

无人机作战平台的智能目标识别方法

吕攀飞*, 王曙光

陆军炮兵防空兵学院, 安徽 合肥 230031

摘要 深入研究了流行的目标识别方法 YOLOv3, 将 Inception 模块融入其特征提取网络 darknet-53 中, 从而得到新网络 darknet-139。相比 YOLOv3 特征提取网络, 新网络具有更好的特征提取能力。采集并制作算法所需的数据集, 分别在 YOLOv3 和本文算法上进行训练并测试。实验结果表明, 相比 YOLOv3, 本文算法的平均识别率提升了约 2%。

关键词 图像处理; 无人机作战平台; 人工智能; 目标识别

中图分类号 TP391.7 文献标识码 A

doi: 10.3788/LOP56.071001

Intelligent Target Recognition Method of Unmanned Aerial Vehicle Combat Platform

Lü Panfei*, Wang Shuguang

Army Artillery and Air Defense Forces Academy, Hefei, Anhui 230031, China

Abstract The popular target recognition method YOLOv3 is deeply studied, and the Inception module is integrated into the feature extraction network darknet-53 to get a new network darknet-139. Compared with YOLOv3, the new network has better ability in feature extraction. The data set required by the algorithm is collected and made, and were trained and tested on YOLOv3 and the proposed algorithm, respectively. The experimental results show that the average recognition rate of the proposed algorithm is about 2% higher than that of YOLOv3.

Key words image processing; unmanned aerial vehicle combat platform; artificial intelligence; target recognition

OCIS codes 100.4999; 100.2960

1 引言

在现代战争中, 装备的智能化与信息化已经成为未来战争克敌制胜的关键。快速获取战场动态信息, 掌握敌方作战意图, 对于我方制定战场决策和抢先进行作战部署具有重大意义。现代战争对制信息权的要求不断提高, 无人化装备因其可以突破人的生理和心理限制, 以及具有快速、高机动性、成本较低、可深入敌后作战的特点而逐渐成为世界各国的重点发展方向^[1]。

无人机在作战中的大量使用, 离不开新技术的发展。人工智能技术的不断成熟, 尤其是智能目标识别的出现, 为无人机装上了“火眼金睛”, 而智能目

标识别的关键技术就是神经网络。传统的目标识别方法中, 多尺度形变部件(DPM)算法的效果较为突出, 获得了 2007—2009 年 Pascal VOC (visual object class) 目标识别冠军。随着基于卷积神经网络的目标识别算法(R-CNN)的出现^[2], 大多数致力于改进 DPM 的研究者纷纷转向深度学习研究领域^[3]。近几年来, 受益于计算机硬件性能的提升和机器学习理论的不完善, 基于卷积神经网络的目标识别方法得到了快速发展^[4-5], 先后出现了基于建议区域的 R-CNN 系列^[6]和端到端的 YOLO(You Only Look Once)系列方法。R-CNN 系列虽然在识别精度上高于 YOLO, 但识别耗时太长, 不适用于无人机作战平台; YOLO 系列方法虽然在识别准

收稿日期: 2018-09-17; 修回日期: 2018-10-17; 录用日期: 2018-10-22

基金项目: 陆军装备部“十三五”预研基金

* E-mail: 1055392772@qq.com

确率上稍逊一筹,但可以实现实时识别^[7]。本文结合当前在目标分类中具有领先地位的 Inception 网络模型,利用其在特征提取方面的独特优势对 YOLOv3 方法进行改进,提高网络的特征提取能力,进而提高训练模型的识别准确率。

2 无人机平台目标识别算法关键技术

无人机用于战场侦察时,主要采用视频摄像头作为战场信息的载入端,对敌方区域的人员与装备等信息进行实时采集^[8]。在传统的无人机侦察技术中,通常将采集到的信息通过通信系统送入地面站的信息中心,在信息中心利用目标识别算法与数据库中已有的目标信息进行匹配,从而确认目标区域装备和人员信息,以节省人员消耗,提高作战效率^[9]。这一过程中最关键的就是目标识别算法的准确性和实时性:若算法的稳健性太差,则会严重影响识别效果,无法发现敌方目标;若算法运行复杂度较高,则会导致识别缺乏实时性,从而失去快速掌握战场信息的主动权。因此,对于目标识别算法的选择和改进是实现无人机战场侦察打击能力的重点。基于卷积神经网络的目标识别算法使战场目标的智能化识别成为可能^[2],通过智能目标识别,可以将地面站对信息的处理过程迁移到无人机平台上,从而真正实现无人化侦察预警。

3 目标识别算法

3.1 基于卷积神经网络的目标识别算法

近年来,基于卷积神经网络的目标识别算法发展迅速,根据目标识别的流程,主要可分为二阶段识别方法和单阶段识别方法^[10]:1) 二阶段的识别方法,如 R-CNN 识别方法,首先通过卷积神经网络进行特征提取,然后将提取到的特征分别送入分类器进行分类和边界框回归网络进行目标定位,从而确定图像中目标的类别和位置信息^[11-12];2) 单阶段识别方法,如 YOLO 算法,将特征提取和目标分类与位置回归作为整体进行训练,相较于二阶段识别方法,单阶段识别方法在速度上有绝对优势^[13]。

YOLOv3 是 Redmon 等^[14]提出的基于卷积神经网络的目标识别方法,该方法的特征提取网络共包含 53 个卷积层,因此也被称为 darknet-53,如图 1 所示。网络中大量使用残差连接,在增加网络深度的同时又降低了训练过程中出现过拟合的风险^[15];网络去掉了最大池化层,而采用步长为 2 的卷积层进行代替,实现了全卷积网络,这样可以最大程度地保

持网络信息传输的完整性;网络中只采用 1×1 和 3×3 的小卷积核,大大降低了网络的计算复杂度,提高了网络执行效率。在目标检测阶段,YOLOv3 首先将最终 13×13 的特征图做多次卷积处理后进行检测,得到第一次检测结果;然后将 13×13 的特征图上采样得到 26×26 的特征图,将这个特征图与之前网络中 26×26 的特征图进行特征融合,形成新的特征图,随后进行多次卷积后送入检测层,得到第二次检测结果;再对 26×26 的特征图进行上采样,并与之前层进行特征融合,得到 52×52 的特征图,然后多次卷积送入检测层,得到第三次检测结果。将检测结果进行非极大值抑制处理后就得到了最终的识别效果。

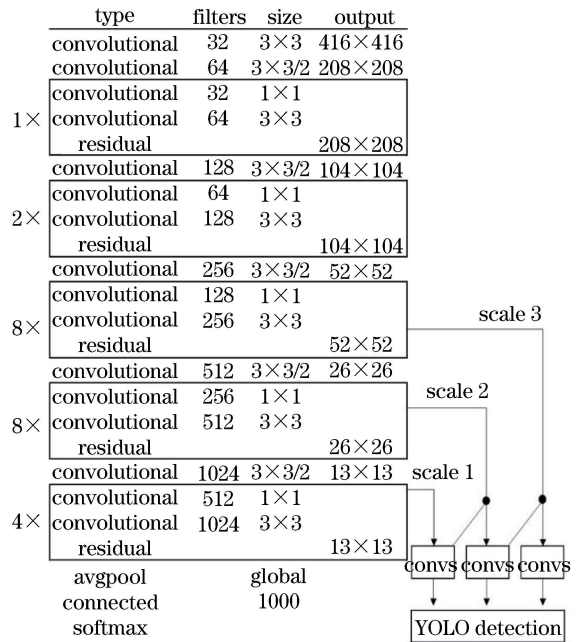


图 1 YOLOv3 特征提取与识别示意图

Fig. 1 Diagram of YOLOv3 feature extraction and recognition

3.2 改进的特征提取网络

尽管 YOLOv3 在目标识别算法中表现优异,但面对一些特征不明显的小目标时仍然缺乏判断力,容易出现错误识别和遗漏目标的问题。本文对 YOLOv3 的特征提取网络 darknet-53 进行改进,提高网络对图像特征的提取能力。

Inception-v4 网络模型是目前分类网络中性能最好的模型之一,这主要得益于其使用的 Inception 模块可以对网络的横向进行拉伸,从而增强网络对上层信息的传递能力^[16]。如图 2 所示,Inception 模块将上层网络信息通过三个连接并行地传递到下一层。相比普通的卷积层,Inception 模块具有更强的获取上层信息的能力。

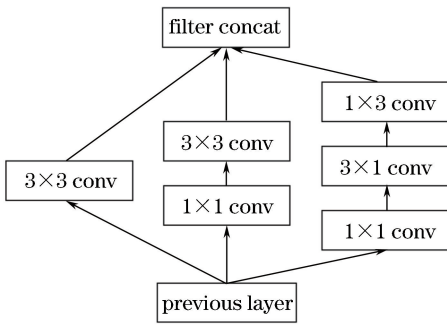


图 2 Inception 模块

Fig. 2 Inception module

对特征提取网络的改进思路主要是将 Inception 模块融入 darknet-53 网络,提高网络的横向宽度,从而提高网络对图像特征的提取能力。图 3 所示为 Inception 模块与 darknet-53 网络中的残差连接结合后形成的新的 Inception-ResNet 模块。首先将上层网络特征图通过三个并行的卷积层进行处理,然后进行特征融合,并通过 1×1 的卷积核改变其输出通道,最后通过短连接与输入特征叠加形成残差连接,得到 Inception-ResNet 模块。图中的 U 为输入的通道数, D 为输出通道数,模块中除了特征融合阶段的 1×1 卷积采用线性激活函数外,其余均采用线性整流函数 (ReLU) 等非线性激活函数。

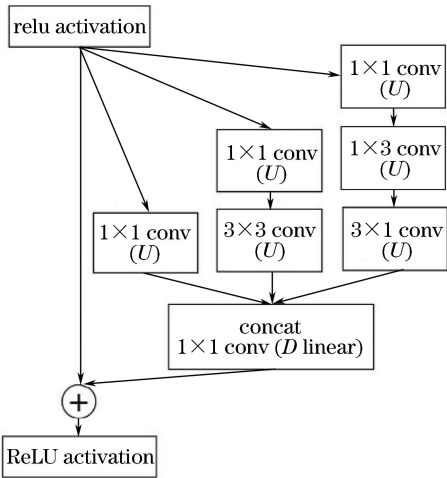


图 3 Inception-ResNet 模块

Fig. 3 Inception-ResNet module

将得到的 Inception-ResNet 模块应用到 darknet-53 特征提取网络中,得到新的特征提取网络 darknet-139,如图 4 所示。

尽管新网络 darknet-139 看起来比 darknet-53 复杂很多,但网络复杂度并没有增加很多。表 1 为两个网络的时间复杂度 $T(n)$ 和空间复杂度 $S(n)$ 的数据对比,可以看出,两个网络的复杂度差距控制

	type	filters	size	output
1 ×	convolutional	32	3 × 3	416 × 46
	convolutional	64	3 × 3/2	208 × 208
	inception	U=32		
	residual	D=64		208 × 208
2 ×	convolutional	128	3 × 3/2	104 × 104
	inception	U=64		
6 ×	residual			104 × 104
	convolutional	256	3 × 3/2	52 × 52
	inception	U=128		
	residual	D=256		52 × 52
6 ×	convolutional	512	3 × 3/2	26 × 26
	inception	U=256		
4 ×	residual	D=512		26 × 26
	convolutional	1024	3 × 3/2	13 × 13
	inception	U=512		
	residual	D=1024		13 × 13
	avgpool		global	
	connected		1000	
	softmax			

图 4 darknet-139 示意图

Fig. 4 Diagram of darknet-139

表 1 网络计算复杂度比较

Table 1 Comparison of computational complexity of different networks

Network	$T(n)$	$S(n)$
darknet-53	2.45×10^{10}	4.05×10^7
darknet-139	2.69×10^{10}	4.81×10^7

在一个数量级内。这是因为 darknet-139 网络中使用的 Inception-ResNet 模块虽然与卷积核效果相同,但计算复杂度小于卷积核,因此整个网络的计算复杂度并没有增加很多。

4 目标识别实验

4.1 目标识别实验

将新的特征提取网络 darknet-139 加载到 YOLOv3 目标识别算法中进行训练。实验设备配置如下:CPU, Intel(R) Xeon(R) E5-2683 v3, 主频 2.00 GHz;内存 32 GB;显卡, 英伟达 Quadro K4200 和英伟达 Quadro M4000 (双显卡);操作系统, Windows 10。实验采用的数据集由无人机采集与网络爬取结合的方式构建,共收集图片 40024 张,主要包含坦克、远火发射车、舰船、自行火炮、高射炮 5 类目标。将收集的图像按照 1:2:7 的比例进行划分,其中 10% 作为测试集,20% 作为验证集,70% 作为训练集,对训练集和验证集图像中的目标进行类别和位置标注后送入网络进行训练。

训练采用的学习率为 0.0001,优化算法采用

Momentum 梯度下降算法,动量参数设为 0.9,每次迭代的批量大小为 64,学习率随迭代轮数更新的下降参数为 0.0005。经过 50000 次迭代后网络最终收敛,收敛时的损失值为 0.050830。图 5 是网络训练过程中的损失变化曲线,可以看出网络损失处于稳定的下降状态并最终收敛。图 6 是 10000~50000 次迭代期间的损失变化,可以看出,迭代次数为 25000 时网络参数已经非常稳定。

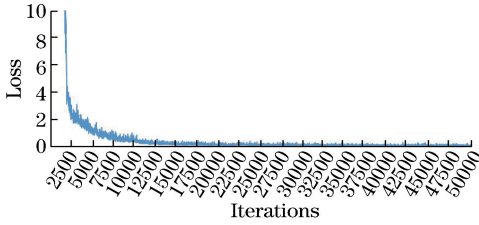


图 5 总体损失曲线

Fig. 5 Curve of total loss

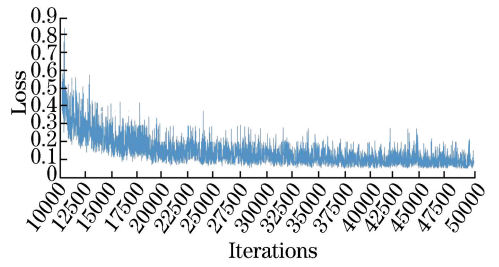


图 6 10000~50000 次迭代的损失曲线

Fig. 6 Loss curve of 10000-50000 iterations

行识别,将部分卷积的中间结果进行可视化,得到如图 7 所示的卷积结果。可以看出,浅层卷积层提取到的特征具有全局性,通常是人眼可判别的物体轮廓。随着卷积层数逐渐加深,深层卷积层提取到的特征越来越抽象,主要是人眼不可判别的图像细节特征。正是由于卷积神经网络可以提取到图像中的高层次语义特征,因此,基于深度学习的目标识别效果要优于传统识别算法。

采用训练好的模型对包含远火发射车的图像进

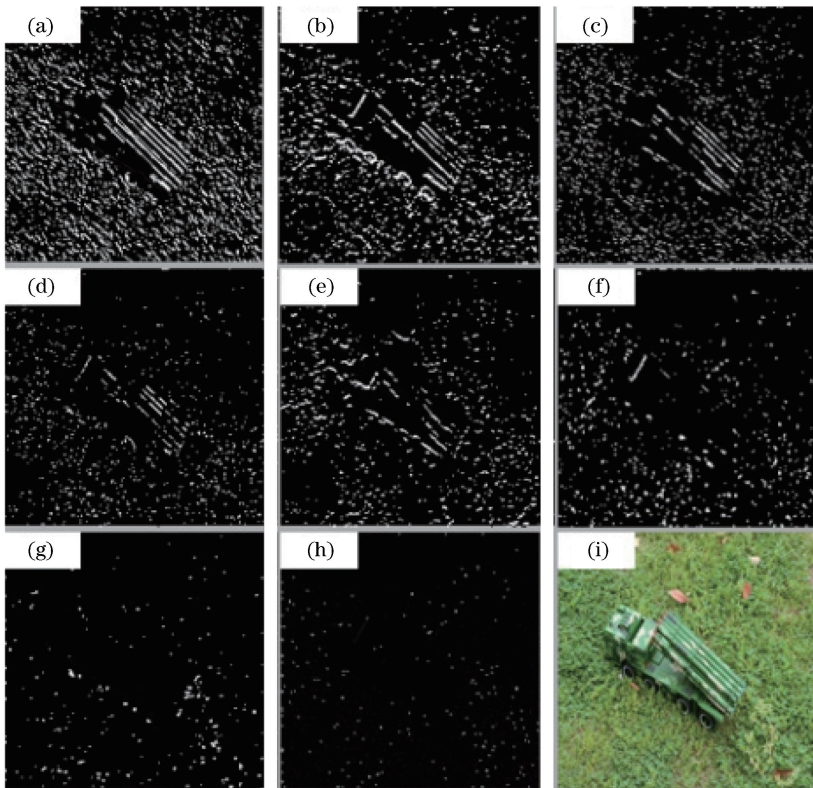


图 7 部分中间卷积结果。(a)第 1 层;(b)第 2 层;(c)第 3 层;(d)第 4 层;(e)第 5 层;(f)第 6 层;(g)第 7 层;(h)第 8 层;(i)原图

Fig. 7 Partial intermediate convolution results. (a) Layer 1; (b) layer 2; (c) layer 3; (d) layer 4; (e) layer 5; (f) layer 6; (g) layer 7; (h) layer 8; (i) original graph

将训练好的网络权重加载到识别网络中进行测试,得到如图 8 所示的识别效果。可以看出,模型对各类目标均能够准确识别,并且能够对目标的类别

进行标注,这样无人机就可以对敌方单位进行智能侦察和打击,实现真正意义上的无人化。

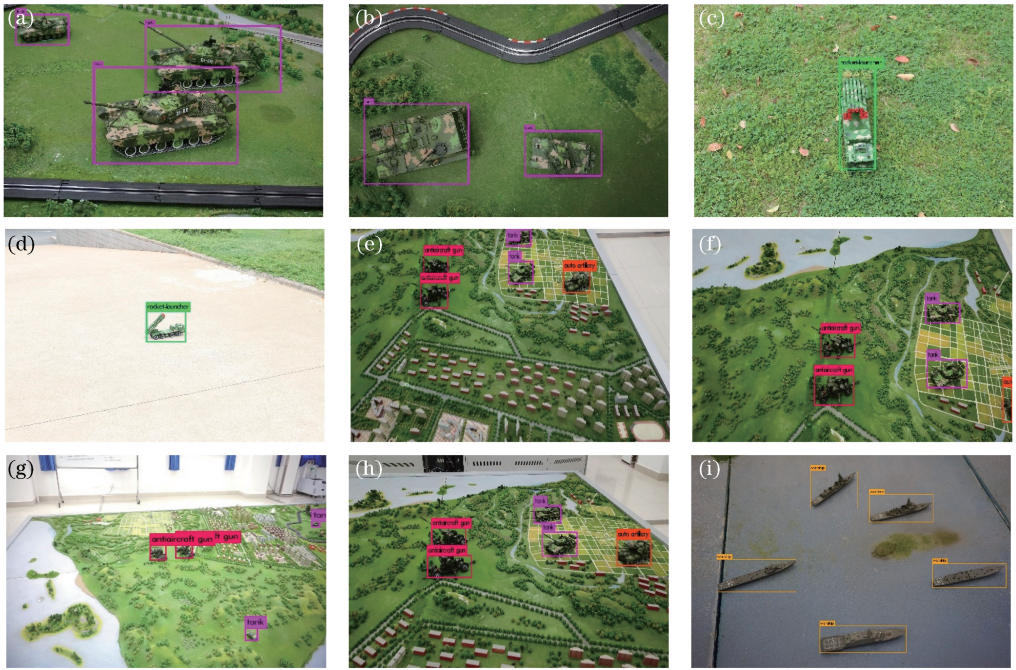


图 8 目标识别效果。(a)(b)坦克目标识别;(c)(d)远程火箭炮识别;(e)~(h)高射炮识别;(i)舰船识别
 Fig. 8 Target recognition effects. (a)(b) Tank target recognition; (c)(d) long-range rocket gun recognition;
 (e)-(h) antiaircraft gun recognition; (i) warship recognition

4.2 目标识别对比实验

为进一步了解新网络对图像特征的提取能力和目标识别准确率的提升效果,将新网络的训练模型与 YOLOv3 的训练模型进行对比测试。算法平均识别率表示为

$$A_p(t) = \frac{1}{c} \sum_c \frac{T_p(t)}{T_p(t) + F_p(t)}, \quad (1)$$

式中: $T_p(t)$ 为正确识别的样本数量; $F_p(t)$ 为错误识别的样本数量; t 为交并比 (IOU) 阈值。将算法对每类样本的识别率相加并求平均值,即可得到算法的平均识别率 A_p ^[17]。表 2 是两种算法在相同测试集上的平均识别率和识别速度,其中 A_p 为 IOU 取均值时的平均识别率, A_{p50} 为 IOU 取 0.5 时的平均识别率。可以看出,新网络相比之前网络在平均识别率上有所提升。 T_m 为算法从读入一幅图片到输出识别结果所花费的时间,与 YOLOv3 识别算法相比,新算法检测每张图片的时间多了 5 ms,但识别速度仍然较快。

表 2 算法性能对比

Table 2 Performance comparison of algorithms

Algorithm	$A_p/\%$	$A_{p50}/\%$	T_m/ms
YOLOv3	85.50	87.46	20
Proposed	87.54	89.73	25

图 9 是两个模型对相同图像的检测结果,其中,

图 9(a)(b)为 YOLOv3 的识别效果,图 9(c)(d)为新模型的识别效果。从图 9 (a)可以看出, YOLOv3 不能准确识别图像中的小型目标;图 9(b)在对坦克目标的识别过程中出现了明显的错误识别,将附近的植被也误识别为坦克;图 9(c)中模型对不同尺寸目标的泛化能力较强,能够将图像中的小型目标识别出来;图 9(d)中新模型将目标从植被中分离出来,实现了目标与背景的准确区分。

5 结 论

首先对无人机作战平台在现代战争中的作用以及实现无人机对敌方目标智能识别的关键技术进行分析,阐明了目标识别算法在自动目标识别中的重要作用;然后以智能目标识别领域具有较高识别准确率和实时性的 YOLOv3 算法为基础,对卷积神经网络目标识别方法进行介绍,并将特征提取能力较强的 Inception 模块融入到 YOLOv3 算法中,提升了算法的识别准确率,取得了较好的识别效果。由于提高算法识别准确率时通常需要增加算法复杂度,这导致本文方法在识别速度上略有降低,而要提高识别速度通常又需要降低算法复杂度;因此,对算法的进一步改进将着重于减少网络的冗余设计,提高模型识别速度。另外,也可以通过提高硬件性能的方式提高算法执行效率。

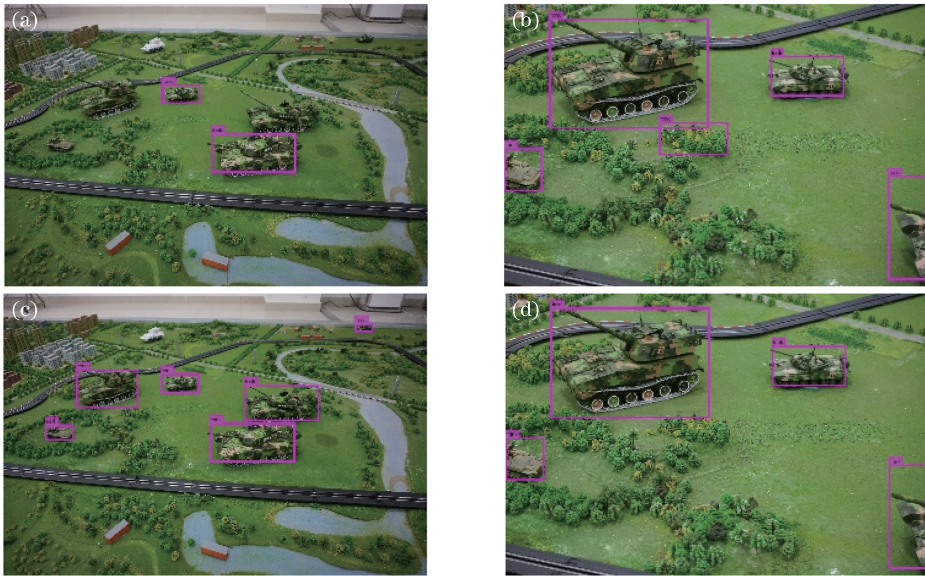


图9 模型识别效果比较。(a)(b) YOLOv3; (c)(d)新模型

Fig. 9 Comparison of model recognition effect. (a)(b) YOLOv3; (c)(d) proposed model

将人工智能技术应用于各类作战平台可以大幅度提高无人化作战效率。例如,将人工智能用于声波分析可以从地表震动分析敌方地面武器装备的型号和动向^[18];在瞄准装置上增加智能瞄准可以更快地锁定敌方单位,减少己方伤亡。随着人工智能技术的发展,智能导航、智能决策等更加先进的技术将大大提高现代战争的作战效率。

参 考 文 献

- [1] Du Y L, Li J Z, Zhang Y, *et al.* Overview of target recognition algorithms in UAV battlefield reconnaissance [J]. *Aerodynamic Missile Journal*, 2016(7): 38-41, 53.
杜玉龙, 李建增, 张岩, 等. 无人机战场侦察中目标识别算法研究综述[J]. *飞航导弹*, 2016(7): 38-41, 53.
- [2] Ou P, Zhang Z, Lu K, *et al.* Object detection in remote sensing images based on convolutional neural networks [J]. *Laser & Optoelectronics Progress*, 2019, 56(5): 051002.
欧攀, 张正, 路奎, 等. 基于卷积神经网络的遥感图像目标检测研究[J]. *激光与光电子学进展*, 2019, 56(5): 051002.
- [3] Zhang H, Wang K F, Wang F Y. Advances and perspectives on applications of deep learning in visual object detection [J]. *Acta Automatica Sinica*, 2017, 43(8): 1289-1305.
张慧, 王坤峰, 王飞跃. 深度学习在目标视觉检测中

的应用进展与展望[J]. *自动化学报*, 2017, 43(8): 1289-1305.

- [4] Gu Y, Xu Y. Architecture design of deep convolutional neural network for SAR target recognition [J]. *Journal of Image and Graphics*, 2018, 23(6): 928-936.
谷雨, 徐英. 面向 SAR 目标识别的深度卷积神经网络结构设计[J]. *中国图象图形学报*, 2018, 23(6): 928-936.
- [5] Russakovsky O, Deng J, Su H, *et al.* ImageNet large scale visual recognition challenge [J]. *International Journal of Computer Vision*, 2015, 115 (3): 211-252.
- [6] Uijlings J R R, van de Sande K E A, Gevers T, *et al.* Selective search for object recognition [J]. *International Journal of Computer Vision*, 2013, 104 (2): 154-171.
- [7] Redmon J, Farhadi A. YOLO₉₀₀₀: better, faster, stronger[C] // *IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 6517-6525.
- [8] Sun H Z, Chang T Q, Wang Q D, *et al.* Image detection method for tank and armored targets based on hierarchical multi-scale convolution feature extraction[J]. *Acta Armamentarii*, 2017, 38(9): 1681-1691.
孙皓泽, 常天庆, 王全东, 等. 一种基于分层多尺度卷积特征提取的坦克装甲目标图像检测方法[J]. *兵工学报*, 2017, 38(9): 1681-1691.
- [9] Lim T R, Guntoro A T. Car recognition using Gabor

- filter feature extraction[C] // Asia-Pacific Conference on Circuits and Systems, 2002: 451-455.
- [10] Chen L L, Zhang Z D, Peng L. Real time detection method based on improved single shot multibox detector [J]. Laser & Optoelectronics Progress, 2019, 56(1): 011002.
陈立里, 张正道, 彭力. 基于改进 SSD 的实时检测方法[J]. 激光与光电子学进展, 2019, 56(1): 011002.
- [11] Girshick R. Fast R-CNN [C] // IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [12] Ren S Q, He K M, 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, 2017, 39(6): 1137-1149.
- [13] Redmon J, Divvala S, Girshick R, *et al.* You only look once: unified, real-time object detection[C] // IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [14] Redmon J, Farhadi A. Yolov3: an incremental improvement[J]. arXiv preprint arXiv: 1804.02767, 2018.
- [15] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition[C] // IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [16] Szegedy C, Ioffe S, Vanhoucke V, *et al.* Inception-v4, inception-Resnet and the impact of residual connections on learning [C] // AAAI Conference on Artificial Intelligence (AAAI-17), 2017: 4278-4284.
- [17] Wei Y M, Quan J C, Hou Y Q Y. Aerial image location of unmanned aerial vehicle based on YOLO v2[J]. Laser & Optoelectronics Progress, 2017, 54(11): 111002.
魏湧明, 全吉成, 侯宇青阳. 基于 YOLO v2 的无人机航拍图像定位研究[J]. 激光与光电子学进展, 2017, 54(11): 111002.
- [18] Hinton G, Deng L, Yu D, *et al.* Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups [J]. IEEE Signal Processing Magazine, 2012, 29(6): 82-97.