，并非在限制本发明的保护范围。

应当理解，本发明中所使用的步骤编号仅是为了方便描述，不作为对步骤执行先后顺序的限定。

应当理解，在本发明说明书中所使用的术语仅仅是出于描述特定实施例的目的而并不意在限制本发明。需要说明的是，在本发明的描述中，术语“一”、“一种”、“所述”，也可包含复数形式。

应当理解，在本发明的描述中，术语“包括”指示所描述特征、模块、结构、元素、操作的存在，但并不排除其他多个特征、模块、结构、元素、操作的存在或添加。

请参阅图2，本发明某一实施例提供了一种基于YOLOv5的人脸口罩检测方法，包括：

步骤一，将通过网络爬取与自行拍摄结合制作的共包含9000张图片的数据集进行进行分类，规范数据集格式，手工标注等预处理。

步骤二，将预处理后的数据进行训练，迭代批量大小设置为32，总迭代次数为600次，初始学习率设置为0.001，采用小批量梯度下降法。随即进行目标定位分类与特征提取。

步骤三，使用Adam优化器更新网络参数。

步骤四，判断是否完成迭代训练并得到最优模型，训练集和测试集所占的比例为8：1。如未得到则返回模型重新训练测试数据。如完成迭代训练并得到最优模型后，则输入测试数据检测目标类别及位置，最终得到检测结果并输出。

在某一实施案例中,需要说明的是，通过网络爬取与自行拍摄相结合采集数据集，其中图片80%来源源于网络，20%来源于实际拍摄，实际拍摄主要获取的是在昏暗条件下的口罩佩戴图片，实验过程中从楼道、室内等光线昏暗的场所进行了图片采集，同时也在傍晚和清晨等光线较弱的环境下进行了拍摄。实验参数如表所示：

|  |  |
| --- | --- |
| 参数 | 数值(张) |
| 总数据集 | 9000 |
| 训练集 | 8000 |
| 测试集 | 1000 |

在某一实施案例中，通过优化损失函数来改进训练效果。

具体地，在对YOLOv5算法模型解释之前，首先对其结构采用的损失函数进行说明。在本发明中，采用二元交叉熵损失函数来计算分类损失和目标置信损失。损失函数对模型的检测效果至关重要。本文使用的损失函数 如下：

其中，K表示网络最后输出的的特征图划分为K×K个格子，M表示每个格子对应的锚框的个数，表示有目标的锚框，表示没有目标的锚框，表示对没有目标锚框的损失系数。

进一步地，经过对比实验，本文采用CIOU Loss作为目标框回归的损失函数，CIOU Loss如下：

其中，*，*，表示预测框与目标框两个中心点的欧式距离，表示最小外接矩形的对角线距离。和分别表示目标框和预测框各自的的宽高比。

在实施案例中，在原始YOLOv5的网络基础上，引入卷积注意力模块CBAM[19](Convolutional Block Attention Module Network)。需要说明的是，CBAM包含两个子模块，分别是通道注意力模块CAM(Channel Attention Module)和空间注意力模块SAM(Spatial Attention Module)。CAM汇总通道注意力信息，CAM为给定的任意中间特征使用基于宽和高的最大池化操作(global max pooling)和全局平均池化(globalaverage pooling)对特征映射在空间维度上进行压缩，得到两个特征图，这两个特征图共享一个两层的神经网络MLP，第一层神经元的个数为C/r(r为减少率)，激活函数是ReLU，第二层的神经元个数是C，然后对MLP输出的两个特征图使用基于元素的加和操作，再经过Sigmoid激活函数进行归一化处理，得到最终的通道注意力特征图。

SAM汇总空间注意力信息，SAM主要关注于目标在图像上的位置信息，它将CAM的输出特征图作为本模块的输出特征图。它首先做一个基于通道的全局最大池化和全局平均池化，分别得两个特征图，然后将这两个特征图串联，基于通道做拼接操作，再经过一个7×7卷积操作生成空间注意力特征。

在某一实施案例中，迭代批量大小优选为32，总迭代次数优选为600次，学习率优选为0.01，来提高训练的效果和完整性。

在某一实施案例中，模型评估指标主要使用平均精度均值(mAP)、召回率(Recall)、准确率(Precision)。需要说明的是，平均精度均值(mAP)，即所有类别的平均精度求和除以数据集中所有类的平均精度，如公式(6)所示，其中AP的值为P-R曲线的面积。

|  |  |
| --- | --- |
|  | (6) |

召回率，即样本中的正确类别被模型预测正确的概率，如公式(7)所示，其中TP表示将正类别预测为正类别的个数，FN表示将正类别预测为负类别的个数。

|  |  |
| --- | --- |
|  | (7) |

准确率，即预测数据集中预测正确的正样本个数除以实际的正样本个数，如公式(8)所示，其中FP表示将负类别预测为正类别的个数。

|  |  |
| --- | --- |
|  | (8) |

在某一实施案例中，模型训练完成后，将得到的模型与参考文献[5]中的方法和AIZOO方法的口罩检测模型进行对比实验，分别在光照强度为30-75Lux(昏暗)，75-250Lux(较昏暗)和250-1000Lux(正常光照)的条件下进行对比实验。需要说明的是，其中，光照强度是指单位面积上所接受可见光的能量，常用于指示光照的强弱和物体表面积被照明程度的量，单位是Lux，光照强度越大，表明光照越强，物体表面被照的越亮。

至此，已经结合附图所示的优选实施方式描述了本发明的技术方案，但是，本领域技术人员容易理解的是，本发明的保护范围显然不局限于这些具体实施方式。在不偏离本发明的原理的前提下，本领域技术人员可以对相关技术特征做出等同的更改或替换，这些更改或替换之后的技术方案都将落入本发明的保护范围之内。

以上所述是本发明的优选实施方式，应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理的前提下，还可以做出若干改进和润饰，这些改进和润饰也视为本发明的保护范围。