

基于改进 YOLO 网络的双通道显著性目标识别算法

段辉军¹, 王志刚², 王彦³

(1. 湖南财经工业职业技术学院电子信息系, 湖南 衡阳 421002; 2. 电子科技大学计算机科学与工程学院, 四川 成都 610054;
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: 由于缺乏目标的先验信息, 实时预警检测系统存在虚警率高、实时性偏低等问题, 限制了实战环境下的广泛应用。为了提升目标检测识别的性能, 本文提出了一种基于改进 YOLO 网络的双通道显著性目标识别算法, 该算法利用红外图像与可见光互补特性进行多尺度融合, 并在融合图像上采用显著性检测获取疑似目标区域, 最后利用改进的识别网络对疑似区域进行多层次目标识别。改进的 YOLO 识别网络增加了一路辅助网络, 改善整个特征提取网络的性能, 并采用注意机制对辅助网络和骨干网络的特征信息融合, 增强有效信息通道, 抑制无效信息通道, 提高网络识别效率。仿真实验结果表明, 本文提出的模型可以有效地提高目标检测与识别精度, 其实时性得到了大大增强。

关键词: 目标识别; 双通道; 图像融合; 显著性; 深度学习; 注意机制

中图分类号: TP37; TN219 文献标识码: A DOI: 10.3969/j.issn.1001-5078.2020.11.014

Two-channel saliency object recognition algorithm based on improved YOLO network

DUAN Hui-jun¹, WANG Zhi-gang², WANG Yan³

(1. Department of Electronic Information, Hunan Financial and Industrial Vocational-Technical College, Hengyang 421002, China;
2. School of Computer Science and Engineering, University of Electronic Science and Technology, Chengdu 610054;
3. Science and Technology on Information Transmission and Dissemination in Communication Networks Laboratory,
The 54th Research Institute of CETC, Shijiazhuang 050081, China)

Abstract: Due to the lack of a priori information of the object, the real-time early warning detection system has problems such as high false alarm rate and low real-time performance, which limits its wide application in actual combat environments. In order to improve the performance of object detection and recognition in real-world combat environment, this paper proposes a two-channel saliency object recognition algorithm based on improved YOLO network, which firstly uses the complementary characteristics of infrared image and visible image for multi-scale fusion, and uses saliency detection to obtain the suspected object area on the fusion image. Finally, the improved recognition network is used to recognize the suspected area in multi-level. The improved YOLO recognition network adds an auxiliary network to improve the performance of the whole feature extraction network, and uses the attention mechanism to fuse the feature information of the auxiliary network and the backbone network, enhances the effective information channel, suppresses the invalid information channel, and improves the network recognition efficiency. Simulation results show that our proposed model can effectively improve the accuracy of object detection and recognition, and the real-time performance is greatly enhanced.

Keywords: object recognition; two-channel; image fusion; infrared saliency; deep learning; attention mechanism

基金项目: 国家自然科学基金项目 (No. 61802270); 装备预研兵器工业联合基金项目 (No. 6141B012858) 资助。

作者简介: 段辉军 (1981 -), 男, 硕士, 讲师, 主要研究方向为计算机技术, 信息化教学。

收稿日期: 2020-04-30; 修订日期: 2020-07-21

1 引言

随着作战模式由平台中心战向一体化联合作战转变,地面装甲突击系统承担的任务日益复杂,迫切需要通过智能化手段提高平台作战效能,以及多平台协同作战的能力。同时,陆战场环境比较复杂,敌方目标善于伪装且机动性大,如何快速检测、识别并锁定目标是战场成败与态势感知的关键^[1]。

图像是自然景物的客观反映,也是环境感知的关键数据^[2]。大多数装备都是利用采集的视频序列实现对外环境探测、观测与瞄准,最终由操作手完成目标的锁定与打击,其识别、锁定目标精度的高低将直接影响打击目标的效果。为了降低操作手的捕获误差,国内外的武器装备集成了各种辅助功能。例如,坦克火控系统利用电子稳像设备使炮长瞄准线的双向稳定,确保坦克瞄得更准^[3];目标指示器具备动目标检测功能,检测出视场中的运动目标,并编号,提升炮手的反应时间^[4];美军的全球鹰无人机集成了基于机器学习的智能增强系统,无需人在回路也能获得高精度的识别效果,但最终也是需要操作手对结果进行判别^[5]。可以看出,这些辅助策略仍然需要人工辅助进行决策,并没有真正意义上实现智能识别。

随着智能信号处理技术的发展,学术上也提出了大量的智能识别算法,但这些算法大都是针对单模态图像进行处理,其应用环境局限性较大。例如,装甲目标在迷彩伪装、夜晚、能见度低的情况下,可见光的目标识别能力有限,而红外图像具有作用距离远,穿透力强,且不受光照影响,但其图像模糊,空间分辨率较低^[6]。众所周知,现有的地面装备光电设备大都集成了红外热像仪与电视摄像机,并输出两路独立的视频信号。光电系统输出的多路视频信号存在大量互补信息和冗余信息,有利于提高目标检测与识别的精度。近年来,国内外专家学者的研究方向大都集中在将不同类型图像的互补信息结合起来,生成信息量更大的合成图像,以此增强同一场景的图像表征能力^[7]。文献[8]提出了基于多尺度变换的红外与可见光融合算法,通过互信息配准基础上对不同尺度特征进行融合,具有较高的融合精度,但易受抖动、噪声的影响。文献[9]在多模融合的基础上对目标进行识别,其识别性能严重依赖于配准与融合的精度。刘佑鑫等人^[10]利用类字典稀疏识别方法

对融合后的可见光-近红外 HSV 图像进行场景分类。由于该方法采用了搜索树特征和分层梯度方向直方图进行特征提取,其场景分类性能仅仅适用简单场景下的典型目标。

随着智能硬件技术的突飞猛进和大数据建模工具的利用,深度学习模型在图像识别领域获得极大的发展,涌现了许多优秀的网络模型,如 Faster R-CNN^[11], Mask R-CNN^[12], YOLO-v3^[13]等。江泽涛等人^[14]使用加权融合算法对红外和可见光图像融合,然后利用改进的深度特征提取融合图像的有效特征向量,再进行分类与识别。Sarfraz 等人^[15]提出了一种跨模态的深度识别算法,该算法使用多模态图像的不同特征形式分别训练四个独立的深度卷积神经网络,分别学习出不同且互补的目标特征信息,最终提高目标的识别精度,其核心模块是引用了特征对齐模块与像素对齐模块。可以看出,基于深度学习的红外-可见光图像的识别算法大都是在图像配准的基础上进行融合,然后采用深度学习算法进行目标识别。虽然深度学习模型提高了双模态图像的互补特征的表征能力,但现有的算法都是在已配准图像上进行独立网络训练与识别,且资源消耗较大。然而,光电系统的红外图像与可见光图像视场大小存在偏差,光轴也并不是完全重合,直接采用红外与可见光图像进行深度识别,其效果不利于军事装备工程应用。因此,本文提出一种基于改进 YOLO 网络的双通道显著性目标识别算法,该算法首先结合双通道网络提取红外与可见光图像的融合特征进行图像融合,然后利用融合图像目标显著性特性可以快速获取疑似目标区域,最后通过改进的 YOLO 模型提高目标识别精度。仿真结果表明,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。

2 相关技术

YOLO 系列网络是目标检测领域常用的一种端到端深度网络,该网络将检测问题转化为回归问题,可以预测边界框坐标的同时,在端到端网络中对目标进行分类,大大提高了检测速度^[13]。与 R-CNN 系列的两级网络相比,这种网络结构思路更明晰,并不需要穷举疑似区域,网络更简洁。

YOLO 网络衍生出许多高效率的改进网络,如 YOLO-v1, YOLO-v2, YOLO-v3, YOLO-v3-tiny 等^[16]。YOLO 系列网络的主干网络模型如图 1 所

示。YOLO 目标检测方法将输入图像划分为 $S \times S$ 个互补重叠的网格,其中每个网格最多预测 K 个边

界框。每个边界框的特定类别置信分数可以用如下等式表示:

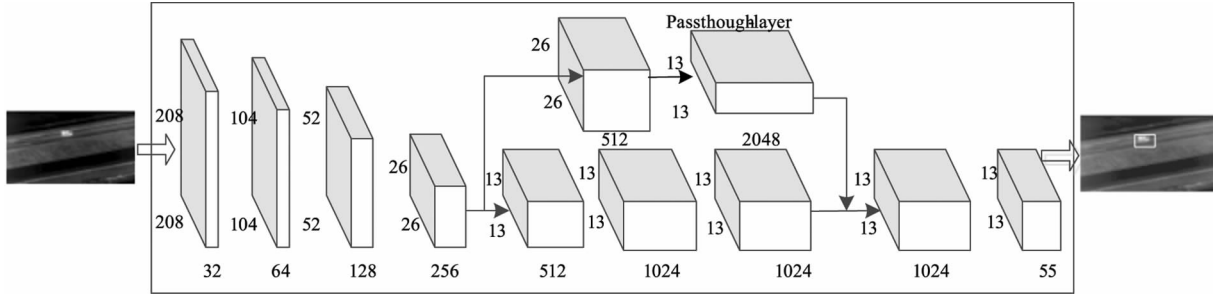


图 1 YOLO 模型检测流程

Fig. 1 Detection process for YOLO

$$C = \Pr(\text{Class}_i) * \text{IoU}_{\text{pred}}^{\text{truth}}$$

$$= \Pr(\text{Class}_i | \text{Object}) * \Pr(\text{Object}) * \text{IoU}_{\text{pred}}^{\text{truth}} \quad (1)$$

其中, $\text{IoU}_{\text{pred}}^{\text{truth}}$ 表示预测框与基准框之间的重叠度; $\Pr(\text{Class}_i) * \text{IoU}_{\text{pred}}^{\text{truth}}$ 表示边界框包含目标 i 的置信度; $\Pr(\text{Class}_i | \text{Object})$ 表示目标属于 i 类的条件概率。最终, YOLO 的预测编码为 $S \times S \times (K \times (5 + C))$ 张量。YOLO 的骨干网络通过类似于 VGG 网络的下采样卷积结构来提取目标特征。当卷积神经网络前向传播时,第 l 层与第 $(l-1)$ 层之间的关系如下所示:

$$x^l = f(y^l) = f(x^{l-1} \otimes w^l + b^l) \quad (2)$$

卷积神经网络中第 l 层的输入表示为 x^l ; 激活函数为 $f(\cdot)$ 。中间变量表示为 $y^l = x^{l-1} \otimes w^l + b^l$, 其中 w^l 是卷积核的权重, b^l 是偏置参数, $\otimes$ 表示卷积。YOLO 的网络可以预测出多个边界框及其类概率, 最终选择输出具有最高 IoU 的检测结果。为了防止输出过多的虚假冗余边界框, YOLO 采用了阈值策略消除了大多数无用的边界框。由于单个目标的预测结果可能同时包含多个边界框, 从而导致不同边界框对应同一个目标。YOLO 采用非极大值抑制算法来组合多个预测边界框, 从而消除预测置信度较低的边界框, 并将具有较高置信度得分的预测边界框选择为目标检测框。最终, 预测边界框的损失函数由四部分组成, 如下所示:

$$L = \text{loss}_1 + \text{loss}_2 + \text{loss}_3 + \text{loss}_4 \quad (3)$$

其中, loss_1 是预测中心坐标的损失; loss_2 是预测边界框的宽度和高度的损失; loss_3 是预测类别的损失; loss_4 是预测的置信度损失。随着卷积网络前向传播, 其损失函数 L 的梯度可以表示为:

$$\delta_{l-1} = \frac{\partial L}{\partial y_{l-1}} = \frac{\partial L}{\partial y_l} \frac{\partial y_l}{\partial y_{l-1}} \quad (4)$$

以上分析可以看出, YOLO 将图像分成固定尺

度的网络, 虽然可以获得较快的检测速度, 但其边界框的定位不准确, 且对小目标检测精度不高。为了解决这个问题, YOLO-v2 引入了锚框 (anchor) 机制的思想, 并使用聚类方法生成合适的先验边界框。随着网络的不断加深, 在训练过程中存在梯度消失或梯度爆炸问题, YOLO-v3 引入了残差结构, 采用 Darknet-53 作为主干网络, 主要由 53 个卷积层组成, 并包含大量 3×3 、 1×1 卷积内核。在进入残差模块之前将特征与残差模块输出的特征进行组合可以提取更抽象的深度特征。

3 红外显著性快速目标检测

现有的光电系统大都如图 2 所示, 集成了电视摄像机和红外热像仪两种探测设备。红外图像可以全天候全天时获取场景图像信息, 但由于辐射成像原理的限制, 红外图像存在对比度低、细节模糊, 目标检测与识别精度低; 而可见光图像分辨率高, 细节丰富, 符合人眼视觉习惯, 但容易遭受烟雾、光线、天气的影响^[17]。红外图像与可见光图像间信息存在冗余和互补, 利用双通道目标融合识别, 可提高目标的识别精度。本文采用文献[18]提出的图像融合算法进行红外与可见光图像融合, 该算法在稀疏编



图 2 光电系统

Fig. 2 Photoelectric system

码和融合过程中,充分考虑了每个局部图像块的信息及其空间上下文信息,并利用鲁棒稀疏表示确保空间相邻图像块之间的局部一致性。实验结果表明,该算法对未经校正的多源图像的融合效果也非常理想。

为了定位到目标的具体位置,大多数检测算法通常会把图像分成许多子块,并提取子块的特征进行识别。目前存在两类子块划分策略:滑动窗口法与选择性搜索法,前者通过穷举整个区域获得大量疑似区域,再进行类别分类,但复杂度太高;后者则有效地去除冗余候多尺度选区域,使得计算量大大的减小,但需要先验信息进行辅助。

为了有效地利用光电系统不同通道的互补数据,提升目标识别的精度与速度,本文提出了一种双通道显著性目标识别算法,如图3所示。首先利用融合策略获取红外图像与电视图像的融合结果,然后采用显著性检测获取红外图像中的疑似目标区域;最后在融合结果基础上利用改进 YOLO 网络对疑似区域进行多层次目标识别。

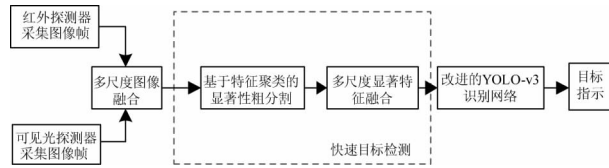


图3 双通道显著性目标识别算法

Fig.3 Two channel saliency target recognition algorithm

3.1 基于特征聚类的显著性粗分割

由于视距内的装甲目标红外特征明显,其目标特征与背景存在一定的差异。即便是沙漠地带,典型目标仍然可以通过显著性检测进行定位。为了实现目标显著性检测,本文采用了一种基于特征聚类的显著性检测算法,该方法逐点提取特征向量,并在先验信息未知情况下利用均值漂移算法进行目标聚类。本文选用的特征向量 x_i 涵盖区域的强度、对比度、空间属性等多种特征,其可以有效地捕获显著性区域中的对比度和空间信息,便于后续特征聚类。

假定 d 维空间 R^d 中存在 n 个样本点 $x_i (i = 1, \dots, n)$, 在 x 点的核密度函数估计可以表示为如下等式:

$$\begin{aligned} f(x) &= \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \\ &= \frac{c}{nh^d} \sum_{i=1}^n K\left(\left|\frac{x - x_i}{h}\right|^2\right) \end{aligned} \quad (5)$$

其中, h 是核密度估计函数的带宽,一般设置为 30; $K(x) = ck(\|x\|^2)$ 是核函数。等式(5)的核密度梯度可以为:

$$\Delta f(x) = \frac{2c}{nh^{d+2}} \sum_{i=1}^n g\left(\left|\frac{x - x_i}{h}\right|^2\right) M_h(x) \quad (6)$$

其中, $c = 3$ 。核函数表征了样本密度分布估计情况,工程应用中常采用高斯核函数作为核密度估计公式中的加权函数。假定 $G(x)$ 是高斯核函数; $w(x_i) (w(x_i) \geq 0)$ 是样本 x_i 所对应的权值,因此均值偏移向量 $M_h(x)$ 可以表示为如下等式:

$$M_h(x) = \frac{\sum_{i=1}^n G\left(\frac{x_i - x}{h}\right) w(x_i) (x_i - x)}{\sum_{i=1}^n G\left(\frac{x_i - x}{h}\right) w(x_i)} \quad (7)$$

因此,给定一个初始点 x ,核函数 $G(x)$,此过程可以通过交替迭代沿着梯度的方向不断移动,并采用 $x^{t+1} = x^t + M_h(x^t)$ 进行更新,最终获得最优的聚类结果。在聚类结果的基础上对图像进行分块,计算每个块的质心 i 。以质心 i 为中心点,计算特征向量 x_i 。若与 x_i 相邻的图像块 x_j 的欧式距离 $d(x_i, x_j)$ 比其他任意的图像块都大时,则该点的特征是显著的。

3.2 多尺度显著特征融合

由于基于特征聚类的显著性粗分割结果中包含大量的虚假目标,大大提高了后续检测的复杂度。因此,为同时均衡检测效率与检测精度,本文采用消除虚假目标,同时利用空间特征筛选出疑似目标。

本文采用多尺度融合对显著性粗分割结果进行优化。首先对图像进行金字塔分解;假定第 i 层图像中质心 p 的归一化到 $[0, 1]$ 区间,则图像的显著性结果可以表示为 $S_{(r,p)}^i$ 。由于不同尺度下同一个显著性区域具有一致性,那么第 $(i-1)$ 层的显著性区域 $S_{(r,p)}^{i-1}$ 经过下采样后得到的结果应该与 $S_{(r,p)}^i$ 具有一致性。因此,本文采用多尺度限制性融合策略对显著性结果进行精细化处理,最终得到最优的显著性结果。

$$\begin{cases} S_{(r,p)}^i = w_p^i S_{(r,p)}^i + w_p^{i-1} S_{(r,p)}^{i-1} \\ w_p^i + w_p^{i-1} = 1 \\ w_p^i S_{(r,p)}^{i-1} = w_p^{i-1} S_{(r,p)}^i \end{cases} \quad (8)$$

经过多尺度融合的显著性区域大大地消减了不必要的背景干扰,其覆盖的区域最能代表目标所在的位置、大小与尺度信息。然而,红外场景内的显著性目标类型多样,且形态各异,仅仅靠多层次显著性

目标差异,很难排除虚假目标的干扰。由于装甲目标具有明显的形状特性,因此本文采用装甲目标的结构特征剔除虚假目标区域。最直观的策略就是采用几何特性进行判别。文献[19]采用长宽比对目标的几何特性进行判断,其车辆的长宽比在 2~3 之间。文献[20]采用 SVM 分类对显著性区域的几何特征进行分类,其中几何特征包括方向梯度、Hu 不变矩等,然后采用 SVM 进行训练与学习,最终可以实现桥梁、车辆、建筑等典型目标的识别。由于本章的目的仅仅是消除虚假区域,缩短后续深度学习的识别时间,并不需要获得准确的车辆显著性目标。因此,本章采用简单的几何特征进行判断,不增加漏检概率的情况下,尽可能的剔除虚假区域,为下一步目标识别提供可靠的目标位置,缩小识别的空间复杂度。

4 改进的目标识别网络

目前,依靠更深的网络层提取目标特征并增强卷积层的语义信息来提高识别精度是一种常见的网络结构设计方法。然而,由于红外与电视融合图像的疑似目标区域已经明确,为了降低处理时间其识别网络的深度也不宜太深,尤其是小目标在多次卷积与池化后,其特征表征能力也在下降。因此,本文提出了一种基于改进 YOLO-v3 的目标识别网络,该网络增加了一路辅助网络,改善整个特征提取网络的性能;并采用注意机制对辅助网络和骨干网络的特征信息融合,增强有效信息通道,抑制无效信息通道,提高网络识别效率。

与单一结构的 YOLO-v3 网络相比,本文增加了一个特征提取辅助网络,其规模比骨干网络略小。主干网的原始残差模块使用 3×3 卷积核提取特征,而辅助网络中的残差模块使用两个连续的 3×3 卷积核进行残差提取,获得 5×5 的感受野,然后将提取的特征融合到骨干网中。辅助网络中使用的感受野大小为 5×5 。特征图上使用大的感受野进行全局特征提取有助于得到目标区域特征,因此辅助网络将获取的位置特征传输到骨干网络。骨干网络结合使用 3×3 局部感受野和辅助网络提供的目标位置信息,可以更准确地学习目标详细特征。因此,辅助网络在一定程度上促进了高级和低级语义特征的融合,从而大大提高了网络的识别性能。YOLO-v3 的原始网络具有很深的深度。如果将辅助网络添加到整个骨干网络,则会引入更多计算,从而导致运行

速度变慢。基于上述原因,改进的网络仅将辅助网络添加到 YOLO-v3 的三个相应尺度检测的特征提取层中。

本文中辅助网络和主干网的连接方式是在两个网络之间添加注意机制。当网络达到一定深度时,特征越容易聚合成针对任务的更高级语义特征。本文增加的辅助模块主要完成对有效特征的处理和传输,对无效特征进行信道抑制。两个网络之间的注意机制的实现使用了 SE 模块(Squeeze-and-Excitation 模块),该模块结构简单,易于部署。添加 SE 模块的目的是重新校准辅助模块的输出特性。首先对特征映射进行压缩,通过平均池化将二维特征通道转化为一维。此时,将特征图的大小转换为 $1 \times 1 \times C$,将二维特征图合并成一维特征图的目的是为了更好地显示该层各通道特征值的分布情况。在特征图压缩完成后,将激发一维特征图,其计算公式为:

$$S_c = F_{ex}(Z, W) = \sigma(W_2 \cdot \text{ReLU}(W_1 Z)) \quad (9)$$

其中, S_c 是激励后的特征图; σ 是 sigmoid 激活函数; Z 表示压缩后的输入一维卷积层,其维度是 $1 \times 1 \times C$; $W_1 Z$ 表示全连接操作, W_1 的维数是 $C/r * C$ 的全连接层; r 是尺度例因子,其功能主要是压缩通道数量以减少参数量; W_2 也是一个全连接层,其尺寸为 $C * C/r$ 。 $W_1 Z$ 输出结果将通过 W_2 连接后输出 $1 \times 1 \times C$ 的特征图,最后将其激活处理获得特征图 S_c 。 S_c 是连接模块的核心,用于指示辅助模块输出的通道权重。通过将不同的信道值乘以不同的权重,可以增加对关键信道域的关注。最后,辅助残差模块通过 SE 模块重定向,然后进入骨干网络以完成特征融合。

5 实验结果与讨论

本文提出的双通道显著性目标识别算法主要分为双模图像融合(double-model fusion, DMF)、显著性检测(Saliency Detection, SD)与目标识别(Object Recognition, OR)三个模块。为了验证所提出模型的性能,尤其是各种模块之间的组合对目标识别性能的影响,本实验主要从消融角度进行分析。本文提出的算法采用 Python 3.5 进行开发,选用 TensorFlow 作为深度学习框架,所有实验均在戴尔服务器上进行,其配置如下: Intel core i9-9900k @ 3.6 GHz(×8), 16 GB ×4(DDR4 3200 MHz), NVIDIA GeForce GTX1050TI 16G ×4, Ubuntu 16.04, 64 位操作系统。值得注意的是,不加优化的情况下,4 个 GPU 对

现有数据集进行 YOLO-v3 训练大约需要一天的时间才能收敛。为了提高优化效率,采用 ADAM 优化算法,其参数设定为: $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$ 和 $\epsilon = 10^{-8}$ 。实际训练过程中每次调整参数前所选取的样本数量,学习率的初始值,随机梯度下降的动量和衰减分别设置为 16、0.01、0.9 和 10^{-6} 。图像融合采用的西电张强教授提出的基于双层次决策规则的图像融合,其算法是开源的;目标显著性检测采用本文第 3 节提出的方法,其 $c = 3$, $h = 30$, 长宽比为 3.1。

5.1 实验数据及其评价指标

为了实现对所提模型进行训练、测试与验证,并进行定性定量分析,实验采用公开数据集与自建数据集进行对比实验,其中公开数据集是国际公开配准红外与可见光图像 TNO 数据;自建数据集则是利用图 2 所示的光电系统进行采集的地面车辆数据,所有数据都时间标记,并根据靶板参数映射到相同的尺寸。实验最终收集了从 8000 对红外与可见光数据对,随机选择了 4500 个用于训练,2450 个用于验证,1050 个用于测试。为了便于性能对比,本文将地面装甲车辆分为履带式装甲、轮式装甲、货车、面包车及其他车辆。

实验选择了 YOLO-v3^[13], DenseNet^[20], ResNet^[22], EfficientDet^[23], R-FCN-3000^[24] 和 ATSS^[25] 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。由于对比算法是对自然影像的检测,为了便于公平的定性定量的对比,所有的深度算法模型都用相同的训练集进行训练。本文采用已被广泛应用于目标识别评估的平均精度均值 (Mean Average Precision, mAP) 和每秒帧率 (Frames Per Second, FPS) 作为衡量模型的标准,其中 mAP 是指多个类别精度均值 (Average Precision, AP) 的平均值,每个类别都可以根据准确率 (Precision) 和召回率 (recall) 绘制一条 PR 曲线。

5.2 训练性能分析

为了便于分析该模型的识别效果,本实验不使用显著性检测的结果进行训练,而是直接采用 YOLO-v3 的选取锚点框的方法进行区域选择,然后进行相关的训练。在训练过程中,可以通过绘制损失曲线来直观地观察训练的动态过程。图 4 显示了本文算法训练期间的平均损失曲线的变换,其中横坐标表示训练迭代次数,纵坐标表示训练期间的损失值。可以看出,本文算法的训练与测试机上的精度与损失度类似,整个网络迭代到 3000 次就实现了收敛,充分表明模型的泛化能力较好,适合复杂战场环

境下的工程应用。

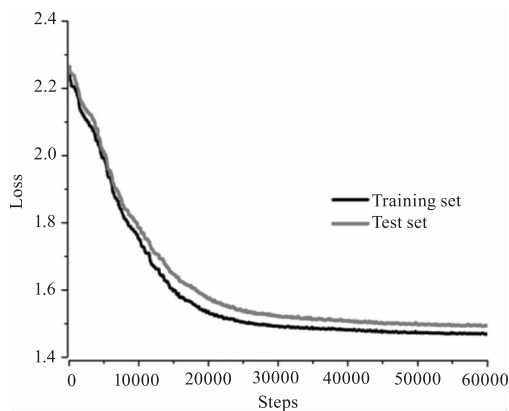


图 4 训练过程

Fig. 4 Training process

5.3 消融实验分析

本文提出的双通道显著性目标识别算法主要应用于光电系统下目标指示与识别,其处理流程是首先对红外与可见光图像进行多模态融合,然后采用显著性检测标注出若干疑似显著性区域,最后采用改进的目标识别算法进行识别。为了对不同模块的性能进行消融分析,本章设计了如下对比算法:改进的识别算法在可见光图像上的识别效果 (CCD-Reco),改进的识别算法红外图像上的识别效果 (IR-Reco),改进的识别算法在融合图像上的识别效果 (Fusion-Reco),改进识别算法在显著性区域上的识别效果 (Sal-reco),以及本文提出的模型,表 1 展示了不同模块下的消融结果。实验数据可以看出,CCD-Reco 与 IR-Reco 的识别准确率仅仅相差 2%,检测结果中红外与 CCD 精度存在互补性。也就是说,红外图像目标不明显的区域,在可见光下的比较明显,最后导致 CCD-Reco 的精度优于 IR-Reco,反之亦然。经过融合后的图像具有二者的互补特性,最终在识别结果上达到了 78.7% 的识别精度。以上三种对比算法主要是验证多模态影像融合对最终识别精度的影响,但其识别速度都在 30 帧左右。在实际应用中,只有较高的识别速度,才能有助于工程应用。从表 1 最后一列数据也可以看出,增加了显著性目标检测的识别算法获得了 51 帧的检测速度,完全适合 50 Hz 的工程应用。本文设计的显著性目标是提升识别速度的关键,通过显著性检测性能对比得知,该算法能覆盖大多数目标,其最小显著性目标可以检测到 15×15 ,但此时肉眼识别目标也比较困难。即便识别算法不能识别出典型目标,但目标指示器也会给出虚框指示,便于操作手判断。若识

别算法给出较高置信度,则可以自动导引对准目标。因此,从消融角度分析可知,本文设计的三种模块都是非常合适,且能提升精度与速度。

表 1 不同模块性能分析

Tab. 1 Performance analysis of different modules

指标	CCD-Reco	IR-Reco	Fusion-Reco	Sal-Reco
mAP	48.12	51.38	78.70	89.11
FPS	27	28	22	51

5.4 定性定量对比分析

本章的选用的对比算法都是具有源码的深度网络模型,主要应用于自然图像的目标检测与识别。针对本文主要应用光电系统的多模态二维灰度图像上的目标识别,因此所有的网络都是在融合后的图数据集上进行训练。表 2 展示了不同的深度模型对所有融合图像的检测识别结果。从定

量测试结果来看,本文提出网络的识别精度高于 ATSS,和 EfficientDet 的识别率几乎相同。YOLO - v3 是单阶段识别网络,但我们的算法的 mAP 仅仅比其略高 0.2 %,但是检测速度确高得多。也就是说,本文所提出的网络在识别精度和识别速度方面具有良好的性能。图 5 展示了本文模型与其他对比算法之间的 PR 曲线以便进一步比较不同算法之间的性能差异。EfficientDet 和 ATSS 都是 CVPR2020 中提出并开源的目标检测网络,前者从网络在特征金字塔网络的基础上,提出一种高效的双向多尺度的自适应加权融合方法,后者则能够利用目标区域的相关统计自动选择合适的锚框作为正样本,在不增加额外计算量和参数的情况下,大幅提升模型的泛化性能。这两种网络是目前最优的深度神经网络,具有较高的精度与速度。

表 2 不同方法的识别准确率

Tab. 2 Comparison for different methods in recognition accuracy

指标	YOLO - V3	DenseNet	ResNet	EfficientDet	R - FCN - 3000	ATSS	Proposed
mAP	88.92	79.67	71.15	85.83	83.99	86.35	89.11
FPS	29	2	27	41	74	45	51

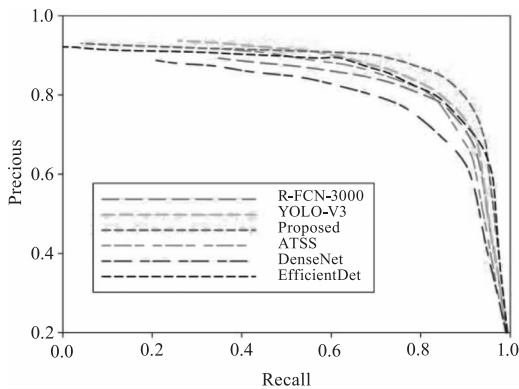


图 5 不同检测识别模型的 PR 曲线

Fig. 5 PR Curves with different models

图 6 是 YOLO - v3 与本文算法的检测识别定性对比,其中图 6(a)与图 6(b)分别是可见光探测器与红外热像仪获得的 CCD 图像与 IR 图像,图 6(c)是红外与可见光图像融合结果,图 6(d)是 YOLO - v3 与本文算法对图 6(c)中融合图像的识别结果。可以看出,本文所提模型对融合图像的识别结果较好,而 YOLO - v3 虽然没有出现漏检,但结果存在定位误差以及误检。改进的识别模型能够很好地识别目标,显著性检测模块已经大大消除了背景杂波的干扰。实际上,训练过程中未加框的区域将默认标

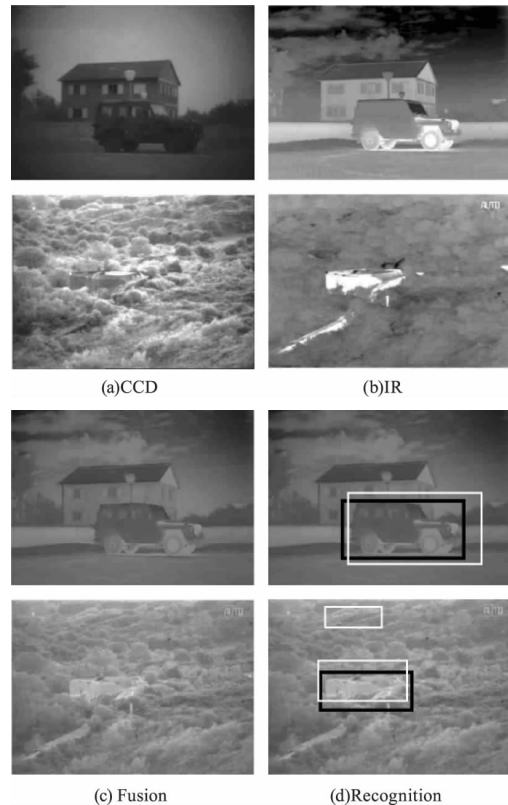


图 6 YOLO - v3 (黑框)与本文算法(白框)的检测识别定性对比

Fig. 6 Qualitative comparison of between YOLO - v3 (White line) and the proposed algorithm(Black line)

记为背景区域,改进的识别算法只在显著性区域进行识别,即便显著性区域存在目标粘连,识别算法也进行区分。显著性检测的目的就是大大降低背景干扰,缩小感兴趣区域的范围,进一步降低复杂度。从这点可以看出,本文改进的 YOLO-v3 可以提供较高的分类精度。

图 7 是不同算法的识别结果对比。可以看出添加辅助网络后,相比原始 YOLO-v3 网络结果,本文模型提高了小目标检测的有效性和准确性。例如,原始的 YOLO-v3 无法检测到图中的较小的装甲车辆,但是经过改进的网络可以有效地检测目标,即使在广域复杂战场环境中,改进的网络仍然可以准确地检测和识别。除了常规的模型指标测试外,本文还添加了干扰状态的单个目标测试。外场采集图像时,人为释放了烟雾,干扰目标的成像质量。受限于篇幅,本文选用了一张图像进行对比分析。识别结果表明,本文改进的识别算法也能才准确识别出目标。主要归咎于结果红外与可见光的融合图像有助于提升目标的抗干扰能力,以及改进网络具有较高的泛化能力。以上分析表明,本文改进网络的边界框位置与基准目标之间的拟合度也比对比算法更为准确,充分验证了结合注意机制的辅助网络有助于提高目标的识别精度。

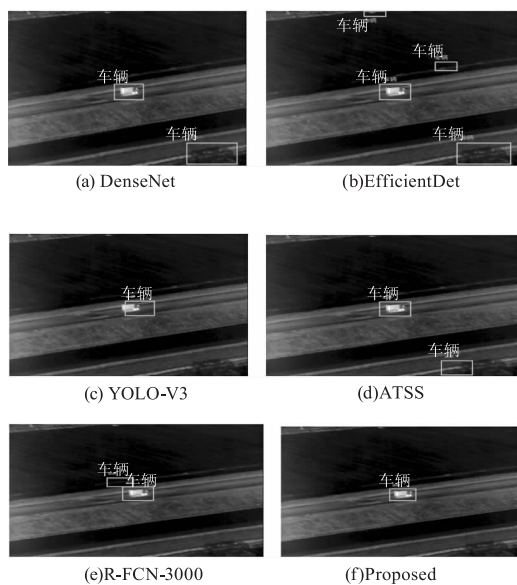


图 7 不同算法的识别结果对比

Fig. 7 Comparison of recognition results for different algorithms

本文的算法首先对红外图像与可见光图像进行融合,然后进行显著性检测快速锁定疑似目标区域;然后利用改进的识别网络对疑似区域进行识别,大大降低了搜索范围,提高了识别速度。实验结果表

明,改进网络其识别准确率比现有最优检测网络略有提升,但实时性得到了大大增强。

6 结 语

本文提出了一种基于改进 YOLO 网络的双通道显著性目标识别算法,该算法利用红外图像与可见光互补特性进行多尺度融合,并在融合图像上采用显著性检测获取疑似目标区域,最后利用改进的识别网络对疑似区域进行多层次目标识别。改进的 YOLO 识别网络增加了一路辅助网络,改善整个特征提取网络的性能,并采用注意机制对辅助网络和骨干网络的特征信息融合,增强有效信息通道,抑制无效信息通道,提高网络识别效率。实验结果表明,本文提出的模型可以有效地提高现有网络的性能。下一步,项目组将进行工程化研制,全面提升装备的智能化水平。

参考文献:

- [1] Chen Ji, Zhang Ruolan. Panoramic imaging technology applied inIRST and status awareness wystem[J]. Infrared Technology, 2016, 38(4): 269-279. (in Chinese)
陈洁, 张若岚. 应用于红外搜索跟踪和态势感知系统的全景成像技术[J]. 红外技术, 2016, 38(4): 269-279.
- [2] Gonzalez-Aguirre D, Asfour T, Dillmann R. Towards stratified model-based environmental visual perception for humanoid robots[J]. Pattern Recognition Letters, 2011, 32(16): 2254-2260.
- [3] Li Yingjuan, Ji Ming, Hou Chunping, et al. Indirect Line-of-sight Stabilization Technique of Vehicular Gunner's Periscopic Sight[J]. Journal of Missiles and Guidance, 2015, 35(6): 123-126. (in Chinese)
李颖娟, 纪明, 侯春屏, 等. 车载炮长镜瞄准线间接稳定技术[J]. 弹箭与制导学报, 2015, 35(6): 123-126.
- [4] Zheng H, Mao X, Chen L, et al. A novel method for quantifying target tracking difficulty of the infrared image sequence[J]. Infrared Physics & Technology, 2015, 72: 8-18.
- [5] Brandon B, Joshua L, J Karl H. Airborne, Autonomous & Collaborative[J]. Mechanical Engineering, 2011, 133(4): 29-37.
- [6] Tang Z, Wang S, Huo J, et al. Bayesian framework with non-local and low-rank constraint for image reconstruction[C]//Journal of Physics: Conference Series. IOP Publishing, 2017, 787(1): 012008.
- [7] Luo C, Sun B, Deng Q, et al. Comparison of Different Level Fusion Schemes for Infrared-Visible Object Tracking: An Experimental Survey[C]//2018 2nd International Confer-

- ence on Robotics and Automation Sciences (ICRAS). IEEE, 2018:1592 – 1609.
- [8] Liu G, Liu S, Liang L, et al. Scene Matching for Infrared and Visible Images with Compressive Sensing SIFT Feature Representation in Bandelet Domain[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2019, (1):2054029.
- [9] 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.
- [10] Liu Jixin, Wei Man. Scene sparse recognition method via intra – class dictionary for visible and near – infrared HSV image fusion[J]. Journal of Computer Applications, 2018, 38(12):3355 – 3359. (in Chinese)
刘佳鑫, 魏曼. 可见光 – 近红外 HSV 图像融合的场景类字典稀疏识别方法[J]. 计算机应用, 2018, 38(12):3355 – 3359.
- [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] Kaiming H, Georgia G, Piotr D, et al. Mask R-CNN[C]//Proceedings of the IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017:2961 – 2969.
- [13] Redmon J, Farhadi A. YOLOv3: An incremental improvement[J]. arXiv preprint arXiv:2018, 1804:02767.
- [14] Jiang Zetao, Wu Hui, Zhou Xiaoling. Infrared and visible image fusion algorithm based on improved guided filter and dual channel pulse emission cortex model[J]. Acta Optica Sinica, 2018, 12(3):238 – 247. (in Chinese)
江泽涛, 吴辉, 周晓玲. 基于改进引导滤波和双通道脉冲发放皮层模型的红外与可见光图像融合算法[J]. 光学学报, 2018, 12(3):238 – 247.
- [15] Sarfraz M S, Stiefelhagen R. Deep perceptual mapping for Cross-Modal face recognition[J]. International Journal of Computer Vision, 2017, 122(3):426 – 438.
- [16] Li T, Song H, Zhang K, et al. Learning Residual Refinement Network with Semantic Context Representation for Real-Time Saliency Object Detection[J]. Pattern Recognition, 2020, 105:107372.
- [17] 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.
- [18] Zhanga Q, Shib T, Wangb F, et al. Robust sparse representation based multi-focus image fusion with dictionary construction and local spatial consistency [J]. Pattern Recognition, 2018, 83(18):299 – 313.
- [19] Jiang Xiurong, Zhu Lin, Hao Yuanhong, et al. Moving target detection method based on RPCA and spatial correlation constraints [J]. Firepower and Command Control, 2018, 43(9):130 – 134. (in Chinese)
蒋秀蓉, 朱林, 郝元宏, 等. 基于 RPCA 与空间相关约束的运动目标检测方法[J]. 火力与指挥控制, 2018, 43(9):130 – 134.
- [20] Chen Tieming, Fu Guangyuan, Li Shiyi, et al. Typical target detection for infrared homing guidance based on Yolo – v3[J]. Progress in Laser and Optoelectronics, 2019, 56(16):155 – 162. (in Chinese)
陈铁明, 付光远, 李诗怡, 等. 基于 YOLO – v3 的红外末制导典型目标检测[J]. 激光与光电子学进展, 2019, 56(16):155 – 162.
- [21] Huang G, Liu Z, Maaten L V D, et al. Densely connected convolutional networks [J]. Computer Era, 2017, 37(3):160 – 164.
- [22] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2016:1841 – 1849.
- [23] Tan M, Pang R, Le Q V. EfficientDet: Scalable and Efficient Object Detection [C]//IEEE Conference on Computer Vision & Pattern Recognition, 2019:612 – 629.
- [24] 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.
- [25] Zhang S, Chi C, Yao Y, et al. Bridging the gap between anchor – based and anchor – free detection via adaptive training sample selection [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:9759 – 9768.
- [26] Yu Zhichao, Zhang Ruihong. Improved siamese network based object tracking combined with the deep contour feature [J]. Journal of Xidian University, 2020, 47(3):40 – 49. (in Chinese)
余志超, 张瑞红. 结合深度轮廓特征的改进孪生网络跟踪算法[J]. 西安电子科技大学学报, 2020, 47(3):40 – 49.