(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 110322423 B
(45)授权公告日 2023.03.31

(21)申请号 201910355408.3

(22)申请日 2019.04.29

(65)同一申请的已公布的文献号
申请公布号 CN 110322423 A

(43)申请公布日 2019.10.11

(73)专利权人 天津大学
地址 300072 天津市南开区卫津路92号

(72)发明人 侯春萍 夏晗 杨阳 莫晓蕾 徐金辰

(74)专利代理机构 天津市北洋有限责任专利代理事务所 12201
专利代理师 程毓英

(51)Int.Cl.
G06T 5/50(2006.01)
G06T 7/00(2017.01)
G06N 3/0464(2023.01)

(56)对比文件
CN 105825491 A,2016.08.03
WO 2018076732 A1,2018.05.03
CN 108090888 A,2018.05.29
JP 2014192743 A,2014.10.06
陈木生.结合NSCT和压缩感知的红外与可见光图像融合.《中国图象图形学报》.2016,(第01期),

审查员 张鹏翼

权利要求书1页 说明书6页 附图5页

(54)发明名称

一种基于图像融合的多模态图像目标检测方法

(57)摘要

本发明涉及一种基于图像融合的多模态图像目标检测方法,包括:1)将预先采集好的红外图像及其可见光图像,制作多模态图像数据集;2)将预处理好的成对图像作为融合模型中生成模型G的输入;生成模型G基于U-Net等全卷积网络,以残差网络为基础的卷积神经网络作为生成网络模型结构,包括收缩过程和扩张过程,收缩路径包括多个卷积加ReLU激活层再加最大池化(Max Pooling)结构,下采样的每一步特征通道数都增加一倍,输出生成的融合图像;融合图像输入融合模型中的判别网络模型;根据训练过程中损失函数的变化,按迭代次数调节学习率训练指标,经训练,基于自有多模态图像数据集,能够得到同时保留红外图像热辐射特征和可见光图像结构性纹理特征的图像融合模型。

CN 110322423 B

1.一种基于图像融合的多模态图像目标检测方法,包括下列步骤:

1)将预先采集好的红外图像及其可见光图像,制作多模态图像数据集,图像格式均为单通道,包含具有结构和纹理特征的可见光图像和具有热成像信息的红外图像,按照图像模态分别制作训练集和测试集;

2)将步骤1)得到的训练集中的红外和可见光的多模态图像对进行包括进行裁剪和旋转平移操作在内的图像预处理,将预处理好的成对图像作为融合模型中生成模型G的输入;

生成模型G基于U-Net的全卷积网络,以残差网络为基础的卷积神经网络作为生成网络模型结构,包括收缩过程和扩张过程,收缩路径包括多个卷积加ReLU激活层再加最大池化Max Pooling结构,下采样的每一步特征通道数都增加一倍,输出生成的融合图像;

3)将步骤2)中得到生成模型G输出的融合图像输入融合模型中的判别网络模型D,判别网络模型D由一组孪生架构的卷积神经网络组成,对生成的融合图像进行无监督的质量评估,包括一组由交叉熵、重建误差、结构误差组成的损失函数,用以对融合图像和训练集中的原图像进行相似度丈量,来确定图像融合任务的完成度;同时优化生成网络G和判别网络D,方法为:输入训练集中原图像,更新G,误差变大;更新D,误差变小;更新G,重建误差变小,最终实现纳什均衡的动态平衡和同时优化;优化方法采用最小二乘的生成对抗方法;

4)重复进行步骤3),根据训练过程中损失函数的变化,按迭代次数调节学习率训练指标;

经训练,基于自有多模态图像数据集,能够得到同时保留红外图像热辐射特征和可见光图像结构性纹理特征的图像融合模型;

5)在步骤1)中构造的测试集中取红外和可见光的多模态图像对,输入到步骤4)中训练得到的融合模型中,继而输出融合图像;

再将得到的测试集,融合串联进基于深度卷积神经网络的检测模型,以在检测行人的红外热信息进行试例分析,得到行人的位置以及置信度。

一种基于图像融合的多模态图像目标检测方法

技术领域

[0001] 本发明属于深度学习、计算机视觉和图像融合领域，涉及一种基于深度神经网络的红外-可见光的多模态图像融合模型和目标检测模型的目标检测方法。

背景技术

[0002] 在自然环境中，物体会辐射出人眼无法看到的不同频率的电磁波，称为热辐射[1]。使用红外传感器所拍摄出的红外图像，能够记录不同物体的热辐射。红外(InfraredImage,IR)图像相较于可见光(Visible Image,VI)图像，具有如下特征:能够减少阳光，烟雾等外部环境的影响[1];对具有明显红外热特性的物体和区域敏感。目前红外图像中的目标检测任务应用较广，包括军事、电力、建筑等方向均有重要应用。然而，红外图像不具有较高的空间分辨率和较丰富的细节和明暗对比度，可见边缘和物体细节等往往不可见。红外传感器仅从一个方面获取信息，因此无法提供所有必需的信息。

[0003] 对于红外图像而言，利用图像融合技术，可以针对同一个目标得到的不同的图像，以最大化提取有用信息为目的，生成包含可见光图像中丰富的细节信息和红外图像中热信息的互补的融合图像[2]，可作为目标检测这一高层视觉任务的基础;也可以应用于医疗影像、电力缺陷等具体应用任务之上。

[0004] 目标检测(ObjectDetection)是模式识别领域中一个基础性的研究课题，作为被检测物体的两种不同信息的表达方式，物体类别的获取和物体位置信息的采集是物体检测任务主要针对的两个问题，主要的评价指标是准确性和实时性。目标检测任务也逐渐由传统方法向深度学习方法转变，特别是Ross B.Girshick教授R-CNN,FastR-CNN,FasterR-CNN这三项里程碑式的工作，开创了把深度学习方法应用于目标检测的先河，学界也开启了新的一轮基于深度学习方法的目标检测技术研究讨论和进展[4]。

[0005] 目标检测包括以下几个步骤，创建样本数据集(一般均包括正样本和负样本)，选择和提取目标特征，训练检测器。图像数据集中包含相当大的数据量，可以处理原始的图像数据来得到和其他相比最符合其类别的特点，即提取其特征。可把维数较高的原始数据空间中表现出的模式，通过非线性特征提取方法进行降维，用于分类和检测。

[0006] 卷积神经网络(ConvolutionalNeuralNetwork,CNN)提供了一种端到端的学习模型，经过训练后的卷积神经网络能够较好的学习到图像中的特征，并且完成对图像特征的提取和分类。

[0007] 现有的红外和可见光图像融合方法，一般根据其采用的基础理论可主要分为七类。即多尺度变换，稀疏表示，神经网络和其他方法。基于多尺度变换的方法是图像融合中最活跃的领域，其假设图像由不同区域中的多个层组成。此类型方法将源图像分解为多个层，使用人工设计的特定规则分别融合相应的层，并相应地通过逆变换重建目标图像。用于图像分解和重建的常用变换，包括小波变换，图像金字塔，曲线波等方法。基于稀疏表示的方法，利用在超完备字典中具有稀疏基础的线性组合的图像的可能表示，来实现其融合。基于神经网络的方法，通过设计人工神经网络(ArtificialNeuralNetwork,ANN)来模仿人脑

的感知行为处理神经信息,具有良好的适应性,容错性和抗噪声能力。综上所述,现有通用性红外和可见光图像融合方法各有优缺,因而综合以上的混合模型着力于具体应用需求,结合了特定方法的优点来提高图像融合性能。除上述外,现今学界亦有基于总变差,模糊理论和信息熵等理论的图像融合方法[3]。

[0008] 参考文献

[0009] [1]Ma J,Ma Y,Li C.Infrared and visible image fusion methods and applications:a survey[J].Information Fusion,2019,45:153-178.

[0010] [2]王峰,程咏梅.基于Shearlet变换域改进的IR与灰度VIS图像融合算法[J].控制与决策,2017(4):703-708.

[0011] [3]Li S,Kang X,Fang L,et al.Pixel-level image fusion:A survey ofthe state ofthe art[J].Information Fusion,2017,33:100-112.

[0012] [4]Elguebaly T,Bouguila N.Finite asymmetric generalized Gaussian mixture models learning for infrared object detection[J].Computer Vision and Image Understanding,2013,117(12):1659-1671.

发明内容

[0013] 本发明的目的是提供一种可以提升小目标检测效果的多模态图像目标检测方法。本方法将图像融合网络作为目标检测模型的前序步骤,提出一种通用的适用于红外图像的目标检测方法,在获取同一目标的红外和可见光图像的基础上,将图像对通过融合模型生成同时具有红外和可见光图像特征的融合图像,再通过基于深度神经网络的目标检测模型对融合图像进行检测,从而克服单一红外传感器的缺少结构特征的问题,对检测结果的提升和实际工程应用有着重要意义。技术方案如下:

[0014] 一种基于图像融合的多模态图像目标检测方法,包括下列步骤:

[0015] 1)将预先采集好的红外图像及其可见光图像,制作多模态图像数据集,图像格式均为单通道,包含具有结构和纹理特征的可见光图像和具有热成像信息的红外图像,按照图像模态分别制作训练集和测试集。

[0016] 2)将步骤1得到的训练集中的红外和可见光的多模态图像对进行包括进行裁剪和旋转平移操作在内的图像预处理,将预处理好的成对图像作为融合模型中生成模型G的输入;

[0017] 生成模型G基于U-Net等全卷积网络,以残差网络为基础的卷积神经网络作为生成网络模型结构,包括收缩过程和扩张过程,收缩路径包括多个卷积加ReLU激活层再加最大池化(Max Pooling)结构,下采样的每一步特征通道数都增加一倍,输出生成的融合图像。

[0018] 3)将步骤2中得到生成模型G输出的融合图像输入融合模型中的判别网络模型D,判别网络模型D由一组孪生架构的卷积神经网络组成,对生成的融合图像进行无监督的质量评估,包括一组由交叉熵、重建误差、结构误差组成的损失函数,用以对融合图像和训练集中的原图像进行相似度丈量,来确定图像融合任务的完成度;同时优化生成网络G和判别网络D,方法为:输入训练集中原图像,更新G,误差变大;更新D,误差变小;更新G,重建误差变小,最终实现纳什均衡的动态平衡和同时优化;优化方法采用最小二乘的生成对抗方法。

[0019] 4)重复进行步骤3,根据训练过程中损失函数的变化,按迭代次数调节学习率训练

指标.

[0020] 经训练,基于自有多模态图像数据集,能够得到同时保留红外图像热辐射特征和可见光图像结构性纹理特征的图像融合模型。

[0021] 5)在步骤1中构造的测试集中取红外和可见光的多模态图像对,输入到步骤4中训练得到的融合模型中,继而输出融合图像。

[0022] 再将得到的测试集,融合串联进基于深度卷积神经网络的检测模型,以在检测行人的红外热信息进行试例分析,得到行人的位置以及置信度。

[0023] 该方法根据红外图像和可见光图像分别具有的热辐射特性和结构化特性,基于图像融合技术,利用深度神经网络的生成对抗模型和深度目标检测算法,通过训练融合图像生成模型和检测模型,能够生成同时具有红外辐射和清晰结构的融合图像,将融合模型和检测模型混合串联,相较于单一红外图像,速度相对较快,并且能够显著提升检测的准确度。

附图说明

[0024] 图1专利流程图

[0025] 图2融合模型架构图

[0026] 图3红外和可见光图像及其融合结果图

[0027] 图4检测结果图

[0028] 具体实施方法

[0029] 为使本发明的技术方案更加清楚,下面结合附图对本发明具体实施方案做进一步的描述。具体实施方案流程图如图1所示。

[0030] 本方案的融合网络工作目标是,基于生成对抗网络的结构,学习一种映射函数,该函数根据多个未标记集合给出的两个输入图像生成融合图像,分别为可见光输入图像v与红外输入图像u。该网络不仅限于两种图像间的图像域转换,而是可以用于未标记的图像集,应用于融合任务。

[0031] 融合图像不仅能够保留红外图像中目标和背景之间高对比度的特性,而且相比于源图像中能够保留更多纹理细节,类似锐化的红外图像,融合图像具有清晰的突出显示的目标和丰富的纹理,模型亦可以能够融合不同分辨率的源图像。

[0032] 判别模型用于判断融合图像间的相似度,生成模型的任务是去产生一个同时包含红外和可见光信息的融合图像。这两个模型一起对抗训练,生成模型产生一张图片去欺骗判别模型,然后判别模型去判断这张图片是真是假,最终在这两个模型训练的过程中,两个模型的能力越来越强,最终达到稳态。

[0033] 1.构建融合图像生成网络模型(G):

[0034] 参考U-Net等全卷积网络,构建以残差网络为基础的卷积神经网络作为生成网络模型结构,包括左边的收缩路径和右边的扩张路径。收缩路径包括多个卷积加RELU激活层再加最大池化的结构,下采样的每一步特征通道数都增加一倍。

[0035] 扩张路径的每一步包括上采样、卷积(减少一半通道数),和相应收缩路径中的剪裁过的特征层的串联以及RELU激活。最后一层用了1×1卷积映射到想要的目标分布。

[0036] 神经元输入与输出:

[0037] $$y=f\left(\sum_{i=0}^{n}w_{i}x_{i}-\theta\right)$$

[0038] 2.构建融合图像判别网络模型(D)：

[0039] 构建用于判断融合图像间的相似性度量的分类网络，来分辨融合图像的真实度，即信息保留度，来引导训练的方向面向图像质量评价指标和图像信息度进行训练。

[0040] 其中判别网络模型的损失函数如下：

[0041] 交叉熵损失函数，用于全局优化：

[0042] $$\min_{G}\max_{D}L(D,G)=\mathop{\mathrm{E}}_{\substack{v\sim p_{data(\mathrm{VI})}\\u\sim p_{data(\mathrm{IR})}}}[\log D]+\mathop{\mathrm{E}}_{z\sim p_{data(\mathrm{VI})}}[\log(1-D(G(z)))]$$

[0043] 采用交叉熵损失函数，是针对的是生成的融合图片质量不高以及训练过程不稳定这两个缺陷进行改进。在判别模型D中的最后一个输出层的节点个数与分类任务的目标数相等，那么对于每一个样例，神经网络得到一个数组作为输出结果，这个数组也就是样本的融合结果，是神经网络的期望输出结果。

[0044] 用于无监督的用于G的相似损失函数：

[0045] $$\min L_{S}=\mathop{\mathrm{E}}_{\substack{u\sim \mathrm{P}_{data}(\mathrm{IR})\\v\sim \mathrm{P}_{data}(\mathrm{VI})}}\left\|\mathrm{G}(u,\mathrm{G}(u,v))\text{-}\mathrm{G}(\mathrm{G}(u,v),v)\right\|$$

[0046] 用于检验信息保留程度的G的重构损失函数：

[0047] $$\min L_{G}=\lambda_{u}\mathop{\mathrm{E}}_{u\sim \mathrm{P}_{data}(\mathrm{IR})}\left\|\mathrm{G}(u,u)\text{-}u\right\|+\lambda_{v}\mathop{\mathrm{E}}_{v\sim \mathrm{P}_{data}(\mathrm{VI})}\left\|\mathrm{G}(v,v)\text{-}v\right\|$$

[0048] 当输入被检测为融合样本对的时候，相似损失函数逐渐减小，相同类型的融合图像会持续在特征空间形成聚类。反之，当网络输入不相似的样本对时，相似损失函数会逐渐变大。通过最小化损失函数，最后可以使正样本对之间距离逐渐变小，负样本对之间距离逐渐变大，从而满足融合任务的需要。

[0049] 3.训练图像融合生成对抗网络

[0050] 生成对抗网络训练需要达到纳什均衡，梯度下降法较难实现，所以训练GAN不够稳定，本技术方法采用如下方法帮助训练：

[0051] 3.1.使用Wasserstein距离稳定收敛

[0052] 根据直线采样，梯度惩罚引入参数λ，去掉鉴别器的批正则化，使用Adam参数设置，引入双面惩罚和二次惩罚。

[0053] $$W\left(P_{r},P_{q}\right)=\sup_{\|f\|_{L}\leq 1}\mathrm{E}_{x\sim P_{r}}\left[f\left(x\right)\right]-\mathrm{E}_{x\sim P_{\theta}}\left[f\left(x\right)\right]$$

[0054] 表征了在最优路径规划下的最小消耗，衡量了原始两个分布之间的距离。

[0055] 3.2.特征匹配方法

[0056] 使用判别器中间层的特征来匹配图像的真伪，并将其作为一个监督信号来训练生成器，使得其生成数据会匹配真实数据的统计特性以及判别器中间层的预期特征值。

[0057] 3.3.小批量方法

[0058] 判别器每次考虑一小批的样本而不是一个单独样本，使得不同样本在空间上有合适的距离。

[0059] 3.4.历史平均方法

[0060] 加入一个惩罚项来惩罚和历史平均权重相差过多的权重值。

[0061] 3.5.输入规范化和批规范化

[0062] 将图像规范化至固定范围,针对真实数据和生成数据构建不同的小批。

[0063] 3.6.自适应矩估计动态收敛方法

[0064] 针对生成模型G使用自适应矩估计法Adam作为优化器,针对判别模型D使用随机梯度下降法SGD作为优化器,并且在生成器的不同层去除输入作为噪声。

[0065] 3.7.控制变量

[0066] 使用控制变量GAN,以及通过人工约束明确G适应的输入图像任务和顺序。聚焦于图像的每一部分,输出特征图而不是只输出单一的值。引入一个部分注意力机制用于对更重要的地方进行融合操作,可以更集中于不同的地方。

[0067] 4.检测网络搭建

[0068] 检测网络的输入是融合图像提取的候选区域,输出为固定长度的特征向量。候选区域输入图像金字塔池化(Spatial Pyramid Pooling)后,将该区域按照N种的尺度划分方法分别划定了N种S×S的划分区域,每种划分后的候选区域共计有S×S个块(Block)。

[0069] 将每种候选区域划分中的每一块(Block)进行最大池化(Max Pooling)下采样,共提取出$\sum_{i=1}^{N} S_i \times S_i$个特征,并组成特征向量进入候选区域网络,该网络对图像中的物体候选区域进行捕获,该方法可以将物体候选区域的捕获——物体候选区域的筛选——区域中该物体的分类完全整合于卷积神经网络方法,实现了在物体检测领域首次完全利用深度学习的方法进行操作。

[0070] 该网络能够利用不同尺度的兴趣点(即Anchor)将由原图映射而来的特征图进行物体位置提取,并且将生成的候选区域输入识别网络,识别网络和候选区域网络互相进行微调优化的过程对物体的位置进行不断逼近,同时获取物体类别信息。

[0071] 人体检测网络结构

[0072]

层级	输出尺寸	层数（共 59）
1	112*112	7*7, 64, stride 2
2	56*56	3*3 max pooling, stride 2
		[1*1, 64; 3*3, 64; 1*1, 256] *3
3	28*28	[1*1, 64; 3*3, 64; 1*1, 256] *3
4	14*14	[1*1, 128; 3*3, 128; 1*1, 512] *4
5	14*14	[1*1, 256; 3*3, 256; 1*1, 256] + 1*1
		[1*1, 256; 3*3, 256; 1*1, 256] *2
6	14*14	[1*1, 256; 3*3, 256; 1*1, 256] + 1*1
		[1*1, 256; 3*3, 256; 1*1, 256] *2
7	1*1	1000 -d, softmax

[0073] 5.检测网络训练

[0074] 首先应用数据增强，其主要分为训练集增强和测试集增强。训练集增强我们主要用了随机裁剪、平行翻转、随机擦除。通过附加测试集增强，主要包括翻转、平移、变化尺度等等。从高斯分布中得到一个随机的权重值，然后将这个权重值除以输入的节点个数的开根号，得到的新的值为权重的初始值。该学习率参数不断减少，对当前网络做快照，之后学习率调大，反复此过程。得到多个模型，最后做融合。不过考虑到复赛的模型数量限制，最后我们的学习率采用了5个训练轮次1e-4，5个训练轮次1e-5，5个训练轮次1e-6的方式。


开始
准备红外/可见光的训练/测试数据
搭建图像融合模型和目标检测模型
训练图像融合模型和目标检测模型
测试集输入图像融合生成融合图像
融合图像输入目标检测模型输出结果

图1

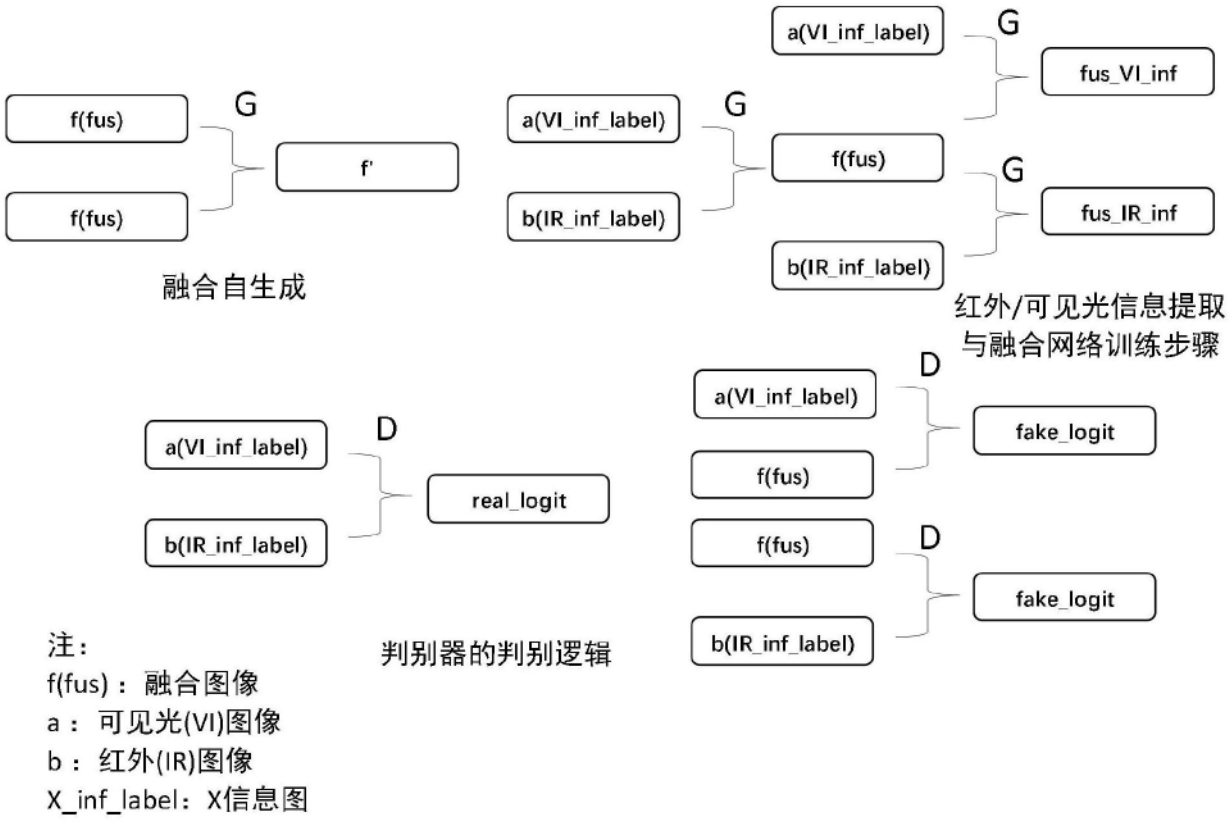
f(fus)
f(fus)
G
f'
融合自生成
a(VI_inf_label)
b(IR_inf_label)
G
f(fus)
a(VI_inf_label)
G
fus_VI_inf
G
fus_IR_inf
b(IR_inf_label)
红外/可见光信息提取
与融合网络训练步骤
a(VI_inf_label)
b(IR_inf_label)
D
real_logit
判别器的判别逻辑
a(VI_inf_label)
f(fus)
D
fake_logit
f(fus)
b(IR_inf_label)
D
fake_logit
注：
f(fus)：融合图像
a：可见光(VI)图像
b：红外(IR)图像
X_inf_label：X信息图

图2

(a) 红外图像

(b) 可见光图像

(c) 融合图像

图3

图4