# 基于改进YOLOv5的小目标检测算法

**技术领域**

本发明涉及小目标检测技术领域，具体涉及一种基于改进YOLOv5的小目标检测算法。

**背景技术**

目标检测作为计算机视觉领域的核心问题之一，结合了目标定位和目标分类两大任务，被广泛应用于人脸识别、自动驾驶、行人检测、智能监控等计算机视觉领域。传统的目标检测算法由3部分组成，分别是区域选择、特征提取和分类器，但由于其存在手工设计的特征鲁棒性差、区域选择策略没有针对性等特点，检测效果并不理想。

小目标检测长期以来就是目标检测中的重点和难点之一，相比于常规目标，小目标在图像中所占的像素数较少，分辨率低，信息量少，特征表达能力弱。早期的目标检测方法是使用手工提取特征，再在此基础上构造模型。使用该方法设计的模型不仅结构复杂，而且难以提升精度。

随着深度学习的发展，人们发现卷积神经网络具有极好的学习特征的能力，目前两个令人瞩目的方法是以R-CNN(Region-based Convolutional Neural Network)系列为代表的基于候选框的两阶段深度学习算法和以YOLO(You Only Look Once)系列为代表的基于回归的单阶段深度学习目标检测算法。因此，Girshck等人将CNN应用到目标检测中，提出了R-CNN模型，使得模型的检测性能有了很大的提升，但其网络只能输入固定尺寸的图像，且R-CNN通过选择性搜索(SS, Selective Search)算法生成的候选区域大量重叠，造成了计算资源的极大浪费。针对模型存在的缺点，人们提出了多种网络模型，从Fast R-CNN模型发展到Faster R-CNN模型，虽然模型的检测速度和精度均有所提高，但是这类两阶段目标检测算法与单阶段目标检测算法在检测速度上相比仍具有一定的差距。

**发明内容**

本发明的目的是提出一种基于改进YOLOv5的小目标检测算法，通过使用Mosaic-8方法进行数据增强，增加一个浅层特征图，调整损失函数, 增强网络对小目标的感知能力；通过修改目标框回归公式，解决训练过程中梯度消失等问题，提升了小目标检测精度。

为了实现上述任务，本发明采用以下技术方案：

一种基于改进YOLOv5的小目标检测方法包括下面步骤：

获取小目标口罩佩戴数据集，输入到训练好的完成模型迭代的最佳小目标检测模型中，输出检测小尺寸口罩佩戴目标的检测结果。

所述小目标检测模型的搭建包括：

在原始YOLOv5的基础上，在Backbone骨干网络和Head网络中，新增尺寸为输入图像尺寸四分之一的特征图，提升对小目标数据的挖掘，采用多尺度反馈以引入全局上下文信息，使用CIOU[18]损失函数，使用Mosaic-8数据增强，从数据集中取出一个批量的图片，从该批量图片中采用8张图片随机裁剪、随机排列、随机缩放，然后组合成一张图片。在YOLOv5骨干网络的基础上对原始输入图片增加一个4倍下采样的过程，原始图片经过4倍下采样后送入到特征融合网络得到新尺寸的特征图。将低层特征图与高层特征图进行信息融合，将特征金字塔网络与路径聚合网络(PAN, Path Aggregation Network)相结合，特征金字塔网络自顶向下传递深层次语义特征，路径聚合网络自底向上传递目标的位置信息。从而建立小目标检测网络。

利用预处理后的数据集训练上述小目标检测网络，完成模型迭代后得到最优检测模型。

进一步地，对所述数据集的预处理包括：

利用标记软件LabelImg对数据集进行YOLO格式的标注，共有两个标记类别，分别是bad(未佩戴口罩) 和good(佩戴口罩)。标注完成后，每一张图片都对应着一个与该图片名称相同的txt文件，txt文件中的每一行表示一个标记实例，共5列，从左到右分别表示标签类别、标记框中心横坐标与图片宽度的比值、标记框中心纵坐标与图片高度的比值、标记框宽度与图片宽度的比值、标记框高度与图片高度的比值。本文数据集来源于WIDER FACE、MAPA(Masked Faces) 这两个公开数据集和网络，从中手动筛选出含有多人场景下的佩戴口罩和未佩戴口罩的人脸图片，最终得到训练集4000张，测试集1320张，共计5320张。

进一步地，实验环境使用Ubuntu20.04 操作系统，使用GeForce GTX 1080Ti显卡进行运算，显存大小为11GB，CPU配置为Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40GHz，CUDA版本为11.4.0，Pytorch版本为1.9.0，Python语言环境为3.7.4。

进一步地，选择CIoU替代GIoU作为目标框回归的损失函数。本实验总迭代次数为140 次，迭代批量大小设置为32，优化器选择SGD。模型训练时学习率使用Warmup[24] 训练预热，在Warmup阶段，偏置层的学习率由0.1下降至0.01，其他的参数学习率由0增加至0.01，Warmup 结束之后，采用余弦退火学习算法对学习率进行更新。

进一步地，对目标框公式进行修改。对真实目标框的预测，通过预测相对位置的方法预测出目标框相对于坐上角的相对坐标。目标框公式如下所示：











通过上述公式最终得到预测目标框的中心坐标、和宽高、。是预测框的置信度，由预测框的概率和预测框与真实框的IoU值相乘得到。对设定阈值，过滤掉置信度较低的预测框，然后再对剩下的预测框用非极大值抑制算法(NMS，Non-Maximum Suppression)得到最终的预测框。

与现有技术相比，本发明具有以下有益效果：

1. 采用Mosaic-8数据增强，丰富了数据集，同时增加了小样本目标，可以提升网络训练速度，在进行归一化操作时，可以一次性计算8张图片，小批量不需要太大，能有效达到降低模型对内存的需求，其中合理引入一些随机噪声，增强网络模型对图像中小目标样本的区分力，提升模型的泛化力。
2. 改进特征提取模型，在YOLOv5骨干网络的基础上对原始输入图片增加一个4倍下采样的过程，原始图片经过4倍下采样后送入到特征融合网络得到新尺寸的特征图，该特征图感受野较小，位置信息相对丰富，可以提升检测小尺寸口罩佩戴目标的检测效果；改进特征融合网络，通过顶向下和自底向上的特征信息融合有利于模型更好的学习到特征，增强模型对小目标和遮挡目标的敏感度。
3. 改进目标框公式，综合考虑了真实框与预测框之间的重叠率、中心点距离、长宽比，能使得目标框回归更加稳定，收敛的精度更高。本文算法在密集人群口罩佩戴检测效果中表现突出，检测精度有明显上升，在人群密集的条件下误判和漏检的效果明显下降，对小目标异常角度、人脸区域存在遮挡的鲁棒性明显提升。

**附图说明**

通过参考附图阅读下文的详细描述，本公开示例性实施方式的上述以及其他目的、特征和优点将变得易于理解。在附图中，以示例性而非限制性的方式示出了本公开的若干实施方式，并且相同或对应的标号表示相同或对应的部分，其中：

**具体实施方式**

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，本领域技术人员应知，本发明所描述的实施例是本发明一部分实施例，而不是全部的实施例。因此，本领域技术人员在没有做出创造性劳动前提下，基于本发明中的实施例所获得的所有其他实施例，都属于本发明保护的范围。

参见图1，本发明公开了一种基于改进YOLOv5的小目标检测算法,包括以下步骤：

将采集到的待测人脸口罩数据集输入至目标YOLOv5算法模型，采用Mosaic-8数据增强，即采用8张图片随机裁剪、随机排列、随机缩放，然后组合成一张图片，同时合理引入一些随机噪声。然后对原始输入图片增加一个4倍下采样的过程，原始图片经过4倍下采样后送入到特征融合网络，然后得到新尺寸的特征图，再将所得不同分类的特征图送入至训练后得到的最佳模型中，检测目标类别及位置，最终得到识别结果。

其中，采用Mosaic-8数据增强流程中从待测的人脸口罩数据集中，取出一个批次的数据集，从抽出的批次数据集中随机抽取8张图片，对这8张图片进行随机裁剪、随机排列、随机缩放，然后组合成一张图片，重复该步骤，在合成的图片中合理增加一些随机噪声，达到增强网络模型对图像中小目标样本的区分力，提升模型的泛化能力的效果。

其中，在特征提取过程中，改进特征提取模型。在改进特征提取模型中，在YOLOv5骨干网络的基础上对原始输入图片增加一个4倍下采样的过程，原始图片经过4倍下采样后送入到特征融合网络得到新尺寸的特征图，该特征图感受野较小，位置信息相对丰富，可以提升检测小尺寸口罩佩戴目标的检测效果。

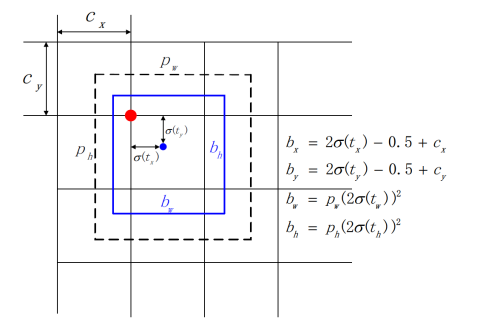
其中，在改进特征提取网络中，将特征金字塔网络与路径聚合网络相结合，特征金字塔网络自顶向下传递深层次语义特征，路径聚合网络自底向上传递目标的位置信息，通过顶向下和自底向上的特征信息融合有利于模型更好的学习到特征，增强模型对小目标和遮挡目标的敏感度。

在一具体实施案例中，选择CIoU替代GIoU作为目标框回归的损失函数。

具体地，CIoU的计算如公式 (6) 所示。其中是一个平衡参数，不参与梯度计算，的定义如公式 (7)所示所示，是用来衡量长宽比一致性的参数，的定义如公式 (8) 所示。

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

在一具体实施案例中，改进目标框回归公式，对真实目标框的预测，通过预测相对位置的方法预测出目标框相对于坐上角的相对坐标。先验框与预测框的关系如下图所示：



需要说明的是，其中，虚线框表示先验框，实线框表示预测框。预测框通过先验框平移缩放得到。将原始图片根据特征图尺寸划分成个网格单元，每个网格单元会预测3个预测框，每个预测框包含4个坐标信息和1个置信度信息。当真实框中某个目标中心坐标落在某个网格中时，就由该网格预测这个目标。

目标框的坐标预测计算公式如下:

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

需要说明的是， 、 、、为网络模型预测得到4个偏移，表示Sigmoid激活函数，用于将网络预测值 、 、、映射到[0,1]之间， 、单元网格中相对于图片左上角的偏移量， 、是先验框宽高。通过上述公式最终得到预测目标框的中心坐标、和宽高、。是预测框的置信度，由预测框的概率和预测框与真实框的IoU值相乘得到。对设定阈值，过滤掉置信度较低的预测框，然后再对剩下的预测框用非极大值抑制算法得到最终的预测框。

在一具体实施案例中，需要说明的是，本发明的测试平台和实验环境为：

实验环境使用Ubuntu20.04 操作系统，使用GeForce GTX 1080Ti显卡进行运算，显存大小为11GB，CPU配置为Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40GHz，CUDA版本为11.4.0，Pytorch版本为1.9.0，Python语言环境为3.7.4。

在一具体实施案例中，需要说明的是，在将待测数据输入基于改进YOLOv5的小目标检测算法网络中之前，需对待测数据进行预处理。

具体地，利用标记软件LabelImg对数据集进行YOLO格式的标注，共有两个标记类别，分别是bad(未佩戴口罩) 和good(佩戴口罩)。标注完成后，每一张图片都对应着一个与该图片名称相同的txt文件，txt文件中的每一行表示一个标记实例，共5列，从左到右分别表示标签类别、标记框中心横坐标与图片宽度的比值、标记框中心纵坐标与图片高度的比值、标记框宽度与图片宽度的比值、标记框高度与图片高度的比值。

在一具体实施案例中，需要说明的是，本文数据集来源于WIDER FACE、MAPA(Masked Faces) 这两个公开数据集和网络，从中手动筛选出含有多人场景下的佩戴口罩和未佩戴口罩的人脸图片，最终得到训练集4000张，测试集1320张，共计5320张。

具体地，本实验总迭代次数为140 次，迭代批量大小设置为32，优化器选择SGD。模型训练时学习率使用Warmup训练预热，减缓模型在初始阶段对小批量数据的过拟合现象，避免模型振荡以便保证模型深层次的稳定性。在Warmup阶段，偏置层的学习率由0.1下降至0.01，其他的参数学习率由0增加至0.01，Warmup 结束之后，采用余弦退火学习算法对学习率进行更新。

需要说明的是，本文评估指标采用平均精度(AP, Average Precision)、平均精度均值(mAP, mean Average Precision)以及每秒检测图片的帧数(FPS, Frames Per Second)这三种在目标检测算法中较为常见的评价指标来评估本文算法的性能。平均精度与精确率(Precision) 和召回率(Recall) 有关，精确率是指预测数据集中预测正确的正样本个数除以被模型预测为正样本的个数；召回率是指预测数据集中预测正确的正样本个数除以实际为正样本的个数。

在一具体实施案例中，为验证本文算法的有效性，将本文算法与参考文献中的方法、AIZOO方法、原始YOLOv5 算法在同一测试集上进行测试，各项性能指标比较，结果如表 2所示。

表 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 算法 | AP/% | | mAP/% | Times/s | FPS |
| bad | good |
| 参考  文献[23] | 83.53 | 84.17 | 83.85 | 0.028 | 35.3 |
| AIZOO | 87.36 | 86.88 | 87.12 | 0.021 | 47.6 |
| YOLOv5 | 89.49 | 91.16 | 90.33 | 0.024 | 41.6 |
| 本文  算法 | 93.21 | 96.54 | 94.88 | 0.033 | 30.3 |

在表 2中，相较于参考文献[23] 方法、AIZOO方法和原始YOLOv5算法，本文算法在密集人群场景下对口罩这个小目标的检测表现效果更好，mAP值可以达到94.88%，在原始YOLOv5的基础上，bad和good类别的AP值分别提高了3.72%和5.38%，mAP值提高了4.55%。本文算法在检测速率上不及其他算法，FPS 为30.3，与原始YOLOv5 相比，FPS下降了11.3，检测单张图片的时间增加了9ms。

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。