(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 113326735 A
(43)申请公布日 2021.08.31

(21)申请号 202110475048.8

(22)申请日 2021.04.29

(71)申请人 南京大学
地址 210023 江苏省南京市栖霞区仙林大道163号
申请人 江苏万维艾斯网络智能产业创新中心有限公司

(72)发明人 霍静 孙宏伟 李文斌 高阳

(74)专利代理机构 南京泰普专利代理事务所(普通合伙) 32360
代理人 窦贤宇

(51)Int.Cl.
G06K 9/00(2006.01)
G06K 9/62(2006.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

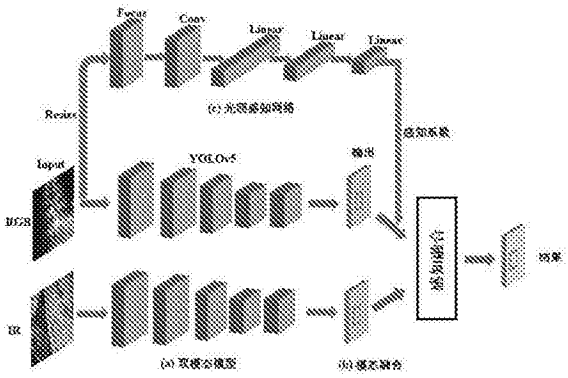
权利要求书2页 说明书3页 附图1页

(54)发明名称

一种基于YOLOv5的多模态小目标检测方法

(57)摘要

本发明公开了一种基于YOLOv5的多模态小目标检测方法，本方法主要解决联合使用红外图像和可见光图像进行目标检测的问题，该方法主要包括：构建一个轻量化的光照感知网络，使用其实现对可见光模态图像感知系数的计算；基于设计的光照感知网络，将红外模态和可见光模态数据在YOLOv5架构下进行多模态融合。本发明使用光照感知网络评估可见光模态图像下光照感知系数，对训练好的双模态目标检测网络在NMS算法中进行感知加权融合，该方法在多模态数据集下取得良好的检测效果，面对夜晚等复杂环境该模型具有很好的鲁棒性。



CN 113326735 A

1.一种基于YOLOv5的多模态小目标检测方法,具体包括如下步骤:

步骤(1)、对需要应用的场景进行数据采集,并做划分,得到训练集和验证集;

步骤(2)、对光照感知网络数据集进行缩放处理,对多模态数据集进行数据增广处理;

步骤(3)、设计光照感知网络,采用二元交叉熵损失单独训练光照感知网络;

步骤(4)、在多模态数据集下,基于YOLOv5检测架构分别对可见光模态和红外模态进行独立训练;

步骤(5)、将独立训练好的光照感知模型、可见光模态模型和红外模态模型集成到定义好的多模态网络中;

步骤(6)、通过光照感知网络计算可见光模态图像感知系数,使用感知系数对可将光模态尾部输出进行加权,最终将双模态输出结果进行融合并输入到非极大值算法中。

2.根据权利要求1所述一种基于YOLOv5的多模态小目标检测方法,其特征在于:所述步骤(1)数据集划分过程中,涉及到两种类型数据集;第一种是光照感知网络数据集,第二种是多模态检测数据集。

3.根据权利要求1所述一种基于YOLOv5的多模态小目标检测方法,其特征在于:步骤(3)设计轻量化的光照感知网络,在具有Conv、Linear结构时,于光照感知网络头部引入Focus结构,Focus结构将输入图像上下做间隔采样,增加输入通道的同时降低图像尺寸,从而有效降低网络计算量。

4.根据权利要求1所述一种基于YOLOv5的多模态小目标检测方法,其特征在于:步骤(4)的多模态模型训练,相比于单模态目标检测,引入红外模态作为互补模态,以此提升复杂环境下的目标检测。

5.根据权利要求1所述一种基于YOLOv5的多模态小目标检测方法,其特征在于:步骤(5)的多模态光照感知融合模型训练策略,基于最新的YOLOv5目标检测算法作为多模态融合架构,并引入光照感知网络,基于光照感知融合的多模态检测算法总体损失函数定义如公式所示:

$$L=\sum_{m\in M}\left(\gamma_0*L_{box}^{m}+\gamma_1*L_{obj}^{m}+\gamma_2*L_{cls}^{m}\right)+L_{aware}$$

其中M为模态集合,该集合包含可见光模态和红外模态两个元素,其中L_{aware}为光照感知网络下的训练损失,双模态损失均是由L_{obj}、L_{cls}、L_{box}三部分组成,γ_0、γ_1、γ_2分别用来平衡三种损失的超参数,光照感知网路损失具体定义如下:

$$L_{aware}=-x'_d*\log(x_d)-x'_n*\log(x_n)$$

其中公式中x_d、x_n分别代表白天和晚上的真实标签,x'_d与x'_n分别代表光照感知网络的输出值;

对于前后背景损失和类别分类损失统一使用交叉熵损失架构定义,同光照感知损失类似,具体定义如下:

$$L_{obj}=L_{cls}=\sum_{i=1}^{n}\left(-w_i*[y_i*log(\sigma(x_i)+(1-y_i)*log(1-\sigma(x_i))]\right)$$

其中n代表样本数量,w_i代表第i个样本的损失权重系数,x_i代表第i个样本点的网络输出,y_i代表第i个样本点的真实标签值,σ(•)是Sigmoid激活函数;

对于位置回归损失使用CIoU loss进行计算，该损失函数定义如下：

$$L_{box} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha * v$$

其中$\rho^2(\cdot)$是欧式距离计算，b，b^{gt}分别代表物体BBox的中心点坐标，c代表BBox与$BBox^{gt}$最小外接矩形的对角线距离，α用来做trade-off参数，v用来衡量长宽比一致性参数。

6.根据权利要求1所述一种基于YOLOv5的多模态小目标检测方法，其特征在于：步骤(6)基于光照感知网络的多模态融合，最终将可见光模态和红外模态下输出的结果集合，根据可见光图像下的光照感知系数进行加权融合。

一种基于YOLOv5的多模态小目标检测方法

技术领域

[0001] 本发明公开了一种基于YOLOv5的多模态小目标检测方法，属于计算机视觉领域。

背景技术

[0002] 越来越多的研究者关注利用多传感器来提升目标检测模型的识别精度。面对复杂环境，研究者通常利用多模态数据互补的特点提升模型效果，这主要因为不同传感器记录信息方式不同，传感器的差异性使得模态之间信息具有互补性。常用传感器有红外相机、激光雷达、深度相机等，则不易受外部环境影响。

[0003] 2015年，Hwang等人在CVPR上发表一篇关于多模态的数据集，该数据集以行人检测为背景，提供可将光和红外两种模态对齐的图像，取名为Kaist。Kaist数据集的提出作为一个Benchmark开启了多模态目标检测领域的大门。Li等学者基于Kaist数据集，提出具有光照感知门融合的多模态互补技术，作者在Faster R-CNN上面做了实验验证，同时对Input Fusion、Early Fusion、Halfway Fusion、Late Fusion等融合架构进行了具体分析。Input Fusion融合是在数据输入层进行融合，可见光模模态图像由红、绿、蓝三个通道组成，红外模态一般是灰度图，即单通道，两个模态图像融合在一起是四个通道，较简单实现；Early Fusion是在骨干网络底层进行融合，一般是实现底层语义特征的融合，该方法缺失对高层语义特征的融合；Halfway fusion是在骨干网路中间层进行融合，中间层比较容易实现特征较好的融合，但训练比较困难；Late Fusion基于网络输出层进行融合，该种方式更侧重于对结果的融合，采用该方式无论是在模型训练还是部署都很容易实现。

[0004] 继Hwang之后，Lu等学者在Li的基础之上，进一步对多模态融合进行了详细分析，作者认为多模态在进行Halfway Fusion时，需要考虑不同模态中物体坐标漂移的问题，对于一个训练好的模型，作者在推理阶段对可将光模态做模拟以为进行验证坐标漂移对模型精度的影响。最后作者先对Kaist数据集两种模态下的物体坐标进行手工纠正，同时提出RFA模块从算法上进一步纠正，以此提升多模态的有效融合，不过RFA的模块引入降低了模型推理速度。Yang等人以SSD作为研究框架，提出基于GFU的多模态融合单元，将多模态融合技术应用到one-stage目标检测框架中。Heng等人提出循环细化融合模块，并引入语义监督损失作为辅助策略，使得特征融合更加平衡。Zhou等人基于Lu的基础上做进一步分析，认为多模态融合时分别受光照和特征两种不平衡因素影响，作者基于SSD检测模型，提出基于电路差分思想的特征融合和光照感知融合两个模块。

[0005] 基于以上研究内容，可知大部分学者均是使用Halfway Fusion的方式进行多模态目标检测融合，这种方式实现起来比较复杂，多模态之间特征域分布的不一致性使得模型训练更加困难，给目标检测模型部署应用带来很大困难。

发明内容

[0006] 本发明专门针对复杂环境中的多模态目标检测提出了一种创新的算法，该算法基于轻量化光照感知网络，将可见光模态和红外模态下检测出的结果做多模态融合，即在

late fusion融合阶段引入可见光模态下的光照感知系数进行加权处理。

[0007] 提供一种基于YOLOv5的多模态小目标检测方法，具体包括如下步骤：

[0008] 步骤(1)、对需要应用的场景进行数据采集，并做划分，得到训练集和验证集；

[0009] 步骤(2)、对光照感知网络数据集进行缩放处理，对多模态数据集进行数据增广处理；

[0010] 步骤(3)、设计光照感知网络，采用二元交叉熵损失单独训练光照感知网络；

[0011] 步骤(4)、在多模态数据集下，基于YOLOv5检测架构分别对可见光模态和红外模态进行独立训练；

[0012] 步骤(5)、将独立训练好的光照感知模型、可见光模态模型和红外模态模型集成到定义好的多模态网络中；

[0013] 步骤(6)、通过光照感知网络计算可见光模态图像感知系数，使用感知系数对可将光模态尾部输出进行加权，最终将双模态输出结果进行融合并输入到非极大值算法中。

[0014] 有益效果：本发明使用光照感知网络评估可见光模态图像下光照感知系数，对训练好的双模态目标检测网络在NMS算法中进行感知加权融合，该方法在多模态数据集下取得良好的检测效果，面对夜晚等复杂环境该模型具有很好的鲁棒性。

附图说明

[0015] 图1基于光照感知网络融合的多模态目标检测。

[0016] 图2基于光照感知网络的多模态融合伪代码。

具体实施方式

[0017] 为细致展示本发明的目的、特征和优点，下面将结合附图和具体的实施案例来对本发明做进一步详细说明。

[0018] 1.基于Focus结构的光照感知网络

[0019] 由于可见光模态下的图像受光照等环境影响比较大，尤其是夜晚环境。从算法模型角度出发，可见光模态下检测到的目标并不是完全可靠的，会存在漏检或者误检问题，因此需要对可见光模态下的图像做一个加权评估系数。

[0020] 本发明借鉴YOLOv5模型中的Focus卷积结构，将其应用到光照感知网络定义中。具体来说，Focus结构由一个Conv卷积网络组成，其中卷积核大小为1×1，对于输入128×128的Focus内部会对图像从横向和纵向两个方向做间隔采样，形成四张64×64的下采样图，最终堆叠一起形成一个具有12通道的输入数据。然后通过2×2大小的池化层做下采样，并采取Dropout方法以0.2概率舍弃神经元节点，最后将得到的特征向量输入到Linear层做预测，同时网络尾部采取softmax函数进行处理。

[0021] 可见光模态光照感知系数计算公式如下：

[0022] $$w' = (1-\mu) * w + \mu * k\#$$
$$\epsilon = w'[0]$$
其中w代表光照感知网络输出向量，该向量由w_1、w_2两个元素组成。μ代表平滑因子，k为标签类别数量，w'为经过平滑后的向量，∈为计算后的感知系数，即取第一个元素赋值。

[0023] 2.基于光照感知系数多模态融合

[0024] 本发明基于最新的YOLOv5目标检测架构实现对多模态信息的融合。如图1所示，基于光照感知的多模态目标检测融合架构由光照感知网路、双模态融合网络组成。

[0025] 首先基于光照感知融合的多模态检测算法总体损失函数定义如下公式所示：

[0026] $$L=\sum_{m\in M}\left(\gamma_0*L_{box}^{m}+\gamma_1*L_{obj}^{m}+\gamma_2*L_{cls}^{m}\right)+L_{aware}\#$$

[0027] 其中visible为可将光模态下的训练损失，lwir为红外模态下的训练损失，L_{aware}为光照感知网络下的训练损失。双模态损失均是由L_{obj}、L_{cls}、L_{box}三部分组成，γ_0、γ_1、γ_2分别用来平衡三种损失的超参数。光照感知网路损失具体定义如下：

[0028] $L_{aware}=-x'_d*\log(x_d)-x'_n*\log(x_n)\#$

[0029] 其中公式中x_d、x_n分别代表白天和晚上的真实标签，x'_d与x'_n分别代表光照感知网络的输出值。

[0030] 对于前后背景损失和类别分类损失统一使用交叉熵损失架构定义，同光照感知损失类似，具体定义如下：

[0031] $$L_{obj}=L_{cls}=\sum_{i=1}^{n}\left(-w_i*\left[y_i*\log(\sigma(x_i)+(1-y_i)*\log(1-\sigma(x_i))\right]\right)\#$$

[0032] 其中n代表样本数量，w_i代表第i个样本的损失权重系数，x_i代表第i个样本点的网络输出，y_i代表第i个样本点的真实标签值，σ(•)是Sigmoid激活函数。

[0033] 对于位置回归损失使用CIoU loss进行计算，该损失函数定义如下：

[0034] $$L_{box}=1-IoU+\frac{\rho^2(b,b^{gt})}{c^2}+\alpha*v\#$$

[0035] 其中$\rho^2(•)$是欧式距离计算，b,b^{gt}分别代表物体BBox的中心点坐标，c代表BBox与$BBox^{gt}$最小外接矩形的对角线距离。α用来做trade-off参数，v用来衡量长宽比一致性参数。

[0036] 基于光照感知网络的多模态融合伪代码如图2所示，首先对于双模态输出的结果集合A与B以及对应的置信度集合R与S，通过光照感知网络以及感知系数计算公式获取可见光图像当前感知系数∈，最终在进行融合之前，将可见光模态下输出的置信度与∈系数做乘法计算，然后再输入到非极大值抑制算法中进行融合。

[0037] 以上详细描述了本发明的优选实施方式，但是，本发明并不限于上述实施方式中的具体细节，在本发明的技术构思范围内，可以对本发明的技术方案进行多种等同变换，这些等同变换均属于本发明的保护范围。

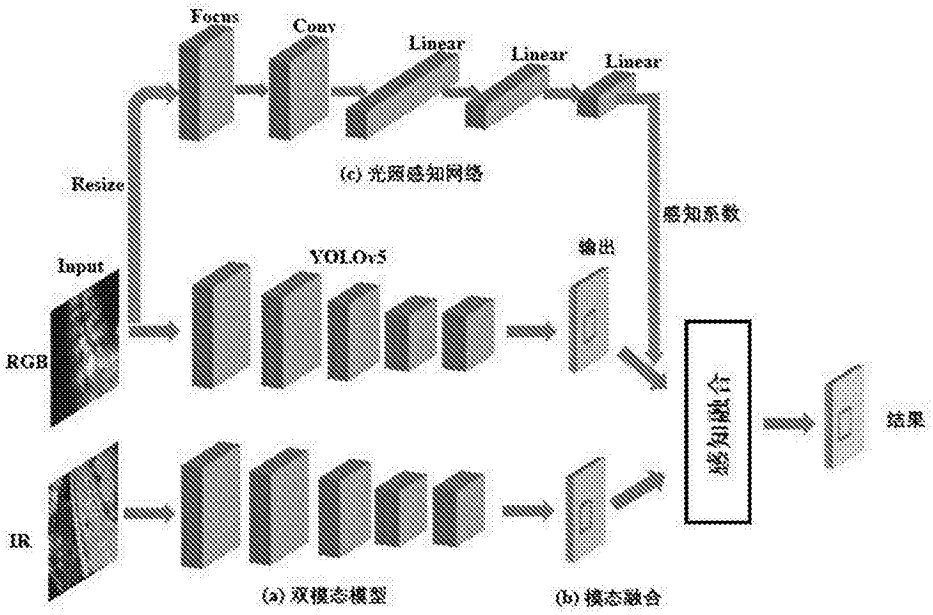


图1

Algorithm 3: 基于光照感知网络的多模态融合算法

Input:

$A = a_1,\ldots,a_M$, $R = r_1,\ldots,r_M$

$B = b_1,\ldots,b_N$, $S = s_1,\ldots,s_N$, N_t

A 是可将光模态初始检测到的 detection boxs

R 是可将光模态每个 box 所对应的置信度

B 是红外模态初始检测到的 detection boxs

S 是红外模态每个 box 所对应的置信度

N_t 是 NMS 算法阈值

ϵ 是光照感知网络计算可见光图像的感知系数

begin

$D \leftarrow \{\}$

$C = A \otimes B$; $T = R \otimes S$

$R = R * \epsilon$

while $C \neq empty$ **do**

$m \leftarrow argmax\, C$

$M \leftarrow c_m$

$D \leftarrow D \cup M$; $C \leftarrow C - M$

for c_i *in* C **do**

if $iou(M, c_i) \geq N_t$ **then**

$C \leftarrow C - c_i$; $T \leftarrow T - t_i$

end

end

end

$return D, T$

end

图2