(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 113065558 A
(43)申请公布日 2021.07.02

(21)申请号 202110432768.6
(22)申请日 2021.04.21
(71)申请人 浙江工业大学
地址 310014 浙江省杭州市下城区潮王路18号
(72)发明人 朱威 王立凯 靳作宝 何德峰 郑雅羽
(74)专利代理机构 杭州赛科专利代理事务所(普通合伙) 33230
代理人 吴琰
(51)Int.Cl.
G06K 9/32(2006.01)
G06K 9/46(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7页 附图5页

(54)发明名称

一种结合注意力机制的轻量级小目标检测方法

(57)摘要

本发明涉及一种结合注意力机制的轻量级小目标检测方法，包括以下步骤：(1)搭建基于YOLOv4的小目标检测网络：构建MSE多尺度注意力模块插入到特征提取网络，同时添加浅层特征图作为预测层，以及SPP模块的改进，增强特征的提取能力；(2)构建小目标数据集，使用数据增强策略对训练集数据进行增强，对锚框进行自定义(3)对模型进行通道剪枝，同时采用知识蒸馏恢复模型精度；(4)输入一幅无人机航拍图像，获取目标分类和定位结果。本发明利用通道注意力机制和模型压缩策略，能够在有效改善小目标错检漏检现象的同时保证模型的实时性。

CN 113065558 A

1.一种结合注意力机制的轻量级小目标检测方法，其特征在于：所述方法包括以下步骤：

(1)搭建基于YOLOv4改进的小目标检测网络；

(2)训练并优化小目标检测网络；

(3)对小目标检测网络进行模型轻量化；

(4)利用已训练的小目标检测网络模型对输入图像进行检测。

2.根据权利要求1所述的一种结合注意力机制的轻量级小目标检测方法，其特征在于：所述步骤(1)包括以下步骤：

(1-1)构建MSE多尺度注意力机制模块，插入到特征提取网络；

(1-2)添加浅层特征图作为预测层；

(1-3)SPP模块改进。

3.根据权利要求2所述的一种结合注意力机制的轻量级小目标检测方法，其特征在于：所述步骤(1-1)包括以下步骤：构建MSE多尺度注意力机制模块，插入到YOLOv4特征提取网络CSPDarknet53的每个CSP模块中Concat层和CBM模块之间，组成新的MSE-CSPUnit模块，得到带有注意力信息的MSE-CSPDarknet53的特征提取网络。

4.根据权利要求2或3所述的一种结合注意力机制的轻量级小目标检测方法，其特征在于：所述步骤(1-1)在SE注意力机制模块基础上构建MSE多尺度注意力机制模块，包括以下步骤：

(1-1-1)将CSP模块的Concat层的输出作为输入特征X，通过不同尺寸的卷积核集成多种尺度的特征图，得到多尺度融合特征输出X_c；卷积核尺寸分别为3×3、5×5、7×7，$X_c=V_{3\times3}X+V_{5\times5}X+V_{7\times7}X$，其中，V代表使用不同尺寸卷积核的卷积操作；

(1-1-2)对X_c进行挤压操作，使用全局平均池化和全局最大池化分别对通道进行挤压得到通道级的特征信息，其中全局平均池化注重全局特征，全局最大池注重化局部特征，

$$X_{avg}=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}X_c(i,j)$$；

$X_{max}=\max(X_c(i,j))$；

其中，X_{avg}为全局平均池化后获取的特征，X_{max}为全局最大池化后获取的特征，i＝1，2，…，H，j＝1，2，…，W，H、W分别为输入高度、输入宽度；

(1-1-3)分别对X_{avg}和X_{max}进行激励操作，并相加、经过归一化操作生成注意力权重信息X_s，FC_1、FC_2为两个不同的全连接层，其中$FC_1\in R^{\frac{C}{r}\times C}$，$FC_2\in R^{C\times\frac{C}{r}}$，C为输入通道，r为降维比例，$FC_1$起到降维的作用，以减少全连接层参数，$FC_2$起恢复原始维度的作用；

$X_a=FC_2(\mathrm{Mish}(FC_1(X_{avg}))$

$X_m=FC_2(\mathrm{Mish}(FC_1(X_{max}))$

$X_s=\mathrm{Softmax}(X_a+X_m)$

其中，Mish为非线性激活函数，Softmax为归一化函数；

(1-1-4)将(1-1-3)生成的X_s与(1-1-1)生成的X_c进行加权操作，得到MSE多尺度注意力模块的输出X_{weight}，$X_{weight}=\mathrm{Scale}(X_c,X_s)$，将$X_{weight}$作为MSE-CSPUnit模块中CBM模块的输入。

5.根据权利要求2所述的一种结合注意力机制的轻量级小目标检测方法，其特征在于：所述步骤(1-2)中，删除FPN和PAN结构输出的19×19大小的特征图，保留FPN和PAN结构原有38×38、76×76的输出特征图；使用FPN和PAN结构将MSE-CSPUnit*2的输出和下方深层特征图上采样的结果进行特征融合，获得152×152大小的浅层特征图；最后得到38×38、76×76、152×152三个不同大小的特征图对不同尺度的目标进行预测。

6.根据权利要求6所述的一种结合注意力机制的轻量级小目标检测方法，其特征在于：所述步骤(1-3)中，在FPN和PAN结构和对应的三个预测层间分别放置SPP模块，SPP模块将输入特征图进行1×1、5×5、9×9、13×13的最大池化操作后，再将生成的不同尺度的特征图进行张量拼接。

7.根据权利要求1所述的一种结合注意力机制的轻量级小目标检测方法，其特征在于：所述步骤(2)包括以下步骤：

(2-1)构建小目标数据集；

(2-2)数据增强，并对图片数据进行多模式随机调整；

(2-3)设置锚框，用于拟合数据集中的目标。

8.根据权利要求1所述的一种结合注意力机制的轻量级小目标检测方法，其特征在于：所述步骤(3)包括以下步骤：

(3-1)通道剪枝

选用BN层的γ作为缩放因子，在损失函数中添加关于BN层的γ的L1正则化项，对网络进行预设轮数次的稀疏化训练后，基于梯度更新后的γ值，对除了上采样层前的卷积层、SPP模块之外的层进行通道剪枝，得到通道剪枝后的模型文件和模型结构配置文件；

(3-2)知识蒸馏恢复网络精度

以未进行剪枝的YOLOv4网络作为教师网络，通道剪枝后的网络作为学生网络；分别计算教师网络和标签值、学生网络和标签值的L2损失，设置偏差范围，当学生网络和标签值的L2损失与教师网络和标签值的L2损失的偏差超过范围w时，在总损失中计入学生网络的L2损失，整体损失函数为

$$L_b(R_s,R_t,y_{reg})=\begin{cases}\left\|R_s-y_{reg}\right\|_2^2, & if\ \left\|R_t-y_{reg}\right\|_2^2-\left\|R_s-y_{reg}\right\|_2^2<w\\ 0, & otherwise\end{cases}$$

$$L_{reg}=(1-v)L_{sL1}(R_s,y_{reg})+vL_b(R_s,R_t,y_{reg})$$

其中，w为预设的偏差范围，y_{reg}是标签值，R_t和R_s分别是教师网络和学生网络的回归输出，L_b为模型蒸馏部分损失，L_{sL1}为学生网络回归输出与标签值的之间的损失，v是L_b和L_{sL1}之间的平衡因子，在网络训练前80％的时间设置在0.1～0.5之间，后20％的训练时间设置在0.6～0.9之间；L_{reg}为网络蒸馏学习时的总损失。

9.根据权利要求1所述的一种结合注意力机制的轻量级小目标检测方法，其特征在于：所述步骤(4)包括以下步骤：

(4-1)输入一帧图像；

(4-2)在读取完一幅图像后，送入训练并优化完成的小目标检测网络中进行目标的定位和分类；将图像输入至带有注意力机制的特征提取网络进行特征的提取，经过SPP模块分别输出3个不同分辨率大小的特征图，对3个特征图进行三种不同尺度目标的检测，设置置

信阈值为0.2～0.6,经过阈值过滤之后,获得目标的分类和定位结果;

(4-3)重复步骤(4-1)至步骤(4-2),直至完成测试集中图片的检测。

一种结合注意力机制的轻量级小目标检测方法

技术领域

[0001] 本发明属于深度学习技术在机器视觉领域的应用，具体涉及一种结合注意力机制的轻量级小目标检测方法。

背景技术

[0002] 目标检测在给定图像中找出特定目标类别及其准确位置，其中小目标检测是目标检测领域的重要研究内容，在遥感影像目标识别、红外成像目标识别、农业病虫害识别等场景都有着重要的应用价值。在目标检测中，通常将目标像素值占整个图像0.12％以下的或者像素值小于32*32的目标称为小目标。由于小尺寸物体的分辨率低和噪声大，往往在多层卷积后提取到的特征不明显，因此检测图像中的小目标是非常困难的。

[0003] 早期的小目标检测主要是通过手工设计的方法来获取目标的特征信息。温佩芝等人将小波变换应用在小目标检测过程中(见温佩芝，史泽林，于海斌，吴晓军.基于小波变换的海面背景红外小目标检测方法[J].光电工程，2004)，利用正交小波分解的多分辨率分析实现频带选择，抑制噪声和背景的干扰，并利用不同方向边缘进行融合，获得候选点，最后根据灰度阈值排除干扰目标。CHEN等人(见C.L.P.Chen,H.Li,Y.Wei,et al.A Local Contrast Method for Small Infrared Target Detection[J]//IEEE Transactions on Geoscience and Remote Sensing,2014,52(1):574-581)受生物视觉机制的启发，使用建议的局部对比度度量获取输入图像的局部对比度图，该度量可表示当前位置及其邻域之间的差异，这样可以同时实现目标信号增强和背景杂波抑制，最后通过自适应阈值分割目标。上述方法从图像的底层特征出发，使用基本图像特征来实现检测任务，操作较简单，但对于复杂背景的小目标检测来说还存在漏检错检以及实时性上的问题。

[0004] 近年来，随着计算机算力的提升和深度学习理论的快速发展，深度学习技术已被广泛用于目标检测。目前流行的目标检测模型大致可以分为两类：一阶段检测算法，分类和定位看作回归任务，代表性的算法有SSD和YOLO；二阶段检测算法，候选框选取和目标分类分离，代表性的算法有R-CNN和Faster R-CNN。其中由于一阶段检测算法将整个检测任务看作是回归操作，因此在实时性方面占有很大优势。

[0005] 利用深度学习技术来提高小目标检测的主要方式有多尺度表示、上下文信息、超分辨率等方法。申请号为CN202010537199.7的专利公开了一种用于图片小目标的检测方法。从待检测图片中获取六个不同尺寸的特征图，采用双线性插值法将六个不同尺寸特征图中的金字塔底层特征图与金字塔高层特征图进行特征融合，得到新的六个不同尺寸的特征图，以新的六个不同尺寸的特征图参与预测。该方法采用多尺度特征图增强了目标特征信息，但易受复杂背景的干扰，误检率较高。申请号为CN202010444356.X的专利公开了一种基于分辨率增强的遥感图像小目标检测方法，对包含小目标的遥感图像进行超分辨处理后再进行目标检测，解决了由于遥感图像中小目标可利用的特征信息少以及小目标区域存在几何形变，采用超分辨处理技术进一步完善小目标的细节特征信息，应用基于区域的可变形卷积网络充分利用了小目标有限的特征信息，提高了对遥感图像中小目标的检测能力。

该方法虽然有较好的准确性，但由于图片分辨率增大导致网络实时性下降，不利于网络的轻量化。

发明内容

[0006] 为了解决现有目标检测方法对于小目标检测存在误检率较高、漏检、实时性差等问题，本发明提供了一种结合注意力机制的轻量级小目标检测方法，所述方法包括以下步骤：

[0007] (1)搭建基于YOLOv4改进的小目标检测网络

[0008] 本发明的小目标检测网络是在一阶段目标检测网络YOLOv4的基础上改进得到，具体的网络结构改进包括以下三个方面：

[0009] (1-1)构建MSE多尺度注意力机制模块，插入到特征提取网络

[0010] 本发明构建的MSE多尺度注意力机制模块是对SE注意力模块进行改进得到的，SE注意力模块是由Hu等人在2017年提出的一种用于计算机视觉领域的轻量级注意力机制模块，它可以方便的插入到特征提取网络的两个网络层之间，通过学习全局信息来选择和强调感兴趣的特征通道，并抑制无关的干扰信息。

[0011] 通过构建MSE多尺度注意力机制模块，插入到YOLOv4特征提取网络CSPDarknet53的每个CSP模块中Concat层和CBM模块之间，组成新的MSE-CSPUnit模块，得到带有注意力信息的MSE-CSPDarknet53的特征提取网络。其中MSE多尺度注意力机制模块的构建具体步骤如下：

[0012] (1-1-1)首先将CSP模块的Concat层的输出作为输入特征图，通过不同尺寸的卷积核集成多种尺度的特征图，并基于多尺度特征图进行下一步的特征提取操作。卷积核尺寸分别为3×3、5×5、7×7，对于使用大尺寸卷积核导致的参数量暴增的情况，使用2层3×3的卷积核代替5×5的卷积核，3层3×3的卷积核代替7×7的卷积核。设输入特征图$X\in R^{C\times H\times W}$，C、H、W分别为输入通道、输入高度、输入宽度，则对于输入特征图使用不同尺寸卷积核进行特征提取的过程如下式：

[0013] $X_c=V_{3\times3}X+V_{5\times5}X+V_{7\times7}X$

[0014] 其中，X_c为多尺度特征图输出，V代表使用不同尺寸卷积核的卷积操作。

[0015] (1-1-2)对X_c进行挤压操作，使用全局平均池化和全局最大池化分别对通道进行挤压得到通道级的特征信息，其中全局平均池化注重特征图的全局特征，全局最大池化注重特征图的局部特征：

[0016] $$X_{avg}=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}X_c(i,j)$$

[0017] $X_{max}=\max(X_c(i,j))$

[0018] 其中，X_c为输入的多尺度特征，X_{avg}为全局平均池化后获取的特征，X_{max}为全局最大池化后获取的特征，i=1,2,…,H,j=1,2,…,W,H、W分别为输入高度、输入宽度。

[0019] (1-1-3)分别对X_{avg}和X_{max}进行激励操作，并相加、经过归一化操作生成通道注意力权重信息X_s。在进行激励操作时，使用Mish激活函数保留通道之间的更多的非线性关系，FC_1、FC_2为两个不同的全连接层，其中$FC_1\in R^{\frac{C}{r}\times C}$，$FC_2\in R^{C\times\frac{C}{r}}$，C为输入通道，r为降维比

例，FC_1起到降维的作用，以减少全连接层参数，FC_2起恢复原始维度的作用。激活和归一化操作如下式：

[0020] $X_a = FC_2(Mish(FC_1(X_{avg}))$

[0021] $X_m = FC_2(Mish(FC_1(X_{max}))$

[0022] $X_s = Softmax(X_a + X_m)$

[0023] 其中，Mish为非线性激活函数，Softmax为归一化函数。

[0024] (1-1-4)将(1-1-3)生成的通道注意力权重信息与(1-1-1)生成的多尺度特征图进行加权操作，得到MSE多尺度注意力模块的输出X_{weight}，将X_{weight}作为MSE-CSPUnit模块中CBM模块的输入。

[0025] $X_{weight} = Scale(X_c, X_s)$

[0026] (1-2)添加浅层特征图作为预测层

[0027] 深层特征具有更强的语义信息，更适合定位；而浅层特征有丰富的分辨率信息，更利于小目标的检测。删除FPN和PAN结构输出的19×19大小的特征图，保留FPN和PAN结构原有38×38、76×76的输出特征图；使用FPN和PAN结构将MSE-CSPUnit*2的输出和下方深层特征图上采样的结果进行特征融合，获得152×152大小的浅层特征图；最后得到38×38、76×76、152×152三个不同大小的特征图对不同尺度的目标进行预测。

[0028] 这里MSE-CSPUnit*2是指两个MSE-CSPUnit模块。

[0029] (1-3)SPP模块改进

[0030] SPP模块可以丰富特征图的表达能力，提供重要的上下文信息。为了提高小目标检测时的性能，因此在38×38、76×76、152×152特征图前分别放置SPP模块，实现局部特征和全局特征的有效融合。SPP模块将输入特征图进行1×1、5×5、9×9、13×13的最大池化操作后，再将生成的不同尺度的特征图进行张量拼接。

[0031] (2)训练并优化小目标检测网络

[0032] 针对具体应用场景，构建小目标检测数据集，通过数据增强，对图片数据进行多模式随机调整对，包括对数据中小目标个数、图片亮度、对比度、饱和度进行随机调整，增强模型的泛化性能。

[0033] 最后设置锚框，用于拟合数据集中的目标；通过Kmeans++算法对目标数据集的锚框重新聚类，得到更适合当前数据集的锚框参数，加快网络的收敛速度。

[0034] (3)对小目标检测网络进行模型轻量化

[0035] (3-1)通道剪枝

[0036] 针对网络的参数冗余，对小目标检测网络进行通道剪枝。使用YOLOv4的卷积模块BN层的γ作为缩放因子，在损失函数中添加关于BN层的γ的L1正则化项，对网络进行预设轮数次的稀疏化训练后，基于梯度更新后的γ值，对γ进行排序，通过设置剪枝阈值，将小于剪枝阈值的γ所在的通道移除，得到剪枝后的轻量级YOLOv4网络。在YOLOv4网络中，除了上采样层前的卷积层和SPP结构，对其它含有BN层的卷积模块进行通道剪枝，得到通道剪枝后的模型文件和模型结构配置文件。对于YOLOv4稀疏训练时，建立的目标损失函数为：

[0037] $$L_{loss} = \sum_{(x,y)} l(f(x,W),y) + \lambda \sum_{\gamma \in \Gamma} g(\gamma)$$

[0038] 其中，x为模型的输入值，y为期望输出值，w为网络中可训练的参数，g(.)为缩放因

子的惩罚项，λ为平衡因子。

[0039] (3-2)知识蒸馏恢复模型精度

[0040] 经过通道剪枝后，虽然移除的通道对于模型输出的贡献微小，但是剪枝后的模型精度还是会有小幅度的下降，因此要恢复模型精度。

[0041] 使用未进行剪枝的YOLOv4网络作为教师网络，通道剪枝后的网络作为学生网络，进行知识蒸馏。YOLOv4的知识蒸馏将进行分类任务和回归任务的学习，对于回归结果的蒸馏，由于回归的输出是无界的，且教师网络的预测结果可能与标签值相反，因此在计算回归损失时，不是直接向教师网络学习。首先分别计算教师网络和标签值、学生网络和标签值的L2损失，设置一个范围w，当学生网络和标签值的L2损失与教师网络和标签值L2损失的偏差超过范围w时，才会在损失中计入学生网络的L2损失。即当学生网络的性能超过教师网络一定的值时，就不计算学生网络的损失。其整体损失函数为：

[0042] $$L_b(R_s,R_t,y_{reg})=\begin{cases}\left\|R_s-y_{reg}\right\|_2^2, & if\ \left\|R_t-y_{reg}\right\|_2^2-\left\|R_s-y_{reg}\right\|_2^2<w\\ 0, & otherwise\end{cases}$$

[0043] $L_{reg}=(1-v)L_{sL1}(R_s,y_{reg})+vL_b(R_s,R_t,y_{reg})$

[0044] 其中，w为预设的偏差范围，y_{reg}是真实标签值，R_t和R_s分别是教师和学生的回归输出，L_b为模型蒸馏部分损失，L_{sL1}为学生网络与真实标签的损失，v是L_b和L_{sL1}之间的平衡因子，在网络训练前80％的时间设置在0.1～0.5之间，后20％的训练时间设置在0.6～0.9之间；L_{reg}为网络蒸馏学习时的总损失。

[0045] (4)利用已训练的小目标检测网络模型对输入图像进行检测

[0046] 输入一帧无人机航拍图像，送入训练并优化完成的小目标检测网络中进行目标的定位和分类。该网络首先将图像输入至带有注意力机制的特征提取网络进行特征的提取，经过SPP模块分别输出3个不同分辨率大小的特征图。使用回归和分类思想对3个特征图进行三种不同尺度目标的检测，经过置信阈值过滤之后，获得目标的分类和定位结果；重复直至完成测试集中图片的检测。

[0047] 与现有技术相比，本发明具有以下有益效果：

[0048] 本发明在端到端的卷积神经网络YOLOv4上进行改进实现轻量级小目标检测网络，同传统的小目标检测方法相比，基于SE设计了MSE注意力模块，并将所设计的注意力模块插入到YOLOv4特征提取网络中，增强网络对于感兴趣区域的关注能力，减少在小目标检测过程中复杂背景的干扰；然后添加浅层特征图作为预测层，使用38×38、76×76、152×152三个不同大小的特征图对不同尺度的目标进行预测；对SPP模块改进，在38×38、76×76、152×152特征图前分别放置SPP模块，实现局部特征和全局特征的有效融合；最后使用通道剪枝和知识蒸馏策略对模型进行压缩优化，在极少精度损失下，实现了模型参数数量的大幅压缩；此外，使用数据增强方式，对数据集中小目标个数、图片亮度、对比度、饱和度进行随机调整，增强模型训练效果。在小目标数据集中，本发明网络具有较好的检测效果和鲁棒性，同时满足了轻量化模型部署的要求。

附图说明

[0049] 图1为本发明的流程图；

[0050] 图2为加入MSE多尺度注意力机制模块后的MSE-CSPUnit模块；

[0051] 图3为本发明的MSE多尺度注意力模块结构；

[0052] 图4为本发明设计的小目标检测网络结构；

[0053] 图5为模型压缩后通道数量的对比，其中深色柱为未剪枝前，浅色柱为剪枝后；

[0054] 图6为本发明小目标检测网络对目标图片的检测效果图，其中(a)、(c)为改进前检测效果，(b)、(d)为对应(a)、(c)的改进后的检测效果。

具体实施方式

[0055] 下面结合实施例和附图来详细描述本发明，但本发明并不仅限于此。本发明目标检测的实施例对象为数据集中各类小目标，本发明选用的处理平台为Intel i9-9900k、NVIDIA RTX2080ti和32G RAM的组合，操作系统为Linux64 Ubuntu18.04。本发明方法选择在深度学习框架Pytorch1.6上进行实现。

[0056] 如图1所示引入注意力机制的轻量级小目标检测方法，包括四个部分：

[0057] (1)搭建基于YOLOv4改进的小目标检测网络；

[0058] (2)训练并优化所述小目标检测网络；

[0059] (3)对小目标检测网络进行模型轻量化；

[0060] (4)利用已训练的小目标检测网络模型对输入图像进行检测。

[0061] 第一部分搭建基于YOLOv4改进的小目标检测网络具体包括：

[0062] (1-1)设计MSE多尺度注意力机制模块，嵌入到特征提取网络

[0063] 构建MSE多尺度注意力机制模块，插入到YOLOv4特征提取网络CSPDarknet53的每个CSP模块中Concat层和CBM模块之间，组成新的MSE-CSPUnit模块，得到带有注意力信息的MSE-CSPDarknet53的特征提取网络，如图2所示，除MSE外的其余模块为YOLOv4特征提取网络CSPDarknet53的常规结构模块。MSE多尺度注意力机制模块的构建过程如下：

[0064] 首先将CSP模块的Concat层的输出作为输入特征图，通过不同尺寸的卷积核集成多种尺度的特征图，并基于多尺度特征图进行下一步的特征提取操作，其中卷积核尺寸分别为3×3、5×5、7×7。对于使用大尺寸卷积核导致的参数量暴增的情况，使用2层3×3的卷积核代替5×5的卷积核，3层3×3的卷积核代替7×7的卷积核。设输入特征图$X\in R^{C\times H\times W}$，C、H、W分别为输入通道、输入高度、输入宽度，则对于输入特征图使用不同尺寸卷积核进行特征提取的过程如下式：

[0065] $X_c=V_{3\times3}X+V_{5\times5}X+V_{7\times7}X$

[0066] 其中，X_c为多尺度融合特征输出，V代表使用不同尺寸卷积核的卷积操作。

[0067] 对X_c进行挤压操作，针对小目标特征信息偏少的特点，使用全局最大池化操作注重特征图的局部信息，同时使用全局平均池化操作着重于特征图的全局特征，池化操作如下式：

[0068] $$X_{avg}=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}X_c(i,j)$$

[0069] $X_{max}=\max(X_c(i,j))$

[0070] 其中，X_{avg}为全局平均池化后获取的特征，X_{max}为全局最大池化后获取的特征，i=1,2,…,H,j=1,2,…,W,H、W分别为输入高度、输入宽度。

[0071] 分别对X_{avg}和X_{max}进行激励操作，并相加、经过归一化操作生成注意力权重信息X_s。在进行激励操作时，使用Mish激活函数保留通道之间的更多的非线性关系。FC_1、FC_2为两个不同的全连接层，其中$FC_1 \in R^{\frac{C}{r} \times C}$，$FC_2 \in R^{C \times \frac{C}{r}}$，C为输入通道，r为降维比例，$FC_1$起到降维的作用，以减少全连接层参数，$FC_2$起恢复原始维度的作用。激活和归一化操作如下式：

[0072] $X_a = FC_2(\mathrm{Mish}(FC_1(X_{avg}))$

[0073] $X_m = FC_2(\mathrm{Mish}(FC_1(X_{max}))$

[0074] $X_s = \mathrm{Softmax}(X_a + X_m)$

[0075] 其中，Mish为非线性激活函数，Softmax为归一化函数。

[0076] 将X_s与第一步生成的多尺度特征图X_c进行加权操作，得到MSE多尺度注意力模块的输出X_{weight}，将X_{weight}作为MSE-CSPUnit模块中CBM模块的输入。

[0077] $X_{weight} = \mathrm{Scale}(X_c, X_s)$

[0078] (1-2)在预测层中添加浅层特征

[0079] 深层特征具有更强的语义信息，更适合定位；而浅层特征有丰富的分辨率信息，更利于小目标的检测。删除FPN和PAN结构输出的19×19大小的特征图，保留FPN和PAN结构原有38×38、76×76的输出特征图；使用FPN和PAN结构将MSE-CSPUnit*2的输出和下方深层特征图上采样的结果进行特征融合，获得152×152大小的浅层特征图；最后得到38×38、76×76、152×152三个不同大小的特征图对不同尺度的目标进行预测。

[0080] (1-3)SPP模块改进

[0081] SPP模块可以丰富特征图的表达能力，提供重要的上下文信息。为了提高小目标检测时的性能，因此在38×38、76×76、152×152特征图前分别放置SPP模块，实现局部特征和全局特征的有效融合。SPP模块将输入特征图进行1×1、5×5、9×9、13×13的最大池化操作后，再将生成的不同尺度的特征图进行张量拼接。

[0082] 第二部分训练并优化所述小目标检测网络具体包括：

[0083] (2-1)构建数据集

[0084] 首先构建小目标数据集，实验选用了无人机航拍数据集VisDrone2019。VisDrone2019数据集由于是无人机航拍方式，因此包含有大量的小物体和密集对象，另外光照变化和对象遮挡也是这个数据集的难点。同时由于无人机图像是垂直拍摄的缘故，待检测对象包含特征较少。例如对于行人检测而言，地面拍摄的图像可能包含人体手臂、腿等特征，而对于无人机图像，则可能只有头顶这一处特征。

[0085] (2-2)数据增强，并对图片数据进行多模式随机调整

[0086] 网络训练时，对数据集采用在线增强的方式提高小目标的训练效果。由于数据集中包含小目标的图片可能比较少，导致会模型在训练的时候会偏向中等目标和大尺寸的目标。数据在线增强通过将小目标在图片中复制多份，通过人工增加小物体在图片中出现的次数，增加小目标被anchor包含的概率，让模型在训练的过程中，也能够有机会得到更多的小目标训练样本。同时对图片进行随机旋转和缩放，同时对亮度、对比度、饱和度进行调整，以增加模型的鲁棒性。

[0087] (2-3)自定义锚框，用于拟合数据集中的目标

[0088] 对于极端尺度对象的目标检测，合适的锚框可以更加准确的拟合数据集中的物

体。对于无人机航拍数据集，通过Kmeans++算法对目标数据集的锚框重新聚类，得到更适合当前数据集的锚框参数。通过Kmeans++算法，获得的锚框参数为(1,4)、(2,8)、(4,13)、(4,5)、(8,20)、(9,9)、(16,29)、(16,15)、(35,42)。

[0089] 第三部分小目标检测网络模型轻量化具体包括：

[0090] (3-1)通道剪枝

[0091] 针对网络的参数冗余，对小目标检测网络进行通道剪枝。使用YOLOv4的卷积模块BN层的γ作为缩放因子，在损失函数中添加关于BN层的γ的L1正则化项，对网络进行预设轮数次，如300轮次的稀疏化训练后，基于梯度更新后的γ值，对γ进行排序，通过设置剪枝阈值，将小于剪枝阈值的γ所在的通道移除，得到剪枝后的轻量级YOLOv4网络。在YOLOv4网络中，除了上采样层前的卷积层和SPP结构，对其它含有BN层的卷积模块进行通道剪枝。通过多次实验选取通道裁剪比例，以达到较好的速度与精度之间的平衡，最终选定裁剪比例为0.7，并得到通道剪枝后的模型文件和模型结构配置文件。

[0092] (3-2)知识蒸馏恢复模型精度

[0093] 经过通道剪枝后，虽然移除的通道对于模型输出的贡献微小，但是剪枝后的模型精度还是会有小幅度的下降，因此要恢复模型精度。

[0094] 使用未进行剪枝的YOLOv4网络作为教师网络，通道剪枝后的网络作为学生网络，进行知识蒸馏。YOLOv4的知识蒸馏将进行分类任务和回归任务的学习，对于回归结果的蒸馏，由于回归的输出是无界的，且教师网络的预测结果可能与真实值相反，因此在计算回归损失时，不是直接向教师网络学习。首先分别计算教师网络和标签值、学生网络和标签值的L2距离，通过多次实验对比，设置偏差范围w＝0.3，当学生网络和标签值的L2距离与教师网络和标签值的偏差超过范围w时，才会在损失中计入学生网络的L2损失。即当学生网络的性能超过教师网络一定的值时，就不计算学生网络的损失。其整体损失函数为：

[0095] $$L_b(R_s,R_t,y_{reg})=\begin{cases}\left\|R_s-y_{reg}\right\|_2^2, & if\ \left\|R_t-y_{reg}\right\|_2^2-\left\|R_s-y_{reg}\right\|_2^2<w\\ 0, & otherwise\end{cases}$$

[0096] $L_{reg}=(1-v)L_{sL1}(R_s,y_{reg})+vL_b(R_s,R_t,y_{reg})$

[0097] 其中，w为预设的偏差范围，y_{reg}是真实标签值，R_t和R_s分别是教师和学生的回归输出，L_b为模型蒸馏部分损失，L_{sL1}为学生网络与真实标签的损失，v是L_b和L_{sL1}之间的平衡因子，在网络训练前80％的时间设置在0.1～0.5之间，后20％的训练时间设置在0.6～0.9之间；L_{reg}为网络蒸馏学习时的总损失。

[0098] 第四部分检测图片小目标具体包括：

[0099] (4-1)输入一幅无人机航拍图像

[0100] (4-2)在读取完一幅无人机航拍图像后，送入训练并优化完成的小目标检测网络中进行目标的定位和分类。该网络首先将图像输入至带有注意力机制的特征提取网络进行特征的提取，经过SPP模块分别输出3个不同分辨率大小的特征图。使用回归和分类思想对3个特征图进行三种不同尺度目标的检测，置信阈值为0.2～0.6，一般设置置信阈值为0.3，经过阈值过滤之后，获得目标的分类和定位结果。

[0101] (4-3)重复步骤(4-1)至步骤(4-2)，直至完成测试集中图片的检测，各类小目标的检测效果如图6所示。

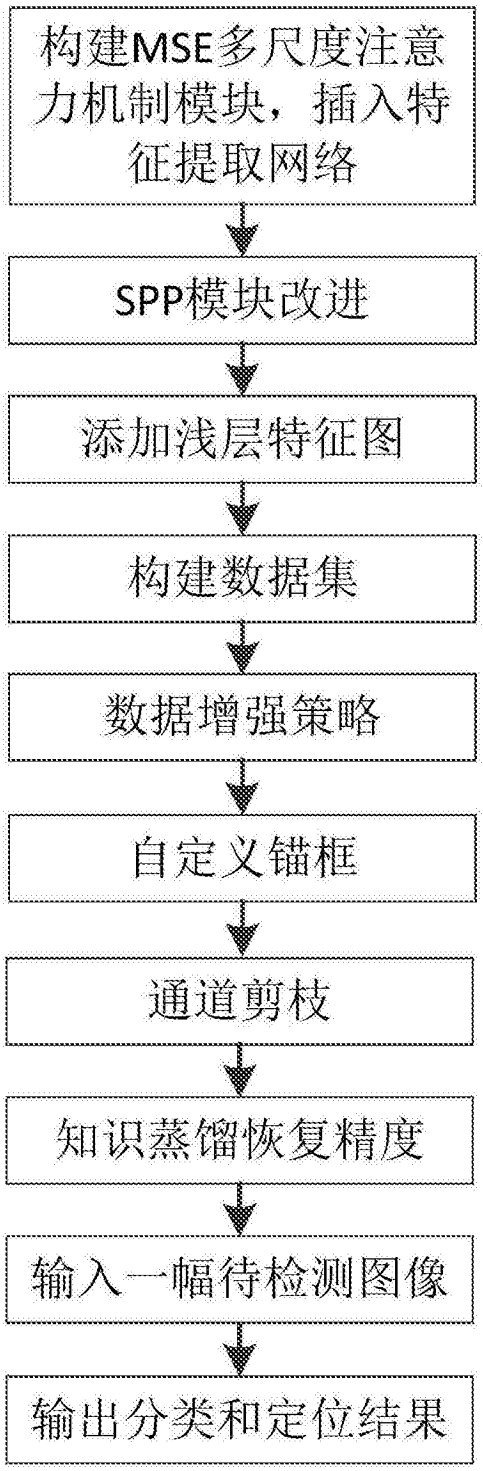
构建MSE多尺度注意力机制模块，插入特征提取网络
SPP模块改进
添加浅层特征图
构建数据集
数据增强策略
自定义锚框
通道剪枝
知识蒸馏恢复精度
输入一幅待检测图像
输出分类和定位结果

图1

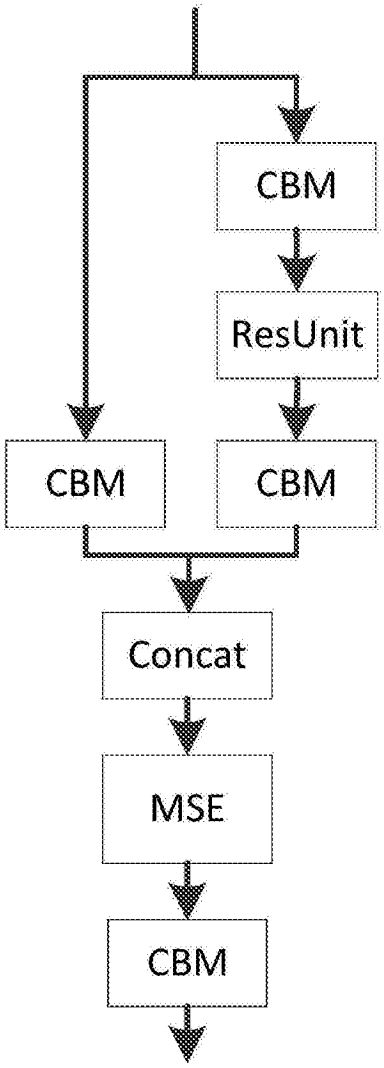

图2

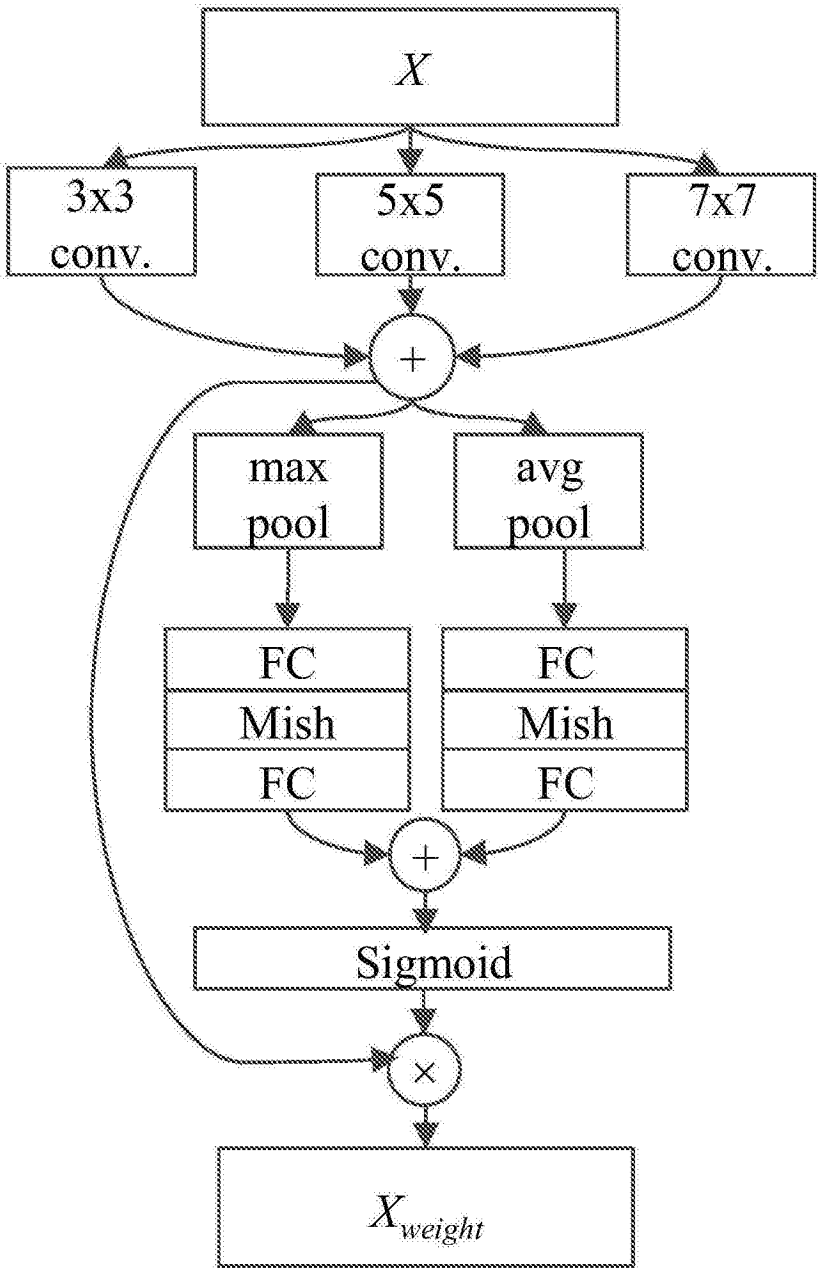
X
3x3 conv.
5x5 conv.
7x7 conv.
+
max pool
avg pool
FC
Mish
FC
FC
Mish
FC
+
Sigmoid
×
X_{weight}

图3

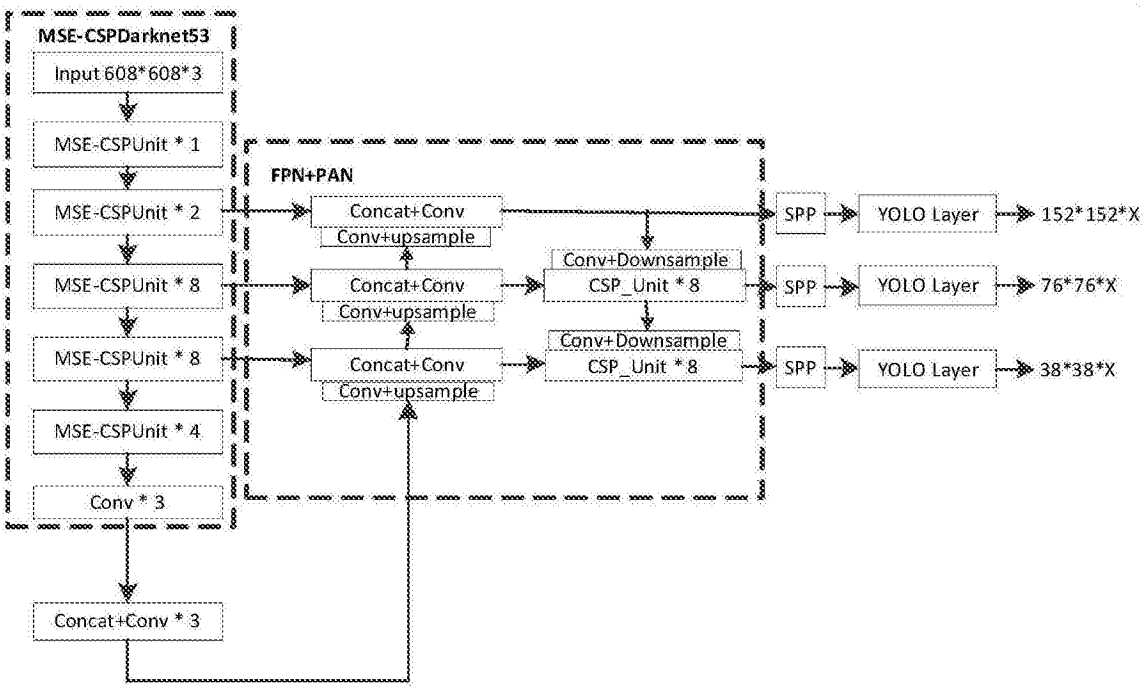
MSE-CSPDarknet53
Input 608*608*3
MSE-CSPUnit * 1
MSE-CSPUnit * 2
MSE-CSPUnit * 8
MSE-CSPUnit * 8
MSE-CSPUnit * 4
Conv * 3
Concat+Conv * 3
FPN+PAN
Concat+Conv
Conv+upsample
Concat+Conv
Conv+upsample
Concat+Conv
Conv+upsample
Conv+Downsample
CSP_Unit * 8
Conv+Downsample
CSP_Unit * 8
SPP
YOLO Layer
152*152*X
SPP
YOLO Layer
76*76*X
SPP
YOLO Layer
38*38*X

图4

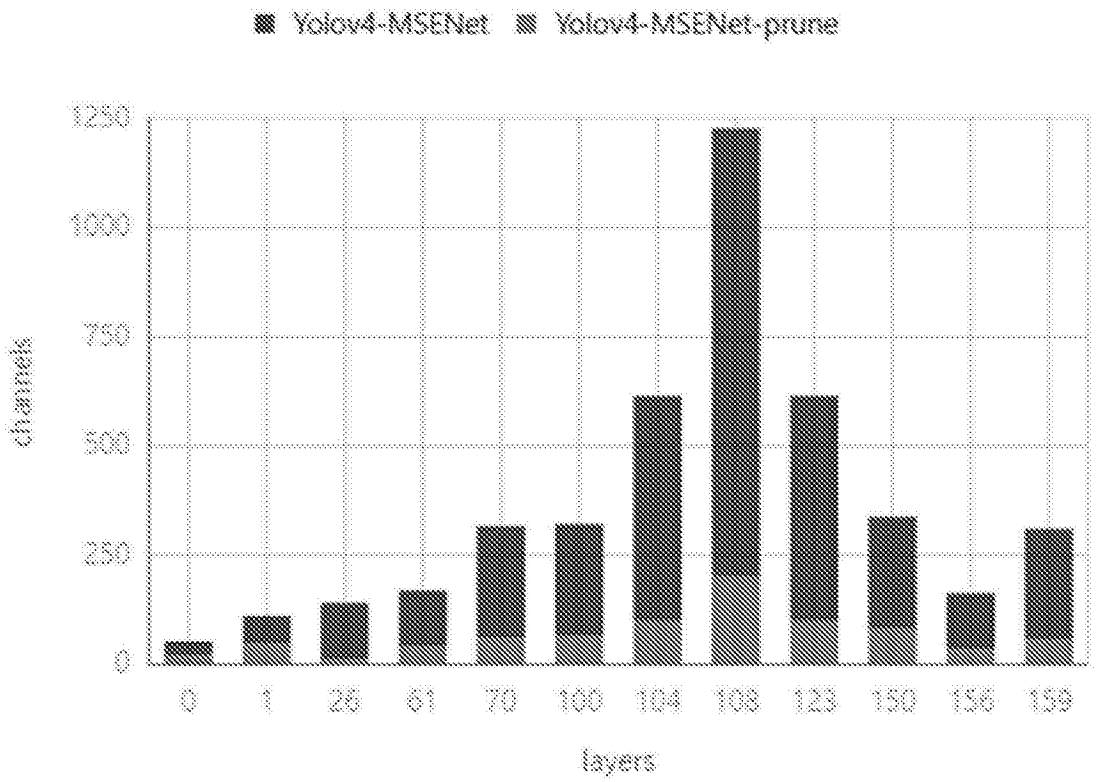
Yolov4-MSENet
Yolov4-MSENet-prune
channels
1250
1000
750
500
250
0
0
1
26
61
70
100
104
108
123
150
156
159
layers

图5

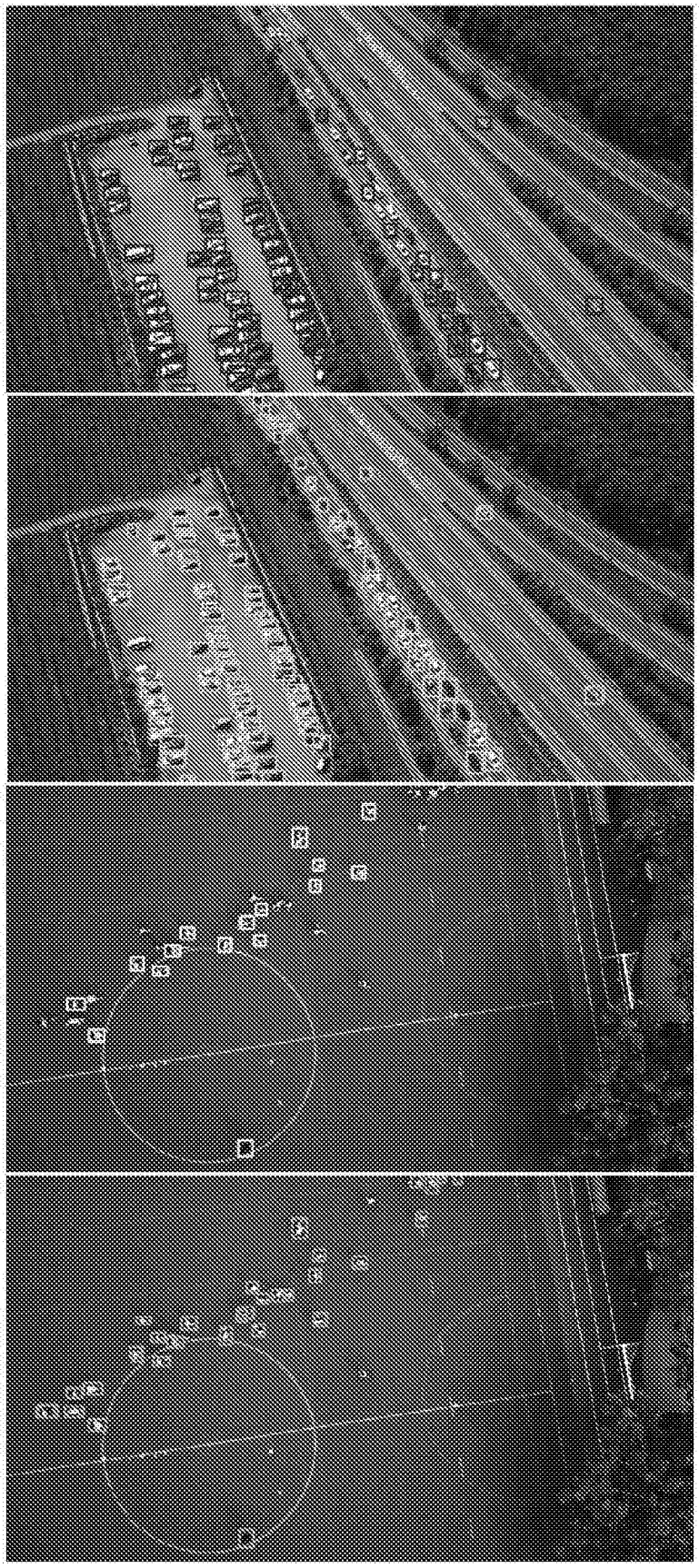

图6