

文章编号:1001-5078(2024)08-1309-07

· 图像与信号处理 ·

基于卷积神经网络的目标检测与识别技术

周宽,耿宇飞,金旭,刘纪洲,任静
(中电科光电科技有限公司,北京 100015)

摘要:为解决传统的目标检测与识别方法在复杂战场环境中的限制,本文提出了一种基于深度卷积神经网络(CNN)的方法,用于提取图像特征和定位目标。该方法综合考虑了传统的形状和纹理特征以及红外图像中的热点信息,通过大规模标注数据集的训练和反向传播算法的优化,提高了目标检测和识别的准确性。相比于传统方法,该方法能够自动学习图像中的特征表示,无需依赖手工设计的特征和分类器。为了验证算法的有效性,本文选用海思Hi3559AV100作为核心处理芯片设计硬件平台,通过将算法移植到该平台上,对收集到的数据样本进行分析测试,实验结果表明,该系统在复杂的背景环境中表现出相对稳定的性能,能够可靠地进行目标检测和识别。

关键词:深度学习;目标检测;卷积神经网络;红外图像

中图分类号:TP391.41;TN976 **文献标识码:**A **DOI:**10.3969/j.issn.1001-5078.2024.08.019

Convolutional neural network based on target detection and recognition technique

ZHOU Kuan, GENG Yu-fei, JIN Xu, LIU Ji-zhou, REN Jing
(CETC Electro-Optics Technology Co. Ltd., Beijing 100015, China)

Abstract: To address the limitations of traditional target detection and recognition methods in complex battlefield environments, a deep convolutional neural network (CNN)-based method for extracting image features and localizing targets is proposed in this paper. Traditional shape and texture features as well as the hotspot information in infrared images are comprehensively considered, and the accuracy of target detection and identification is improved through the training of large-scale labeled datasets and the optimization of the back propagation algorithm. Compared with traditional methods, the method can automatically learn the feature representations in images without relying on manually designed features and classifiers. In order to verify the effectiveness of the algorithm, this paper selects the Hathi Hi3559AV100 as the core processing chip to design the hardware platform, and by porting the algorithm to this platform, the collected data samples are analyzed and tested, and the experimental results show that the system exhibits relatively stable performance in complex background environments, and is able to reliably perform target detection and recognition.

Keywords: deep learning; target detection; convolutional neural networks; infrared images

1 引言

随着现代军事技术的不断进步和战争形态的演

变,红外传感技术在军事领域中的应用日益广泛。红外热成像成为获取目标信息的重要手段,可用于

目标检测、识别与跟踪等任务,为作战决策提供准确、实时的情报支持^[1]。传统的红外观瞄系统中只提供了图像显示功能,而缺乏自动识别和跟踪目标的能力。随着无人作战需求的增加,目标识别跟踪技术在自动驾驶、监控等领域得到广泛运用^[2]。通过将红外热成像与识别跟踪功能相结合,并利用图像处理 and 模式识别技术对红外热像仪采集到的图像进行分析和处理,系统能够自动检测和跟踪感兴趣的目标,从而实现自动瞄准和锁定目标的功能。这极大地提高了作战系统的效率和精确度,同时减少了对人工干预的需求,实现了自主决策和行动的能力。

传统的目标检测与识别方法通常采用精心设计的手工特征描述符(如 SIFT、HOG)描述图像,然后训练强判别力的分类器(如 Adaboost、SVM)从背景中区分感兴趣目标^[3]。然而,这些方法在复杂场景中的表现受限,难以应对目标变化和背景干扰较大的情况,限制了其在实战中的应用。为了克服传统方法的局限性,许多研究者提出了针对红外图像的各种深度学习模型和算法,通过参考前人方法^[4-8],本文提出了一种基于深度卷积神经网络的红外图像目标检测与识别方法,相比于现有方法具有以下优点:

(1)在网络结构方面,设计了一种全新的卷积神经网络,分别用于特征提取和目标分类定位。特征提取网络采用深度残差网络(ResNet)作为基础结构,分类定位子网络采用基于锚点(anchor)的思想,无需训练额外的网络用于提取候选窗口,能够有效提高整个系统的计算效率;

(2)在训练方法方面,采用神经网络框架 PyTorch 实现网络训练。使用 ResNet-50-FPN 在公开数据集 ImageNet 1K 上进行图像分类任务的训练,利用大规模标注数据集训练网络,采用反向传播算法优化参数,通过迭代实现目标特征学习,提高检测和识别的准确率;

(3)在前向计算方面,使用不同于网络训练框架的其他前向计算框架提高前向计算的效率。输入红外图像通过网络实现快速目标类别和位置解析,满足战场环境下快速响应和高准确性的要求。

相较于现有方法,该方法在网络设计、训练方法等方面都做了一定创新,提高了整体系统性能。为

了验证算法的有效性,本文选用国产海思 Hi3559AV100 为核心处理芯片搭建硬件平台,通过将算法移植到硬件平台上进行实际环境测试,更加真实地评估出了算法在实际应用中的性能,该技术的成功应用将为红外热像仪的发展和军事作战情报的支持奠定基础。

2 深度卷积神经网络

2.1 网络结构设计和优化

本文采用单阶段的卷积神经网络来实现目标的检测与识别,网络结构如图 1 所示。

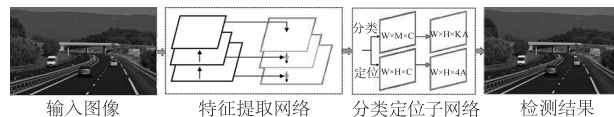


图 1 网络结构示意图

Fig. 1 Network architecture diagram

输入图像经过特征提取网络和分类定位子网络,即可获得准确的检测结果,无需训练额外的网络用于提取候选窗口,从而有效提高整个系统的计算效率。网络结构包括特征提取网络和分类定位子网络两部分内容。特征提取网络是网络结构的首要步骤,通过层层堆叠的卷积层、池化层以及激活函数等构件,从原始数据中提取具有代表性的特征。这些特征能够捕捉数据中的关键信息,使得模型能够更好地理解和表达输入数据。特征提取网络的设计和优化直接影响着网络的性能和泛化能力,在图像识别和目标检测等任务中发挥着至关重要的作用^[9]。与此同时,分类定位子网络构成了网络结构的后续部分,通过一系列的全连接层和 softmax 函数等构件,将提取到的特征映射到具体的类别或位置^[10]。分类定位子网络通过学习数据中的关联规律和特征之间的关系,能够对输入数据进行准确的分类和定位,实现对不同类别和位置的精确识别。

2.1.1 特征提取网络

特征提取网络采用了深度残差网络作为基础结构,并结合特征金字塔网络(FPN)从单幅输入图像中提取多尺度的丰富图像特征。不同尺度的图像特征可以用于检测不同尺度的感兴趣目标。

深度残差网络采用了经典的 ResNet-50 结构,它由浅到深逐渐提取中高层语义特征,并引入了残差学习来防止网络过拟合,为后续的目标分类和定位提供了良好的基础^[11]。根据实际任务和训练数

据的规模,可以调整深度残差网络的层数,便于在计算效率和检测精度之间取得平衡。

特征金字塔网络以深度残差网络的第3到第5个残差模块(即C3、C4、C5)的输出为基础,构建了一个5层的特征金字塔。这些特征金字塔分别表示步长为 $\{8, 16, 32, 64, 128\}$ 的下采样特征图。特征金字塔使系统能够有效处理在战场上目标随着探测距离的远近而发生较大尺度变化的挑战。同时,它也将多尺度检测的问题融入整个网络的学习过程中,具有更强的灵活性。

2.1.2 分类定位子网络

分类定位子网络采用基于锚点的思想,以特征提取网络输出的特征图为输入,通过若干卷积层,为特征图中每一个位置做固定数量的预测。如图2所示,每一个预测都是以特征图某个位置为中心不同尺度、不同长宽比的矩形框,由于每个预测矩形框与其中心位置的相对偏移是固定的,因此将这种预测矩形框称为锚点。对每一个锚点,分类子网络输出的是与类别相关的分类结果,若目标识别的类别数量为 K ,则其输出为长度为 K 的向量,其中每个元素表示该锚点所在图像区域属于特定类别的概率。定位子网络输出的是一个长度为4的向量,表示该锚点相对于真正的目标矩形框在横纵坐标和长宽4个维度上的偏移。若锚点数量为 A ,则对于一个大小为 $W \times H$ 的特征图,分类子网络的输出为 $W \times H \times KA$ 的张量,定位子网络的输出为 $W \times H \times 4A$ 的张量,后续经过相应后处理即可获得最终检测结果。

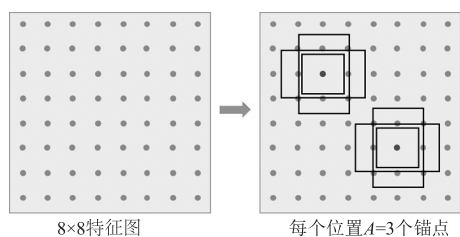


图2 锚点示意图

Fig. 2 Schematic diagram of anchor point

2.2 训练方法和数据集介绍

深度学习算法的训练数据对于网络性能至关重要。卷积神经网络的训练通常需要大量的训练数据支持^[12]。尽管通过精细的网络结构和目标函数设计可以减轻其对大数据的依赖,但增加训练数据规模和提高数据标注的精度仍然有助于提高

网络的最终输出结果。通过充分利用大规模训练数据和适当的预训练策略,网络可以学习到更丰富的特征表示,包括不同场景、不同天气条件和不同目标等,从而提高战场目标检测与识别任务的性能。

2.2.1 训练数据

针对战场环境的目标检测与识别任务,训练数据可以表示为一定数量的图像与其对应的标注。对于训练数据的标注,每幅图像都会有对应的目标矩形框标注,来表示感兴趣目标的位置和大小。这些目标矩形框可以通过标注工具进行人工标注获得。为了减少训练数据和标注图像的工作量,本文采用预训练和微调策略,将已有的公开数据集用于网络的初始化和预训练,从而加快网络的收敛速度以提高其性能。

具体地,首先使用 ResNet-50-FPN 在公开数据集 ImageNet 1K 上进行图像分类任务的训练,获得对应网络层的参数用于对特征提取网络进行初始化,其他卷积层采用高斯分布进行初始化。在公开数据集 MS-COCO 2017 上面向通用目标检测与识别任务对整个网络进行预训练,保留除分类定位子网络最后一层卷积层之外的所有网络参数。

面向战场目标检测与识别任务,在自主收集和标注的数据集上对网络进行微调,同时使用数据扩增方法提高数据的多样性。根据数据集的规模和初步训练结果,可选择只微调分类定位子网络以降低训练难度。

2.2.2 网络优化

本文采用神经网络框架 PyTorch 实现网络训练。网络参数的优化方法采用小批量随机梯度下降(mini-batch SGD),并设置 0.0001 的权重衰减和 0.9 的动量加权。迭代次数和学习率根据训练数据规模确定,并使用退火策略根据迭代次数逐步降低学习率。

网络训练的目标函数由分类误差和定位误差加权求和组成。分类误差表示网络对每一个锚点的类别预测是否与标注相符合,使用交叉熵(Cross Entropy)结合 softmax 函数可以计算得到。定位误差表示网络对每一个锚点的真实位置与大小预测是否与标注相符合,使用 Smooth L1 损失函数计算得到。为了确定锚点预测的矩形框与标注的目标矩形框之间

的对应关系,采用以下匹配方法:

(1) 针对每个锚点,计算该锚点与目标矩形框之间的重叠率,选取与该锚点重叠率最高的目标矩形框,若重叠率高于 0.5,则认为两者匹配,该锚点负责预测此目标矩形框对应的目标,可计算对应的分类与定位误差;

(2) 若该锚点与所有目标矩形框之间的重叠率都低于 0.4,则认为该锚点负责预测背景,可计算分类误差,定位误差为 0;

(3) 若该锚点与所有目标矩形框之间的重叠率都在区间 $[0.4, 0.5]$ 内,则认为该锚点不利于检测与分类,在训练时忽略该锚点。

通过以上匹配方法,可以将每个锚点与相应的目标矩形框进行对应,从而计算出分类误差和定位误差,用于训练网络。

2.3 前向计算

本文采用的深度卷积神经网络是一个全卷积神经网络,不包含全连接层。网络的前向计算是指将输入数据通过网络,逐层传递并计算输出结果的过程。在战场目标检测与识别任务中,前向计算的目的是获取网络对输入图像进行目标检测和识别的结果。

前向计算的过程即是將一幅图像输入神经网络,经过特征提取网络和分类定位子网络,获得网络输出结果的过程。为了获得最终的检测结果,需要对网络输出结果进行一定的后处理操作,具体操作如下:

针对特征金字塔中的每一层特征图,分类子网络的输出张量表示该层特征图中所有锚点预测的矩形框的类别得分,定位子网络的输出张量表示该层特征图中所有锚点预测的矩形框的位置偏移。

为了提高计算效率,首先对所有类别得分大于 0.05 的预测矩形框进行排序,只保留得分最高的前 1000 个预测矩形框。然后,根据定位子网络输出的位置偏移,计算这些预测矩形框的最终位置和大小。接下来,将所有尺度的特征图上的预测矩形框进行非极大值抑制操作。通过设置一定的重叠率阈值(典型值为 0.5),过滤掉过度聚集的矩形框,从而获得最终的检测结果。

此外,为了保证系统的实时性,可以使用与网络

训练框架不同的前向计算框架(如 Tensor RT、PaddlePaddle 等),以提高前向计算的效率。

3 实验结果分析

3.1 硬件平台

为了验证本文算法的有效性,并进一步评估其性能,本文选用国产海思 Hi3559AV100 作为核心处理芯片,并搭建了硬件平台来支持目标检测与识别算法的运行,其硬件原理框图如图 3 所示。

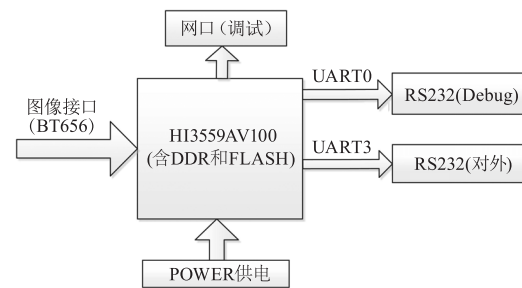


图 3 HI3559 识别跟踪电路功能原理框图

Fig. 3 Functional block diagram of HI3559 identification tracking circuit

该硬件平台主要功能包括接收红外热像仪发送的 BT656 图像数据,并根据指令选择跟踪目标,在选择目标后,平台会运行目标跟踪算法,并返回跟踪结果。图像数据通过输入接口直接与 HI3559AV100 芯片相连。此外,硬件平台还配备了两个存储器:FLASH 芯片用于存放操作系统启动和运行的代码,DDR 芯片则用于临时存储 ARM 处理器算法运行过程中的运算数据。

3.2 数据集的准备和预处理

为了进行数据采集和实验验证,本文使用实验室自制的中/长波双波段红外探测器组件,并搭载双色红外成像系统进行数据采集。

实验环境配置:本文提出的算法在深度学习服务器上实现,服务器配置为双核 ARM Cortex A73@1.6 GHz + 双核 ARM Cortex A53@1.2 GHz,支持 Neon 加速,集成 FPU 处理单元;GPU 双核 ARM Mali G71@900 MHz,256 kB cache 提供视觉计算处理能力,双核 NNIE@84 0MHz 神经网络加速引擎^[13]。

实验数据集介绍:通过使用双色红外成像系统成功收集了 100 个包含红外图像的数据样本。为了更好地理解和利用这些数据,本文对每个样本进行人工标注,标注的过程包括目标的位置和类别信息的标记。通过人工标注的方式将样本与其真实标签进行对比,从而确定算法的准确率。

3.3 评价指标

根据统计样本特征,可以计算出一些重要的指标值,如 FN、FP、TP 和 TN。其中,TP(True Positive)表示模型正确预测为正类的样本数,FP(False Positive)表示模型错误预测为正类的样本数(即模型误判),FN(False Negative)表示模型错误预测为负类的样本数(即模型漏判),TN(True Negative)表示模型正确预测为负类的样本数(由于模型不存在判断反例的能力,因此该指标在模型中不存在)。为了直观地评估检测效果,本文采用了三种常用的评价指标,分别是准确率、召回率和 F1 值^[14],其定义如下:

(1) 准确率 (Accuracy): 准确率是指分类正确的样本数与总样本数之比。它衡量了模型正确分类的能力,准确率的计算公式如下:

$$\text{准确率} = \frac{TP + TN}{Total} \times 100 \% \quad (1)$$

(2) 召回率 (Recall): 召回率是指被正确分类为正类的样本数与实际正类样本数之比。它衡量了模型对正类样本的识别能力,召回率的计算公式如下:

$$\text{召回率} = \frac{TP}{TP + FP} \times 100 \% \quad (2)$$

(3) F1 值 (F1 score): F1 值综合考虑了准确率和召回率,是二者的调和平均。F1 值越高,表示模型在准确率和召回率之间取得了更好的平衡^[15], F1 值的计算公式如下:

$$F1 \text{ 值} = \frac{2 \times (\text{准确率} \times \text{召回率})}{\text{准确率} + \text{召回率}} \quad (3)$$

3.4 实验结果分析

图 4 展示了本文算法在不同区域的检测识别效果。检测结果显示,100 个样本中共有 11 个样本被误判,15 个样本被漏判。

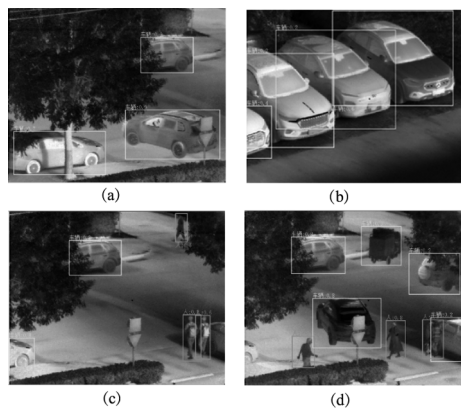


图 4 真实场景检测图

Fig. 4 Real scene inspection chart

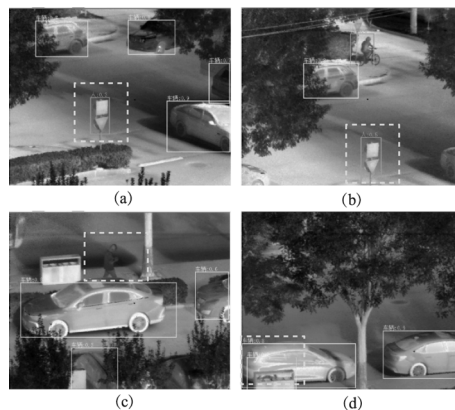


图 5 部分被误判分类的样本

Fig. 5 Partially misclassified samples

图 5 展示了其中一部分被误判、漏判分类的样本,图中用虚线方框标示了这些误判和漏判的目标。在图 5 中的图 (a) 和图 (b) 中,能够观察到误判的正样本中包括将路牌误认为行人标记,这种误判情况相对较高,可能是由于路牌的形状、颜色或纹理与行人目标相似,导致算法错误地将其分类为行人。在图 (c) 中,可以看到由于背景较暗,车辆上方的行人目标没有被正确识别出来。这可能是由于光照条件不足或行人与背景之间的对比度较低,导致算法难以准确地检测和识别行人。在图 (d) 中,观察到左下角的车辆目标与垃圾桶重叠,并被错误地判定为车辆目标。这种情况可能是由于目标之间的遮挡或重叠导致算法难以准确地分离它们。

通过对这些误判和漏判的情况进一步分析,可以识别出算法在特定场景或特定目标上可能存在的局限性。接下来引用准确率、召回率和 F1 值三种评价指标对算法的识别和分类能力进行衡量考核,如表 1 所示为统计的样本数中的 FN、FP、TP 和 TN 指标值。根据表 1 给定的数据计算得出的准确率、召回率和 F1 值如表 2 所示。

表 1 样本指标值

Tab. 1 Sample indicator values

样本数量	TP	FP	FN	TN
20	15	2	3	0
40	29	4	7	0
60	43	7	10	0
80	57	10	13	0
100	74	11	15	0

表 2 评价指标值
Tab. 2 Evaluation indicator values

样本数量	准确率/%	召回率/%	F1 值/%
20	75.00	83.33	78.91
40	72.50	80.56	76.33
60	71.70	81.10	76.12
80	71.25	81.40	75.94
100	74.00	83.10	78.31

为了更直观地分析算法的性能,图 6 和图 7 分别给出了样本指标曲线变化图,从图中可以看出,在样本数量较少的情况下,随机性对算法准确率的影响较大。随着样本数量的增加,准确率和召回率逐渐趋于平衡,可以看出算法的性能相对较好。

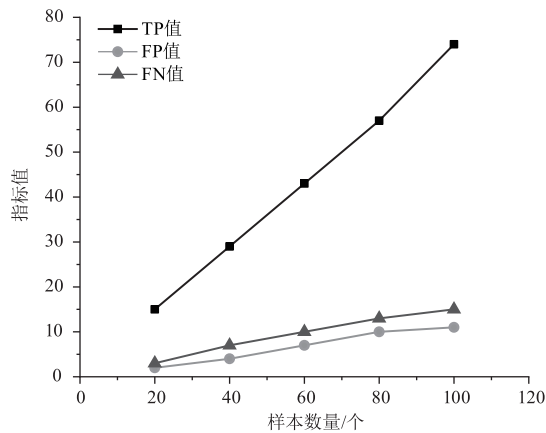


图 6 样本指标值

Fig. 6 Sample indicator values

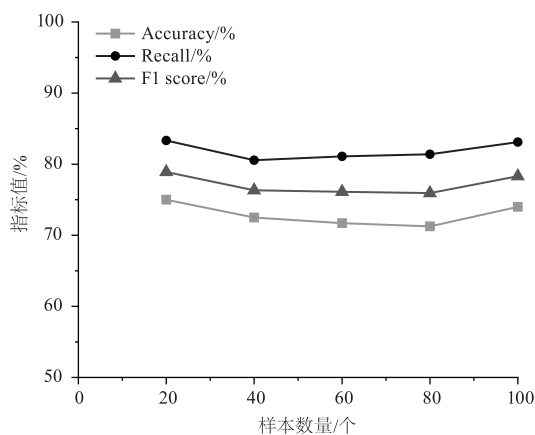


图 7 评价指标值

Fig. 7 Evaluation indicator values

准确率在不同样本数量下的范围大约在 71 % 到 75 % 之间,这意味着识别分类算法在样本数量较少的情况下具有相对较高的准确性。召回率在不同样本数量下的范围大约在 80 % 到 83 % 之间,

这表示识别分类算法对于识别正类样本的能力相对较强,尽可能准确地捕捉到了真实目标。F1 值综合了准确率和召回率,是这两个指标的调和平均值。F1 值在不同样本数量下的范围大约在 76 % 到 79 % 之间,这表明识别分类算法在样本数量较少的情况下具有较好的平衡性能。

4 结论

本文提出了一种基于深度卷积神经网络的红外图像目标检测与识别方法,该方法通过综合考虑传统的形状和纹理特征以及红外图像中的热点信息,能够自动学习图像中的特征表示,无需依赖手工设计的特征和分类器。通过大规模标注数据集的训练和反向传播算法的优化,提高了目标检测和识别的准确性。通过在国产海思 Hi3559AV100 核心处理芯片搭建的硬件平台上进行真实环境测试,实验结果表明,在样本数量从 20 增加到 100 的过程中,准确率在 71 % ~ 75 % 之间,召回率在 80 % ~ 83 % 之间,F1 值在 76 % ~ 79 % 之间。根据统计特征,数值的波动情况随着数值偏大,逐渐趋向于稳定,系统的性能相对较好,具有很好的环境适应性,能够可靠地进行目标检测和识别,为解决红外热像仪在复杂环境中难以准确识别目标提供了基础。

参考文献:

- [1] Bingrong Z. The research on infrared ship recognition and track technology under the sea clutter [D]. Changsha: National University of Defense Technology, 2011. (in Chinese) 卓炳荣. 海面背景下的红外舰船目标识别跟踪技术研究 [D]. 长沙:国防科技大学, 2011.
- [2] Keqi C, Zhiliang Z, Xiaoming D, et al. A review of deep learning research on multiscale target detection [J]. Journal of Software, 2021, 32(4): 1201 - 1227. (in Chinese) 陈科圻, 朱志亮, 邓小明, 等. 多尺度目标检测的深度学习研究综述 [J]. 软件学报, 2021, 32(4): 1201 - 1227.
- [3] Guan Xiaobin, Li Zhanming. Vehicle detection algorithm based on video SIFT and HOG feature fusion [J]. Computer and Digital Engineering, 2021, 49(6): 1113 - 1117. (in Chinese) 关晓斌, 李战明. 基于 SIFT 和 HOG 特征融合的视频车辆检测算法 [J]. 计算机与数字工程, 2021, 49(6): 1113 - 1117.
- [4] Liu K J, Ma R S, Ma Z M, et al. Design of infrared target

- recognition system with optimized convolutional neural network[J]. *Optics and Precision Engineering*, 2021, 29(4):822–831. (in Chinese)
- 刘可佳,马荣生,唐子木,等. 采用优化卷积神经网络的红外目标识别系统[J]. *光学精密工程*, 2021, 29(4):822–831.
- [5] Wang K, Li S, Niu S, et al. Detection of infrared small targets using feature fusion convolutional network[J]. *IEEE Access*, 2019, 7:146081–146092.
- [6] Zhang L, Han P, Xi J, et al. Infrared small target detection based on a temporally-aware fully convolutional neural network[J]. *Remote Sensing*, 2023, 15(17):4198.
- [7] Wang W T, Qin H, Cheng W, et al. Small target detection in infrared image using convolutional neural networks [C]//AOPC 2017:Optical Sensing and Imaging Technology and Applications. SPIE, 2017, 10462:1335–1340.
- [8] Tongtong J, Jinyong C, Wenpeng L. Target recognition based on multilayer feature extraction of convolutional neural network[J]. *Computer Systems & Applications*, 2017, 26(12):64–70. (in Chinese)
- 江彤彤,成金勇,鹿文鹏. 基于卷积神经网络多层特征提取的目标识别[J]. *计算机系统应用*, 2017, 26(12):64–70.
- [9] Li J, Du J, Zhou N. Image caption algorithm based on an attention image feature extraction network[J]. *Journal of Nanjing University of Information Science & Technology*, 2019, 11(3):295–301. (in Chinese)
- 李金轩,杜军平,周南. 基于注意力特征提取网络的图像描述生成算法[J]. *南京信息工程大学学报*, 2019, 11(3):295–301.
- [10] Han G, Xu H, Duong T Q, et al. Localization algorithms of wireless sensor networks: a survey[J]. *Telecommunication Systems*, 2013, 52:2419–2436.
- [11] Liu X Y, Liu J M, Liu C, et al. Novel botnet DGA domain detection method based on character level sliding window and deep residual network[J]. *Acta Electronica Sinica*, 2022, 50(1):250–256. (in Chinese)
- 刘小洋,刘加苗,刘超,等. 融合字符级滑动窗口和深度残差网络的僵尸网络 DGA 域名检测方法[J]. *电子学报*, 2022, 50(1):250–256.
- [12] Renganathan S A, Maulik R, Ahuja J. Enhanced data efficiency using deep neural networks and Gaussian processes for aerodynamic design optimization[J]. *Aerospace Science and Technology*, 2021, 111:106522.
- [13] Wang K, Li S, Niu S, et al. Detection of infrared small targets using feature fusion convolutional network[J]. *IEEE Access*, 2019, 7:146081–146092.
- [14] Che B, Zhou D. Tag recommendation method combining network structure information and text content[J]. *Journal of Computer Applications*, 2021, 41(4):976–983. (in Chinese)
- 车冰倩,周栋. 融合网络结构信息及文本内容的标签推荐方法[J]. *计算机应用*, 2021, 41(4):976–983.
- [15] Shan Wei, Wang Jiangtao, Chen Debao, et al. Human detection in infrared image based on deep convolution neural network[J]. *Laser & Infrared*, 2020, 50(5):634–640. (in Chinese)
- 单巍,王江涛,陈得宝,等. 基于深度卷积神经网络的红外图像行人检测[J]. *激光与红外*, 2020, 50(5):634–640.