

基于深度卷积神经网络的红外小目标检测

吴双忱, 左峥嵘*

(华中科技大学 人工智能与自动化学院 多谱信息处理技术国家级重点实验室, 湖北 武汉 430074)

摘要:提出了一种新的解决红外图像小目标检测问题的深度卷积网络,将对小目标的检测问题转化为对小目标位置分布的分类问题;检测网络由全卷积网络和分类网络组成,全卷积网络对红外小目标进行增强和初步筛选,实现红外图像的背景抑制,分类网络以原始图像和背景抑制后的图像为输入,对目标点后续筛选,网络中引入 SEnet (Squeeze-and-Excitation Networks) 对特征图进行选择;实验验证了整个检测网络相对于传统小目标检测算法的优势,所提出的基于深度卷积神经网络的小目标检测方法对复杂背景下低信噪比且存在运动模糊的小目标具有很好的检测效果.

关键词:模式识别与智能系统;红外小目标检测;深度卷积网络;低信噪比;运动模糊;SEnet

中图分类号:TP753 **文献标识码:**A

Small target detection in infrared images using deep convolutional neural networks

WU Shuang-Chen, ZUO Zheng-Rong*

(National Key Laboratory of Science & Technology on Multi-spectral Information Processing, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: A new deep convolutional network for detecting small targets in infrared images is proposed. The problem of small targets detection is transformed into the classification of small targets' location distribution. First, a Fully Convolutional Networks is used for enhancing and initially screening the small targets. After that, the original image and the background suppressed image are selected as the inputs for classification network which is used for the follow-up screening, and then the SEnet (Squeeze-and-Excitation Networks) is used to select the feature maps. The experimental results show that the detection network is superior to multiple typical infrared small target detection methods and has an excellent result on different signal-to-noise ratio, different scenes and motion blur targets.

Key words: pattern recognition and intelligent systems, small Target Detection in infrared images, deep Convolutional Networks, low signal-to-noise ratio, motion blur, SEnet

PACS: 84.40.Xb

引言

红外图像小目标检测技术一直以来是红外图像处理中所研究的重点^[1-3]. 但是在实际场景中,一方面,红外小目标通常被淹没在复杂的背景和噪声中;另一方面,红外小目标缺少具体的形状和纹理,因此

检测过程中缺少显著的特征和纹理信息. 因此红外小目标检测问题是一个具有较大挑战性的课题^[4]. 现有的红外小目标检测技术可以分为两大类:基于单帧空域和基于时域序列图像的检测方法. 然而基于时域序列图像的检测方法需要更多的先验信息并且通常需要借助于单帧的空域检测方法. 因此对

收稿日期:2018-08-25,修回日期:2019-03-11

Received date: 2018-08-25, revised date: 2019-03-11

基金项目:国家自然科学基金(61773389)

Foundation items: Supported by the National Natural Science Foundation of China (61773389)

作者简介 (Biography): 吴双忱 (1995-), 男, 湖北襄阳人, 硕士研究生, 主要研究领域为红外目标检测、模式识别与智能系统. E-mail: m201772486@hust.edu.cn

* 通讯作者 (Corresponding author): E-mail: zhrzuo@hust.edu.cn

于预警问题来说,基于单帧空域的检测方法显得非常重要.传统的基于单帧空域的检测方法有以 max-mean 滤波^[5],形态学 top-hat 滤波^[6],高通滤波^[7]和小波变换^[8]等为代表的一系列方法.然而当目标信噪比较低时,这些方法通常会导致数量较多的检测虚警以及较差的检测性能.基于偏微分方程^[9]的方法通过图像中像素点的梯度函数对图像背景进行非线性预测达到对背景更好的抑制作用;基于 MRF^[10]理论的检测算法利用背景与目标空间分布的差异,通过定义新的势函数对小目标进行检测,然而上述方法存在着人为调整的参数较多并且需要在线自适应计算的缺点.

利用其强大的特征提取能力以及相对于传统检测算法更强的泛化能力,深度学习近些年来在对图像中物体的检测方面都取得了相当大的进步,无论在检测速度还是准确度上都达到了非常好的水平,现在以 faster rcnn^[11],Yolo^[12],SSD^[13]等为代表的物体检测网络,在各个数据集上取得了很好的成绩,然而在利用深度学习对红外图像中的小目标的检测上,目前还没有太多的研究.本文提出了一种可以高性能地对红外图像中小目标进行单帧检测的深度卷积神经网络结构.我们检测方案的出发点在于深度卷积神经网络通过对训练样本的学习,可以区分出背景图像结构与红外小目标结构特性,进而达到对小目标检测的目的.针对小目标信噪比太弱难以检测的问题,设计了一种基于全卷积的增强网络对小目标区域进行初步筛选,其后的分类网络利用原始图像中的结构信息来抑制检测中的虚警点,同时利用背景抑制图像中小目标的局部特性提高小目标的检测率,因此该网络采用了原始图像与背景图像的双输入进行目标的判别.在网络结构的设计上,我们采用 SEnet^[14](Squeeze-and-Excitation Networks)对双输入网络的特征图进行筛选以提高目标在运动模糊时的检测率.最后得出了一种可以很好地区分目标点和虚警点,保持高检测率,低虚警率的网络,我们同时也介绍了我们训练集的制作方法;经过实验验证,我们所设计的网络结构对不同信噪比,不同复杂程度的背景中的小目标检测上相对于传统的检测算法以及以 PDE, MRF 为代表的自适应检测算法都有相当高的检测率,同时虚警率也相对很低.

1 检测网络结构

1.1 检测方案

深度学习技术应用于红外图像小目标检测的难

点在于目标形状微小,其大小通常只有 3、4 个像素,缺乏轮廓特征信息,而当前基于深度学习技术的目标检测方法利用了目标的各类特征信息,其中,轮廓特征对于整个网络学习十分重要,轮廓特征的缺乏对小目标检测网络的设计带来一定挑战;再者,由于小目标所处背景的多样性和复杂性,小目标的自身特性的起伏变化(如亮度,大小等)也加大了检测难度.

在单帧图像上要利用深度学习对小目标进行检测,主要利用的是小目标的在局部图像中的特征,即在一块较小的区域内,其辐射亮度在局部范围内较为突出;当前,在时间上和性能上均取得令人满意的效果的深度神经网络结构有以 faster rcnn, Yolo2, SSD 等.然而对于只有几个像素的小目标的检测,其网络层数较多,造成不必要的运算时间,本文经过实验表明即使较浅层的卷积网络也可以对小目标的特征有很好的提取作用,同时由于要满足实时性的要求,上述网络都是将整幅图像作为输入,然后通过选取候选窗口的方式,对物体进行检测,由于红外图像中的小目标往往只有几个像素,因此如果要通过上述网络对小目标进行检测,就需要选取足够小的候选窗口,这样反过来就造成了候选窗口选取数目变得非常多,从而在时间上丧失了这些方法的主要优势.因此,本文提出了一种与上述的网络结构不同的解决单帧红外图像中小目标检测问题的网络结构,该网络相对于上述网络结构更为简单直接,同时也保证了网络检测的时间.

由于红外小目标形状微小,在一幅较大尺寸的图像中所占像素数目比例非常低,因此利用深度卷积神经网络直接对一副较大图像进行小目标检测时会造成网络在训练过程中负样本数目远大于正样本数目的情况,增加了训练难度和准确度.为了改善这一情况,本文在提取候选目标点时采用与滑动窗口类似的方法,不同点在于,滑动步长设为窗口尺寸,同时为了改善目标被窗口边界切割的情况,将步长设为(窗口尺寸-1),这样可以保留目标的大部分.图 1 是这种滑动方法的示意图:

通过对窗口大小的选择,可以控制对整幅图像的搜索次数,搜索计算方式为:

$$\frac{W - S + 1}{stride} \times \frac{H - S + 1}{stride}, \quad (1)$$

在式(1)中 W, H 为图像长宽, S 为窗口大小, stride 为搜索步长.我们选择了 21 * 21 的窗口,步长为 20,图像大小为 288 * 384,因此一幅图需要的遍历

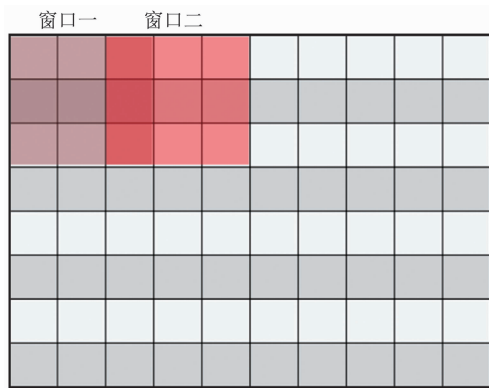


图1 窗口滑动示意图(若窗口大小为3,则步长选择为2)
Fig. 1 Schematic of sliding window (if the window size is 3, the step length is 2)

次数为266次,远小于逐像素遍历所需要的110592次,因此在检测效率上有很大的提高.经过这种滑动方法得到266个图像后,接下来需要对目标进行检测,我们最终想要得到的是目标的具体位置,而目标在某幅图像中可能有随机分布在某一位置上.图2表示的是目标可能的三种不同类别的分布形式:

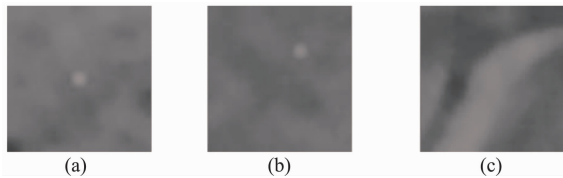


图2 目标在图像块中可能的三种分布形式
Fig. 2 The three possible distributions of the target in the image block

从图2中可以看到可以将分布形式可以分为三类:(a)含有目标点且目标点在该图像的中心位置,(b)含有目标点但目标点不在该图像的中心位置,(c)完全不含有目标点.基于此,我们将目标的检测问题转换成目标位置分布的分类问题,当分类结果属于第一类,即目标在图像的中心位置时此时图像的中心坐标即为目标检测的具体位置.这样做的好处是省去了对目标位置的回归网络,不需要事先标定目标所在区域的坐标进行回归训练,大大简化了网络的设计结构.

1.2 检测网络结构

利用小目标的特性可以将小目标的检测问题转换成分类问题.而当将输入图像分为第二类,即含有目标且目标不在中心位置,此时需要确定目标在图中的具体位置,一种可以采用的方法是:对这个 21×21 的图像所在的原图区域再进行一次滑动窗口

进行分类直至目标刚好在图像中心,然而这种方法增加了检测的次数,对于一个 21×21 的图像区域,最坏情况下需要441次检测,增加了检测时间;为此,我们设计了目标增强网络来解决这个问题,增强网络的设计思想与对物体进行分割所采用的全卷积网络相同,该网络将输入为 21×21 大小的图像转换为幅 21×21 大小的概率图像,该概率图像中每个像素点的像素值大小表示输入图像中对应像素点属于目标点的概率值大小;得到此图像后,进行目标检测的算法流程如下:

- 1) 对概率图像按概率大小进行排序,选择概率在前5%大小的图像位置;
- 2) 根据选择的位置获得分类网络的输入图像数据;
- 3) 利用分类网络对输入图像进行分类.

整体检测算法的检测思路如图3所示.在流程1中5%大小的比例选取阈值是通过实验结果确定的.

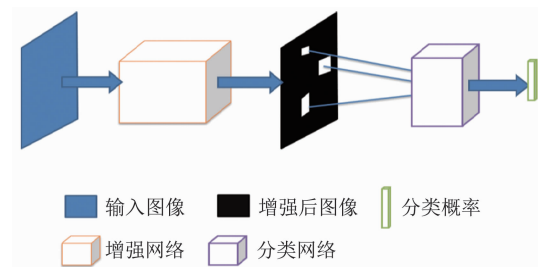


图3 检测网络结构
Fig. 3 Detection network structure

1.2.1 增强网络

增强网络的主要目的是得到概率最大的目标候选位置,从而减少搜索次数,其作用是对目标进行一个粗筛选;增强网络的设计思想如图4所示,利用目标的特性,卷积神经网络能够通过卷积提取目标特征并通过多层卷积对目标的形成一个分类能力,得出目标点的概率大小;对于小目标问题,我们认为不同卷积核对图像区域进行不同的模板匹配,以此捕捉不同信噪比,不同背景的目标特征,然后将所得的概率值再通过卷积进行层层筛选,最后输出得到一个目标特性越强,对应概率值越大的概率分布图像;由于小目标大小网络只有几个像素,据此,我们网络中选取的卷积核大小为 3×3 ,这个模板大小基本覆盖小目标所在区域.增强网络的具体结构在图5中,其采用的设计结构来自于用与目标分割中的全卷积神经网络^[15].可以看到我们只采用了三层卷积,最

后一层采用反卷积^[16]将输出通道变为 1. 增强网络的输入为经过去中值滤波得到的图像 S, 即输入给增强网络的图像 S 为式(2)所示, 其中 F 表示原始图像:

$$S = F - \text{median filter}(F) \quad (2)$$

得到 S 后为了便于网络的训练, 我们对其进行归一化操作, 我们采用的归一化操作公式如式(3):

$$S = \frac{S - \min(S)}{\max(S) - \min(S)} \times \max(S) \quad (3)$$

我们之所以选取图像 S 作为输入的好处在于对原图 F 进行中值滤波相当于对原图的背景进行一个估计, 通过减去背景达到是目标特性突出, 从而使得增强网络对目标的特性有更好的学习能力. 在对增强网络进行训练的时候, 作为增强的目标图像是一幅二值图像, 在该二值图像中, 标记目标点所在位置为 1, 其余均为 0.

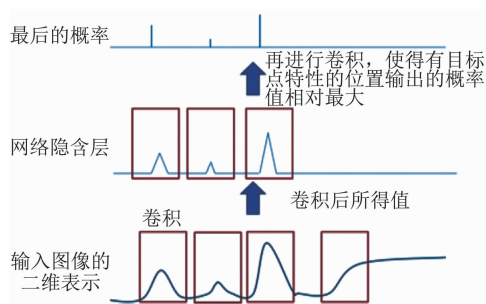


图 4 增强网络设计思想
Fig. 4 Design idea of enhance network

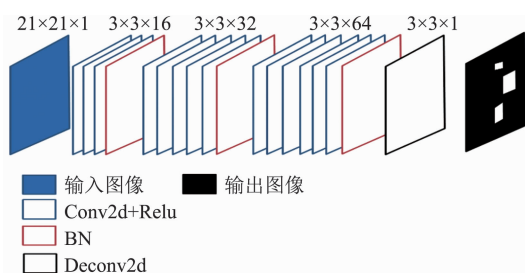


图 5 增强网络具体结构
Fig. 5 Specific structure of enhance network

为了减少训练和检测时间, 我们对卷积层的层数进行了考虑, 实验部分中比较了不同层数的增强网络在测试集上的表现, 最终选取了图 5 所示的网络结构. 为了观察增强网络对小目标的实际增强效果, 在输入图像包含目标的情况下, 我们依次选取了网络中各个隐含层的部分特征图进行观察, 结果如表 1 所示. 同时为了进一步探究增强网络对小目标特性的捕捉能力, 我们采取与文献^[17]中的方法可视

化隐含层权重的特征响应情况. 方法如下: 对于某一个特定的卷积核, 我们以其输出特征图经过激活函数后平均值作为其损失值, 以表示该卷积核特征图, 则得到一个卷积核可视化特征响应的算法为:

- 1) 用高斯分布随机初始化一个 21 * 21 大小的输入图像;
- 2) 计算输入图像在卷积核中产生的特征图的输出均值;
- 3) 以均值作为网络的损失, 即:

$$\text{loss} = \frac{1}{N * M} \sum F_w \quad (4)$$

N, M 为对应特征图的大小

- 4) 计算梯度并更新输入图像, 其中 α 为更新系数, 我们取为 1:

$$\text{gradient} = \frac{\partial(\text{loss})}{\partial(I)}$$

$$I = I + \alpha * \text{gradient} \quad (5)$$

- 5) 重复上述步骤若干次至输入图像 I 保持不变.

我们选取了增强网络中第四层的卷积核作为可视化目标, 我们选了了其中特征响应最大的 16 个卷积核, 最终得到的可视化图像如图 6 所示.

表 1 增强网络隐含层特征图
Table 1 Feature maps of hidden layers in enhance network

输入	第一层卷积	第二层卷积	第三层卷积

从表 1 中可以看到, 输入图像中心的目标点的特征随着层数增加越来越明显, 到第三层已经足够大, 再经过一层反卷积输出, 可以取得很好的增强效果.

图 6 中所示的是使第四层卷积核有最大响应值时的输入图像, 从图像中可以观察到, 不同卷积核所捕捉的特征的共同点都是具有许多点状的小目标,

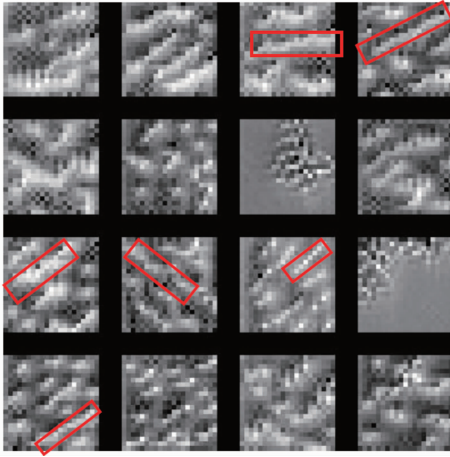


图6 增强网络第四层卷积核可视化图像
Fig. 6 Visual image of the fourth-level convolution kernel in enhance network

这表明卷积核与点状的小目标进行卷积产生了最高的特征响应值,从而达到对小目标的增强作用,并且不同的特征图对不同位置分布和亮度的小目标具有不同的捕捉能力.从而可以实现目标在不同位置 and 不同信噪比的捕捉.但如图中红色方框内标记所示,当输入图像有较为明亮的轮廓线时,可能会对增强网络造成影响,使其误判为许多目标的分布在一条线段上.基于上述分析,需要后续的分类网络对提取的候选目标进行进一步分类.在增强网络的损失函数的设计上,采用的是基于灰度互相关的损失函数,由于目标图像是只有极少的点是1,其余都为0的二值图像,因此采取灰度互相关损失可以获得比均方差损失更大的损失梯度,因此损失函数 L 包含 L_1 , L_2 两部分, L_1 计算的是网络输出图像与目标图像的均差,其目的是使输出图像与目标图像在均值上相近, L_2 计算的是两幅图像之间的灰度互相关系数,目的是使输出图像与目标图像在像素点的变化上保持一致,因此采用 L_1 与 L_2 相加的方式,损失函数 L 使得输出图像与目标图像尽可能在所有的像素上达到一致,误差计算公式如下:

$$L = L_1 + L_2$$

$$L_1 = -\frac{1}{batch_size} \sum_{i=1}^{batch_size} abs(G(i) - S(i))$$

$$L_2 = -\frac{1}{batch_size} \log\left(\sum_{i=1}^{batch_size} \frac{(S(i) + 0.01) * (G(i) + 0.01)}{\sqrt{(S(i) + 0.01)^2 * (G(i) + 0.01)^2}}\right), \quad (6)$$

式(6)中 G 表示的是网络输出图像, S 表示目标图

像,为了防止 $\log(0)$ 的情况出现,所以每一项都加上了平滑系数 0.01. $batch_size$ 表示每次训练的批量大小,我们设置为 32.

1.2.2 分类网络

分类网络主要用来对目标的分布形式进行分类,其设计好坏直接影响到小目标检测率以及虚警率的大小;分类网络的输入来自于增强网络得到的候选目标位置;直观上,有以下两种输入方式:

1) 在原始图像 F 上以候选目标位置为中心得到 $21 * 21$ 大小的图像块输入;

2) 经过背景抑制后的图像 S 作为输入.

对于输入方式 1,原始图像 F 中的图像结构信息在分类得到了运用从而可以帮助抑制虚警,但同时其复杂的背景也会对目标的分类造成干扰;对于输入方式 2,在背景抑制后的图像 S 中目标特性相对明显,但由于图像结构信息的丢失,会在检测中引入许多虚警点;实验部分比较了两种不同的输入方式所得到的检测率以及虚警率情况,实验结果验证了上述分析,为了弥补两种输入方式的缺点并且结合其优点,我们最终采取将原始图像 F 和背景抑制后的图像 S 同时输入作为输入方式,如图 7 所示.

图 7 中的分类网络在分类时将两个卷积分支得到的特征图展开拼接为一个特征向量输入到全连接网络,为了对所得特征向量更为细致的分类,我们采用了三层全连接层,两两之间加入 Dropout 层.

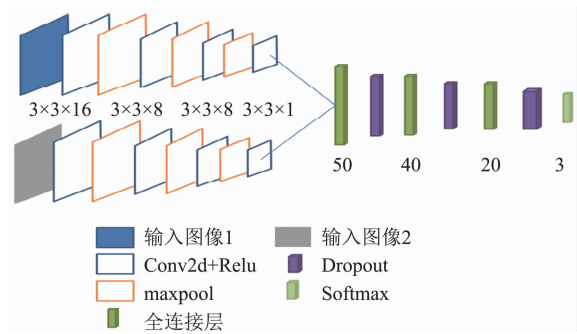


图7 分类网络结构
Fig. 7 Classification network structure

分类网络的分类结果见实验部分.由于训练集中目标和实际场景中目标的差异,当目标产生不规则的运动模糊时,分类网络在这种情况下检测率有所下降.图 8 所示为未产生运动模糊的目标点和运动模糊目标点在原始图像和背景抑制后的图像中的对比,从图像相似性来看,由于运动模糊的产生,图像(b)中的目标周围产生了较多的模糊亮斑,导致目标点在(a),(b)中的特性差异较大.在对图像进

行背景抑制之后,由于运动模糊产生的是低频信息,而背景抑制对低频信息有一定的抑制作用,因此在背景抑制后的图像中运动模糊对目标特性造成的影响相对较小,如图像(d)所示,经过背景抑制后许多在(b)中产生的模糊亮斑此时更接近于背景灰度,并且目标特性相对于背景更为突出,与(c)中的目标特性差异相对于(a),(b)来说更小,并且目标周围背景此时灰度分布也较为均匀,因此此时采用S进行分类判断会更加合理,然而有些时候需要利用F中物体轮廓的连续性和结构性来抑制虚警;因此,F和S同时输入时,不同情况下F和S的特征对分类的重要性不同,也就是要求分类网络需要能够对两组特征有一个选择性.对两个卷积网络特征图直接进行拼接的方法用式(7)表示:

$$R = F([H, V]) \quad (7)$$

式(7)中R表示为输出结果,F表示卷积神经网络后的全连接网络,H,V表示不同的输入得到的特征图.通过对图8的分析,我们希望对来自不同输入的特征图施加不同的权重后再进行分类,就相当于对特征图进行了有条件的筛选,可以提高在不同情况下的分类效果,可以表示为式(8)所示:

$$R = F([W_1 * H, W_2 * V]) \quad (8)$$

通过对 W_1, W_2 的调整,可以使网络对最后的选择有一个适应性. SEnet (Squeeze-and-Excitation Networks) 网络能够通过特征压缩提取对不同特征图进行权重赋值,使得所需要的特征图能够得到增强.我们的设计是利用SEnet对来自于两个网络的特征图进行选择,从两组特征图中选择出我们所需要的特征图,相对于此前的直接将不同卷积输出的特征图进行拼接,这种做法针对性更强,更容易加强期望的特征图.

SEnet网络最初设计目的是在网络训练过程中使对最后分类有更大作用的特征图得到增强,从而加快网络的训练速度,同时增加整个网络对特征的选择和捕捉能力,SEnet广泛地与残差^[18]以及inception网络^[19]进行结合使用.图9是SEnet网络的结构图:

SEnet的映射关系可以表示为:

$$Z = F_{sq}(U_c) = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H U_c(i, j)$$

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z))$$

$$X_c = F_{scale}(U_c, s) = s \cdot U_c \quad (9)$$

式(9)中的 F_{sq} 和 F_{ex} 分别表示对特征图进行压缩和特征提取操作,压缩所采取的方式是将原始特

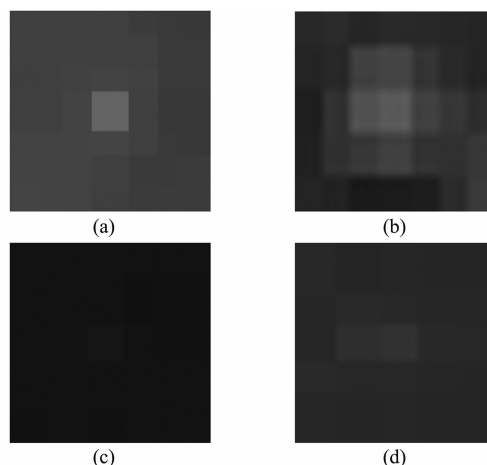


图8 产生运动模糊时F和S对比(a)训练集原始图像F,(b)产生运动模糊的图像G,(c)F背景抑制后图像S,(d)G背景抑制后图像S'

Fig. 8 Comparison of F and S with motion blur (a) Training set original image F, (b) Motion blur image G, (c) Background suppression image S of image F, (d) Background suppression image S' of image G

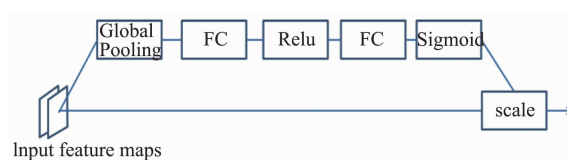


图9 SEnet结构图

Fig. 9 Structure diagram of SEnet

征图 U_c 的均值作为其特征表示,从而达到数据的降维的目的,对应于图8中的Global_Pooling,将多通道的 $W * H$ 大小的二维特征图转化为压缩特征向量Z后,通过全连接层的权重 W_1, W_2 ,训练出对特征图的选择能力的全连接网络,在式(9)中g表示的是激活函数Relu,最后通过 σ (sigmoid)激活函数,输出一个不同特征图相对应的权重向量,并与原始特征图对应相乘得到最终输出 X_c .

如图10所示式已加入了SEnet的网络结构,网络的卷积层和全连接层与图7中的网络结构完全相同,实验部分比较了加入SEnet前后的网络在目标检测率和虚警率上的提升,我们采用其作为最终分类网络.

2 网络训练

2.1 数据集制作

训练数据集的质量直接影响着网络最终泛化能力的强弱.因此,在制作训练集的需要考虑红外目标检测问题中的多样性和复杂性.首先,在训练集中我们加入了局部信噪比从1到7变化的目标;目标本

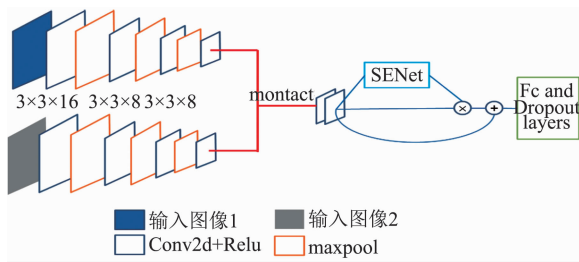


图 10 加入 SENet 后的分类网络

Fig. 10 Classification network after joining SENet

身的特性可以近似为服从高斯分布的点扩散;因此我们采用方差为的二维高斯分布来近似目标点扩散效应;目标区域的灰度计算可以用公式表示如下:

$$I(x, y) = I(0, 0) \times e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (10)$$

其中 $I(x, y)$ 表示在图像区域内相对于中心坐标为 (x, y) 处的点的灰度值, $I(0, 0)$ 为区域中心点的灰度值, 其大小可以根据不同的信噪比进行计算;为了定量添加 SNR 不同的目标点, 我们选取了目标点为中心 11×11 的邻域对局部信噪比进行计算, 计算公式为:

$$snr = \frac{I(0, 0)}{\sigma_{bp}^2} \quad , \quad (11)$$

其中 σ_{bp}^2 表示目标所在的 11×11 像素大小的背景区域的方差, 这样我们通过上述计算就得到了不同 snr 下的目标点。

我们将目标背景大概分为地面背景, 天空背景以及天地交接三种类型, 图 10 是不同背景的图片示例;对于每一类的背景, 选取差别较大的 200 幅添加目标点, 图像之间的差异通过对其进行 PCA 分解得到的特征向量的差值来衡量;在每幅图像上添加 30 个目标点并记录下相应位置;随机选取其中 500 幅作为训练, 其余 100 幅以及含有实际运动目标点的 300 幅图像作为测试图像. 根据添加的目标点位置, 选取以目标点为中心的 21×21 的图像作为目标点在中心的类别. 对于目标点不在图像中心的情况, 每个位置出现的概率可以认为其服从均匀分布 $U(0, mums-1)$;在 21×21 的图像中, 去掉中心点之后, 其分布的范围是 $0 \sim 420$;因此按照均匀分布 $U(0, 420)$ 产生一个一维坐标, 换算成图像坐标后, 在图像块上加入目标点, 这样产生第二类训练集;对于没有目标点的情况, 采用随机选择背景图像的方法. 最后得到了不同背景以及不同信噪比下的训练图像共 50 000 幅. 图 12 展示了部分训练集图像.

在图 12 的训练集图像中, 第一行中的图像展示

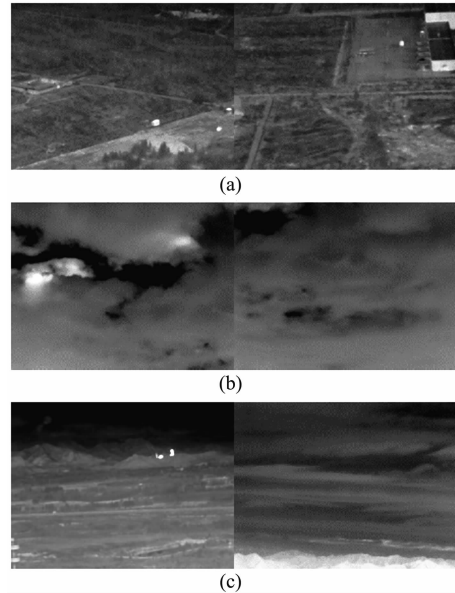


图 11 不同背景示例 (a) 地面背景 (b) 天空背景 (c) 天地交接背景

Fig. 11 Different background examples (a) Ground background, (b) Sky background, (c) sky-ground background

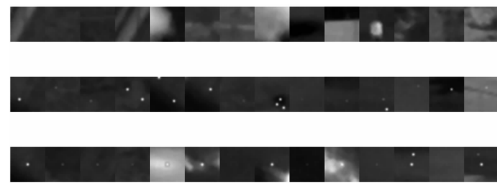


图 12 部分训练集图像

Fig. 12 Part of training set images

的是没有目标点的情况, 第二行是目标不在中心的情况, 第三种是有目标点在中心的情况.

2.2 网络训练方法

在网络训练中, 我们首先对增强网络进行训练; 然后将对所制作的分类训练集输入到分类网络中训练, 为了提高模型的泛化能力, 在训练中所采用的均为提前中止^[20]的方法. 即在每次训练完一个 epoch 之后, 对测试集进行测试, 在迭代 500 个 epoch 之后保存的是在测试集上面表现最好的一次网络模型, 并记录下此时的测试误差;接着再对训练集和测试集一起进行训练, 直到训练的误差小于等于之前所最记录的最小的测试误差;在训练中所取的 dropout^[21]的系数为 0.4;正则项系数为 0.01;虽然在网络中引入了 BN (Batch Norm)^[21]网络, 但是实验中发现 dropout 和正则项的引入对实验结果改善还是较为明显. 由于我们所设计的两个网络之间并没有直接连接, 所以训练完之后直接按照图 3 中检测步骤对目标点进行检测.

3 实验结果

在增强网络的设计中,我们比较了含有不同卷积层数的网络在测试集上的表现,我们按照 4: 1 的比例划分训练集和测试集,选取了含有 3, 4, 5 层卷积的增强网络进行实验对比,其测试误差曲线如图 13 所示. 从测试误差曲线中可以看到卷积层数为 3 时,其最终的测试误差与更深的网络层数相比仍相对较高,含有 5 层卷积的网络虽然误差收敛速度最快,但 4 层卷积网络最终也达到了几乎相同的测试误差,并且收敛更加稳定,因此出于检测时间上的考虑,我们最终选取了 4 层卷积网络.

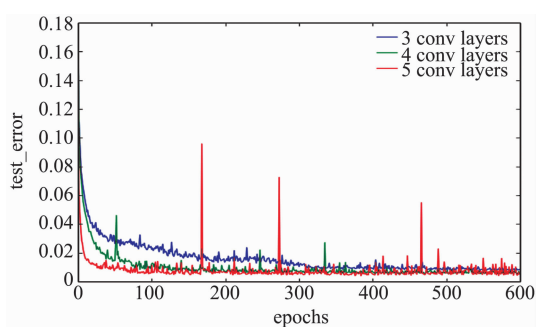


图 13 不同卷积层数的测试误差曲线

Fig. 13 Test error curves for different convolution layers

为了验证所设计的增强网络的增强效果和必要性,对增强网络的有无进行了一次实验对比,对比实验中去掉了第一部分的增强网络,而候选目标点位置是由背景抑制后的图像 S 得到的. 即检测算法变为对 S 按照灰度值大小排序,选择其中前 5% 的点,然后通过分类网络分类;有无增强网络的分类结果如下,其中分类网络采用最终设计的模型:

表 2 有无增强网络对目标检测率影响

Table 2 Effect on the target detection rate of enhance network

SNR	1	2	3	4	5	6	7
有	92.9%	94.3%	95.3%	96.2%	97.0%	98.1%	98.2%
无	32.2%	46.5%	53.3%	66.7%	73.3%	76.8%	81.3%

表 3 有无增强网络对虚警个数影响(平均每幅图像)

Table 3 Effect on the number of false alarms of enhance network(average value of each image)

SNR	1	2	3	4	5	6	7
有	7	7	8	8	8	9	8
无	11	9	10	8	11	10	8

在分类网络中采用的输入为双输入和单输入的

情况进行的检测率和虚警率的实验对比,实验是比较它们在不同信噪比的目标下的实际检测效果,其中 S 表示输入为经过中值滤波后得到的背景抑制后的图像的情况, F 表示输入为原始图像的检测情况, S + F 表示将其同时输入时的检测情况,表 4 展示了在目标 SNR 不同的情况下不同输入方式的检测率的实验对比情况,表 5 展示了在目标 SNR 不同的情况下不同输入方式的虚警点个数实验对比情况,所采用的分类网络是图 6 中所示的分类网络,我们检测的图像大小均为 288 384 个像素:

表 4 分类网络对不同输入情况下目标检测率

Table 4 Target detection rate of the classification network under different input conditions

SNR	1	2	3	4	5	6	7
S	93.7%	94.6%	96.0%	97.9%	98.1%	98.3%	98.8%
F	89.1%	90.3%	91.1%	92.4%	93.6%	94.5%	96.2%
S + F	90.2%	91.0%	91.5%	93.7%	95.3%	96.4%	96.7%

表 5 分类网络在不同输入情况下虚警个数

Table 5 False alarms' numbers of the classification network under different input conditions

SNR	1	2	3	4	5	6	7
S	21	24	26	28	28	27	28
F	5	6	6	6	7	7	7
S + F	7	8	9	8	8	9	9

从表 4 和表 5 可以看出将 S 和 F 同时作为输入时在检测和虚警抑制方面相对于单幅图像输入有一个比较好的均衡,然而在目标检测问题上,我们希望能够尽最大可能地将目标检测到同时不增加虚警,因此我们加入了 SEnet 改进分类网络,得到的实验对比结果如表 6,表 7 所示:

表 6 有无 SEnet 对目标检测率影响

Table 6 SEnet's Effect on Detection Rate

SNR	1	2	3	4	5	6	7
SEnet	92.9%	94.3%	95.3%	96.2%	97.0%	98.1%	98.2%
None	90.2%	91.0%	91.5%	93.7%	95.3%	96.4%	96.7%

表 7 有无 SEnet 对虚警个数影响

Table 7 SEnet's Effect on False alarms' numbers

SNR	1	2	3	4	5	6	7
SEnet	7	7	8	8	8	9	8
None	7	8	9	8	8	9	9

部分对实际红外场景检测结果图对比:

在上面的真实场景中,均只有一个真实目标点,其中蓝色实线框代表真实目标点位置且检测到,蓝

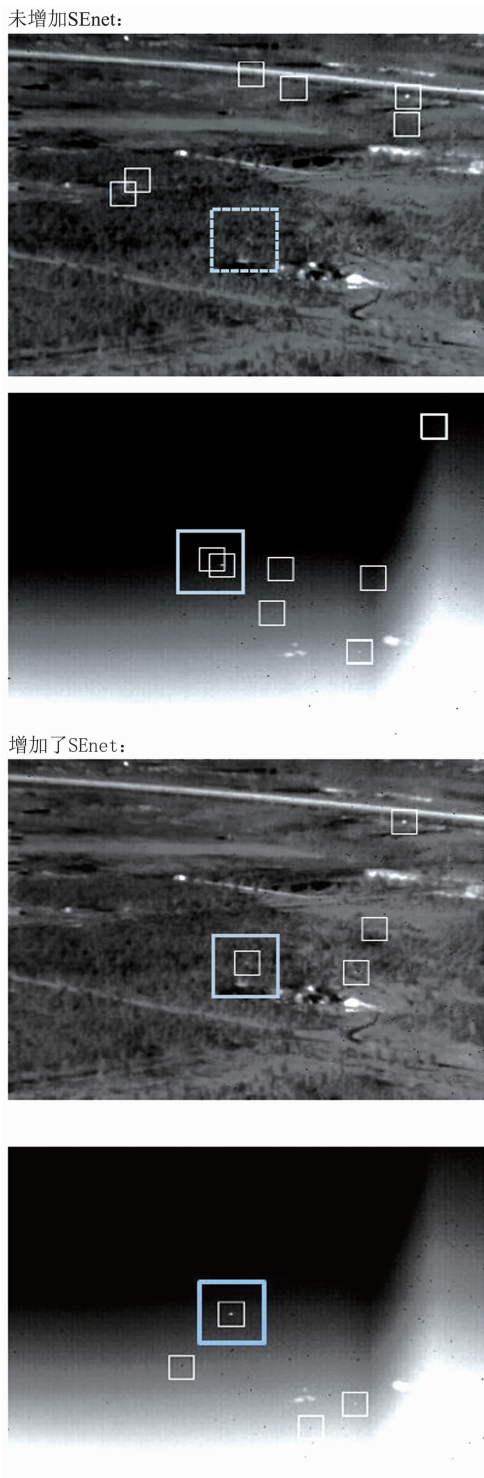


图 14 真实场景检测图
Fig. 14 Detection results in real scene

色虚线框表示未检测到,第一幅地面背景的图片表示目标产生运动模糊时的检测情况,第二种情况下展示了图像存在部分的干扰点时,由此产生的虚警情况.改进后的网络结构还是取得了更好的检测效果.

为了比较我们所设计的基于深度卷积网络的单帧检测算法相对于传统单帧检测算法在检测性能上的优势,我们比较了最终所设计的模型与目前常用的 Top-hat, 环形去中值滤波, 单阈值 PDE 以及 MRF 算法在测试集上的检测率和虚警率上的实验结果, 如表 8 和表 9 所示, 表中的 A, B, C, D 分别表示 Top-hat, 环形去中值滤波, 单阈值 PDE 以及 MRF 算法.

表 8 不同检测算法在测试集上检测效果比较

Table 8 Comparison of detection rate among the detection algorithms

SNR	1	2	3	4	5	6	7
A	52.6%	63.3%	75.4%	85.7%	87.6%	89.1%	90.6%
B	51.4%	63.1%	74.8%	81.2%	86.6%	87.3%	89.1%
C	79.8%	84.6%	89.0%	93.4%	95.7%	96.3%	97.0%
D	82.3%	85.6%	91.2%	93.6%	96.0%	96.5%	97.7%
Ours	92.9%	94.3%	95.3%	96.2%	97.0%	98.1%	98.2%

表 9 不同检测算法在测试集上虚警个数比较

Table 9 Comparison of false alarms among the different detection algorithms

SNR	1	2	3	4	5	6	7
A	14	16	14	12	12	13	12
B	12	15	13	14	12	12	13
C	9	10	11	11	12	11	11
D	8	8	10	8	9	10	9
Ours	7	7	8	8	8	9	8

从统计结果中可以看到我们所设计的算法在检测率和虚警率上均优于对比的方法, 而且克服了对比算法中所需的图像分割中的阈值选取以及 PDE 和 MRF 中参数的选择问题. 我们所设计的网络具有良好的检测率和对虚警率的抑制, 对目标信噪比的变化具有相当高的鲁棒性. 我们所设计的训练集制作方法也为用深度学习进行小目标检测的训练集设计上提供了参考.

4 结论

提出了一种对复杂背景下低信噪比且存在运动模糊的小目标具有高性能检测能力的深度卷积神经网络结构. 提出首先采用增强网络对目标点进行初步筛选的方法, 该网络对目标点进行初步判别, 使只有判别概率较大的候选目标点进入到后续的分类网络中, 减少了分类网络对目标点的搜索次数. 在对目标位置分布进行分类时, 我们考虑到小目标特性在原始图像以及去中值滤波背景抑制后的图像上的差异, 采用原始图像和背景抑制后图像双输入的方式, 该方式与单输入的方式相比, 提高了网络检测性能;

针对实际场景中目标运动引起的运动模糊问题,在双输入网络的特征图融合阶段增加了SENet对特征图筛选,使网络对不同情况下的目标有不同的特征侧重,该结构改善了网络在实际场景中对运动模糊目标的漏检情况。最终设计的网络结构在测试结果上令人满意,可以作为一种利用深度卷积网络在空域上对红外小目标进行检测的参考方法。

References

- [1] Chen C L P, Li H, Wei Y, *et al.* A local contrast method for small infrared target detection [J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2013, **52**(1): 574–581.
- [2] Yang C, Ma J, Qi S, *et al.* Directional support value of gaussian transformation for infrared small target detection [J]. *Applied Optics*, 2015, **54**(9): 2255–2265.
- [3] Dehghani A, Pourmohammad A. Small target detection and tracking based on the background elimination and Kalman filter [C]. International symposium on artificial intelligence and signal processing, 2015:328–333.
- [4] Qi H, Mo B, Liu F, *et al.* Small infrared target detection utilizing local region similarity difference map [J]. *Infrared Physics & Technology*, 2015, **71**: 131–139.
- [5] Deshpande S D, Er M H, Venkateswarlu R, *et al.* Max-mean and max-median filters for detection of small targets [C]. SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, 1999, **3809**:10.
- [6] Bai X, Zhou F. Analysis of new top-hat transformation and the application for infrared dim small target detection [J]. *Pattern Recognition*, 2010, **43**(6): 2145–2156.
- [7] Wang P, Tian J W, Gao C Q. Infrared small target detection using directional highpass filters based on ls-svm [J]. *Electronics Letters*, 2009, **45**(3): 156–158.
- [8] LI L, TANG Y Y. Wavelet-hough transform with applications in edge and target detections [J]. *International Journal of Wavelets Multiresolution & Information Processing*, 2006, **4**(03): 567–587.
- [9] Wang T, Chen F S, Xiao-Feng S U, *et al.* Infrared dim target detection based on partial differential equation [J]. *Laser & Infrared*(王涛, 陈凡胜, 苏晓峰, 等. 基于偏微分方程的红外弱小目标检测技术研究. *激光与红外*), 2016, **46**(7):885–889.
- [10] XUE Y H, RAO P, FAN S W, *et al.* Infrared dim small target detection algorithm based on generative Markov random field and local statistic characteristic [J]. *Journal of Infrared & Millimeter Waves*(薛永宏, 饶鹏, 樊士伟, 等. 基于生成 MRF 和局部统计特性的红外弱小目标检测算法. *红外与毫米波学报*), 2013, **32**(5): 431–436.
- [11] Ren S, He K, Girshick R, *et al.* Faster r-cnn: towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2017, **39**(6): 1137–1149.
- [12] Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection [C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016:779–788.
- [13] Liu W, Anguelov D, Erhan D, *et al.* Ssd: Single shot multibox detector [C]. European conference on computer vision, 2016:21–37.
- [14] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [J]. *arXiv preprint arXiv:170901507*, 2017, 7.
- [15] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015:3431–3440.
- [16] Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation [C]. IEEE International Conference on Computer Vision, 2015:1520–1528.
- [17] Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps [J]. *arXiv preprint arXiv:13126034*, 2013.
- [18] He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition [C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016:770–778.
- [19] Szegedy C, Ioffe S, Vanhoucke V, *et al.* Inception-v4, inception-ResNet and the impact of residual Connections on Learning [C]. AAAI, 2017, **4**:12.
- [20] Srivastava R K, Greff K, Schmidhuber J. Training very deep networks [C]. Advances in neural information processing systems, 2015:2377–2385.
- [21] Srivastava N, Hinton G, Krizhevsky A, *et al.* Dropout: a simple way to prevent neural networks from overfitting [J]. *Journal of Machine Learning Research*, 2014, **15**(1): 1929–1958.

(上接第 350 页)

- [8] Kim S, Neikirk D P. Compact equivalent circuit model for the skin effect [C]. *IEEE MTT-S International Microwave Symposium Digest*, June 1996.

- [9] Mei S, Ismail Y I. Modeling skin and proximity effects with reduced realizable RL circuits [J]. *IEEE Transactions on very large scale integration (VLSI) systems*, 2004, **12**(4): 437–447.