



# (12) 发明专利申请

(10) 申请公布号 CN 113159215 A

(43) 申请公布日 2021.07.23

(21) 申请号 202110504503.2

(22) 申请日 2021.05.10

(71) 申请人 河南理工大学

地址 454000 河南省焦作市高新区世纪大道2001号

(72) 发明人 刘群坡 王淇璟 刘尚争 王满利  
张建军 吴中华 魏萍

(74) 专利代理机构 焦作市科彤知识产权代理事务  
所(普通合伙) 41133

代理人 杨晓彤

(51) Int.Cl.

G06K 9/62 (2006.01)

G06K 9/46 (2006.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

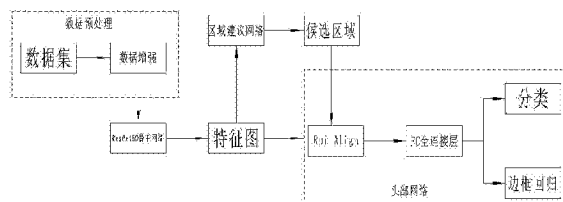
权利要求书2页 说明书9页 附图2页

(54) 发明名称

一种基于Faster Rcn的小目标检测识别方法

(57) 摘要

本发明涉及一种基于Faster Rcn的小目标检测识别方法,属于计算机视觉识别系统技术领域,具体包括配置环境,搭建基于Faster Rcn的小目标检测识别模型以及模型训练平台,小目标检测识别模型包括数据预处理;骨干网络、区域建议网络和头部网络。本申请的有益效果为:用resnet50作为主干网络提取图像特征,以及配合使用Roi Align方法,减小了网络的量化误差,还可以精确获得了原图中物体的位置坐标信息,采用DIOU算法,解决了在计算传统Iou时两框包含与分离时的值不变的问题。



1.一种基于Faster Rcnm的小目标检测识别方法,其特征在于:包括配置环境,搭建基于Faster Rcnm的小目标检测识别模型以及模型训练平台,所述小目标检测识别模型包括四个部分:

1)、数据预处理:对训练数据集中的原始图片进行数据增强,并对数据增强后的训练数据集中图像小目标进行标注,形成小目标数据集;

2)、骨干网络:采用带有残差边的ResNet网络提取小目标数据集中图像特征;

3)、区域建议网络:在原始图像上提取出目标可能存在的候选区域,并在候选区域生成多个锚框,采用DIOU处理锚框并同时锚框进行软非极大值抑制操作;

4)、头部网络:根据区域建议网络的结果,在特征图中将相应的区域转化为固定尺度的特征图,然后再进行精确的目标分类和目标定位;头部网络包含Roi Align和两个分支网络,在Roi Align中使用双三次插值来精确的寻找每个锚框块对应的特征,并从特征图中提取出固定长度的特征向量;然后把特征向量送入FC全连接层,最后再送入两个分支网络中,其中一个分支网络为分类网络,进行分类预测,输出具体的类别;另一个分支网络为边界回归网络,进行回归预测,输出目标精确的位置;每一个锚框块区域都定义了多任务的损失函数: $L=L_{\text{cof}}+L_{\text{lof}}$ ,其中 $L_{\text{cof}}$ 为分类损失函数,对应分类网络, $L_{\text{lof}}$ 为定位损失函数,对应边界回归网络。

2.根据权利要求1所述的基于Faster Rcnm的小目标检测识别方法,其特征在于:原始图片进行数据增强的方式包括随机的翻转、错切、缩放、旋转、对比度拉伸、裁剪的一种或多种组合。

3.根据权利要求1所述的基于Faster Rcnm的小目标检测识别方法,其特征在于:所述ResNet网络为ResNet50。

4.根据权利要求1所述的基于Faster Rcnm的小目标检测识别方法,其特征在于:区域建议网络根据卷积神经网络生成特征图,先在原始图片上生成多个锚框;区域建议网络利用了滑动窗口机制,窗口每滑动一次可以产生k个锚框;利用DIOU定位锚框位置,利用软非极大值抑制操作确定最终的检测框;

DIOU公式如下:

$$DIOU=IOU-\frac{d^2}{c^2} \quad (1)$$

IOU公式如下:

$$IOU=\frac{A \cap B}{A \cup B} \quad (2)$$

式(1)和式(2)中,A代表预测框(锚框),B代表真实框(标签值), $\cap$ 代表取交集, $\cup$ 代表取并集,d表示预测框和真实框中心点的距离,c表示外包框的对角线距离,IOU表示交并比,DIOU表示距离交并比;

所述软非极大值抑制操作利用目标框和真实框的重叠率排序,然后选出其中重叠率最大的目标框并把其余的目标框删除,选出的目标框就是最终的检测框,公式为:

$$F=\begin{cases} F, & DIOU(M, b_i) < N_t \\ F(1-DIOU(M, b_i)), & DIOU(M, b_i) \geq N_t \end{cases} \quad (3)$$

式(3)中:M为当前得分最高的框, $b_i$ 为待处理的框,F为最终的得分, $N_t$ 为事先设定的阈值,当重叠率越大,F的值越高,重叠率越小,F的值越低,但是并不会为0;

在选出目标框后还需要对目标框进行微调操作,微调公式如下:

$$f(P_x, P_y, P_w) = (Z_x, Z_y, Z_w, Z_h) \quad (4)$$

$$Z_x = P_x + P_w d_x(P), Z_y = P_y + P_h d_y(P) \quad (5)$$

$$Z_w = P_w e^{d_w(P)}, Z_h = P_h e^{d_h(P)} \quad (6)$$

式(4)(5)(6)中,  $(P_x, P_y, P_w, P_h)$  代表目标框的左上角坐标  $(P_x, P_y)$  和框的长  $(P_w)$  和高  $(P_h)$ ,  $(Z_x, Z_y, Z_w, Z_h)$  代表真实框的左上角坐标  $(Z_x, Z_y)$  和框的长  $(Z_w)$  和框  $(Z_h)$ ,  $d_x, d_y, d_w, d_h$  是四个映射函数,由以下公式得到:

$$t_x = \frac{G_x - P_x}{P_w}, t_y = \frac{G_y - P_y}{P_h} \quad (7)$$

$$t_w = \log\left(\frac{G_w}{P_w}\right), t_h = \log\left(\frac{G_h}{P_h}\right) \quad (8)$$

$$d_s(P) = W_s^T \phi_s(P) \quad (9)$$

$$W_s = \operatorname{argmin}_{W_s} \sum_i^n (t_s^i - W_s^T \phi_s(P^i))^2 + \gamma \|W_s\|^2 \quad (10)$$

式(7)-(10)中,  $t_s$  代表预测值和真实值的差距,  $d_s(P)$  是得到的预测值, 对应  $d_x$  或  $d_y$  或  $d_w$  或  $d_h$  四个映射函数,  $\phi_s(P)$  是输入的特征向量,  $W_s$  是要学习的参数,  $t_x$  是x轴方向上的预测框与真实框的差距,  $t_y$  是y轴方向上的预测框与真实框的差距,  $G_x$  是真实框左上角x轴坐标,  $G_y$  是真实框左上角y轴坐标,  $P_x$  是目标框左上角x轴坐标,  $P_y$  是目标框左上角y轴坐标,  $P_w$  是目标框的长,  $P_h$  是目标框的高,  $G_w$  是真实框的长,  $G_h$  是真实框的高,  $t_w$  是目标框与真实框的长的差距,  $t_h$  是目标框与真实框的高的差距,  $P^i$  是输入的特征向量,  $\gamma$  是随机参数,  $W_s^T$  是  $W_s$  的转置矩阵。

5. 根据权利要求1所述的基于Faster Rcnnet的小目标检测识别方法,其特征在于:头部网络中Roi Align先遍历每一个候选区域,并且保持浮点数;然后将候选区域分割成 $k \times k$ 个单元,最后利用双三次插值计算出每个单元中固定的四个坐标位置,并进行最大池化操作;双三次插值公式为:

$$W(x) = \begin{cases} (a+3)|x|^3 - (a+3)|x|^2 + 1 & |x| \leq 1 \\ a|x|^3 - 5a|x|^2 + 8a|x| - 4a & 1 < |x| < 2 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

$$B(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} * W(i) * W(j) \quad (12)$$

式(11)和(12)中,  $a$  取值为-0.5,  $W(i)$  表示横坐标权重,  $W(j)$  表示纵坐标权重,  $x$  代表某个像素点的像素值,  $a_{ij}$  代表第*i*行*j*列的像素点。

## 一种基于Faster Rcn的小目标检测识别方法

### 技术领域

[0001] 本发明涉及计算机视觉识别系统技术领域,特别涉及一种基于Faster Rcn的小目标检测识别方法。

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

### 背景技术

[0002] 目标检测是计算机视觉领域的一大基本任务,被广泛应用于人脸识别、自动驾驶、视频监控、军事侦察等领域。近年来,随着计算机信息技术的发展,深度学习被大量应用于目标检测任务中,使得目标检测完成了从手动提取人工设计的特征到应用卷积神经网络让计算机自动提取特征的转变,极大的提高了速度和准确度,使得基于深度学习的目标检测算法成为图像处理领域研究的主流。

[0003] 基于深度学习的目标检测算法主要分为目标定位和目标识别这两个基础任务,目的就是图像中的每一个目标进行精确的分类并用包围盒框出每一个目标具体的位置。目标比较流行的目标检测方法主要分为两类,其中一类是YOLO系列算法(YOLOv3、YOLOv4、YOLOv4\_tiny),该系列算法是one-stage的,主要步骤是直接对输入图像进行卷积操作,提取图像特征并在特征图上进行边框回归;另一类是R-CNN系列算法(R-CNN, Fast R-CNN, Faster R-CNN),该系列算法是two-stage的,主要步骤是先通过各种图像处理技术生成可能包含目标的候选区域集合,再利用卷积神经网络对候选区域提取特征,最后进行精确的目标分类和边框回归。YOLO系列算法速度快,准确度要低一点,R-CNN系列算法速度慢,准确度要高一点。

[0004] 尽管上述两类算法经过不断改进,在速度和精度上逐渐提升,但是对自然场景中的小型目标的检测依然存在精度不高或漏检等问题。小目标一般是指尺寸小于32\*32像素的目标,由于小目标的像素数在原图中占比小,携带的特征少,纹理、形状、颜色等外观信息缺乏,再经过卷积神经网络中的下采样操作,特征信息逐步丢失,使得小目标的检测难度增大,检测精度普遍偏低。

[0005] 综上所述,现有技术中针对小目标检测识别方法还多有不足,还需要进行改进。

### 发明内容

[0006] 本发明的目的在于提供一种基于Faster Rcn的小目标检测识别方法,以解决上述背景技术中提出的问题。

[0007] 为实现上述目的,本发明采用如下技术方案:一种基于Faster Rcn(Faster Region-based ConvolutionalNetwork method,基于区域的更快卷积神经网络)的小目标检测识别方法,包括配置环境,搭建基于Faster Rcn的小目标检测识别模型以及模型训练平台,所述小目标检测识别模型包括四个部分:分部分介绍各个模块

[0008] 1)、数据预处理:对训练数据集中的原始图片进行数据增强,并对数据增强后的训练数据集中图像小目标进行标注,形成小目标数据集;数据预处理对网络模型是十分重要的,它直接影响到网络训练的结果,因为一般成功的神经网络都需要大量的参数,而这些参

数都需要从数据中提取出来,但是有时训练数据集可能没有这么多数据,这就需要数据预处理来提高图片的数量。由于小目标在图中所占比例小,所以这里使用随机的翻转、错切、缩放、旋转、对比度拉伸、裁剪的一种或多种组合对原始图片进行数据增强的方法来提高小目标在图像中的多样性,给网络提供更多的特征数量。

[0009] 2)、骨干网络:采用带有残差边的ResNet网络提取小目标数据集中图像特征;ResNet网络优选为ResNet50,其是由Conv Block(卷积块)和Identity Block(标识块)两个基本块组成,其中Conv Block(卷积块)中运用了直连通道传递低层特征,保证小目标特征不会严重丢失,改变了网络的维度。Identity Block(标识块)没有直连通道,只是加深网络的深度,精细提取物体特征,提高了分类精度。

[0010] 3)、区域建议网络:在原始图像上提取出目标可能存在的候选区域,并在候选区域生成多个锚框,采用DIOU处理锚框并同时锚框进行软非极大值抑制操作;

[0011] 区域建议网络是根据卷积神经网络生成特征图的,先在原始图片上生成多个锚框;然后对锚框分别进行分类和回归,得到锚框中是否有物体和锚框的具体位置信息。区域建议网络利用了滑动窗口机制,窗口每滑动一次可以产生k个锚框,然后进行有无物体的分类和位置信息的回归。分类分为正类和负类两个类,可以得到2k个分类输出。回归可以得到锚框的左上角和右下角的位置坐标信息,有4k个输出。

[0012] 在区域建议网络中,我们需要对大量的锚框进行处理,比如对锚框正负样本的定义和对多余锚框的抑制(NMS,non maximum suppression,非极大值抑制)。

[0013] 在现有的Faster Rcn中,当预测框和真实框处于包含关系和分离关系时,现有算法并不能明确表示出两框的具体位置关系,为了解决这个问题,本申请利用DIOU处理锚框,DIOU尤其对像素小的目标,提高了定位精度,使得对微小的偏差都有很强的鲁棒性。

[0014] 具体的,DIOU公式如下:

$$[0015] \quad DIOU = IOU - \frac{d^2}{c^2} \quad (1)$$

[0016] IOU公式如下:

$$[0017] \quad IOU = \frac{A \cap B}{A \cup B} \quad (2)$$

[0018] 式(1)和式(2)中,A代表预测框(锚框),B代表真实框(标签值), $\cap$ 代表取交集, $\cup$ 代表取并集,d表示预测框和真实框中心点的距离,c表示外包框的对角线距离,IOU表示交并比,DIOU表示距离交并比;

[0019] 在用DIOU处理锚框的基础上,还需要对得出的锚框进行非极大值抑制操作。由于Faster Rcn是两阶段的检测算法,会得到大量的目标框,这些目标框都可以检测到目标,所以要从这些目标框中找出最符合实际的目标框,传统的方式是用非极大值抑制方法。

[0020] 非极大值抑制的原理是利用目标框和真实框的重叠率排序,然后选出其中重叠率最大的目标框并把其余的目标框删除,选出的目标框就是最终的检测框。但是这种方法可能会导致误删目标框,因此,本申请对其进行了改进,采用soft\_nms(软非极大值抑制),软非极大值抑制操作利用目标框和真实框的重叠率排序,然后选出其中重叠率最大的目标框并把其余的目标框删除,选出的目标框就是最终的检测框,公式为:

$$[0021] \quad F = \begin{cases} F, & DIOU(M, b_i) < N_t \\ F(1 - DIOU(M, b_i)), & DIOU(M, b_i) \geq N_t \end{cases} \quad (3)$$

[0022] 式(3)中:M为当前得分最高的框,  $b_i$ 为待处理的框,  $F$ 为最终的得分,  $N_t$ 为事先设定的阈值,当重叠率越大,  $F$ 的值越高,重叠率越小,  $F$ 的值越低,但是并不会为0,不会导致目标框误删除的结果;

[0023] 在选出目标框后还需要对目标框进行微调操作,看出目标框和真实框是有一定差距的,为了使得最终的精度更高,就需要微调目标框,得到最终框。微调公式的过程如下:

$$(P_x, P_y, P_w, P_h) = (Z_x, Z_y, Z_w, Z_h) \quad (4)$$

$$[0024] \quad Z_x = P_x + P_w d_x(P), Z_y = P_y + P_h d_y(P) \quad (5)$$

$$[0025] \quad Z_w = P_w e^{d_w(P)}, Z_h = P_h e^{d_h(P)} \quad (6)$$

[0026] 式(4)(5)(6)中,  $(P_x, P_y, P_w, P_h)$ 代表目标框的左上角坐标  $(P_x, P_y)$ 和框的长  $(P_w)$ 和高  $(P_h)$ ,  $(Z_x, Z_y, Z_w, Z_h)$ 代表真实框的左上角坐标  $(Z_x, Z_y)$ 和框的长  $(Z_w)$ 和框  $(Z_h)$ ,  $d_x, d_y, d_w, d_h$ 是四个映射函数,由以下公式得到:

$$[0027] \quad t_x = \frac{G_x - P_x}{P_w}, t_y = \frac{G_y - P_y}{P_h} \quad (7)$$

$$[0028] \quad t_w = \log\left(\frac{G_w}{P_w}\right), t_h = \log\left(\frac{G_h}{P_h}\right) \quad (8)$$

$$[0029] \quad d_*(P) = W_*^T \Phi_5(P) \quad (9)$$

$$[0030] \quad W_* = \operatorname{argmin}_{W_*} \sum_{i=1}^n (t_*^i - W_*^T \Phi_5(P^i))^2 + \gamma \|W_*\|^2 \quad (10)$$

[0031] 式(7)-(10)中,  $t_*$ 代表预测值和真实值的差距,  $d_*(P)$ 是得到的预测值,对应  $d_x$ 或  $d_y$ 或  $d_w$ 或  $d_h$ 四个映射函数,  $\Phi_5(P)$ 是输入的特征向量,  $W_*$ 是要学习的参数,  $t_x$ 是x轴方向上的预测框与真实框的差距,  $t_y$ 是y轴方向上的预测框与真实框的差距,  $G_x$ 是真实框左上角x轴坐标,  $G_y$ 是真实框左上角y轴坐标,  $P_x$ 是目标框左上角x轴坐标,  $P_y$ 是目标框左上角y轴坐标,  $P_w$ 是目标框的长,  $P_h$ 是目标框的高,  $G_w$ 是真实框的长,  $G_h$ 是真实框的高,  $t_w$ 是目标框与真实框的长的差距,  $t_h$ 是目标框与真实框的高的差距,  $P^i$ 是输入的特征向量,  $\gamma$ 是随机参数,  $W_*^T$ 是  $W_*$ 的转置矩阵。

[0032] 4)、头部网络:根据区域建议网络的结果,在特征图中将相应的区域转化为固定尺度的特征图,然后再进行精确的目标分类和目标定位;头部网络包含Roi Align和两个分支网络,在Roi Align中使用双三次插值来精确的寻找每个锚框块对应的特征,并从特征图中提取出固定长度的特征向量。然后把特征向量送入FC全连接层,最后再送入两个分支网络中,其中一个分支网络为分类网络,进行分类预测,输出具体的类别;另一个分支网络为边界回归网络,进行回归预测,输出目标精确的位置;每一个锚框块区域都定义了多任务的损失函数:  $L = L_{\text{cof}} + L_{\text{lof}}$ ,其中  $L_{\text{cof}}$ 为分类损失函数,对应分类网络,  $L_{\text{lof}}$ 为定位损失函,对应边界回归网络,损失函数为现有技术,本申请不进行详细描述,其可以采用现有技术中。

[0033] 头部网络中Roi Align先遍历每一个候选区域,并且保持浮点数;然后将候选区域分割成  $k \times k$ 个单元,最后利用双三次插值计算出每个单元中固定的四个坐标位置,并进行最大池化操作;双三次插值是一种复杂的插值方法,该方法利用待采样点周围确定的16个点的值进行三次插值,不仅考虑到了相邻点的灰度值的影响还考虑到了相邻点的灰度值变化率的影响。虽然加大了运算量,但是可以得到更接近高分辨率图像的效果,使得小目标的信

息丢失较少。计算双三次插值的方法是先构造BiCubic(双三次)函数,从函数中获得周围16个像数点的权重值,然后再根据这到目标点的距离和权重值计算出最终的插值结果。

[0034] 双三次插值公式为:

$$[0035] \quad W(x) = \begin{cases} (a+3)|x|^3 - (a+3)|x|^2 + 1 & |x| \leq 1 \\ a|x|^3 - 5a|x|^2 + 8a|x| - 4a & 1 < |x| < 2 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

$$[0036] \quad B(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} * W(i) * W(j) \quad (12)$$

[0037] 式(11)和(12)中,a取值为-0.5,W(i)表示横坐标权重,W(j)表示纵坐标权重,x代表某个像素点的像素值, $a_{ij}$ 代表第i行j列的像素点。

[0038] 本申请的有益效果为:用resnet50作为主干网络提取图像特征,以及配合使用RoI Align方法,减小了网络的量化误差,还可以精确获得了原图中物体的位置坐标信息,采用DIOU算法,解决了在计算传统IOU时两框包含与分离时的值不变的问题。实验结果表明,本申请的小目标检测识别方法平均识别精度高,消除了漏检问题,对很模糊的小目标也能有很好的识别精度。

## 附图说明

[0039] 图1为本发明实施例1的模型结构。

[0040] 图2为本发明对比例2的原理图。

[0041] 图3为实施例1、对比例1、对比例2分别对BCCD血细胞数据集进行目标检测识别的输出结果图。

## 具体实施方式

[0042] 下面通过对实施例的描述,对本发明作进一步详细说明,目的是帮助本领域的技术人员对本发明的构思、技术方案有更完整、准确和深入的理解,并有助于其实施。

### [0043] 一、测试平台和环境

[0044] 测试平台配置如下:Intel®Core™ i9-9900KF CPU@3.60GHZ,32G内存Windows10操作系统,显卡型号为RTX2080Ti.程序运行环境如下:Python3.6,keras版本为2.1.5,CUDA版本为10.0;选用的数据集为BCCD血细胞数据集。BCCD血细胞数据集是公知的数据集,该数据集共有三类364张图像:WBC(白细胞),RBC(红细胞)和Platelets。3个类别中有4888个标签(有0个空示例)。

[0045] 为了提高对目标的检测精度和模型的泛化能力,在数据量有限的情况下,对原始的细胞数据集依次采取如下的几种数据增强方法,得到不同情况下的不同数据,增加数据的多样性。1)缩放:将图像随机缩小或放大;2)旋转:把图像按随机的角度进行旋转;3)随机裁剪:从图像中随机裁剪出一块区域。最终得到的数据集有2000张有效图片,12944个物体标注实例,并严格按照Pascal Voc2007的格式进行模型的训练。

### [0046] 二、实验细节及评价标准

[0047] 本文网络优化采用SGD随机梯度下降法,对网络进行40000次迭代,其计算式如式



(13)。其中每一批的训练数据包含128张图片,初始化学习率为0.0001,每4000次迭代后学习率衰减为原始学习率的一半,最后设置误差反向传播的动量因子为0.9。

$$[0048] \quad \theta_j = \theta_j - \alpha \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)} \quad (13)$$

[0049] 式13中: $\theta_j$ 为算法中的第j个参数, $\alpha$ 为学习率, $m$ 为训练的样本数量, $x^{(i)}$ 为第i组的x, $y^{(i)}$ 为第j组的y。

[0050] 实验采用AP平均精度对实验结构进行衡量:

$$[0051] \quad A = \sum_{n=1}^M P/M \quad (14)$$

$$[0052] \quad Precision = \frac{TP}{TP+FP} \quad (15)$$

$$[0053] \quad Recall = \frac{TP}{TP+FN} \quad (16)$$

[0054] 式14-16中:A表示某一类的平均精度,M为全部样本中的正例数量,P为不同Recall下的最大Precision,Recall是召回率,代表分类器认为是正类并且确实是正类的部分占有确实是正类的比例,Precision是精确度,代表分类器认为是正类并且确实是正类的部分占有分类器认为是正类的比例,TP是分类器认为是正样本而且确实是正样本的例子,FP是分类器认为是正样本但实际上不是正样本的例子,FN是分类器认为是负样本但实际上不是负样本的例子。

### [0055] 三、实施例1

[0056] 一种基于Faster Rcn (Faster Region-based Convolutional Network method, 基于区域的更快卷积神经网络)的小目标检测识别方法,包括配置环境,搭建如图1所示的基于Faster Rcn的小目标检测识别模型以及搭建模型训练平台,所述小目标检测识别模型包括四个部分:

[0057] 1)、数据预处理:还是分点介绍每个部分,内容与发明技术一致,部分措辞变了片进行数据增强,并对数据增强后的训练数据集中图像小目标进行标注,形成小目标数据集;数据预处理对网络模型是十分重要的,它直接影响到网络训练的结果,因为一般成功的神经网络都需要大量的参数,而这些参数都需要从数据中提取出来,但是有时训练数据集可能没有这么多数据,这就需要数据预处理来提高图片的数量。由于小目标在图中所占比例小,所以这里使用随机的翻转、错切、缩放、旋转、对比度拉伸、裁剪的一种或多种组合对原始图片进行数据增强的方法来提高小目标在图像中的多样性,给网络提供更多的特征数量。在本实施例中,数据集采用(一)测试平台和环境所提供的增强数据集;

[0058] 2)、骨干网络:采用带有残差边的ResNet网络提取小目标数据集中图像特征;ResNet网络优选为ResNet50,其是由Conv Block(卷积块)和Identity Block(标识块)两个基本块组成,其中Conv Block((卷积块))中运用了直连通道传递低层特征,保证小目标特征不会严重丢失,改变了网络的维度。Identity Block((标识块))没有直连通道,只是加深网络的深度,精细提取物体特征,提高了分类精度。

[0059] 3)、区域建议网络:在原始图像上提取出目标可能存在的候选区域,并在候选区域生成多个锚框,采用DIOU处理锚框并同时锚框进行软非极大值抑制操作;

[0060] 区域建议网络是根据卷积神经网络生成特征图的,先在原始图片上生成多个锚



框;然后对锚框分别进行分类和回归,得到锚框中是否有物体和锚框的具体位置信息。区域建议网络利用了滑动窗口机制,窗口每滑动一次可以产生k个锚框,然后进行有无物体的分类和位置信息的回归。分类分为正类和负类两个类,可以得到2k个分类输出。回归可以得到锚框的左上角和右下角的位置坐标信息,有4k个输出。

[0061] 在特征提取网络提取出来的特征图上,进行候选区域生成,提取出候选框。首先在特征图上经过滑窗操作,得到256维的特征,然后在这256维特征上经过2个FC全连接层,会在256维特征的每个点上画出预先设定的9个不同大小的框,并输出每个框的中心点坐标,框的长高和每个框中是否有物体的得分,如果有物体,则为1分,没物体则为0。

[0062] 在区域建议网络中,我们需要对大量的锚框进行处理,比如对锚框正负样本的定义和对多余锚框的抑制(NMS,non maximum suppression,非极大值抑制)。

[0063] 在现有的Faster Rcn中,当预测框和真实框处于包含关系和分离关系时,现有算法并不能明确表示出两框的具体位置关系,为了解决这个问题,本申请利用DIOU处理锚框,DIOU尤其对像素小的目标,提高了定位精度,使得对微小的偏差都有很强的鲁棒性。

[0064] 具体的,DIOU公式如下:

[0065] 
$$DIOU = IOU - \frac{d^2}{c^2} \quad (1)$$

[0066] IOU公式如下:

[0067] 
$$IOU = \frac{A \cap B}{A \cup B} \quad (2)$$

[0068] 式(1)和式(2)中,A代表预测框(锚框),B代表真实框(标签值), $\cap$ 代表取交集, $\cup$ 代表取并集,d表示预测框和真实框中心点的距离,c表示外包框的对角线距离,IOU表示交并比,DIOU表示距离交并比;

[0069] 在用DIOU处理锚框的基础上,还需要对得出的锚框进行非极大值抑制操作。由于Faster Rcn是两阶段的检测算法,会得到大量的目标框,这些目标框都可以检测到目标,所以要从这些目标框中找出最符合实际的目标框,传统的方式是用非极大值抑制方法。

[0070] 非极大值抑制的原理是利用目标框和真实框的重叠率排序,然后选出其中重叠率最大的目标框并把其余的目标框删除,选出的目标框就是最终的检测框。但是这种方法可能会导致误删目标框,因此,本申请对其进行了改进,采用soft\_nms(软非极大值抑制),软非极大值抑制操作利用目标框和真实框的重叠率排序,然后选出其中重叠率最大的目标框并把其余的目标框删除,选出的目标框就是最终的检测框,公式为:

[0071] 
$$F = \begin{cases} F, & DIOU(M, b_i) < N_t \\ F(1 - DIOU(M, b_i)), & DIOU(M, b_i) \geq N_t \end{cases} \quad (3)$$

[0072] 式(3)中:M为当前得分最高的框, $b_i$ 为待处理的框,F为最终的得分, $N_t$ 为事先设定的阈值,当重叠率越大,F的值越高,重叠率越小,F的值越低,但是并不会为0,不会导致目标框误删除的结果;

[0073] 在选出目标框后还需要对目标框进行微调操作,看出目标框和真实框是有一定差距的,为了使得最终的精度更高,就需要微调目标框,得到最终框。微调公式的过程如下:f

( $P_x, P_y, P_w$ ) = ( $Z_x, Z_y, Z_w, Z_h$ ) (4)

[0074]  $Z_x = P_x + P_w d_x(P), Z_y = P_y + P_h d_y(P)$  (5)

具体解释  
公式及其  
公式中参  
数

$$[0075] \quad Z_w = P_w e^{d_w(P)}, Z_h = P_h e^{d_h(P)} \quad (6)$$

[0076] 式(4) (5) (6) 中,  $(P_x, P_y, P_w, P_h)$  代表目标框的左上角坐标  $(P_x, P_y)$  和框的长  $(P_w)$  和高  $(P_h)$ ,  $(Z_x, Z_y, Z_w, Z_h)$  代表真实框的左上角坐标  $(Z_x, Z_y)$  和框的长  $(Z_w)$  和框  $(Z_h)$ ,  $d_x, d_y, d_w, d_h$  是四个映射函数, 由以下公式得到:

$$[0077] \quad t_x = \frac{G_x - P_x}{P_w}, \quad t_y = \frac{G_y - P_y}{P_h} \quad (7)$$

$$[0078] \quad t_w = \log\left(\frac{G_w}{P_w}\right), \quad t_h = \log\left(\frac{G_h}{P_h}\right) \quad (8)$$

$$[0079] \quad d_*(P) = W_*^T \phi_5(P) \quad (9)$$

$$[0080] \quad W_* = \operatorname{argmin}_{W_*} \sum_i^n (t_i^i - w_*^T \phi_5(P^i))^2 + \gamma \|w_*\|^2 \quad (10)$$

[0081] 式(7) - (10) 中,  $t_*$  代表预测值和真实值的差距,  $d_*(P)$  是得到的预测值, 对应  $d_x$  或  $d_y$  或  $d_w$  或  $d_h$  四个映射函数,  $\phi_5(P)$  是输入的特征向量,  $W_*$  是要学习的参数,  $t_x$  是  $x$  轴方向上的预测框与真实框的差距,  $t_y$  是  $y$  轴方向上的预测框与真实框的差距,  $G_x$  是真实框左上角  $x$  轴坐标,  $G_y$  是真实框左上角  $y$  轴坐标,  $P_x$  是目标框左上角  $x$  轴坐标,  $P_y$  是目标框左上角  $y$  轴坐标,  $P_w$  是目标框的长,  $P_h$  是目标框的高,  $G_w$  是真实框的长,  $G_h$  是真实框的高,  $t_w$  是目标框与真实框的长的差距,  $t_h$  是目标框与真实框的高的差距,  $P^i$  是输入的特征向量,  $\gamma$  是随机参数,  $W_*^T$  是  $W_*$  的转置矩阵。

[0082] 4)、**头部网络**: 根据区域建议网络的结果, 在特征图中将相应的区域转化为固定尺度的特征图, 然后再进行精确的目标分类和目标定位; 头部网络包含 Roi Align 和两个分支网络, 针对区域建议网络 RPN 生成的大量的候选框, Roi Align 结合实际的预选框的位置坐标, 把候选框中得分为 1 的区域使用双三次插值来精确的寻找每个锚框块对应的特征, 得到尺寸固定, 且更贴和物体实际大小的特征区域, 并从特征图中提取出固定长度的特征向量。然后把特征向量送入 FC 全连接层, 最后再送入两个分支网络中, 其中一个分支网络为分类网络, 进行分类预测, 输出具体的类别; 另一个分支网络为边界回归网络, 进行回归预测, 输出目标精确的位置; 每一个锚框块区域都定义了多任务的损失函数:  $L = L_{\text{cof}} + L_{\text{lof}}$ , 其中  $L_{\text{cof}}$  为分类损失函数, 对应分类网络,  $L_{\text{lof}}$  为定位损失函, 对应边界回归网络, 损失函数为现有技术, 本申请不进行详细描述, 其可以采用现有技术中。

[0083] 头部网络中 Roi Align 先遍历每一个候选区域, 并且保持浮点数; 然后将候选区域分割成  $k \times k$  个单元, 最后利用双三次插值计算出每个单元中固定的四个坐标位置, 并进行最大池化操作; 双三次插值是一种复杂的插值方法, 该方法利用待采样点周围确定的 16 个点的值进行三次插值, 不仅考虑到了相邻点的灰度值的影响还考虑到了相邻点的灰度值变化率的影响。虽然加大了运算量, 但是可以得到更接近高分辨率图像的效果, 使得小目标的信息丢失较少。计算双三次插值的方法是先构造 BiCubic (双三次) 函数, 从函数中获得周围 16 个像数点的权重值, 然后再根据这到目标点的距离和权重值计算出最终的插值结果。

[0084] 双三次插值公式为:

$$[0085] \quad W(x) = \begin{cases} (a+3)|x|^3 - (a+3)|x|^2 + 1 & |x| \leq 1 \\ a|x|^3 - 5a|x|^2 + 8a|x| - 4a & 1 < |x| < 2 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

[0086] 
$$B(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} * W(i) * W(j) \quad (12)$$

[0087] 式(11)和(12)中,a取值为-0.5,W(i)表示横坐标权重,W(j)表示纵坐标权重,x代表某个像素点的像素值, $a_{ij}$ 代表第i行j列的像素点。

#### [0088] 四、对比例1

[0089] 本对比例采用现有技术的Faster Rcn在(一)测试平台和环境基础上进行小目标检测识别。

[0090] Faster Rcn的步骤为:

[0091] (1) 输入测试图像;

[0092] (2) 将整张图片输入CNN,进行特征提取;

[0093] (3) 用RPN先生成一堆Anchor box,对其进行裁剪过滤后通过softmax判断anchors属于前景(foreground)或者背景(background),即是物体or不是物体,所以这是一个二分类;同时,另一分支bounding box regression修正anchor box,形成较精确的proposal(注:这里的较精确是相对于后面全连接层的再一次box regression而言)

[0094] (4) 把建议窗口映射到CNN的最后一层卷积feature map上;

[0095] (5) 通过RoI pooling层使每个RoI生成固定尺寸的feature map;

[0096] (6) 利用Softmax Loss(探测分类概率)和Smooth L1 Loss(探测边框回归)对分类概率和边框回归(Bounding box regression)联合训练。

#### [0097] 五、对比例2

[0098] 本对比例采用现有技术中的YOLOv3方法在(一)测试平台和环境基础上进行小目标检测识别。

[0099] YOLOv3的原理图如图2所示,其属于现有技术,本对比例不再进行详细描述。

#### [0100] 六、实验对比结果分析

[0101] 实施例1、对比例1、对比例2在(一)测试平台和环境、(二)实验细节及评价标准的规则下,对BCCD血细胞数据集进行训练、目标检测识别和结果评价,目标检测识别输出结果如图3所示,评价结果如下表1所示:

[0102] 表1:实施例1、对比例1、对比例2的评价结果

[0103]

	白细胞(%)	红细胞(%)	血小板(%)	mAP(%)
实施例1	99	80	96	91.86
对比例1	99	75	81	85.23
对比例2	95	79	69	80.91

[0104] 由表1可见,本申请方法对属于小目标的血小板的检测精度最高,并且申请方法的总体精度达到了91.86%,远远高于现有的YOLOv3方法和Faster Rcn方法;本申请的方法比现有的Faster Rcn方法平均精度提升了7%,小目标的检测提升了15%,并且速度还没有降低,说明本申请的方法在小目标的检查上有很好的鲁棒性。

[0105] 此外,在复杂密集的背景下检测目标时,不管是对大目标还是小目标,都会出现大量的漏检现象,效果很不好。图3中显示了实施例1、对比例1和2在细胞数据集中的测试输出结果,图3(a)为实施例1的小目标检测识别结果,可以看出该方法明显消除了漏检问题,对很模糊的小目标也能有很好的识别精度。图3(b)为对比例1的小目标检测识别结果,从中可

以看出有不少漏检现象。图3(c)为对比例2的小目标检测识别结果,从中可以看出该方法存在大量的漏检现象,在小目标的检测上效果稍微好一点,但是也是牺牲了大目标的检测精度换来的,只能检测出很小的一部分,而且检测的置信度也不高。

[0106] 七、普适性验证

[0107] 为了验证本申请的技术手段在其他数据集识别中的普适性,实施例1、对比例1、对比例2分别在一个新的遥感数据集上进行了对比实验。新的数据集是NWPU VHR-10,是由西北工业大学标注的航天遥感小目标检测数据集,该数据集有飞机、舰船、油罐、棒球场、网球场、篮球场、田径场、港口、桥梁、车辆10个类别,大部分图片都是小目标。实施例1、对比例1、对比例2的评价结果如表2所示。

[0108] 表2:实施例1、对比例1、对比例2的评价结果

[0109]		F1 (%)	Recall (%)	mAP (%)
	实施例1	79.11	91.2	89.21
	对比例1	69.6	82.6	81.62
	对比例2	74.8	71.04	79.56

[0110] 从表2来看,本申请的精确明显高于对比例1和对比例2的方法,该对比实验验证了本申请的方法并不局限于专门的数据集,对其他数据集具有很好的普适性,可以用于大多数的小目标检测算法中。

[0111] 以上对本发明进行了示例性描述,只要是采用了本发明的方法构思和技术方案进行的各种非实质性的改进;或未经改进,将本发明的上述构思和技术方案直接应用于其它场合的,均在本发明的保护范围之内。

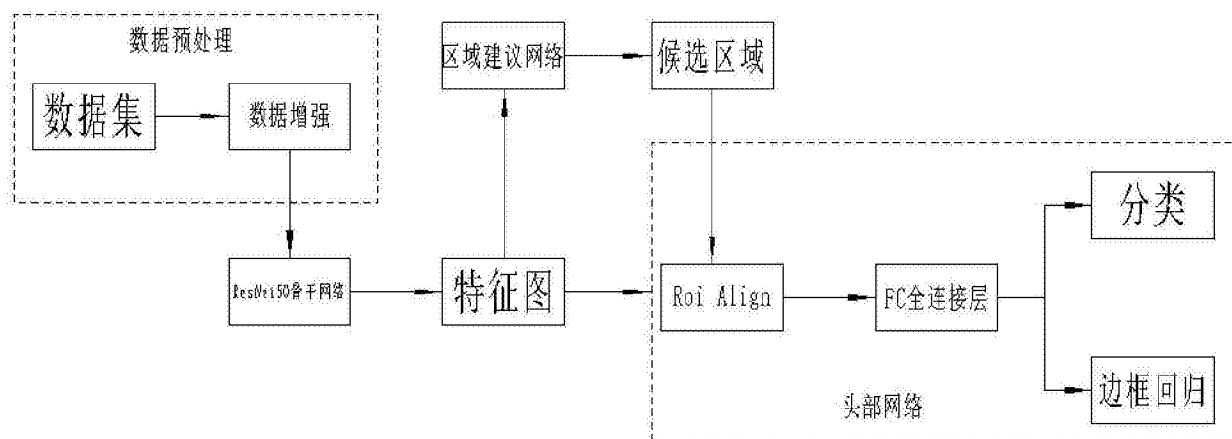


图1

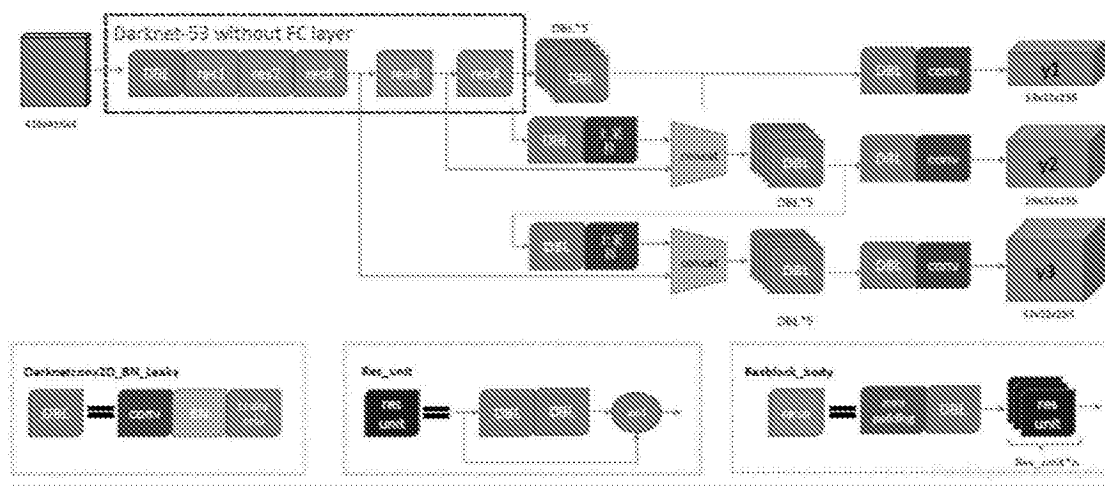


图2

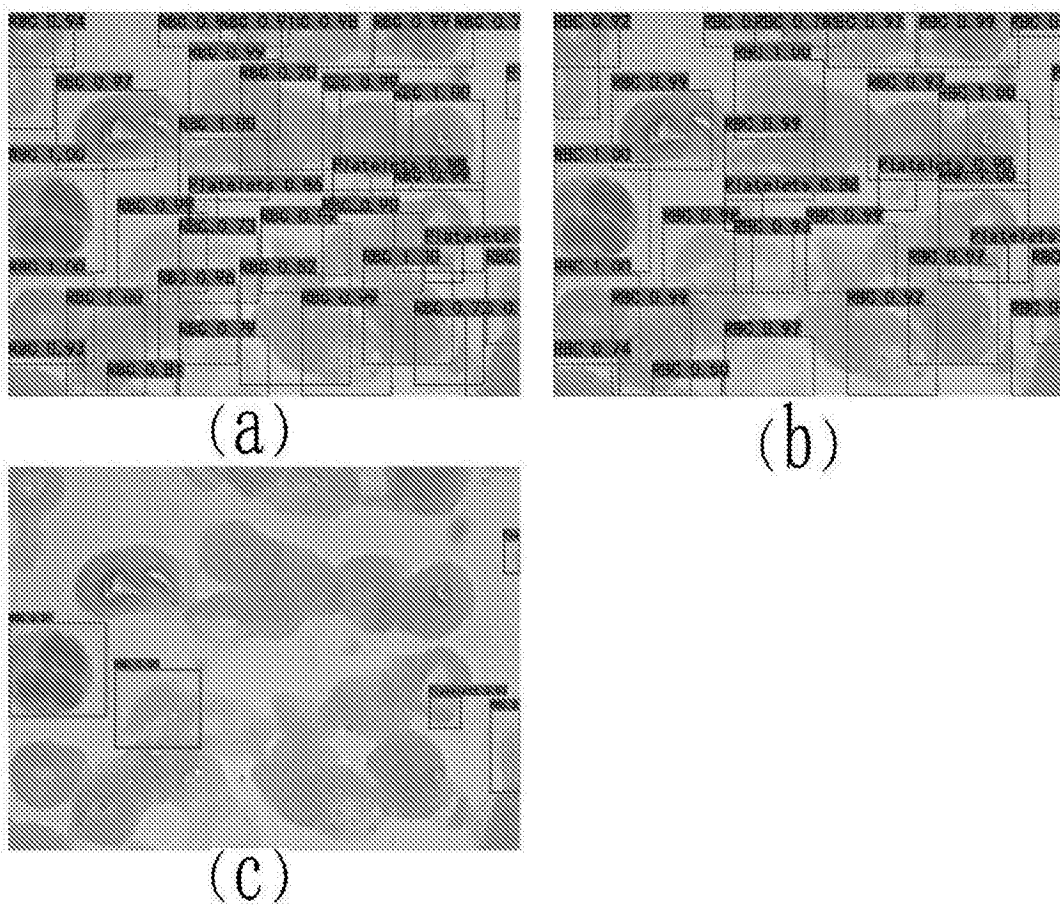


图3