

# 结合红外显著性目标导引的 改进 YOLO 网络的智能装备目标识别研究

侯毅苇<sup>1,2</sup>, 李林汉<sup>2</sup>, 王彦<sup>3</sup>

(1. 河北金融学院, 大数据科学学院, 河北 保定 071051;

2. 河北金融学院, 金融创新与风险管理研究中心, 河北 保定 0710051;

3. 中国电子科技集团公司第五十四所 信息传输与分发技术重点实验室, 河北 石家庄 050081)

**摘要:** 为了提升实际作战环境下目标检测识别的性能, 本文提出了一种基于红外显著性目标导引的改进 YOLO (You Only Look Once) 网络的智能装备目标识别算法, 该算法利用红外图像提供目标可能的位置引导可见光图像中的深度自主学习, 提升检测与识别的实时性。改进 YOLO-V3 识别网络是以 Darknet-53 为基础网络架构, 利用 Dense Net 代替具有较低分辨率的原始转移层, 同时采用分类网络预训练、多尺度检测网络训练等措施增强特征传播, 复用和融合的性能。仿真实验结果表明, 本文提出的模型可以有效地提高现有目标检测与识别的性能。

**关键字:** 目标识别; 红外显著性; 目标导引; 深度学习; YOLO-V3; 智能装备

中图分类号: TN219 文献标识码: A 文章编号: 1001-8891(2020)07-0644-07

## Intelligent Equipment Object Recognition Based on Improved YOLO Network Guided by Infrared Saliency Detection

HOU Yiwei<sup>1,2</sup>, LI Linhan<sup>2</sup>, WANG Yan<sup>3</sup>

(1. School of Big Data Science, Hebei Finance University, Baoding 071051, China;

2. Financial innovation and Risk Management Research Center, Hebei Finance University, Baoding 071051, China;

3. Key Laboratory of Information Transmission and Distribution Technology, The 54th Research Institute of CETC, Shijiazhuang 050081, China)

**Abstract:** To improve the performance of object detection and recognition in a real-world combat environment, an improved intelligent object recognition algorithm based on infrared saliency object guidance is proposed. It uses the object information in an infrared image to guide deep self-learning in vision images. The improved YOLO-V3 recognition network is based on the Darknet-53 network architecture, using DenseNet instead of the original transfer layer with lower resolution. Classification network pretraining, multiscale detection network training, and other measures are used to enhance feature propagation and reuse and fusion performance. Simulation results show that the proposed model can effectively improve the performance of existing object detection and recognition networks.

**Key words:** object recognition, infrared saliency, object guidance, deep learning, YOLO-V3, intelligent equipment

## 0 引言

随着新军事变革的不断推进和发展, 战场形态将走向智能化与信息化, 主要作战方式也朝着整个武器装备体系间的对战发展, 未来装备对于目标自主捕获功能的需求越来越迫切<sup>[1]</sup>。然而, 现有装甲装备火控

系统大多数采用人为指定目标, 启动跟踪器实施对特定目标的跟踪, 无法对可能出现的威胁目标进行检测与识别。因此, 如何提高火控系统的目标自主/半自主识别性能将是未来智能装甲装备的发展方向<sup>[2]</sup>。

目标自主检测与识别是指利用机器学习及人工智能方法在图像中自动获取目标类别和位置。现有的

收稿日期: 2019-09-08; 修订日期: 2020-07-06.

作者简介: 侯毅苇 (1980-), 女, 汉族, 山东菏泽人, 硕士, 讲师, 研究方向: 计算数学、图像处理、智能计算应用、模式识别等。E-mail: hyiwei1983@126.com.

基金项目: 河北省自然科学基金青年科学基金 (A2015410006); 2018 年度河北省科学技术厅软科学研究专项项目 (18454227); 河北金融学院应用数学优秀基础学科基金项目 (20190235A)。

方法大都是对典型目标,如飞机、来袭导弹、舰船、桥梁等进行识别。在实战环境下,实时预警检测系统需要具备复杂背景中广域目标探测与识别<sup>[3]</sup>。由于缺乏目标的先验信息,现有算法存在虚警率高、实时性偏低等问题,限制了实战环境下的广泛应用。迄今为止,国内外广域目标检测与识别项目仍然处于试验阶段,还没有军品上可靠的型号产品,迫切要求研究大范围远距离成像条件下的稳健目标检测识别算法。

众所周知,复杂背景下自主识别能力和实时性成为制约人工智能技术实用性的关键。现有的大多数算法是利用低层次特征进行支持向量机(Support Vector Machine, SVM)分类。文雄志等人提出了一种基于河流先验信息的桥梁识别方法,该方法利用桥梁大概率位于河流之上的先验信息,通过提取河流上的疑似区域特征,然后通过分类器的判断,实现桥梁的检测<sup>[4]</sup>。Yao等人利用机场跑道特有特征,设计了一种多尺度模式分类方法,能够从大范围搜索区域中识别目标<sup>[5]</sup>。虽然这些方法已经能够较高精度地识别典型目标,但该类方法对图像的质量与目标特性要求较高。一旦目标的形状改变,识别率大大降低。因此,现有的装备还并不具备完全依靠系统自身识别能力进行打击。航天九院的出口型“彩虹”无人机仍然是利用数据链由人在回路进行目标打击,其识别系统主要用于辅助识别。目前,以卷积神经网络为代表的深度学习方法已经在可见光图像识别领域取得了不错的成绩,从RCNN<sup>[6]</sup>、SPP-Net<sup>[7]</sup>、Fast-RCNN<sup>[8]</sup>、Faster-RCNN<sup>[9]</sup>、YOLO<sup>[10]</sup>、SSD<sup>[11]</sup>、YOLO-v2<sup>[12]</sup>到YOLO-v3<sup>[13]</sup>正在逐步刷新目标检测与识别的精度和速度。与传统人工设计特征不同,深度网络通过非线性网络结构逐层学习潜在特征,获得目标最本质的特征信息。

由于地面装备的特殊性,直接将现有模型应用到目标检测与识别中,效果不太理想。首先,车载装备需要实时的对目标进行识别,而神经网络模型复杂度太高,很难满足实时性的要求;其次,车辆行驶过程

的烟尘严重影响成像质量,导致目标识别率不高。因此,针对深度学习模型很少直接应用于装甲光电系统,且实时性较差的问题,本文提出一种结合红外显著性目标导引的改进YOLO网络的智能装备目标识别系统,该方法利用红外与电视的互补特征,通过均值漂移聚类快速地获取疑似目标,并通过改进的YOLO模型进行可见光目标识别。实验仿真结果表明,本文提出的方法对地面目标识别精度较高,能够用于战场环境下态势感知、区域监控及目标打击应用。

## 1 YOLO 网络

YOLO-v3网络是目前目标识别领域较好的一种深度学习模型,该网络是从YOLO和YOLO-v2网络演变而来<sup>[13]</sup>。与基于候选区域的深度学习网络相比,YOLO网络将检测问题转化为回归问题,该网络不需要穷举候选区域,直接通过回归生成目标的置信度和边界框坐标。与Faster-RCNN网络相比,大大提高了检测速度。

YOLO检测模型如图1所示。网络将训练集中的每个图像分成 $S \times S$  ( $S=13$ ) 网格。如果真实目标的中心落入网格中,则该网格负责检测目标的类别。在每个网格中预测出来多个边界框,并且要为每个预测出来的边界框评分,以便表示该边界框完全包含目标的置信度(Confidence),其定义如下:

$$C = P_r(\text{object}) \times \text{IoU}_{\text{pred}}^{\text{truth}}, P_r(\text{object} \in \{0, 1\}) \quad (1)$$

式中:  $P_r(\text{object})$ 表示边界框中包含目标的概率(若存在目标  $P_r(\text{object})=1$ ,反之等于0);  $\text{IoU}_{\text{pred}}^{\text{truth}}$ 则表示预测结果和基准边框之间的重叠度(Intersection over Union, IoU)。置信度反映了网格是否包含目标以及预测边界框的准确性。当多个边界框检测到同一目标时,YOLO使用非最大抑制(Non-Maximum Suppression, NMS)方法选择最佳边界框。

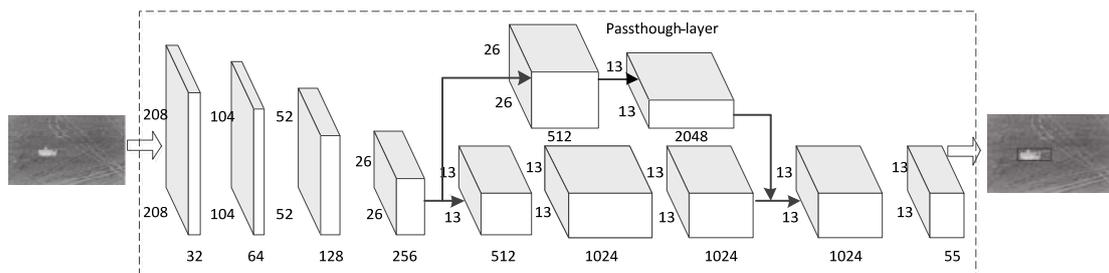


图1 YOLO模型识别流程

Fig.1 Recognition process in YOLO model

虽然 YOLO 获得了更快的检测速度,但它的检测准确率不如 Faster R-CNN。为了解决这个问题, YOLO-v2 改进了网络结构,并使用卷积层替换 YOLO 输出层中的完全连接层。同时, YOLO-v2 还引入了批量归一化、维度聚类、细粒度特征、多尺度训练等策略,与 YOLO 相比大大提高检测精度的其他方法。YOLO-v3 是 YOLO-v2 的改进版,通过使用多尺度预测来检测最终目标,其网络结构比 YOLO-v2 更复杂。YOLO-v3 可以预测不同尺度的边界框,相比 YOLO-v2 能更有效地检测小目标。

## 2 红外显著性快速目标导引

现有的地面装备光电系统将电视摄像机和红外热像仪集成于光电平台上,通过平台的转动对指定区域进行图像采集,然后送出图像处理单元进行分析,实现目标检测与跟踪,为作战人员提供精确指示,进而实现对目标打击<sup>[14]</sup>。红外热像仪可实现广域远距离目标搜索,但是获取的远距离目标图像信噪比偏低,不利于目标识别;电视摄像机可以获取丰富的纹理细节的目标图像,有利于虚假目标的剔除以及真目标识别。因此,结合红外和可见光探测器互补思想,利用目标热特性实现目标定位,再由可见光图像进行识别,降低 YOLO 深度识别网络对疑似区域识别时间,其系统框图如图 2 所示。

大多数深度识别网络在训练前需要设定初始的目标尺度,近似的尺度将获得更加准确的位置,使得模型更加容易收敛。目前存在两类先验框计算方法:第一种是直接对尺寸大小进行预测;第二种是锚点框(anchor box)候选模板<sup>[12]</sup>。这两种方法都是在训练过程进行尺度微调,但前者受误差影响较大容易往更大尺度的边界框变化,后者则不能保证先验框就是最优尺度,容易陷入局部最小。

为了提取目标,需要从红外图像中提取目标像素点,并对像素点进行特征描述,实现特征聚类。为了简化运算,本文采用均值分割对疑似目标进行粗分割,然后采取均值漂移聚类进行目标定位,最后利用目标的结构特性筛选出待识别的目标。

### 2.1 基于均值漂移聚类的目标定位

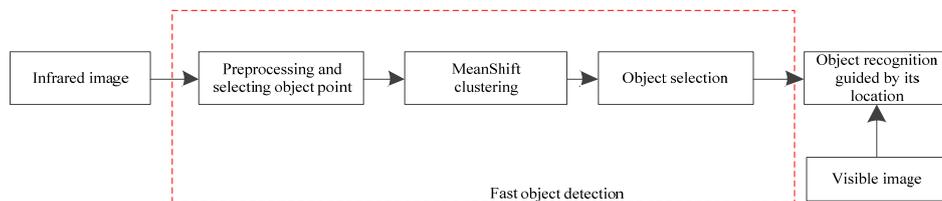


图2 红外目标位置引导下的深度学习目标识别算法框架

Fig.2 Framework of deep learning based object recognition guided by the location of infrared object

为了实现目标检测并定位,需要对疑似目标区域进行聚类分析。在未知目标先验信息的情况下,本文采用均值漂移算法进行聚类。均值漂移算法(Mean Shift)是一种非参数概率密度估计方法<sup>[15]</sup>,通过逐步密度梯度偏移实现最优聚类,其偏移量定义为概率密度  $f(x)$  的局部极大值,也就是概率密度的梯度  $\nabla f(x)$  为 0 的点。假定  $d$  维空间  $R^d$  中存在  $n$  个样本点  $x_i, i=1, \dots, n$ , 在  $x$  点的均值漂向量定义如下:

$$M_h(x) = \frac{\sum_{i=1}^n G\left(\frac{x_i - x}{h}\right) w(x_i) (x_i - x)}{\sum_{i=1}^n G\left(\frac{x_i - x}{h}\right) w(x_i)} \quad (2)$$

式中:  $G(x)$  表示高斯核函数;  $w(x_i)$  是采样点  $x_i$  的权值,  $w(x_i) \geq 0$ ;  $h$  是带宽,一般设置为 30, 主要依赖于目标的最小可识别尺寸。由于均值漂移向量  $M_h(x)$  指向概率密度梯度方向,其本质是在指定带宽范围内寻找最大概率密度函数梯度的收敛点。等式(2)经过变换可重写为如下等式:

$$M_h(x) = \frac{\sum_{i=1}^n G\left(\frac{x_i - x}{h}\right) w(x_i) x_i}{\sum_{i=1}^n G\left(\frac{x_i - x}{h}\right) w(x_i)} - x \quad (3)$$

因此,给定一个初始点  $x$ , 核函数  $G(x)$ , 允许误差  $\varepsilon = 5$ , 若先假定  $M_h(x) = \sum_{i=1}^n G\left(\frac{x_i - x}{h}\right) w(x_i) x_i /$

$\sum_{i=1}^n G\left(\frac{x_i - x}{h}\right) w(x_i)$ , 均值漂移算法可以采用交替迭代沿着概率密度梯度的方向不断移动,最终获得最优聚类中心。因此,通过对所有像素点进行协同的聚类分析,得到不同的类集合。

### 2.2 基于空间结构特性的目标筛选

由于粗分割与定位获得了大量疑似目标区域,为了降低识别网络处理的复杂度,本文采用目标空间结构特性剔除虚假目标。红外目标空间结构特性往往与形状特征、大小特征、位置布局特征等有密切关系,是实现主观视觉判读和机器解译分析的主要参考依据。本文采用长宽比与矩形度作为目标的空间几何特征进行目标筛选。

对典型目标的统计分析表明大多数目标在长宽比与矩形度上符合某个范围约束,例如:车辆的长宽比一般在2~3,军用舰船一般大于5,因此结合长宽比能很快排除一些背景干扰,抑制虚警。目标的矩形度用来描述目标形状的复杂程度,其值越小,表明目标越接近矩形。大多数地面典型目标的形状都是接近于一个矩形。

通过上述目标几何结构特征的分析,可以快速剔除虚假目标,为下一步目标识别提供可靠的目标位置,缩小搜索范围,降低处理时间。

### 3 改进的YOLO-v3识别网络

#### 3.1 密集连接神经网络

由于YOLO网络中存在大量卷积和下采样操作,在训练神经网络的同时降低了特征图的数量,造成特征信息的损失。因此,为了增强目标特征的代表能力,本文提出采用密集连接神经网络(Dense Net)<sup>[16]</sup>来更有效地捕获特征信息,该策略是利用前馈模式将每个层信息连接到其他层。也就是说,第*l*层接收前面*l*-1层的所有特征图作为输入:

$$x_l = H_l(x_0, x_1, \dots, x_{l-1}) \quad (4)$$

式中: $x_0, x_1, \dots, x_{l-1}$ 是前*l*-1层的特征映射的级联, $H_l$ 则是用于处理级联特征的函数。密集网络可以减轻梯度消失问题,增强特征传播,促进特征复用,并大大减少参数数量。虽然密集连接块的结构使得特征图得到了重用,但带来利用率高的同时也导致了越到深层的网络,特征图的数量也就越多,计算的内存需求也逐步提升,因此本文采用文献[16]提出的优化策略进行。

#### 3.2 识别模型

本文提出的改进YOLO-v3识别网络仍然是以Darknet-53为基础网络架构,并使用DenseNet代替具有较低分辨率的原始转移层,以增强特征传播,复用

和融合,其模型结构如图3所示。YOLO-v3并没有采用Softmax函数作为最终的预测分类器,而是采用独立的逻辑回归函数(sigmoid函数)来预测每个边界框的多标签分类。也就是说,红外导引的每个边界框可以属于多个类别,如掩体和坦克,此操作对于复杂战场环境下多目标并存场景是非常有用的。为了满足多目标识别的需要并验证算法的有效性,本文对网络的末端进行了修改,将目标类别的数目改为五类(履带装甲、轮式装甲、人、掩体、靶标)。所有的输入图像首先调整为512×512像素,代替原有的256×256像素图像。然后,改进网络中的32×32和16×16原始转移层与下采样层被DenseNet结构取代。在本文中,传递函数 $H_l$ 使用函数BN-ReLU-Conv(1×1)-BN-ReLU-Conv(3×3),它是卷积算子(Conv),批量归一化(Batch Normalization, BN),线性整流函数(Rectified Linear Unit, ReLU)的组合。 $H_l$ 通过对 $x_0, x_1, \dots, x_{l-1}$ 层的数据非线性变换,缓减梯度消失,其中 $x_i$ 由64个特征提取层组成,每层的分辨率为32×32。特征逐渐前向传递,最终得到大小为16×16×1024的多层次深度特征。

在训练阶段,当图像特征被转移到较高分辨率层时,后一特征层将在密集网络中接收其前面的所有特征层的特征,从而减少特征损失。另外,通过这种方式,可以在低分辨率的卷积层之间实现特征复用,提高特征的代表能力。

### 4 实验结果与讨论

为了验证提出的结合红外显著性目标引导的改进YOLO网络的目标识别模型,本章将从改进的YOLO模型性能与识别精度两方面进行分析。本文实验环境为: Intel 酷睿 i9-9900k @ 3.6 GHz (×8), 16 GB×4 (DDR4 3200 MHz), NVIDIA TESLA P100 16G×2, Ubuntu 16.04, 64位操作系统。

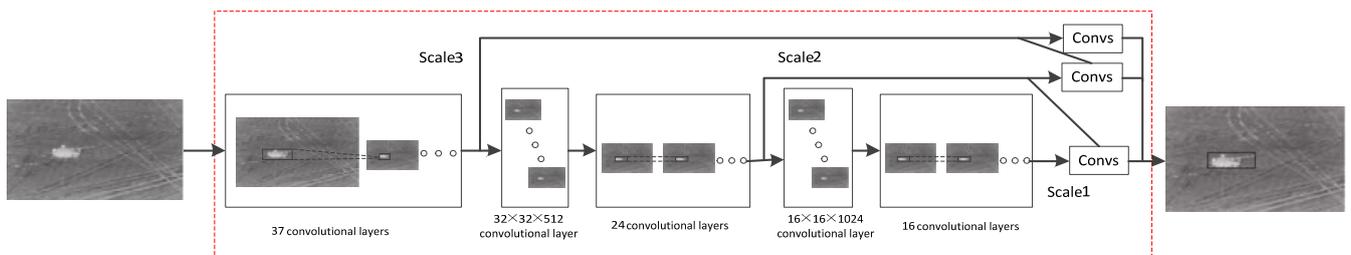


图3 改进的YOLO v3识别网络

Fig.3 Improved YOLO v3 recognition networks

#### 4.1 实验数据及其评价指标

为了定性定量地评价本节所提出的识别模型的有效性,采用自建数据集和 PASCAL VOC 公共数据集进行深度学习训练和测试。自建数据集是利用中海外九洲(陕西)防务科技有限公司研制的 GD/PD-2801A 光电探测设备,其红外图像分辨率是  $640 \times 512$ , 电视图像分辨率是  $1280 \times 720$ 。为确保不同探测器获取图像场景一致,红外与电视视场大小调到相同大小,且光轴重合。试验从外场采集了 46280 张红外及其电视图像,该数据集主要以地面环境下车辆及人员目标的可见光及红外图像为主,目标类别数为 10。选取 32150 幅图像作为训练样本,14130 幅图像作为测试样本。标注数据主要采用耶鲁大学的 Autolable 工具<sup>[17]</sup>,实验所需训练图像均按照 PASCAL VOC 2012 数据集格式进行了人工标注。PASCAL VOC 公共数据集是图像识别和分类领域优秀的数据集,被用来训练本文提出的 YOLO 模型,并验证模型的收敛性能。

实验选择了 YOLO-V3<sup>[13]</sup>, Cascaded RCNN<sup>[18]</sup>, R-FCN-3000<sup>[19]</sup>和 RNOD<sup>[20]</sup>作为对比算法,所有的对比算法都采用作者给出的源代码或可执行文件,并且都用相同的训练集进行训练。本次实验将从算法的均值平均精度(Mean Average Precision, mAP)、帧率

(Frames Per Second, FPS)、IoU 三个方面进行分析。

#### 4.2 改进的深度模型的性能分析。

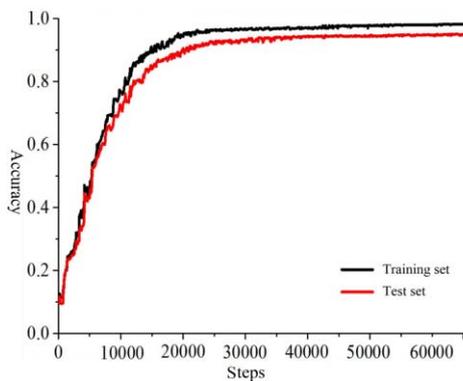
Faster RCNN、FCN 和 SSD 使用 Inception Resnet-v2 作为特征提取网络,而本文提出的改进 YOLO-V3 识别网络是以 Darknet-53 为基础网络架构,其网络初始化参数如表 1 所示。为了提高模型的检测精度,输入图像被调整为  $512 \times 512$  像素以适应 Darknet 框架。动量、初始学习率、权重衰减正则化等参数与 YOLO-V3 模型中的原始参数一致;学习速率初始化设置为 0.001,然后在训练到第 40000 步后降至 0.0001,在 50000 步后降至 0.00001。训练过程中的准确度和损失变化如图 4(a)和图 4(b)所示。训练集与测试集实验结果表明,本文改进的基于改进 YOLO 识别算法具有较高的收敛速度与识别精度。

#### 4.3 定性定量识别性能分析

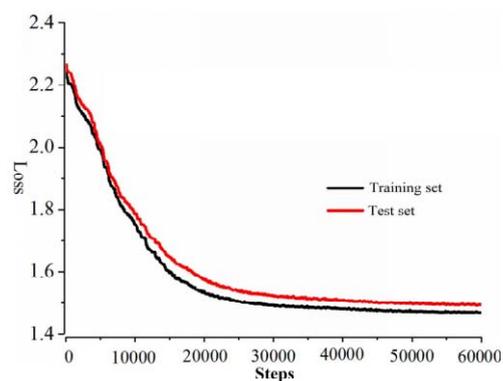
为解决复杂地面环境下低对比度目标检测问题,本文提出了一种基于红外显著性目标引导的改进 YOLO 网络的智能装备目标识别方法,该方法利用了红外与可见光图像的互补特性,通过疑似目标检测、多层卷积层特征提取、多尺度置信度模型完成检测与识别任务。表 2 是不同的深度模型对所有测试图像的定量指标结果,其中 mAP 是评价检测算法对所有类别物体的检测性能,即所有类的平均正确率(Average

表 1 初始网络参数 Table 1 Initialization network parameters

Size	Batch	Momentum	Learning rate	Decay	Training steps
$512 \times 512$	8	0.91	0.001	0.0005	70000



(a) 准确度 (a) Accuracy



(b) 损失变化 (b) Loss changes

图 4 训练过程

Fig.4 Training process

表 2 不同方法的识别准确率 Table 2 Different methods of recognition accuracy

Indexes	YOLO-V3	Cascaded RCNN	R-FCN-3000	RNOD	Our
mAP	80.6	89.6	81.1	73.8	88.9
IoU	0.59	0.72	0.62	0.48	0.68
FPS	27	2	30	6	74

Precision, AP) 的均值。可以看出, Cascaded RCNN 通过级联几个检测网络达到不断优化预测结果, 其检测网络是基于不同 IOU 阈值进行训练, 其精度是所有模型中较高的, 但实时性太差; RNOD 是两个全连接层和 NMS 模块引入目标语义模块中, 通过关联分析提升识别的精度, 但该模型容易引起误判, 尤其是针对户外采集的低质量的图像, 其识别精度较低; R-FCN-3000 是提出了解耦分类支路实现多目标分类, 在保证速度 (30FPS) 的情况下将 R-FCN 的分类类别数延伸至 3000 类。由于本文类别设置不多, 其识别精度与 YOLO-V3 相当; 本文的算法首先对红外图像进行显著性快速目标导引, 利用目标几何特点, 聚类计算目标框尺度; 然后使用改进的 YOLO 网络实现目标检测与识别。实验结果表明本文提出的模型的识别准确率比 YOLO V3 略有提升, 但实时性得到了很大的提升。在相同分辨率的情况下, 帧频接近 74。

图 5 是 YOLO 与本文算法的检测识别性能的视觉效果, 实验选用了一张具有代表性的图像进行分析。

图 5(a)为 YOLO 目标检测结果, 没有加人红外位置引导; 图 5(b)为本文算法的装甲目标检测结果, 该方法利用红外显著性快速目标导引提供的目标质心位置与目标尺度, 右图是在同分辨率可见光图像以及在目标位置导引下的深度网络检测结果。可以看到本文提出的方法具有较好的性能。为了分析本文所提的算法对户外场景的识别效果, 图 6 是不同识别算法的定性分析结果。本文提出的算法能够识别场景中的大部分目标, 尤其是针对土堆旁边的士兵也能准确定位, 但也存在将工事识别成城墙的情况; Cascaded RCNN 的识别精度较高, 主要依赖于该方法对 IOU 的自适应分析, 目标定位精度高, 但也存在识别不全的情况, 尤其是将多个坦克识别成一个。RNOD 算法对孤立目标识别精度较高, 但对遮挡目标差异较大; R-FCN-3000 能识别图像的大多数目标, 由于关联性的分析, 容易把目标识别成多个目标, 例如将装甲上的附着物识别成人。

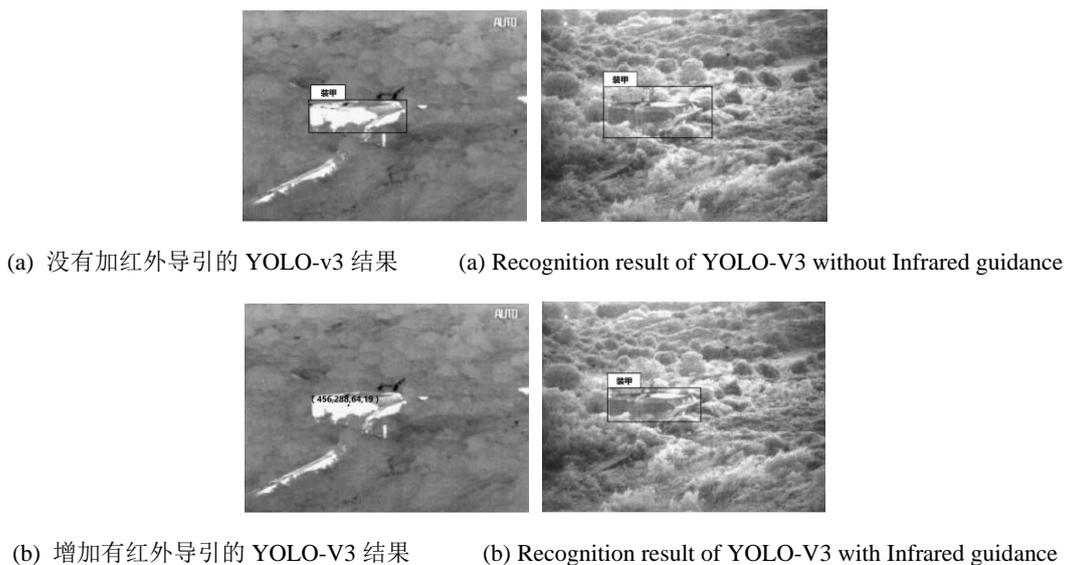


图 5 YOLO-V3 与本文算法的检测识别定性对比

Fig.5 Qualitative comparison of detection and recognition performance between YOLO-V3 and the proposed algorithm

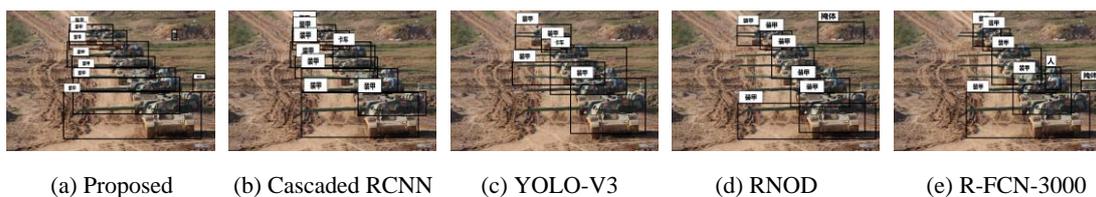


图 6 不同算法的识别结果对比

Fig.6 Comparison of recognition results for different algorithms

## 5 结语

为了提升作战环境下目标检测识别的性能,本文提出了一种基于红外显著性目标引导的改进 YOLO 网络的智能装备目标识别算法,该算法利用红外图像提供目标可能的位置引导可见光图像中的深度自主学习目标检测,从而加速检测的速度。本文提出的改进 YOLO-v3 识别网络是以 Darknet-53 为基础网络架构,并使用 DenseNet 代替具有较低分辨率的原始转移层,以增强特征传播,复用和融合。大量定性定量的实验结果表明,本文提出的模型可以有效地提高现有目标检测与行为识别网络的性能。

本文提出的算法仅仅是利用红外引导下进行自然图像识别,虽然测试数据集的识别性能较好。然而,本文算法适用范围较小,不具备全天候全天的态势感知、区域监控及目标打击应用。下一步,项目组将融合红外与可见光的互补特征,提升算法的全方位泛化能力。

### 参考文献:

[1] 张鑫,李续武,路艳丽. 双向 S-优势粗集及其在意图识别中的应用[J]. 弹箭与制导学报, 2017, 37(3): 160-164.  
ZHANG Xin, LI Xuwu, LU Yanli. Two-direction S-dominance Rough Sets and Its Application in Intention Recognition[J]. *Journal of Projectiles, Rockets, Missiles and Guidance*, 2017, 37(3): 160-164.

[2] 陈文英,张兵志,史力晨,等. 新型智能装甲作战系统体系贡献率评估研究[J]. 兵工学报, 2018, 39(9): 1841-1849.  
CHEN Wenying, ZHANG Bingzhi, SHI Lichen, et al. Research on Evaluation of Contribution Rate of a New Intelligent Armoured Combat System to Army Weapon System-of-systems[J]. *Acta Armamentarii*, 2018, 39(9): 1841-1849.

[3] 刘从军,嵇金荣,杨辉. 智能监控中实时预警系统[J]. 计算机系统应用, 2014, 23(6):65-68.  
LIU Congjun, JIN Jinrong, YANG Hui. Real-time Early Warning System in Intelligent Monitoring[J]. *Computer Systems*, 2014, 23(6): 65-68.

[4] 文雄志,陈学工,王晶,等. 基于红外图像的桥梁目标自动识别方法[J]. 计算机与现代化, 2015(5): 62-66.  
WEN Xiongzi, CHEN Xuegong, WANG Jing, et al. Automatic Recognition Method of Bridge Target Based on Infrared Image[J]. *Computer and Modernization*, 2015(5): 62-66.

[5] YAO X, HAN J, GUO L, et al. A coarse-to-fine model for airport detection from remote sensing images using target-oriented visual saliency and CRF[J]. *Neuro computing*, 2015, 164: 162-172.

[6] Ross Girshick, Jeff Donahue, Trevor Darrell, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//*IEEE Conference on Computer Vision and Pattern*

*Recognition*, 2014: 580-587.

[7] HE K, ZHANG X, REN S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2014, 37(9): 1904-16.

[8] Girshick R. Fast r-cnn[C]//*Proceedings of the IEEE international conference on computer vision*, 2015: 1440-1448.

[9] 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 and Machine Intelligence*, 2015, 39(6): 1137-1148.

[10] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]//*2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015: 779-788.

[11] LIU W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[J]. *European Conference on Computer Vision*, 2015: 21-37.

[12] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]//*2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017: 6517-6525.

[13] Redmon J, Farhadi A. YOLOV3: An incremental improvement[R/OL]. *arXiv preprint arXiv: 1804.02767*, 2018.

[14] 黄伟,曹宇剑,徐国明. 基于 Faster R-CNN 模型的低空平台偏振高光谱目标检测[J]. 红外技术, 2019, 41(7): 600-606.  
HUANG Wei, CAO Yujian, XU Guoming. Polarized Hyperspectral Object Detection with Faster R-CNN for Low-Altitude Platforms[J]. *Infrared Technology*, 2019, 41(7): 600-606.

[15] Aguilar W G, Luna M A, Moya J F, et al. Pedestrian detection for UAVs using cascade classifiers with meanshift[C]//*2017 IEEE 11th international conference on semantic computing (ICSC)*, 2017: 509-514.

[16] HUANG G, LIU Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//*Proceedings of the IEEE Conference On Computer Vision and Pattern Recognition*. 2017: 4700-4708.

[17] WU, Fei, Wang, ZHANG Zhongfei, et al. Weakly Semi-Supervised Deep Learning for Multi-Label Image Annotation[J]. *IEEE Transactions on Big Data*, 1(3):109-122.

[18] CAI Z, Vasconcelos N. Cascade r-cnn: Delving into high quality object detection[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018: 6154-6162.

[19] Singh B, Li H, Sharma A, et al. R-FCN-3000 at 30fps: Decoupling Detection and Classification[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 1081-1090.

[20] HU H, GU J, ZHANG Z, et al. Relation networks for object detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 3588-3597.