

基于双通道的快速低空无人机检测识别方法

马旗*, 朱斌**, 程正东, 张杨

国防科技大学电子对抗学院脉冲功率激光技术国家重点实验室, 安徽 合肥 230037

摘要 以 YOLOv3 的架构为基础, 提出了一种基于双通道的快速低空无人机检测识别方法(Dual-YOLOv3)。该方法将红外与可见光的无人机图像同时输入到深度残差网络中进行特征提取, 对所提取的特征图进行融合以增强特征的表达能力, 利用多尺度预测网络对无人机目标进行类别判断和位置回归, 得到检测识别结果。在真实采集的双波段无人机数据集上进行对比实验, 结果表明, 采用平均融合的 Dual-YOLOv3-D 在 mAP(mean of average precision)上较单一数据源的 YOLOv3 提升了约 6.1%, 检测速度约为 27 s^{-1} 。

关键词 图像处理; 双通道卷积神经网络; 低空无人机; 特征融合; 目标检测

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/AOS201939.1210002

Detection and Recognition Method of Fast Low-Altitude Unmanned Aerial Vehicle Based on Dual Channel

Ma Qi*, Zhu Bin**, Cheng Zhengdong, Zhang Yang

State Key Laboratory of Pulsed Power Laser Technology, College of Electronic Engineering,
National University of Defense Technology, Hefei, Anhui 230037, China

Abstract Using the YOLOv3 architecture, we propose a recognition method for fast low-altitude unmanned aerial vehicle (UAV) detection based on the dual channel (Dual-YOLOv3). In this method, the infrared and visible UAV images are simultaneously input into the deep residual network for feature extraction, and the extracted feature maps are fused to enhance the expression ability of the features. Then, the multi-scale prediction network is used to determine the classification and the position regression of the UAV targets. Finally, we obtain the detection and recognition results. Comparison experiments are conducted on the real collected dataset of dual-band UAVs. The results show that the mAP (mean of average precision) of Dual-YOLOv3-D with average fusion is improved by 6.1% as compared with that of YOLOv3 with the single data source; the detection speed is approximately 27 s^{-1} .

Key words image processing; dual-channel convolutional neural network; low-altitude UAV; feature fusion; object detection

OCIS codes 100.4996; 100.3008; 100.2000

1 引言

近年来,随着自动控制技术的快速发展和生活水平的不断提高,民用小型无人机逐渐进入社会的各个领域,并深受大众的喜爱。无人机在推动社会进步的同时,也对个人隐私、社会安全、军事安全等领域构成了严重威胁。例如,一些不法分子低价购置小型无人机设备进行非法拍摄活动,或者进行简单的改装后挂载危险物品(有毒物品、爆炸物等)来实施恐怖袭击^[1]。因此,快速高效的低空无人机检测是应对无人机威胁的必备条件。

目前,光电探测是应对威胁的常用手段,并且主

要以可见光探测设备为主。同时,由于无人机目标的机身外壳及飞行状态下的电池都具有红外热辐射特性,将可见光与红外探测两种手段相结合能够发挥更强的效果。但通常情况下,两者的工作方式是相互独立,互不干扰的,检测方式也是对单一信道下的图像进行处理,检测过程中只用到了—种图像信息,不能充分利用红外与可见光的目标特征。因此,鉴于此种考虑,将两种通道的信息相结合有助于提高光电探测下的检测水平。

彩色(RGB)图像与红外热(Thermal)图像相结合(RGB-T)已被广泛应用于显著性检测、语义分割、物体检测、目标跟踪等领域。例如, Li 等^[2]提出

收稿日期: 2019-07-03; 修回日期: 2019-07-23; 录用日期: 2019-08-23

* E-mail: 905303927@qq.com; ** E-mail: zhubinee@163.com

了一种加权低秩分解的方法用于前景目标检测,在公共数据集上效果较好,并且也提出了一种新的显著性目标检测方法^[3],推动了该领域研究的进展。Ha 等^[4]提出了一种面向自动驾驶车辆的实时语义分割方法,结合卷积神经网络结构提升了分割的精度与速度。Zhai 等^[5]利用相关滤波的框架进行目标跟踪,提升了精度和效率。因此,利用 RGB-T 技术能够提升对低空无人机的检测识别。

随着人工智能的发展,目标检测方法逐渐转向了深度学习技术。这是由于卷积神经网络能够分析图像中不同层次的特征,模拟人脑对事物特征表现出的理解和表达能力,而且在检测识别方面表现出了强大的信息处理能力,在处理的性能与速度上也能较好地满足大数据时代的要求。目前,深度学习技术在目标检测识别领域应用广泛,涌现出很多优秀的方法,如 Faster R-CNN^[6]、R-FCN^[7]、SSD^[8]以及 YOLO^[9]等。对低空无人机的检测识别不仅有精度要求,而且有速度要求。而检测速度能够达到实时的典型方法为 SSD 和 YOLO。文献[10]中表明,在标准数据集上 YOLOv3 的检测精度和速度都优于 SSD 方法。

此外,现阶段研究的主要方向是对无人机的检测技术^[11-13],并没有对其类别型号进行识别。因此,通过以上分析,将红外与可见光图像信息相结合,构建双通道的卷积神经网络来提升现阶段对低空无人机的检测识别能力是十分可行的。本文以 YOLOv3 的架构为基础,提出一种基于双通道的快速低空无人机检测识别方法(Dual-YOLOv3),该方法将红外与可见光图像的特征融合,使检测识别性能得到了提升,并通过实验对不同阶段的融合与不同融合方式进行比较,结果表明,采用平均融合搭建的 Dual-YOLOv3-D 网络,在保持快速检测的同时检测识别性能也得到了提升。

2 YOLOv3 的原理

YOLOv3 作为一种能够直接预测到目标位置和类别的检测识别方法,与两阶段的方法存在较大差异。同时,YOLO 从第一代版本到第三代经历了很多次改进,主要的改进是主干网络的变化:首先利用深度残差网络提取图像的深度特征,加强了训练过程与网络收敛;然后利用特征金字塔的多尺度预测网络与锚点机制对预测位置进行精修,提升了对小目标的检测效果,且在检测速度上具有很大的优势。

2.1 深度残差网络

对图像进行特征提取是通过卷积神经网络中的卷积和池化之间的搭配操作完成的。根据网络的特性可知,深度越深,检测效果越好。但随着网络深度的继续加深,训练会更加困难,并且会出现梯度消失现象,阻碍网络的收敛,效果变差。而利用残差学习^[14]构建深度残差网络能够很好地解决这个问题。

深度残差网络是由一系列串联的残差块组成的,每个残差块结构类似,是由两个权重层之间的线性堆叠和跳过连接组成的,具体结构示意图如图 1 所示,图中带加号的实心圆圈表示跳过连接,对应的公式为

$$\mathbf{x}_{l+1} = \mathbf{x}_l + F(\mathbf{x}_l, W_l), \quad (1)$$

式中: \mathbf{x}_l 和 \mathbf{x}_{l+1} 分别是第 l 个残差块的输入向量与输出向量; $F(\mathbf{x}_l, W_l)$ 表示转换函数, W_l 为 \mathbf{x}_l 对应的权重。由这种残差块构成的深度网络有助于信息的流动,并且容易训练。

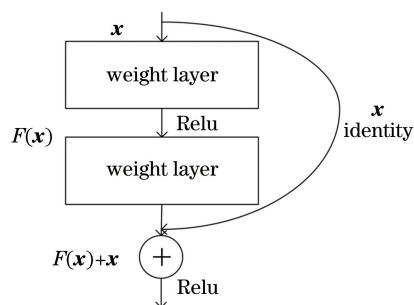


图 1 残差块结构示意图

Fig. 1 Schematic of residual block structure

YOLOv3 采用的是由 1×1 的卷积层(Convolutional layer)、 3×3 的卷积层与跳过连接层(Residual layer)堆叠构成的残差块,如图 2(a)所示。其中,卷积层为了加快网络训练时的收敛速度,在每个普通的 2D 卷积层上都添加了批归一化层(BN layer)^[15]与 leakyReLu 层,如图 2(b)所示。该卷积层进行卷积操作时步长(stride)默认为 1,扩展(padding)默认为“same”模式。

2.2 多尺度预测网络

由于深层的特征语义信息比较丰富,常见的目标检测方法都是只采用深层特征进行预测,但其预测的目标位置比较粗略。反观浅层特征,对应的语义信息较少,但是目标位置准确。因此,通过构建金子塔型^[16]的多尺度预测网络,能够对不同大小的特征层进行独立预测,进而适应尺寸较小的目标,并且预测位置的精度也能得到提高。

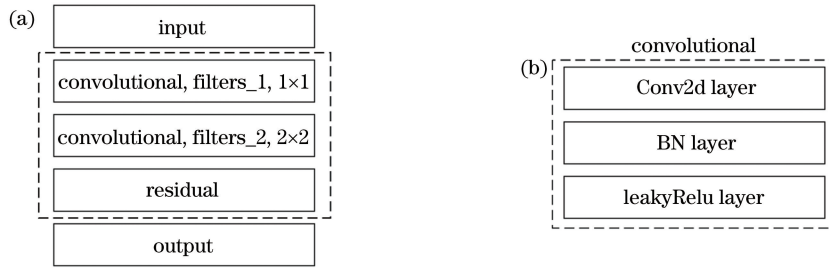


图 2 残差块结构与卷积层结构。(a)残差块结构;(b)卷积层结构

Fig. 2 Residual block structure and convolutional layer structure. (a) Residual block structure; (b) convolutional layer structure

图 3 为多尺度预测网络的特征融合思想,左侧部分为自下而上的特征提取网络,右侧为自上而下的预测网络。假设左侧最上端得到的特征图大小为 $A \times A$,传入右侧顶端的特征图大小也是 $A \times A$ 。之后,经过 2 倍大小的最近邻上采样,第二层与第三层的特征图大小变为 $2A \times 2A$ 与 $4A \times 4A$,将这两层特征与左侧得到的浅层特征进行融合,提升预测位置精度。同时,采用锚点机制^[17],选用 3 种尺度 9 个锚点,即每种尺度对应 3 个锚点。

在预测网络中,每种尺度对应的结构称为预测模块,每个模块都是由卷积层叠加而成的。如图 4 所示为 YOLOv3 采用的预测模块结构,前 6 层由 1×1 与 3×3 的卷积核堆叠构成,卷积核个数与输入特征图相匹配;最后一层为 1×1 大小的 2D 卷积,该层卷积核的个数为 $k \times (4 + 1 + s)$,其中的参数表示为每个模块对应 k 个锚点($k = 3$),4 个预测边界框的偏移量,1 个目标置信度以及 s 个目标类型。

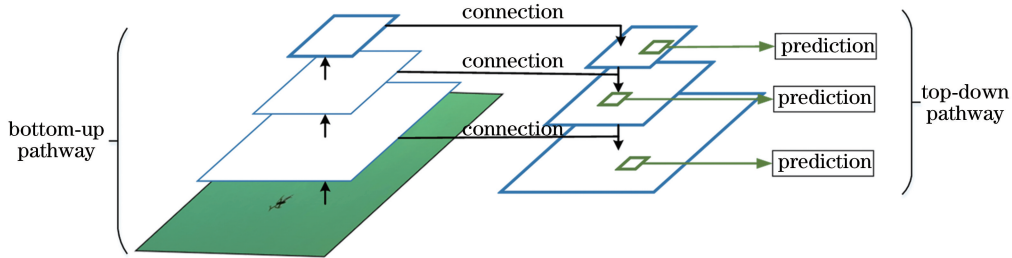


图 3 多尺度预测网络的特征融合思想

Fig. 3 Feature fusion idea of multi-scale prediction network

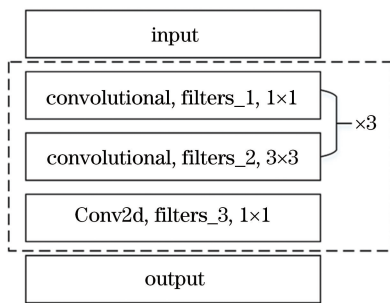


图 4 预测模块结构图

Fig. 4 Schematic of prediction module structure

3 Dual-YOLOv3 方案设计及网络结构

常用的双通道光电探测设备能够同时得到红外与可见光两种类型的图像。可见光具有丰富的纹理细节与场景信息,而红外能够通过热辐射获得目标的轮廓信息与突出的目标信息。两种图像的信息各

具特色,互为补充,将两者进行深度特征提取并加以融合,能够提升网络对特征图的描述能力与表达能力,进而提升模型的检测性能。同时,空中飞行的无人机具有目标小、尺度变化大等特点,利用多尺度预测网络结合上下文语义信息,能够较好地适应低空无人机目标的特性。再加上检测速度快的特性,YOLOv3 的主体架构适用于双通道检测识别网络。如图 5 所示为本文方法的总体设计方案。

3.1 总体设计方案

首先将红外与可见光图像进行预处理及图像配准操作得到配准后的图像(每一组配准图像称为一个图像对),然后将所得图像对同时送入 Dual-YOLOv3 网络中进行训练,最终得到用于低空无人机的检测识别模型。其中,Dual-YOLOv3 网络总体分为 3 个部分,第一个部分由 CNN_VIS 与 CNN_IR 两个结构相同的深度残差网络构成,用于两种图

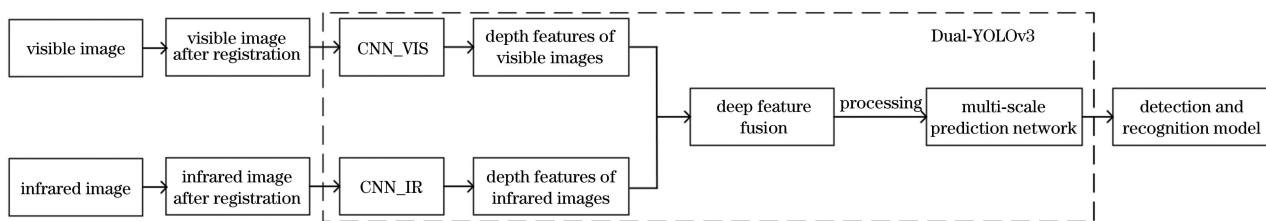


图 5 本文方法的总体设计方案

Fig. 5 Overall design scheme of proposed method

像的特征提取;第二部分是对提取到的深度特征进行特征融合以及更深层次的卷积操作,得到与后续预测网络相匹配的特征;第三部分则搭建金字塔型的多尺度预测网络对图像中不同尺寸目标进行类型识别与位置回归,输出最终的检测识别结果。

3.2 双波段无人机数据集的构建

由于本文方法使用的是基于双通道卷积神经网络的检测识别方法,同类方法较少,且没有公开的或标准的双波段无人机数据集。因此,针对此问题进行了数据集的采集及构建。

1) 数据的采集

通过基于可见光和红外探测器搭建双波段分光轴的图像采集系统,具体的采集系统实物如图 6 所示。

根据多源图像配准误差产生的原理,当探测系统参数固定时,配准误差与目标和探测器间的距离有关,距离越近,配准误差越大,二者呈反比例趋势。

同时,在采集无人机数据时,目标与复杂背景的图像会不断地产生配准误差,对大批量的图像配准造成困难。

因此,为了简化配准,以验证检测识别方法的有效性为主要目的,限定拍摄背景为纯净的天空背景并进行数据采集,选取了 3 种不同的民用无人机(如图 7 所示),包括大疆-精灵 3(标签为 DJ-3)、大疆-御 Pro(标签为 DJ-Pro)和大疆-S900(标签为 DJ-S900)。这 3 类无人机的类型分别为小型四旋翼无人机、中型四旋翼无人机以及大型六旋翼无人机,包含了常见的无人机类型,具有一定的代表性。无人机飞行时离设备的距离在 50~100 m 处变化,并且保持设备焦距固定。同时,为提升数据的多样性,采集过程中充分考虑了无人机飞行中的各种姿态,包括平稳飞行、盘旋、急速升降等。最终得到 3600 对可见光图像与红外图像,且目标像素大小的变化为 15×15 至 80×80。

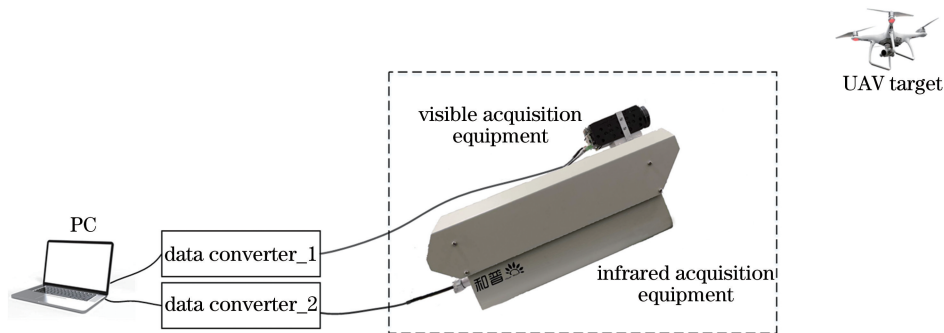


图 6 红外与可见光分光轴采集系统

Fig. 6 Optical axis acquisition system for infrared and visible light



图 7 无人机的类型。(a)大疆-御 Pro;(b)大疆-精灵 3;(c)大疆-S900

Fig. 7 Types of UAVs. (a) Dji-MAVIC Pro; (b) Dji-PHANTOM 3; (c) Dji-S900

2) 数据预处理

首先对采集的数据进行筛选。为确保数据的有效性,对无人机飞到图像边缘造成目标不完整的数据进行剔除,最终得到 3350 对红外与可见光图像。然后对数据进行配准处理。由于两种图像源的分辨率大小不同,需要对图像进行尺寸调整,将调整后的图像对进行配准处理,得到配准后的图像。最后对配准后的图像进行人工标注。由于数据已经配准,只需对任意一种数据源进行边界框标注即可得到所有数据的标签信息。最终,将所有数据按 4:1 的数量比例分为训练集和测试集。

3) 数据扩增

根据先前的经验,可以通过数据扩增的方式提升模型检测识别精度。因此,本文采取了两类操作对训练集进行数据扩增。数据增强操作:采取通用的图像增强方式,包括亮度、对比度的随机调整,对训练集进行处理。数据扩充操作:由于无人机飞行过程中姿态各异,包括平飞、倾斜等,因此扩充手段为 $\pm 10^\circ$ 、 $\pm 20^\circ$ 的旋转以及翻转。同样,如果数据经过扩增后,在图像中靠近边缘的目标出现残缺或完全丢失情况,则去除该样本数据。

4) 数据集构建

经过对无人机数据的采集、处理及扩增操作后,得到了最终的数据集 UAV_Dual。UAV_Dual 的训练集由经过预处理后的图像及其扩增的样本构成,共包含 50000 对红外与可见光样本数据。而测试集则选用步骤 2) 划分得到。

3.3 Dual-YOLOv3 的网络结构

根据不同阶段特征的特征融合,深度融合网络^[18]可分为浅层融合、中层融合和深层融合,设计了 3 种网络结构(分别为 Dual-YOLOv3-S、Dual-YOLOv3-M 与 Dual-YOLOv3-D)进行对比实验,最终通过测试得到性能最优的模型。

图 8、9、10 分别展示了 3 种不同的 Dual-YOLOv3 的网络结构图。图中 Res 块表示一个残差块, $\times 1$ 、 $\times 2$ 、 $\times 8$ 、 $\times 8$ 、 $\times 4$ 表示一种残差块串联的个数,顶部及侧边的数值为特征图的尺寸。Conv 块为预测模块,包含卷积、上采样和特征融合操作,构成了 3 种尺度的金字塔型预测网络。NMS 块为表示非极大值抑制操作,用于剔除重复预测的边界框,提高预测精度。feature fusion 与 concat 表示特征张量的融合与连接。下面对网络结构进行介绍。

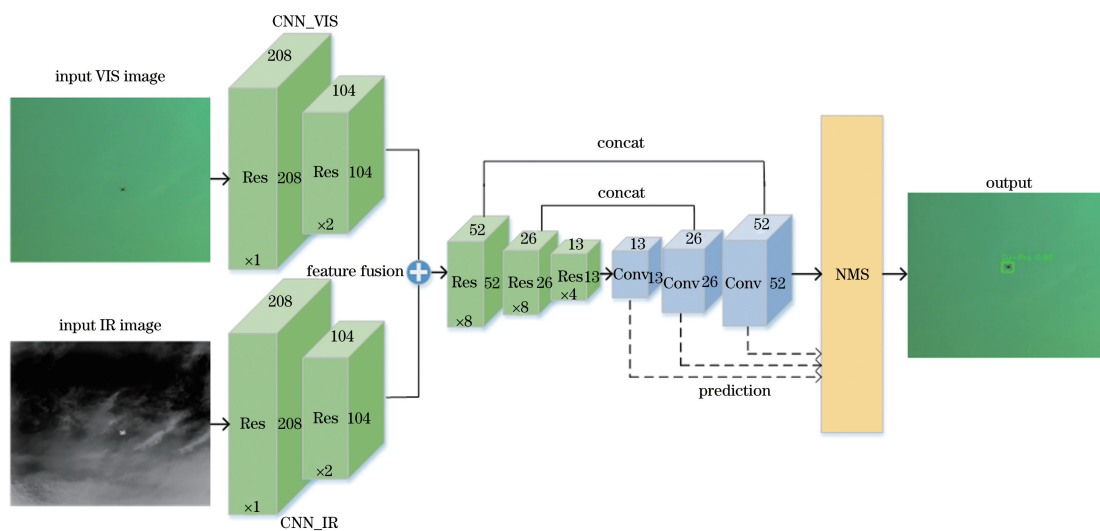


图 8 Dual-YOLOv3-S 网络结构图

Fig. 8 Network structure diagram of Dual-YOLOv3-S

根据图 2 的残差块结构,以每个网络中第一个残差块为例进行说明。残差块输入为 $208 \times 208 \times 32$ 的张量,第一层卷积用滤波数(filters_1)为 16、大小为 1×1 的卷积核实现;之后,把第一层的输出张量输入到第二层,用滤波数(filters_2)为 32、大小为 3×3 的卷积核实现滤波。以这两个卷积层为主,通过线性堆叠及跳过连接最终形成一个完整的残差

块。此外,双通道深度残差网络还需要其他操作来辅助构建。在每个残差块之前加入大小为 3×3 、步长为 2 的卷积核进行滤波,替代传统网络中池化的功能。例如,当输入大小为 416×416 的图像对时,会得到 5 种分辨率不同的特征图,分别为 208×208 , 104×104 , 52×52 , 26×26 , 13×13 。其中, 52×52 , 26×26 , 13×13 这三种分辨率的特征图将与多

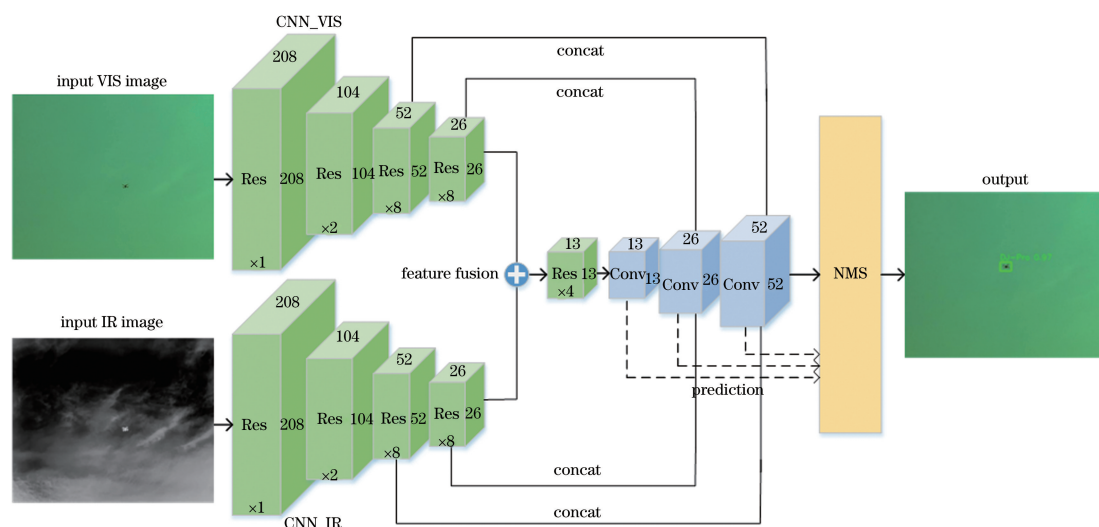


图 9 Dual-YOLOv3-M 的网络结构图

Fig. 9 Network structure diagram of Dual-YOLOv3-M

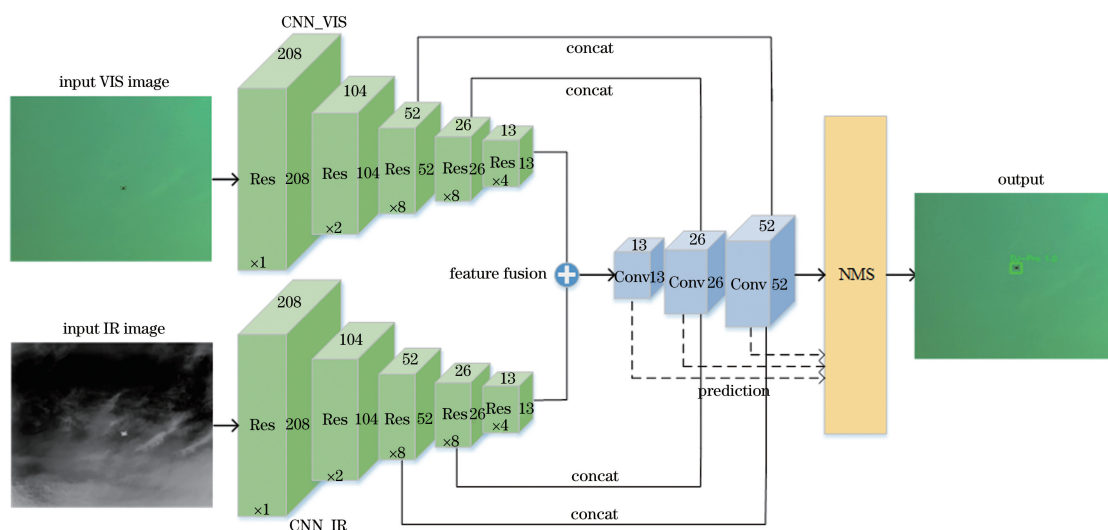


图 10 Dual-YOLOv3-D 网络结构图

Fig. 10 Network structure diagram of Dual-YOLOv3-D

尺度预测网络中对应尺寸的特征图相融合构成特征金字塔以提升无人机目标的位置回归精度。

Dual-YOLOv3 用多尺度预测网络对无人机的位置和类别进行回归预测。在每个网络结构图的右侧部分,共分成 3 种不同分辨率的预测模块,分别为 13×13 , 26×26 和 52×52 ,每个模块对无人机目标独立预测。为提升网络对特征的代表能力,对分辨率为 26×26 与 52×52 的特征图执行 2 倍大小的最近邻上采样,并将上采样的特征图与特征提取网络中对应的特征图进行连接融合。这样能够充分利用图像中不同层次的语义信息,也能够共享从特征提取网络中得到的特征,进而提升模型对不同大小目标的检测精度。

在预测阶段,将配准的红外与可见光图像同时

输入到训练好的模型中,经过预测会得到一个三维的张量,包括无人机目标的类别、边界框和置信度,并将结果显示在可见光(红外)图像上(两幅图像已经配准)。其中,Dual-YOLOv3 将特征张量划分成 $C \times C$ 的网格单元(不同大小的特征图对应不同的 C ,取值分别为 13、26 和 52),每个网格单元根据预设锚点预测 3 个不同的边界框,每个边框对应 4 个边界框的偏移量,1 个目标置信度以及 3 个目标类型,最终张量表示为 $C \times C \times [3 \times (4 + 1 + 3)]$ 。

假设进行特征融合的可见光与红外特征张量大小都为 $w \times w \times d$,且有 3 种融合方式,如图 11 所示。图 11(a)表示将两个特征张量取平均值,得到融合后张量的大小为 $w \times w \times d$;图 11(b)表示可见光特征张量在前,红外特征张量在后,两者级联融合

后张量的大小为 $w \times w \times 2d$; 而图 11(c) 的融合顺序与图 11(b) 相反。同时, 不同的方式会导致输出

的特征张量大小不同, 对应网络中卷积核的数目需要进行相应的调整匹配。

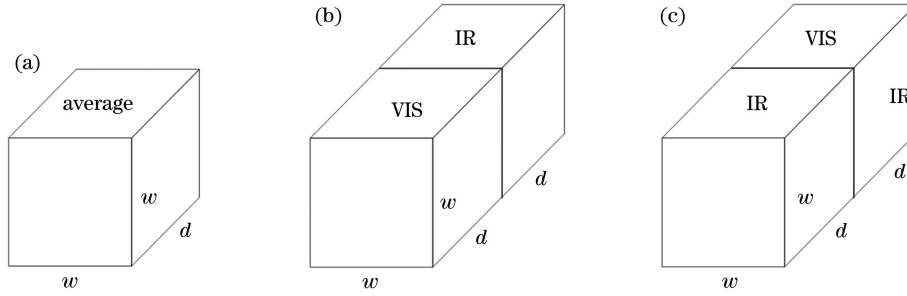


图 11 不同的特征融合方式。(a)平均融合-1;(b)级联融合-2;(c)级联融合-3

Fig. 11 Different feature fusion modes. (a) Average fusion-1; (b) concatenated fusion-2; (c) concatenated fusion-3

4 实验及结果分析

4.1 Dual-YOLOv3 网络的训练

本文以 UAV_Dual 数据集为标准进行了两次实验。第一次实验对 Faster R-CNN、SSD、原始 YOLOv3 与 Dual-YOLOv3 进行对比实验(特征融合方式使用平均融合-1); 第二次实验为最优网络结构下不同特征融合方式的对比实验。实验环境配置如下: 深度学习框架选择 Tensorflow+Keras, 服务器平台配置为 Ubuntu 16.04 操作系统, Intel(R) Xeon(R) CPU E5-2673 v3 @ 2.40 GHz, 内存 32 GB, 并搭载 GeForce GTX 1080Ti/PCIe/SSE2 显卡。

对于 UAV_Dual 训练集, 利用 K-means 聚类算法^[19]对锚点进行更新。考虑到网络参数量对检测速度的影响, Dual-YOLOv3 选用 3 种尺度 9 个锚点进行网络的训练, 分别为 (17×16) , (20×19) , (23×21) , (27×24) , (31×28) , (37×32) , (48×38) , (63×45) , (79×65) , 其平均重叠率为 90.37%, 较原始的 YOLOv3 提高了 15.72%。图 12 展示了 13×13 尺度的示意图, 其中右图为局部放大图, 实线框为一个方格单元, 虚线框为对应的锚点框。

此外, 在整个实验中, 训练过程及参数设置保持一致, 共分两个阶段进行训练。第一阶段, 迭代周期(max epochs)为 80 次, 批量大小(batch size)设置为 8, 初始学习率(learning rate)为 10^{-4} ; 第二阶段, 迭代接着第一阶段进行, 批量大小设置为 4, 起始学习率为 10^{-4} , 每隔 3 个周期进行 0.1 的比例衰减。并设置 loss 检测回调函数, 当 10 个周期内 loss 值没有降低则停止训练。同时, 权重衰减(decay)和动量(momentum)分别设置为 0.9 和 0.0005。

4.2 测试结果及评价

1) 实验一: 原始 YOLOv3 网络与 Dual-

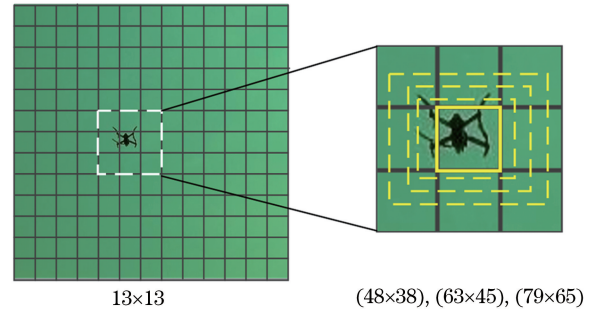


图 12 13×13 尺度的锚点示意图

Fig. 12 Anchor boxes schematic of 13×13 scale

YOLOv3 网络的对比实验

测试阶段利用 UAV_Dual 的测试集对不同网络通过相同的数据训练得到的模型进行评价。重点比较模型对不同类型低空无人机的检测识别能力, 并对结果进行分析。

在对测试集进行检测识别时, 将测试图像输入到训练好的模型, 进行目标识别与边界框的回归。当预测的目标边界框与标注的边界框的交并集比 (IOU) 大于等于 0.5 且无人机的类别判断准确时, 则认为结果正确, 否则为错误。同时, 采用平均准确率 (AP)、平均 AP (mAP) 和模型检测速度作为评价指标, 该指标能够表征模型的准确率、召回率以及实时性。最终, 不同方法在测试集上的评价结果如表 1 所示。

Dual-YOLOv3 的三种方法在 mAP 上都优于其他三种方法, 且 YOLO 系列整体优于 Faster R-CNN 与 SSD。Dual-YOLOv3-S、Dual-YOLOv3-M 和 Dual-YOLOv3-D 的 mAP 分别达到了 95.41%、97.09% 和 98.85%, 最好的 Dual-YOLOv3-D 相较于原始的 YOLOv3 提升了 6.08%。可以看出, 随着特征融合阶段的加深, 准确率得到了大幅的提升。因此, 红外与可见光图像中特征的融合能够

表 1 不同方法在测试集上的评价结果

Table 1 Evaluation results of different methods on test dataset

Method	Image source	Detection speed /s ⁻¹	mAP /%	AP		
				DJ-3	DJ-Pro	DJ-S900
Faster R-CNN	Visible	0.429	86.79	85.23	87.12	88.01
Faster R-CNN	Infrared	0.422	86.77	85.12	86.99	88.21
SSD300	Visible	22.42	81.51	80.21	81.32	82.99
SSD300	Infrared	22.54	81.43	80.08	81.34	82.87
YOLOv3	Visible	30.12	92.77	92.58	91.50	94.23
YOLOv3	Infrared	30.23	92.78	92.41	91.77	94.17
Dual-YOLOv3-S-1	Visible+infrared	27.94	95.41	96.89	93.21	96.14
Dual-YOLOv3-M-1	Visible+infrared	27.59	97.09	98.12	95.94	97.21
Dual-YOLOv3-D-1	Visible+infrared	27.17	98.85	99.83	97.85	98.88

提升模型的检测性能。

单一数据源对应的 Faster R-CNN、SSD 与 YOLOv3 在红外图像源与可见光图像源上的检测速度分别保持在 0.4, 22, 30 s⁻¹ 左右, 而双波段数据源对应的 Dual-YOLOv3 检测速度约为 27 ~ 28 s⁻¹。同样, Dual-YOLOv3-D 网络的参数数量的增加较其他网络多, 在检测速度上略有下降, 但依然保持了较快的水平。

对不同方法进行分析可知, Faster R-CNN 为两阶段模型, 模型的整体复杂度较高, 速度较慢, 在检测精度上有一定优势, 但对无人机这类小目标的检测精度表现略差。SSD、YOLOv3 与 Dual-YOLOv3 都为单阶段模型, 模型复杂度低于两阶段的, 在速度上优势较为明显, 但 SSD 检测精度略低。这是由于 YOLO 结合了多尺度预测网络, 提升了对小目标的检测精度。而 Dual-YOLOv3 将两种图像信息进行深度融合, 提升了对目标的表征能力, 但在模型复杂度上略高于单一通道的 YOLOv3 模型, 使得检测速度有一定的下降。

在测试集中随机抽取 YOLOv3 与 Dual-YOLOv3 方法的检测结果, 如图 13 所示。每张检

测结果展示了三类信息: 预测的无人机位置(以边界框形式给出)、预测类别与置信度(在边界框上方显示)。可以看出, 每种方法都能将低空无人机目标的类别和位置检测出来, 满足基本的检测识别。但仔细比较发现, 在 YOLOv3 的检测结果中, 类别的置信度低于 Dual-YOLOv3-D 方法的结果。这表示当需要抑制虚警(误检)而将置信度阈值升高时, 单一图像源对应方法的检测结果将会减少, 即出现漏检, 对无人机的预警探测十分不利。综上所述, Dual-YOLOv3-D 在检测识别性能上优于单一数据源对应的 YOLOv3, 且稳健性较好。虽然检测速度稍有降低, 但依然具备较快的检测能力。

2) 实验二: Dual-YOLOv3-D 对不同特征融合方式的对比实验

按照另外两种特征融合方式修改网络, 利用相同方式进行训练和测试, 得到在测试集上的评价结果, 如表 2 所示。

由表 2 可以看出, 在第一种特征融合方式(即平均融合)下模型的 mAP 最高, 速度最快。而采用其他两种级联融合方式得到的结果差异不大, 且低于平均融合方式。因此, 采用平均特征融合的方式

表 2 Dual-YOLOv3-D 对不同融合方式在测试集上的评价结果

Table 2 Evaluation results of different fusion modes for Dual-YOLOv3-D on test dataset

Method	Detection speed /s ⁻¹	mAP /%	AP		
			DJ-3	DJ-Pro	DJ-S900
Dual-YOLOv3-D-1	27.17	98.85	99.83	97.85	98.88
Dual-YOLOv3-D-2	24.31	95.19	96.41	94.02	95.13
Dual-YOLOv3-D-3	24.47	95.21	96.34	94.24	95.04

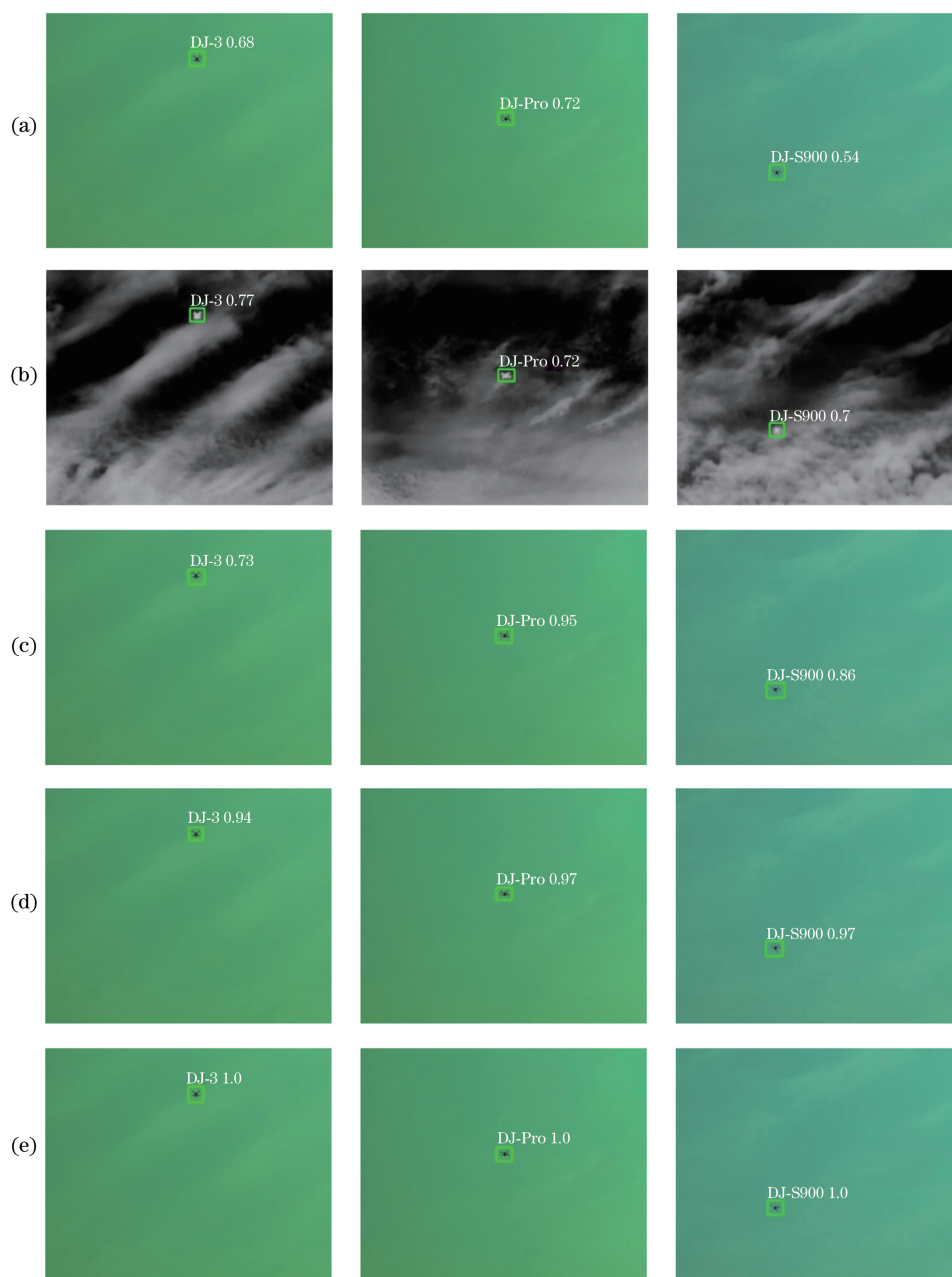


图 13 YOLOv3 与 Dual-YOLOv3 在测试集中部分检测结果对比图(第一列为大疆-精灵 3,第二列为大疆-御 Pro,第三列为大疆-S900)。(a) YOLOv3(可见光数据集);(b) YOLOv3(红外数据集);(c) Dual-YOLOv3-S;(d) Dual-YOLOv3-M;(e) Dual-YOLOv3-D

Fig. 13 Comparison of partial detection results of YOLOv3 and Dual-YOLOv3 on test dataset (first category is Dji-PHANTOM 3, second category is Dji-MAVIC Pro, and third category is Dji-S900). (a) YOLOv3 (visible dataset); (b) YOLOv3 (infrared dataset); (c) Dual-YOLOv3-S; (d) Dual-YOLOv3-M; (e) Dual-YOLOv3-D

能够使两种信息进行混合,而不是简单的堆叠,使得 mAP 得到提升。而级联融合后特征的数量增加,导致网络参数量增加,检测速度有所下降。

5 结 论

分析了现阶段光电探测的工作方式及检测机制,发现可见光与红外两通道的信息不能充分利用。

针对此种问题,为进一步提升检测识别能力,以 YOLOv3 网络的架构为基础,充分利用残差结构与多尺度预测网络,提出了一种基于双通道的快速低空无人机检测识别方法(Dual-YOLOv3)。并在双波段的无人机数据集上进行了对比实验,结果表明,对两种图像采用深度特征提取与平均特征融合的方式,能够使信息得到充分利用,Dual-YOLOv3-D-1

网络训练得到的模型 mAP 比单一数据源的 YOLOv3 提升了 6.08%，检测速度保持在 27 s^{-1} ，整体能力得到提升。

虽然本文方法比单一图像源的方法在检测识别性能上有较大提升，但仍有一些不足之处。受分光轴采集系统的限制，复杂背景下的图像配准误差不断变化，导致配准困难，不能对此类数据进行验证。同时，实验中无人机的类别较少，更多类别的无人机数据可能会是新的挑战。此外，无人机数据均在光照良好的条件下采集得到，而夜间或光线昏暗条件下的性能有待验证和进一步地改进，这些将是未来深入研究的内容。

参 考 文 献

- [1] Zhao Y Y, Chang J D. Detection and handling of drones in low altitude airspace [J]. *Electronic Warfare*, 2016(4): 44-49.
赵云雨, 常晋聃. 对低空无人机的探测与管控新技术综述[J]. *电子对抗*, 2016(4): 44-49.
- [2] Li C L, Wang X, Zhang L, *et al.* Weighted low-rank decomposition for robust grayscale-thermal foreground detection [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2016, 27(4): 725-738.
- [3] Li C L, Wang G Z, Ma Y P, *et al.* A unified RGB-T saliency detection benchmark: dataset, baselines, analysis and a novel approach[J/OL]. (2017-01-11) [2019-07-02]. <https://arxiv.org/abs/1701.02829>.
- [4] Ha Q S, Watanabe K, Karasawa T, *et al.* MFNet: towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes[C]//2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), September 24-28, 2017, Vancouver, BC, Canada. New York: IEEE, 2017: 5108-5115.
- [5] Zhai S L, Shao P P, Liang X Y, *et al.* Fast RGB-T tracking via cross-modal correlation filters [J]. *Neurocomputing*, 2019, 334: 172-181.
- [6] Feng X Y, Mei W, Hu D S. Aerial target detection based on improved faster R-CNN[J]. *Acta Optica Sinica*, 2018, 38(6): 0615004.
冯小雨, 梅卫, 胡大帅. 基于改进 Faster R-CNN 的空中目标检测[J]. *光学学报*, 2018, 38(6): 0615004.
- [7] Dai J F, Li Y, He K M, *et al.* R-FCN: object detection via region-based fully convolutional networks [C]//Advances in Neural Information Processing Systems, December 5-10, 2016, Barcelona, Spain. Canada: NIPS, 2016: 379-387.
- [8] Xie J R, Li F M, Wei H, *et al.* Enhancement of single shot multibox detector for aerial infrared target detection[J]. *Acta Optica Sinica*, 2019, 39(6): 0615001.
谢江荣, 李范鸣, 卫红, 等. 用于空中红外目标检测的增强单发多框检测器方法[J]. *光学学报*, 2019, 39(6): 0615001.
- [9] Redmon J, Divvala S, Girshick R, *et al.* You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 779-788.
- [10] Redmon J, Farhadi A. Yolov3: an incremental improvement[J/OL]. (2018-04-08) [2019-07-02]. <https://arxiv.org/abs/1804.02767>.
- [11] Rozantsev A, Lepetit V, Fua P. Flying objects detection from a single moving camera [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 4128-4136.
- [12] Du L, Gao C Q, Feng Q, *et al.* Small UAV detection in videos from a single moving camera [M]//Yang J F, Hu Q H, Cheng M M, *et al.* Computer vision. CCCV 2017. Communications in computer and information science. Singapore: Springer, 2017, 773: 187-197.
- [13] Wang J Y, Wang X Y, Zhang K, *et al.* Small UAV target detection model based on deep neural network [J]. *Journal of Northwestern Polytechnical University*, 2018, 36(2): 258-263.
王靖宇, 王霰禹, 张科, 等. 基于深度神经网络的低空弱小无人机目标检测研究[J]. *西北工业大学学报*, 2018, 36(2): 258-263.
- [14] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition [C]//The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, USA. New York: IEEE, 2015: 770-778.
- [15] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]//Proceedings of the 32nd International Conference on International Conference on Machine Learning, July 6-11, 2015, Lille, France. USA: MIT Press, 2015, 37: 448-456.
- [16] Ren Z J, Lin S Z, Li D W, *et al.* Mask R-CNN object detection method based on improved feature pyramid [J]. *Laser & Optoelectronics Progress*, 2019, 56(4): 041502.
任之俊, 蔺素珍, 李大威, 等. 基于改进特征金字塔的 Mask R-CNN 目标检测方法[J]. *激光与光电子学*

- 进展, 2019, 56(4): 041502.
- [17] Redmon J, Farhadi A. YOLO9000: better, faster, stronger [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 6517-6525.
- [18] Wang J D, Wei Z, Zhang T, *et al.* Deeply-fused nets [J/OL]. (2016-05-25)[2019-07-02]. <https://arxiv.org/abs/1605.07716>.
- [19] Wei Y M, Quan J C, Houyu 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.