

# 激光与光电子学进展

## 改进 YOLOv4 网络的无人机光通信实时捕获定位技术

陈廷祚, 倪小龙, 白素平\*, 于信

长春理工大学光电工程学院, 吉林 长春 130022

**摘要** 信标光斑位置检测技术广泛应用于基于视觉的光通信粗对准领域中,而检测算法的优劣直接影响捕获定位的精度。针对基于阈值分割搜寻信标光斑的算法易受背景强光影响的缺陷,建立了基于深度学习算法的无人机光通信实时捕获定位系统。首先,改进了 YOLOv4(You only look once, v4)网络,采用能增强浅层特征信息提取的特征图通道拼接方式设计了四个简化模块和一个上采样模块,极大提升了网络的速度。然后,用改进后的网络、原始 YOLOv4 网络及其简化网络在 PASCAL VOC 数据集上进行训练。最后,采集和训练信标光斑数据集,在无人机上运行改进 YOLOv4 网络并输出摄像头视频帧的信标光斑位置。基于比例积分微分算法调节云台进行位置闭环控制,从而实现光通信的实时捕获和定位对准。实验结果表明,改进 YOLOv4 网络在信标光斑测试集上的精确率为 99.6%,召回率为 99.8%,在 NVIDIA Jetson Xavier NX 嵌入式计算机平台上的帧率为 42 frame/s,满足无人机光通信实时捕获定位的要求。

**关键词** 光通信; 无人机; YOLOv4 网络; 比例积分微分算法; 捕获定位

中图分类号 O439

文献标志码 A

DOI: 10.3788/LOP202259.1106006

## Real-Time Acquisition and Positioning Technology of Unmanned Aerial Vehicle Optical Communication Based on Improved YOLOv4 Network

Chen Tingzuo, Ni Xiaolong, Bai Suping\*, Yu Xin

*School of Electro-Optical Engineering, Changchun University of Science and Technology, Changchun 130022, Jilin, China*

**Abstract** The beacon spot position detection technology is widely used in the field of vision-based optical communication coarse alignment, and its detection algorithm directly affects the accuracy of acquisition and positioning. Aiming at the defect that the algorithm of searching beacon spot based on threshold segmentation is easy to be affected by background strong light, a real-time acquisition and positioning system of unmanned aerial vehicle optical communication based on deep learning algorithm is established in this paper. First, the YOLOv4 (You only look once, v4) network is improved, four simplified modules and one up-sampling module are designed using the feature map channel splicing method that can enhance the extraction of shallow feature information, which greatly improves the speed of the network. Then, the improved network, the original YOLOv4 network and its simplified network are trained on PASCAL VOC data set. Finally, collect and train the beacon spot data set, and run the improved YOLOv4 network on the unmanned aerial vehicle to output the beacon spot position of the camera video

收稿日期: 2021-08-12; 修回日期: 2021-08-23; 录用日期: 2021-08-31

基金项目: 中国博士后科学基金(2017M621179)、吉林省科技发展计划(20170521001HJ, 20200401054GX)、长春理工大学青年基金(XQNJJ-2019-01, XJJLG-2018-20)

通信作者: \*2541533443@qq.com

frame. Based on proportion integration differentiation algorithm, the gimbal is adjusted for position closed-loop control, so as to realize real-time acquisition and positioning for optical communication. Experimental results show that the accuracy rate of the improved YOLOv4 network on the beacon spot test set is 99.6%, the recall rate is 99.8%, and the frame rate on the NVIDIA Jetson Xavier NX embedded computer platform is 42 frame/s, which meets the requirements of real-time acquisition and positioning for unmanned aerial vehicle optical communication.

**Key words** optical communication; unmanned aerial vehicle; YOLOv4 network; proportion integration differentiation algorithm; acquisition and positioning

## 1 引言

随着科学技术的迅速发展,无人机(UAV)在当代航空领域扮演着越来越重要的角色。目前,轻小型无人机平台在社会民生中的实时监测和评估方面具备明显优势,在街景拍摄、农业植保、监控巡查以及灾害救助、环保检测、电力巡查等领域也具有广阔的应用前景<sup>[1]</sup>。激光通信能以无损压缩的方式实时回传海量原始数据,传输效率高,成为相关领域的研究热点<sup>[2]</sup>。无人机光通信实时捕获定位技术是实现激光通信的重要环节,也是实现稳定传输的重要保障<sup>[3]</sup>。基于光电传感器探测和全球定位系统(GPS)的传统光通信捕获定位系统结构复杂、部署时间较长,不利于在特殊或紧急情况下使用。而基于视觉的捕获定位系统结构简单、搭建速度较快,是传统光通信捕获定位系统的一个重要补充。现有信标光斑的捕获定位技术主要有基于阈值分割的Otsu算法(大津法)<sup>[4]</sup>和质心算法<sup>[5-6]</sup>,但这类算法易受背景光强的干扰,捕获定位精度不高。而基于神经网络的识别算法能很好地解决该问题,从而提高捕获定位的精度。

为了建立基于视觉的无人机光通信实时捕获定位系统,首先需要在复杂的环境背景下对信标光斑进行检测,确定信标光斑的位置信息。目标检测是目前计算机视觉领域的一个重要研究方向<sup>[7]</sup>。随着深度学习技术的不断突破,计算机的性能获得了大幅提升,图像目标识别与分类的准确度和速度也得到了极大提升。目标检测领域的深度学习算法主要有 one-stage 与 two-stage 两类,前者的速度更快,后者的精度更高<sup>[8]</sup>。one-stage 算法有单步多框检测(SSD)<sup>[9]</sup>、YOLO(You only look once)系列算法<sup>[10-13]</sup>。two-stage 算法有快速区域卷积神经网络(R-CNN)、Faster R-CNN 等系列算法<sup>[14-15]</sup>。其中,YOLO 系列算法能同时在精度和速度上达到较好的效果,最新的 YOLOv4 网络能在确保定位精度的情况下满足实时性要求。

本文将改进的 YOLOv4 网络用于信标光斑的识别与定位,直接对摄像头读取的内容进行实时检测,在识别出信标光斑后,用比例积分微分(PID)算法控制云台转动,使摄像头对准信标光斑中心。实验结果表明,基于改进 YOLOv4 网络的无人机光通信实时捕获定位系统具有结构简单、捕获率高、定位稳定性好等特点。

## 2 系统的构成及工作原理

### 2.1 系统的硬件构成

图 1 为基于改进 YOLOv4 网络的无人机光通信实时捕获定位系统的原理图。该系统主要由图像采集系统、图像检测识别系统和 PID 算法控制系统三部分组成。图像采集系统由 USB 3.0 摄像头和 USB 3.0 数据传输线构成;图像检测识别系统由 NVIDIA Jetson Xavier NX 嵌入式计算机平台以及在该平台上安装的一系列深度学习框架和模块构成;PID 控制系统由通用异步收发传输器(UART)通信模块以及云台构成,通信模块使用了 USB 转晶体管逻辑(TTL)串口线。进行捕获定位时,由图像采集系统通过无人机采集视频图像,由图像检测识别系统确定信标光斑的有无及位置信息,由 PID 控制系统控制云台转动使摄像头始终对准信标光斑的中心。

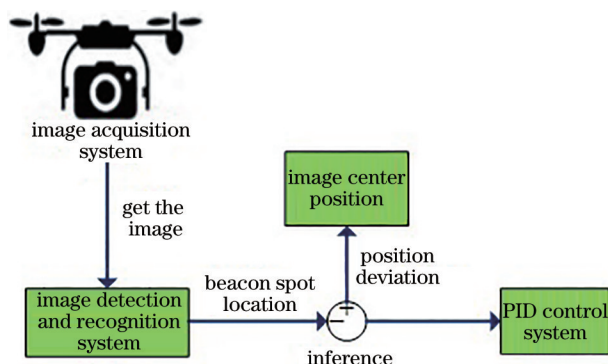


图 1 系统的原理图

Fig. 1 Schematic diagram of the system

## 2.2 系统的工作原理

图 2 为系统的结构示意图。首先,通过无人机云台吊舱内型号为 MV-SUA133GC-M 的 USB 3.0 工业相机采集视频图像。然后,将采集的视频图像通过 USB 3.0 数据线传送给具有 GPU 且安装了一系列深度学习框架的 NVIDIA Jetson Xavier NX 嵌入式计算机平台。嵌入式计算机平台运行改进的 YOLOv4 网络对视频图像进行检测,如果图像中有

信标光斑,则用矩形框标出图像中的信标光斑并输出矩形框的中心位置坐标。最后,计算图像正中心坐标与矩形框中心位置坐标在 X、Y 轴的偏差,并将其通过 UART 通信模块传送给云台。云台上配套三环(位置环、速度环、电流环)PID 控制系统,通过 PID 算法控制横滚轴和俯仰轴转动的角度,使摄像头始终对准信标光斑的中心,实现位置闭环控制。

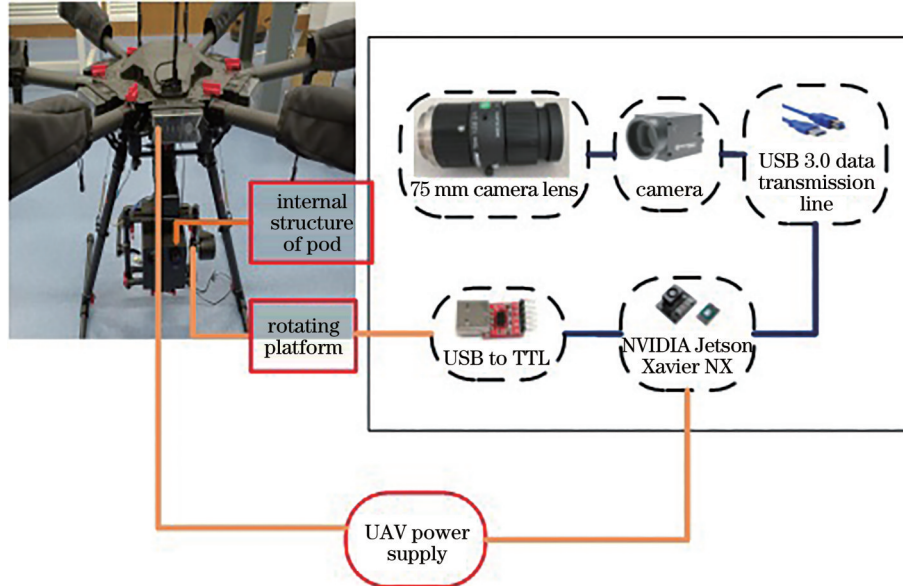


图 2 系统的结构

Fig. 2 Structure of the system

## 3 系统涉及的关键算法

### 3.1 目标检测算法 YOLOv4

YOLO 算法是一种基于深度学习的神经网络目标识别算法<sup>[16]</sup>。相比现有的 R-CNN 系列算法, YOLO 算法不需要先生成候选区域,而是通过网络直接预测输入图像中目标的位置信息和类别概率,将目标识别问题转化为回归问题,从而实现更快的检测速度<sup>[17]</sup>。YOLOv4 是 YOLO 系列的第四代,检测精度较高,但由于网络层过深,运行速度相对较低。

YOLOv4 网络结构将输入图像划分成  $S \times S$  个网格单元,然后以这些网格单元为中心,给出  $N$  个候选框,如图 3 所示。每个预测框都包含了边界框的位置大小信息(候选框的左上角相对图像 X、Y 轴的坐标  $t_x$ 、 $t_y$  和候选框相对图像的长和宽  $t_w$ 、 $t_h$ )、目标概率 ( $P_0$ ) 和类别概率 ( $P_1 \sim P_n$ ),  $t_x$ 、 $t_y$ 、 $t_w$ 、 $t_h$  可分别表示为

$$t_x = \frac{X_{\text{absolute}_x}}{X_{\text{image\_width}}}, \quad (1)$$

$$t_y = \frac{X_{\text{absolute}_y}}{X_{\text{image\_height}}}, \quad (2)$$

$$t_w = \frac{X_{\text{absolute\_width}}}{X_{\text{image\_width}}}, \quad (3)$$

$$t_h = \frac{X_{\text{absolute\_height}}}{X_{\text{image\_height}}}, \quad (4)$$

式中,  $X_{\text{absolute}_x}$  和  $X_{\text{absolute}_y}$  分别为单元格左上方像素点的坐标,  $X_{\text{image\_width}}$  和  $X_{\text{image\_height}}$  分别为图像的宽度和高度,  $X_{\text{absolute\_width}}$  和  $X_{\text{absolute\_height}}$  分别为预测框的宽度和高度。  $t_x$ 、 $t_y$ 、 $t_w$ 、 $t_h$  的取值范围均为 (0.0, 1.0] 内的浮点值,即对象相对于输入图像分辨率的相对坐标值。

目标概率为预测框是否包含目标以及候选框准确标出目标的概率,可表示为

$$P = P_{\text{object}} \times \eta_{\text{IOU}}, \quad (5)$$

式中,  $P_{\text{object}}$  表示该候选框内是否包含目标对象,如果存在对象,  $P_{\text{object}} = 1$ ; 反之,  $P_{\text{object}} = 0$ 。  $\eta_{\text{IOU}}$  为预测边界框与目标标注真实值的重叠交叉比。每个框特定类别的置信度得分可表示为

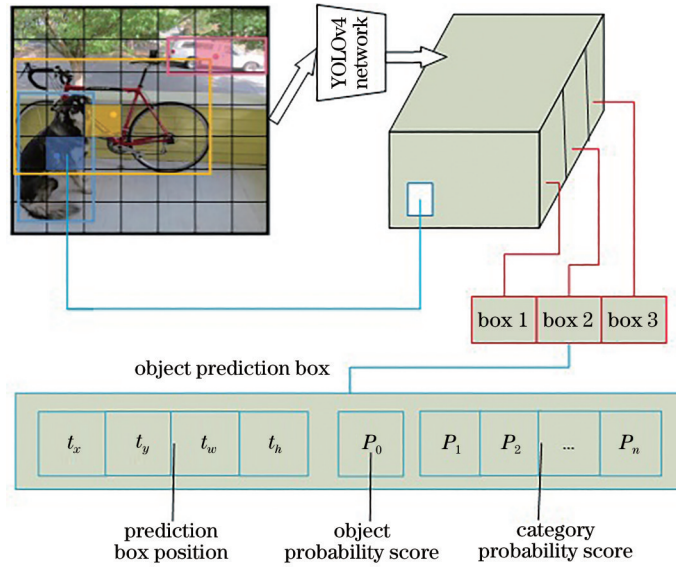


图3 YOLOv4网络预测的特征图属性

Fig. 3 Feature map attributes predicted by the YOLOv4 network

$$P_{\text{class\_scores}} = P_{\text{class}} \times P, \quad (6)$$

式中,  $P_{\text{class}}$  为预测条件的类别概率, 可表示单元格内存在物体且属于某一类的概率。在测试时,  $P_{\text{class\_scores}}$  表示该类别出现在预测框中的概率以及预测框标出对象的准确程度。最后, 对每个单元格输出的概率值进行非极大值抑制处理, 以找出最佳的预测框。

### 3.2 YOLOv4网络的改进与简化

改进 YOLOv4 网络的简化结构如图 4 所示, 该结构在 YOLOv4 网络的基础上用能增强浅层特征信息提取的特征图通道拼接网络模块代替残差网络模块。该网络结构共有 25 个卷积层 (Conv)、3 个最大池化 (Maxpool) 层、7 个通道连接 (Concatenate) 层和 1 个上采样 (Upsample) 层。相比原始 YOLOv4 网络 (100 多层), 该网络在满足精度需求的同时极大提升了网络的速度。其中, predict one 和 predict two 分别表示预测输出层 1 和预测输出层 2, 其功能是确定预测框、输出目标类别概率和位置信息。

相比原始的 YOLOv4 网络结构, 改进后的网络结构主干借鉴了跨阶段部分 (CSP) 连接的处理思想, 融合了一次性聚合 (OSA) 模块<sup>[18]</sup> 的简化思路, 设计了四个网络连接模块和一个上采样模块, 分别对应图 4 中的 block 1~block 5。用这五个简化模块代替原始 YOLOv4 网络中冗余复杂的网络连接模块, 可大幅提高网络的运行速度。每个卷积层都保留了小批量标准化 (CmBN) 处理, 除最后一层网络使用 Linear 激活函数外, 其他网络层都使用了带泄漏修正线性单元 (Leaky ReLU) 激活函数。

在 YOLOv4 网络中, 主干跨阶段连接部分采用 CSPDarknet53 结构<sup>[19]</sup>, 通过残差网络的连接方式堆叠大量的网络层, 避免加深网络时出现的梯度消失问题, 从而提升网络的识别精度。嵌入式平台 NVIDIA Jetson Xavier NX 的计算性能有限, 采用 CSPDarknet53 结构时网络的运行时间较长, 无法保证系统的实时性。因此, 提出了三种能加快网络运行速度的连接方式, 如图 5 所示。假设输入特征图的通道数为  $N$ , 则连接输出后的特征图通道数为  $2N$ , 实现了与 CSPDarknet53 相同的跨阶段连接效果, 即主干特征图逐渐变小, 特征图通道数翻倍增加。这三种结构连接方式堆叠的网络层较少, 运行速度较快。

图 6 为拼接 (Concatenate) 特征融合方式和相加 (Add) 特征融合方式的原理, 其中, 拼接方式是对各个通道进行拼接, 而相加方式是对相应通道内的像素做加法运算。密集卷积网络 (DenseNet)<sup>[20]</sup> 的训练结果表明, 拼接融合比相加融合方式提取的特征效果更好。能量和 GPU 计算效率高的检测骨干网络 (VoVNet)<sup>[21]</sup> 进一步指出了 DenseNet 的不足, 并提出了 OSA 模块, 提高了网络的计算效率。借鉴 OSA 模块的思路设计了三种连接方式, 采用拼接特征融合方式, 丢弃了 YOLOv4 网络中大量的相加特征融合方式, 原因是拼接特征融合方式有利于对浅层特征信息的提取。当特征图的尺寸发生变化时, 在部分网络结构中加入  $2 \times 2$  的最大池化层进行下采样, 如图 4 中的三个 Maxpool 层。将这种块与块

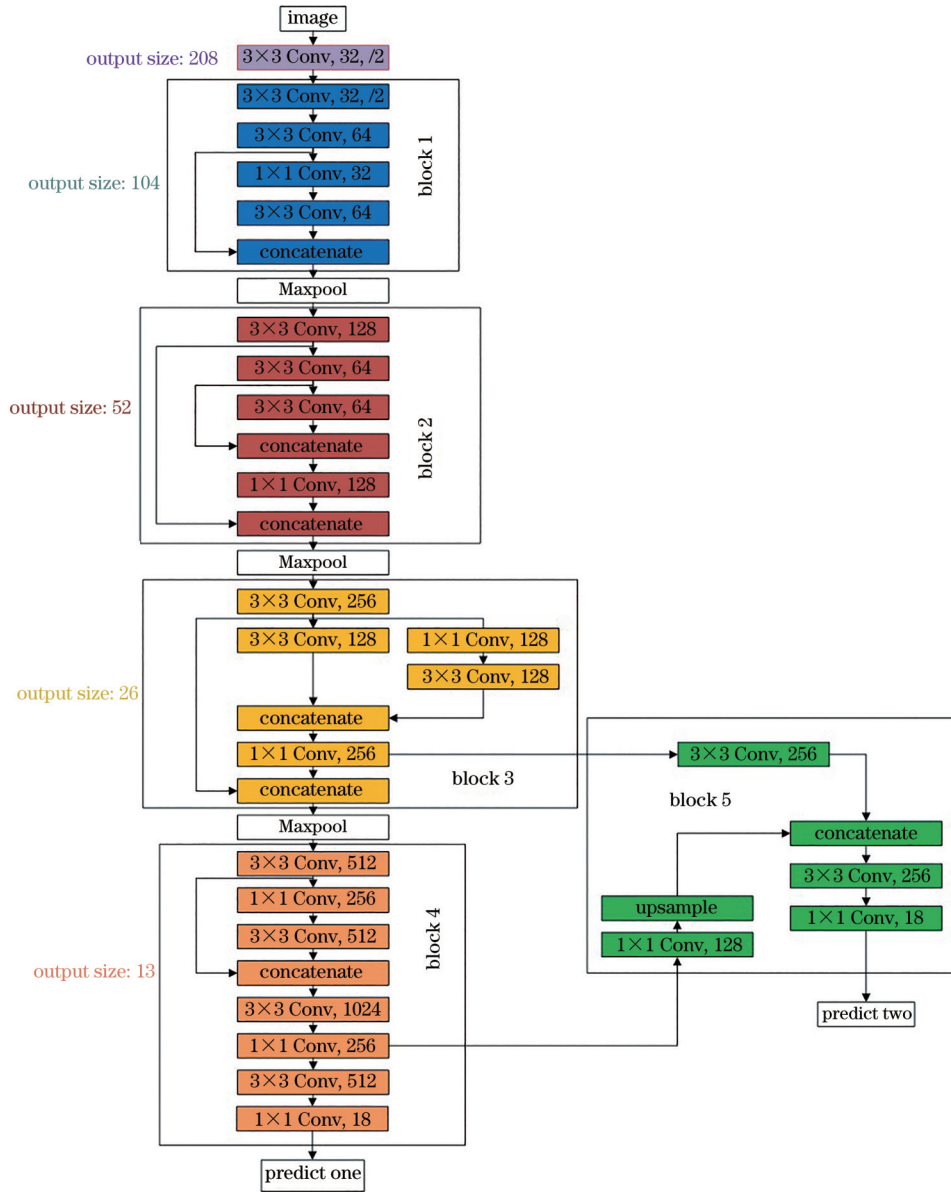


图 4 改进 YOLOv4 网络的简化结构

Fig. 4 Simplified structure of the improved YOLOv4 network

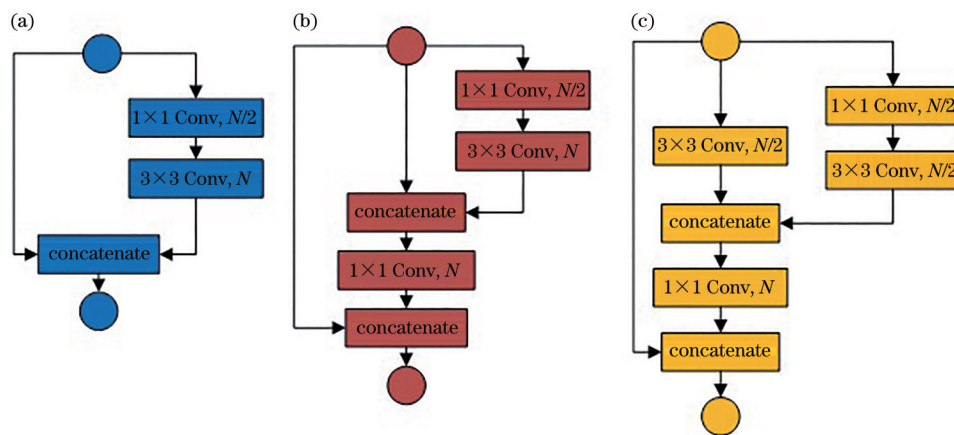


图 5 三种网络结构的连接方式。(a)连接方式 1; (b)连接方式 2; (c)连接方式 3

Fig. 5 Connection mode of three network structures. (a) Connection mode 1; (b) connection mode 2; (c) connection mode 3

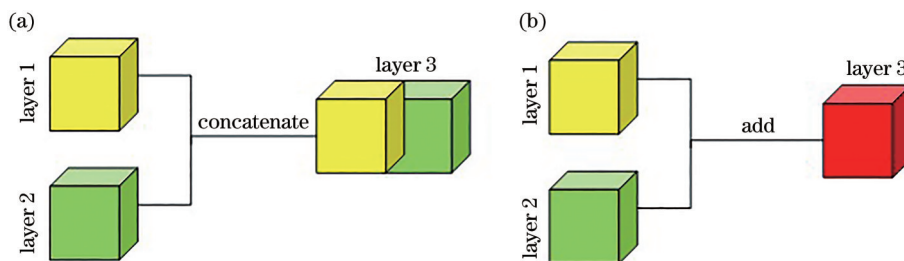


图 6 两种特征融合方式的原理。(a) Concatenate; (b) add

Fig. 6 Principles of two feature fusion modes. (a) Concatenate; (b) add

之间进行衔接的层称为过渡层。

### 3.3 PID 算法的实现

摄像头安装在云台上,在信标光斑实时捕获与对准时,云台的作用就是控制摄像头移动,实现位置闭环。云台上安装有电机,通过控制电机的角度实现云台俯仰轴和横移轴的转动。PID 算法的输出即电机转动的角度,可表示为

$$u(t) = K_p e(t) + K_i \int_0^t e(t') dt' + K_d \frac{de(t)}{dt}, \quad (7)$$

式中,  $e(t)$  为图像中心与目标中心在  $X$ 、 $Y$  轴的差值,单位为 pixel;  $K_p$ 、 $K_i$ 、 $K_d$  分别为比例增益、积分增益和微分增益,  $t$  为时间,  $t'$  为时间积分变量。PID 算法中的比例部分用来控制云台上的摄像头,使其跟随目标移动;积分部分通常用来消除系统产生的稳定误差,确保目标始终在视频中央;微分部分可控制偏差的变化率,防止目标在图像中心抖动较大。图 7 为 PID 算法的流程图。首先,用计算机视觉库 (OpenCV) 从工业相机采集的视频中读取图像帧;然后,对改进的 YOLOv4 网络进行检测,获取目标的位置信息;最后,计算预测框中心与图像中心在  $X$  轴和  $Y$  轴的坐标差值,并将该差值传递给 PID 控制系统,从而实现目标位置闭环。

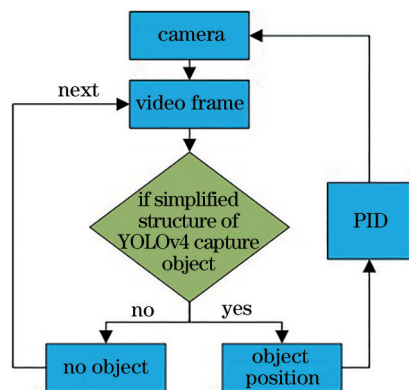


图 7 PID 算法的流程图

Fig. 7 Flow chart of the PID algorithm

## 4 实验结果与分析

### 4.1 数据集的制备

为了识别信标光斑,首先需要构建图像数据集,具体步骤如下。

1) 采集图像:用型号为 MV-SUA133GC-M 的 USB 3.0 工业相机对信标光斑进行拍摄,分别在早上、中午和下午不同时间段对不同位置采集信标光斑图像;拍摄过程中,捕捉不同高度和角度的信标光斑视角,确保数据集中图像的多样性和代表性。此外,还拍摄了不包含目标的负样本,图 8 为采集的部分图像。

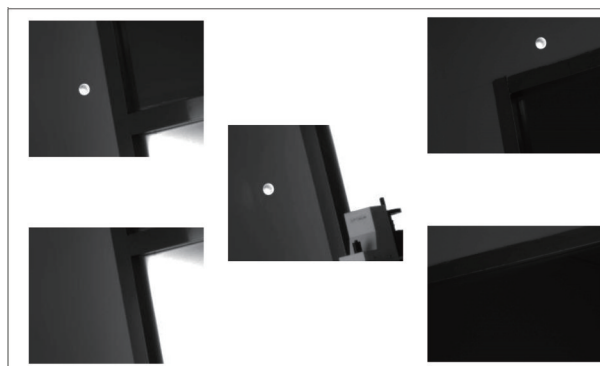


图 8 数据集中的部分图像

Fig. 8 Some images in the data set

2) 标注图像:图像采集完成后,利用开源 LabelImg 软件对图像进行人工标注,标出信标光斑的标签和位置信息;标记完成后,共获得 2072 张包括正负样本的数据集,并按照 3:1 的比例将图像样本划分成训练集和测试集。

3) 数据集扩充:为了训练出更好的泛化模型,防止模型因图像数量过少出现过拟合现象,需要对数据集进行扩充,如图像增强和扩充处理。图像增强主要包括亮度和对比度的图像处理,扩充处理主要是对训练集进行  $\pm 10^\circ$  的旋转操作。

## 4.2 训练结果

采用迁移学习的方式对信标光斑图像数据集进行训练。训练环境:计算机服务器平台为 Ubuntu18.04 操作系统,配备两张 GeForce RTX 3090 显卡,显存容量为 24 G,深度学习训练框架为 Darknet。

基于改进 YOLOv4 网络的预训练参数进行训练,设置的最大迭代次数为 5000 次,动量为 0.9,权重衰减为 0.0005,处理批量为 64,初始学习率为 0.0013,迭代次数达到 4000 次时将学习率降为 0.00013,迭代次数达到 4500 次时将学习率降为 0.000013,训练结果如图 9 所示。对训练出的模型性能进行评估,统计得到模型的识别结果如表 1 所

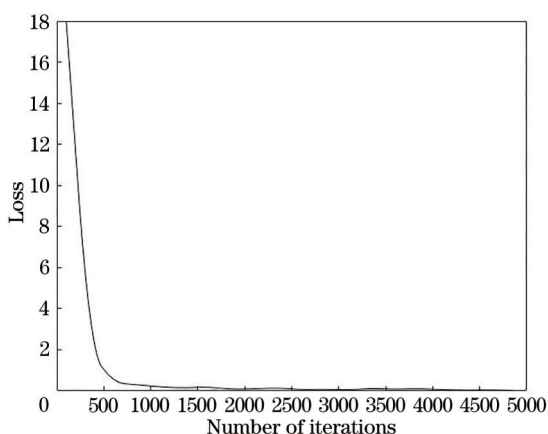


图 9 信标光斑数据集的训练结果

Fig. 9 Training results of the beacon spot data set

表 1 测试集的识别结果

Table 1 Recognition results of the test set

Count	Category	$X_{TP}$	$X_{FP}$	$X_{FN}$	$X_{TN}$
518	beacon spot	503	2	1	13

表 2 PASCAL VOC 2007 和 PASCAL VOC 2012 数据集的统计信息

Table 2 Statistics of PASCAL VOC 2007 and PASCAL VOC 2012 data sets

Data set	Training set		Validation set		Training set + validation set		Test set		Total data set	
	Image	Object	Image	Object	Image	Object	Image	Object	Image	Object
PASCAL VOC 2007	2501	6301	2510	6207	5011	12608	4952	12032	9963	24640
PASCAL VOC 2012	5717	13609	5823	13841	11540	27450	11540	27450	23080	54900
Total	8218	19910	8333	20148	16551	40058	16492	39482	33043	79540

用 Darknet 作为训练框架,将训练图像的分辨率缩小为 608 pixel  $\times$  608 pixel,设置的最大训练迭代次数为 80000 次,初始学习率为 0.001,迭代次数达到 64000 次时将学习率降为 0.0001,迭代次数达到 72000 次时将学习率降为 0.00001。将 PASCAL VOC 2007 的训练集、验证集以及 PASCAL VOC

2012 的训练集、验证集合并生成一个新的训练集,用改进的 YOLOv4 网络、原始 YOLOv4 网络和 YOLOv4. tiny 网络在该训练集上进行训练,并在 PASCAL VOC 2007 测试集中测试网络的性能,结果如图 10 所示。

精确率  $P$  和召回率  $R$  可表示为

$$P = \frac{X_{TP}}{X_{TP} + X_{FP}}, \quad (8)$$

$$R = \frac{X_{TP}}{X_{TP} + X_{FN}}. \quad (9)$$

测试结果表明,改进 YOLOv4 网络的精确率为 99.6%,召回率为 99.8%。用该模型在 NVIDIA Jetson Xavier NX 嵌入式计算机平台上的运行速度能达到 42 frame/s,相比原始 YOLOv4 模型 (8 frame/s) 有较大提升。

## 4.3 改进 YOLOv4 网络的评估

为了评价改进 YOLOv4 网络的性能,在 PASCAL VOC 2007 和 PASCAL VOC 2012 数据集上对比了 YOLOv4. tiny (YOLOv4 网络的官方简化版)、原始 YOLOv4 网络以及改进 YOLOv4 网络的性能,评价指标包括平均精度均值 (mAP) 和在 NVIDIA Jetson Xavier NX 平台上的视频检测速度。PASCAL VOC 数据集是计算机视觉领域中分类识别和检测的一个基准测试,具有丰富的目标标注文件,为各类深度学习算法提供了训练数据库和标准的评估系统。PASCAL VOC 数据集包含了 4 大类目标:人;动物(狗、羊、鸟、猫、马、牛);交通工具(船、飞机、自行车、公共汽车、摩托车、小轿车、火车);家居物品(电视、瓶子、盆栽植物、椅子、餐桌、沙发)。PASCAL VOC 挑战赛中使用的图像质量高,标注信息完善,非常适合作为评定识别算法性能的标准。PASCAL VOC 2007 和 PASCAL VOC 2012 数据集的统计信息如表 2 所示。

2012 的训练集、验证集合并生成一个新的训练集,用改进的 YOLOv4 网络、原始 YOLOv4 网络和 YOLOv4. tiny 网络在该训练集上进行训练,并在 PASCAL VOC 2007 测试集中测试网络的性能,结果如图 10 所示。

表 3 为改进 YOLOv4 网络、原始 YOLOv4 网络

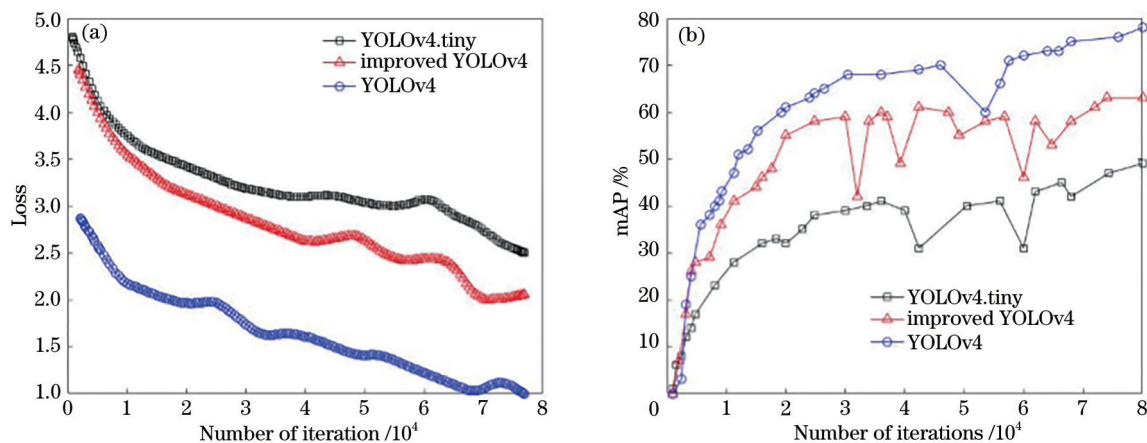


图 10 不同网络的训练结果。(a)损失函数;(b) mAP

Fig. 10 Training results of different networks. (a) Loss function; (b) mAP

表 3 不同网络在 PASCAL VOC 2007 测试集上的检测结果  
Table 3 Detection results of different networks on the PASCAL VOC 2007 test set

Network	Train	mAP / %	FPS
YOLOv4	PASCAL VOC 2007+ PASCAL VOC 2012	76	8
YOLOv4.tiny	PASCAL VOC 2007+ PASCAL VOC 2012	49	47
Improved YOLOv4	PASCAL VOC 2007+ PASCAL VOC 2012	63	42

和 YOLOv4.tiny 网络训练模型的性能,包括不同网络在 PASCAL VOC 2007 测试集的 mAP 和在 NVIDIA Jetson Xavier NX 上的每秒运行帧数 (FPS)。可以发现:相比 YOLOv4.tiny 网络,改进 YOLOv4 网络的速度虽然略有下降,但在

PASCAL VOC 2007 测试集上的 mAP 却提高了 14 个百分点;原始 YOLOv4 网络的识别精度较高,但运行速度过低,无法满足实时性要求。综合评价结果表明,改进 YOLOv4 网络更适合在无人机光通信实时捕获和定位系统上使用。

#### 4.4 捕获与定位对准实验

图 11 为无人机光通信用 NVIDIA Jetson Xavier NX 作为机载 GPU 嵌入式计算机系统的捕获与定位对准结果。可以发现,改进 YOLOv4 网络通过捕获信标光斑的位置使其一直保持在图片中心。图 11(a)为无人机在室内捕获对准的信标光斑,图 11(b)为无人机在背景强光干扰条件下捕获对准的信标光斑,图 11(c)为无人机起飞后捕获对准的信标光斑。NVIDIA Jetson Xavier NX 嵌入式

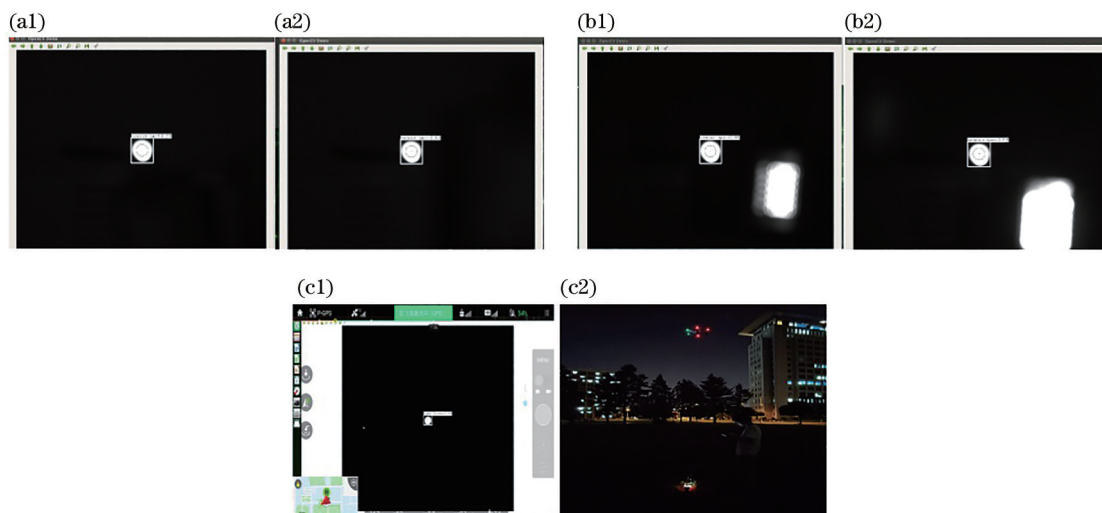


图 11 UAV 的捕获对准结果。(a)室内环境;(b)背景强光干扰环境;(c)飞行状态

Fig. 11 Captured alignment result of the UAV. (a) Indoor environment; (b) background glare interference environment; (c) flight status



计算机系统的参数如表 4 所示。无人机识别出信标光斑的位置信息后通过 PID 算法控制能达到稳定

的捕获对准效果,其控制脱靶量在 105  $\mu$ rad 的范围内波动。

表 4 NVIDIA Jetson Xavier NX 嵌入式系统的参数

Table 4 Parameters of NVIDIA Jetson Xavier NX embedded system

Parameter	NVIDIA Jetson Xavier NX embedded system
GPU	NVIDIA Volta architecture with 384 NVIDIA CUDA <sup>®</sup> cores and 48 Tensor cores
CPU	6-core NVIDIA Carmel ARM <sup>®</sup> v8. 2 64-bit CPU 6 MB L2 + 4 MB L3
Memory	8 GB 128-bit LPDDR4x @ 51.2 Gbit/s
Storage	microSD (128 G)
Mechanical	103 mm×90.5 mm×34.66 mm

## 5 结 论

综合考虑无人机通信领域中的大数据高传输需求以及传统光通信对准和跟踪系统结构复杂、易受背景强光影响、建立稳定链接时间长的问题,提出了一种基于深度学习的实时捕获定位算法。首先,介绍了无人机光通信采用 YOLOv4 网络实时捕获定位系统的结构和工作原理。然后,结合精度以及速度的需求对 YOLOv4 网络进行改进和简化。最后,给出了改进 YOLOv4 网络及对比网络的识别精度和速度。实验结果表明,改进 YOLOv4 网络在信标光斑测试集上的精确率为 99.6%、召回率为 99.8%,在 NVIDIA Jetson Xavier NX 嵌入式计算机平台上的检测速度可达到 42 frame/s,用云台现有配套的三环 PID 控制系统能达到的对准脱靶量范围为 105  $\mu$ rad。实验重点研究的是对信标光斑的识别捕获过程,在后续研究工作中还需进一步提高 PID 控制系统的精度。

### 参 考 文 献

[1] Hossain S, Lee D J. Deep learning-based real-time multiple-object detection and tracking from aerial imagery via a flying robot with GPU-based embedded devices[J]. Sensors, 2019, 19(15): 3371-3396.

[2] 付强,姜会林,王晓曼,等.空间激光通信研究现状及发展趋势[J].中国光学,2012,5(2):116-125.  
Fu Q, Jiang H L, Wang X M, et al. Research status and development trend of space laser communication [J]. Chinese Optics, 2012, 5(2): 116-125.

[3] 时振磊,孟文文,申景诗,等.无人机激光无线能量传输 APT 系统跟踪设计[J].激光技术,2019,43(6):809-814.  
Shi Z L, Meng W W, Shen J S, et al. Tracking design of APT system of laser wireless energy transmission for unmanned aerial vehicle[J]. Laser

Technology, 2019, 43(6): 809-814.

[4] Cao Y, Zhao M F. Research of image processing based on high frame-rate beacon for free space optical communication[J]. Applied Mechanics and Materials, 2012, 236/237: 1015-1019.

[5] Zhu Y F, Huang Y, Li H Y, et al. Research on APT spot detection algorithm in space optical communication[J]. IOP Conference Series: Materials Science and Engineering, 2018, 452: 042184.

[6] Zhu J, Xu Z J, Fu D L, et al. Laser spot center detection and comparison test[J]. Photonic Sensors, 2019, 9(1): 49-52.

[7] 陶磊,洪韬,钞旭.基于 YOLOv3 的无人机识别与定位追踪[J].工程科学学报,2020,42(4):463-468.  
Tao L, Hong T, Chao X. Drone identification and location tracking based on YOLOv3[J]. Chinese Journal of Engineering, 2020, 42(4): 463-468.

[8] 李汉冰,徐春阳,胡超超.基于 YOLOv3 改进的实时车辆检测方法[J].激光与光电子学进展,2020,57(10):101507.  
Li H B, Xu C Y, Hu C C. Improved real-time vehicle detection method based on YOLOv3[J]. Laser & Optoelectronics Progress, 2020, 57(10): 101507.

[9] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.

[10] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.

[11] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6517-6525.

- [12] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2021-05-04]. <https://arxiv.org/abs/1804.02767>.
- [13] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2021-05-04]. <https://arxiv.org/abs/2004.10934>.
- [14] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1440-1448.
- [15] 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.
- [16] 周兵, 李润鑫, 尚振宏, 等. 基于改进的 Faster R-CNN 目标检测算法[J]. 激光与光电子学进展, 2020, 57(10): 101009.  
Zhou B, Li R X, Shang Z H, et al. Object detection algorithm based on improved Faster R-CNN[J]. Laser & Optoelectronics Progress, 2020, 57(10): 101009.
- [17] Jung T H, Cates B, Choi I K, et al. Multi-camera-based person recognition system for autonomous tractors[J]. Designs, 2020, 4(4): 54.
- [18] Lee Y, Hwang J W, Lee S, et al. An energy and GPU-computation efficient backbone network for real-time object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 16-17, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 752-760.
- [19] 罗晖, 贾晨, 李健. 基于改进 YOLOv4 的公路路面病害检测算法[J]. 激光与光电子学进展, 2021, 58(14): 1410025.  
Luo H, Jia C, Li J. Road surface disease detection algorithm based on improved YOLOv4[J]. Laser & Optoelectronics Progress, 2021, 58(14): 1410025.
- [20] Huang G, Liu Z, Laurens V, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017:2261-2269.
- [21] Lee Y, Hwang J W, Lee S, et al. An energy and GPU-computation efficient backbone network for real-time object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). July 16-17, 2019, Long Beach, CA, USA. New York: IEEE, 2019: 752-760.