**技术领域**

针对在光线昏暗条件下口罩佩戴检测准确率较低的问题，提出了将注意力机制与YOLOv5网络模型相结合的口罩佩戴检测方法。

**背景技术**

自2019年新冠肺炎疫情暴发以来，口罩佩戴检测成为疫情防控的必备操作。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。

针对口罩佩戴检测这一特定问题，目前国内外有多位学者进行了研究。邓黄潇等人[4]提出了运用迁移学习和ReinaNet网络对口罩佩戴进行检测，验证集下AP达到86.5%。肖俊杰等人[5]将YOLOv3和YCrCb方法结合，正确佩戴口罩的识别率的达到82.5%。牛作东等人[6]对RetinaFace算法进行了相应的优化，同时引入了自注意力机制，实验结果表明检测效果较好。AIZOO也提出了一种基于Fast-RCNN和YOLOv3的目标检测方法，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。

**发明内容**

本文主要针对在可见度不高、光线昏暗的场景下，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，同时对YOLOv5网络的损失函数进行相应的改进，提高模型在昏暗条件下的鲁棒性。

在YOLOv5的v5.0版本中，使用了新的激活函数SiLU()来替换先前版本中的激活函数LeakReLU()和Hardwish()，使得网络中的任何地方都只使用SiLU()激活函数。

同时，YOLOv5中删减了先前版本中的BottleneckCSP中的部分Conv模块，经过改进后的BottleneckCSP称为C3模块。C3模块的详细结构见图 1。由于C3模块移除了每个瓶颈结构中的一个卷积，导致新版本的YOLOv5能够得到更小的模型，且推理速度也有少许提升。

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)。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，SAM汇总空间注意力信息。

与通道注意力不同，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。

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。添加了CBAM的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。

实验结果表明，在可见度不高、光照强度不大的昏暗条件下，与参考文献[5]和AIZOO的方法相比，本文方法使用图像增强能改善图片的质量，然后通过注意力机制能更加准确的提取人脸口罩关键点特征，从而使得检测的准确率更高，具有较强的鲁棒性和扩展性，基本能够达到视频图像实时性的要求。

**具体实施方案**

针对在光线昏暗条件下口罩佩戴检测准确率较低的问题，提出了将注意力机制与YOLOv5网络模型相结合的口罩佩戴检测方法。首先对训练集图片使用图像增强算法进行预处理，然后将图片送入到引入了注意力机制的YOLOv5网络中进行迭代训练，完成训练后，将最优权重模型保存并在测试集上测试。

本文采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，分别如公式(3)和公式(4)所示。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。

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)。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，SAM汇总空间注意力信息。

与通道注意力不同，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。SAM的结构如图 5所示。

因此实验通过网络爬取与自行拍摄相结合制作数据集，数据集中80%来源源于网络，20%来源于实际拍摄。在昏暗条件下的口罩佩戴图片相对缺乏，实际拍摄主要获取的是这类图片，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。

本文数据集存在轻微类间不平衡[20]的问题，如图 7a所示，这是物体检测的常见问题。模型训练时，当某一类别的样本数量较少时，可能会使模型对该类别的关注较小，模型会主要关注样本数量较多的类别，模型网络参数也会主要根据样本数量较多的类别进行调整，从而导致模型对样本数量较少的类别的识别准确率降低。为了解决该问题，本文使用图像增强技术，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，图像增强后的数据集类别分布如图 7 b所示。

实验结果表明，在可见度不高、光照强度不大的昏暗条件下，与参考文献[5]和AIZOO的方法相比，本文方法使用图像增强能改善图片的质量，然后通过注意力机制能更加准确的提取人脸口罩关键点特征，从而使得检测的准确率更高，具有较强的鲁棒性和扩展性，基本能够达到视频图像实时性的要求。