

基于 YOLOv3 的稻穗识别方法研究

彭 祺,屠礼芬*,程瑞敏,吴 涛

(湖北工程学院 物理与电子信息工程学院,湖北 孝感 432000)

摘 要:为了实现准确、实时农田稻穗识别,引入深度学习技术,设计了一套基于 YOLOv3(you only look once)物体识别算法的农田稻穗识别系统,为植保无人机、耕种机器人等提供基础数据。该系统采用大疆无人机对稻田进行中空俯拍,采集图像信息;对采集到的农田图像进行存储、组织管理,构建基于 web 的图像数据库,且具备图像预处理功能;对数据库中的图像进行筛选、分割,并按照 YOLOv3 算法要求进行图像标注,制作训练样本;最后基于 YOLOv3 深度学习算法对农田中的稻穗进行识别,获得精确的定位框和较好的识别效果。

关键词:农田实时实地;深度学习;物体识别;大疆无人机;样本标注

中图分类号:TP391 **文献标志码:**A **文章编号:**2095-4824(2021)06-0056-05

精确、实时地获取农田的各项信息和农作物的生长动态,是实现精准农业的前提^[1-2]。无人机遥感技术,是一种综合应用遥感、空间定位、无人驾驶飞行、遥测等技术^[3],可以无接触获取农田空间信息,已经在精准农业领域实现了小规模实际应用。但无人机的工作主要在一个较大的范围,在小范围的精细化操作方面还无法代替农民的工作。以图像领域为例,无人机在图像大范围采集方面具有先天优势,相关算法对图像中大物体的识别,如行人、汽车、动物等均取得较好的效果,但对于图像中的小物体识别,如农田中的稻穗,因为目标尺寸较小,采用无人机进行检测的方法很少。稻穗的数量与水稻产量直接相关,其表型特征在水稻育种研究中,是十分重要的量化指标,主要包括:一次枝梗的长度、籽粒数和结实率,每穗籽粒数,稻穗骨架结构等,因此,实时准确的农田稻穗识别是很重要的。

深度学习技术在目标检测和分类领域,表现出比传统图像分析方法更为优秀的效果^[4-8],目前,常用的基于深度学习的目标检测方法主要分

为两类:一类是基于回归的一阶检测器,应用较为广泛的有 YOLO、SSD (Solid - state Drive) 和 FPN (Feature Pyramid Networks)等^[9-11];另一类是基于区域建议的二阶检测器,应用较为广泛的有 R - CNN (Regions with CNN feature)、SPP - Net (Spatial Pyramid Pooling Networks)、Fast R - CNN 和 Faster R - CNN 等^[12-15]。一阶检测器的优点是测速快,缺点是在模型训练时可能出现前景和背景类别不均衡^[16],难以提取小目标特征信息,从而导致部分目标检测效果不佳^[17]。二阶检测器的检测精度相对较高,但由于其使用到了候选区域生成方法和相对复杂的网络结构,所以运算量大,导致检测速度较慢,无法达到实时检测的目的^[18]。

为了实现准确实时的农田稻穗识别,本文利用深度学习的目标检测 YOLOv3 算法训练图像样本数据,构建性能较好的模型,从而实现对稻穗的识别。YOLOv3 算法所需的大量图像原始数据由无人机在农田实时实地采集,通过建立图像数据库来存储、组织管理图像数据,编写了一套拥有

收稿日期:2021-09-22

基金项目:湖北省教育厅科学技术研究项目(D20192701);湖北工程学院教学研究项目(2020A30)

作者简介:彭 祺(1983-),男,湖北武汉人,湖北工程学院物理与电子信息工程学院副教授,博士。

屠礼芬(1986-),女,湖北孝感人,湖北工程学院物理与电子信息工程学院讲师,博士,本文通信作者。

自主知识产权的标注软件来对图像进行预处理和制作训练样本。

1 材料与方法

1.1 硬件集成

为了获取标注的样本图像,首先进行图像采集,无人机采用遥控器来操纵,通过地面遥控器定位对其进行实时控制,旋翼飞行时产生的向下气流有助于压倒水稻叶片,从而能更好地拍摄稻穗,同时远距离操控不会对水稻的生长产生影响,后期可对图片、视频进行图像处理,得到有效的稻穗图像。

大疆经纬 M100 无人机对农田进行俯拍采集图像。DJI Matrice 100 是一款功能强大,可灵活拓展的四旋翼无人机,开放的 SDK 提供了灵活易用的接口,外部设备可以灵活接入,对无人机进行控制与监控,实现硬件集成。其抗风性能良好,飞行时俯拍十分稳定,在无人机高速飞行的状态下,配合云台相机减震系统,在空旷农田有风环境下,依然能采集到清晰的图像。经过多次实验发现,该平台能采集到清晰可用的农田稻穗图像数据。本系统利用四旋翼机 M100 搭载云台相机,可进行灵活的视频监控任务。无人机模型图如图1所示。



图1 无人机模型图

1.2 图像预处理

为了将研究成果更快地应用于生产实践,对采集到的农田图像构建了图像数据库进行组织管理,配合大疆经纬 M100 对农田中水稻图像、视频进行实验数据采集,对视频每帧进行图像预处理得到实验图片,图像预处理的比例切割图如图2所示。

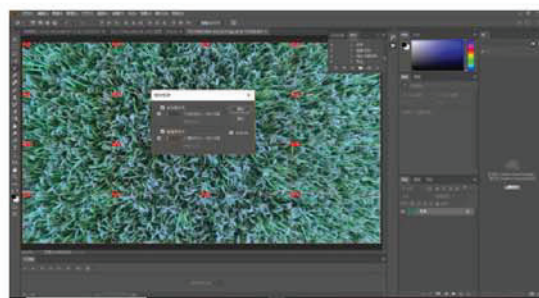


图2 比例切割图

试验数据的预处理过程,使用了 Adobe 公司的 Premiere 和 Photoshop 软件,分别用于从视频中导出图像数据和切割图像数据。筛选过程则是在导出图片数据后将无用的模糊图片进行清理。

1.3 软件集成

由于水稻的稻穗图像与其他图像相比,背景更加复杂,常规样本标注软件因没有预处理功能,制作出来的样本图像训练效果不佳。为了能够准确地识别稻穗,对水稻稻穗图像提取 HSV 色彩特征进行掩模处理,将复杂背景剔除,保留单纯的稻穗目标图像,在此基础上进行标注,扩大样本数据,提高基于 YOLOv3 的深度学习算法样本训练效果。软件运行流程图如图3所示。



图3 软件运行流程图

该软件对样本图像进行 HSV 色彩模型转换,提取 H 通道特征,以此制作掩模,滤除杂乱背景干扰,保存为新的样本图像,增扩样本数据;对已完成样本制作的图像标签信息进行增加、删除修改;标签信息保存为文本文档,也能转化为标准 XML 格式;展示样本标注范例,帮助指导非专业人员进行图像样本标注工作。

2 结果与分析

YOLO 的 loss 函数采用均方和误差函数计算。一共由 3 部分构成,即分类误差、IOU 误差和坐标误差,loss 函数用式(1)计算。

$$loss = \sum (coordErr + iouErr + clsErr) \quad (1)$$

Yolo 在训练过程中 loss 函数用式(2)计算。

$$\begin{aligned} & \lambda_{coord} \sum_{i=0}^{\hat{I}} \sum_{j=0}^S l_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \\ & + \lambda_{coord} \sum_{i=0}^{\hat{I}} \sum_{j=0}^B l_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] + \sum_{i=0}^{\hat{I}} \sum_{j=0}^B l_{ij}^{obj} (C_i - \hat{C}_i)^2 + \lambda_{coord} \sum_{i=0}^{\hat{I}} \sum_{j=0}^B l_{ij}^{obj} (C_i \\ & - \hat{C}_i)^2 + \sum_{i=0}^{\hat{I}} l_i^{obj} \sum_{c \in classes} (p_i(c) - \hat{p}_i(c))^2 \end{aligned} \quad (2)$$

首先是对预测的中心坐标做损失,再对预测边界框的宽高做损失,然后是对预测的类别做损失,最后是对预测的置信度做损失。

Yolo v3 的边界框预测用式(3)计算。

$$b_x = \sigma(t_x) + c_x$$

$$b_y = \sigma(t_y) + c_y$$

$$b_w = p_w e^{t_w}$$

$$b_h = p_h e^{t_h} \quad (3)$$

网络实际的预测值为 t_x, t_y, t_w, t_h , 根据式(3)分别计算获得预测框的中心点坐标和宽高坐标 b_x, b_y, b_w, b_h 。其中, c_x, c_y 为当前目标框相对于左上角目标框的偏移量。 $\sigma()$ 是 logistic 函数,作用是将坐标进行归一化处理,使结果分布在 0-1 之间。 b_x, b_y 为采用归一化处理后的相对于目标框的 cell 值。 p_w, p_h 为与真实框重合度最大的 anchor 框的宽和高。实际操作时,为了将 b_w, b_h 也归一化到 0-1 之间, p_w, p_h 采用 anchor 宽、高与 featuremap 宽、高的比值。获得的 b_w, b_h 为归一化处理后的值。

整个 YOLO v3 架构中没有池化层和全连接层。在进行前向传播时,张量的尺寸变换,通过改变卷积核的步长来得到。在 YOLO v2 中,要经历 5 次缩小处理,那么原输入图像的尺寸会变成实际尺寸的 1/25。可以对比 YOLO v2 和 YOLO v3 的 backbone,如图 4 所示。

Yolo v2 base net			
Type	Filters	Size/Stride	Output
Convolutional	32	3 × 3	224 × 224
Maxpool		2 × 2/2	112 × 112
Convolutional	64	3 × 3	112 × 112
Maxpool		2 × 2/2	56 × 56
Convolutional	128	3 × 3	56 × 56
Convolutional	64	1 × 1	56 × 56
Convolutional	128	3 × 3	56 × 56
Maxpool		2 × 2/2	28 × 28
Convolutional	256	3 × 3	28 × 28
Convolutional	128	1 × 1	28 × 28
Convolutional	256	3 × 3	28 × 28
Maxpool		2 × 2/2	14 × 14
Convolutional	512	3 × 3	14 × 14
Convolutional	256	1 × 1	14 × 14
Convolutional	512	3 × 3	14 × 14
Maxpool		2 × 2/2	7 × 7
Convolutional	1024	3 × 3	7 × 7
Convolutional	512	1 × 1	7 × 7
Convolutional	1024	3 × 3	7 × 7
Convolutional	512	1 × 1	7 × 7
Convolutional	1024	3 × 3	7 × 7
Convolutional	1000	1 × 1	7 × 7
Avgpool		Global	1000
Softmax			

Yolo v3 base net			
Type	Filters	Size	Output
Convolutional	32	3 × 3	256 × 256
Convolutional	64	3 × 3/2	128 × 128
Convolutional	32	1 × 1	
Convolutional	64	3 × 3	
Residual			128 × 128
Convolutional	128	3 × 3/2	64 × 64
Convolutional	64	1 × 1	
Convolutional	128	3 × 3	
Residual			64 × 64
Convolutional	256	3 × 3/2	32 × 32
Convolutional	128	1 × 1	
Convolutional	256	3 × 3	
Residual			32 × 32
Convolutional	512	3 × 3/2	16 × 16
Convolutional	256	1 × 1	
Convolutional	512	3 × 3	
Residual			16 × 16
Convolutional	1024	3 × 3/2	8 × 8
Convolutional	512	1 × 1	
Convolutional	1024	3 × 3	
Residual			8 × 8
Avgpool		Global	
Connected		1000	
Softmax			

图 4 YOLO v2 和 YOLO v3 的 backbone 对比图

YOLO v3 采用逻辑回归的方法,来获得每个边界框分数值,通过计算真实框与先验边界框的重叠度来预测结果,当该重合度得分最高时,认为是最佳框。完全重合时得分为 1,但实际操作中发现,边界框通常不能达到最佳,当与真实对象的重叠超过某个阈值(这里是 0.5),那么就忽略这

次预测。YOLO v3 为每个真实对象分配唯一一个边界框,当真实对象与先验边界框不吻合时,则不生成类别预测损失,只生成目标预测损失。另外,本次稻穗识别模型使用 logistic 函数作为分类器,损失函数采用 binary cross-entropy。

制作训练样本后,训练实现 30 次迭代得到新

的权值文件,运行新的权重,输入图片名称即可对图片进行识别,实验前期效果见图 5,识别效果如图 6 所示。

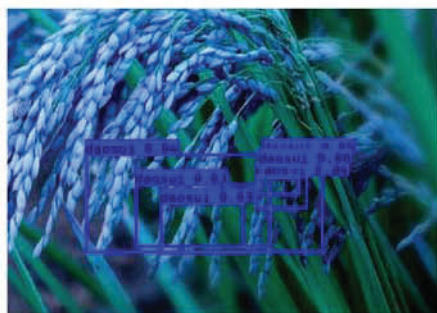


图 5 实验前期效果图



图 6 识别效果图

该方法可实现对稻穗的识别,由图 5 和图 6 可见,相比实验前期有很大的提高,标注的标准逐渐统一,识别效果图精度有所提高,置信度也有所提高,对于 VOC 自己的样本,前景和背景区分度较大,还能成功分割,而农田中的稻穗和背景区分度小、杂乱、难以分割。我们使用无人机对稻田图像进行俯拍,再对图像进行分割筛选、标注,识别的精确度有较大的提高。

3 讨论

近年来,通过图像处理对叶片进行分类识别的方法有很多,但应用于农田复杂环境下的目标识别方法是较少的,为保证实验的严谨性,研究中农田图像均由本团队在农田中实地取得,后期进行图像处理得到 19000 余张图片,其中 9000 张作为实验图像数据,剩余部分取样作为验证图像数据,验证识别结果时需要每张水稻图像具有良好的处理效果,以保证数据的真实性。实际应用过程中,运行程序后只需要输入图像相应名称即可对图像中的稻穗进行识别,YOLO v3 算法应用于

实际生产中,可以实时有效地对水稻成熟状况进行监控,在出现水稻倒伏等状况时可采取一定措施进行补救,此目标识别技术也可扩展到农业各个方面的果实识别。

[参 考 文 献]

- [1] PIERCE F J, NOWAK P. Aspects of precision agriculture[J]. *Advances in Agronomy*, 1999, 67 (1): 1-85.
- [2] ZHANG N, WANG M, WANG N. Precision agriculture—a worldwide overview[J]. *Computers & Electronics in Agriculture*, 2002, 36(2/3): 113-132.
- [3] 范承啸,韩俊,熊志军,等.无人机遥感技术现状与应用[J]. *测绘科学*, 2009, 34(5): 214-215.
- [4] XIE X, WANG C, CHEN S, et al. Real-time illegal parking detection system based on deep learning[C]//*Proceedings of the 2017 International Conference on Deep Learning Technologies*. Chengdu: ACM, 2017: 23-27.
- [5] TSEHAY Y K, LAY N S, ROTH H R, et al. Convolutional neuralnetwork based deep-learning architecture for prostate cancer detection on multiparametric magnetic resonanceimages[C]//*Medical Imaging 2017*. Orlando: Computer-Aided Diagnosis. International Society for Optics and Photonics, 2017: 1013405.
- [6] ZHANG X, CHN G, SARUTA K, et al. Deep convolutional neural networks for all-day pedestrian detection[C]//*International Conference on Information Science and Applications*. Singapore: Springer, 2017: 171-178.
- [7] SUN X, WU P, HOI S C H. Face detection using deep learning: an improved faster RCNN approach[J]. *Neurocomputing*, 2018, 299: 42-50.
- [8] JIANG H, LEARNED-MILLER E. Face detection with the faster R-CNN[C]//*Automatic Face Gesture Recognition*. Washington: IEEE, 2017: 650-657.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. Las Vegas: IEEE, 2016: 779-788.
- [10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//*European Conference on Computer Vision*. Amsterdam: Springer, 2016: 21-37.

- [11] LIN T Y, DOLLAR P, GIRSHICK R B, et al. Feature pyramid networks for object detection [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117 – 2125.
- [12] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Columbus: IEEE, 2014: 580 – 587.
- [13] 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, 2015, 37(9): 1904 – 1916.
- [14] GIRSHICK R. Fast r-cnn[C]// Proceedings of the IEEE international conference on computer vision. Santiago: IEEE, 2015: 1440 – 1448.
- [15] 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.
- [16] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2017, 99: 2999 – 3007.
- [17] ZHANG S, ZHU X, LEI Z, et al. S³FD: single shot scale-invariant face detector[C]//Proceedings of the IEEE international conference on computer vision. Venice:IEEE Computer Society, 2017: 192 – 201.
- [18] REN Y, ZHU C, XIAO S. Object detection based on fast/faster RCNN employing fully convolutional architectures [J]. Mathematical Problems in Engineering, 2018:1 – 7.

Research on Rice Ear Recognition Based on YOLOv3

Peng Qi, Tu Lifen*, Cheng Ruimin, Wu Tao

(School of Physics and Electronic Information Engineering, Hubei Engineering University, Xiaogan, Hubei 432000, China)

Abstract: In order to achieve real-time and accurate field ear recognition, the deep learning technology was introduced, and a set of farmland rice ear recognition system based on YOLOv3 (you only look once) object recognition algorithm was designed to provide basic data for plant protection eppo unmanned aerial vehicle (UAV) and cultivation robot. Dji UAV was used in the system to take aerial photos of the rice field at medium and low altitude to collect image information. The image database based on web was built to store, organize and manage the collected farmland images, and had the function of image preprocessing. The images in the database were screened and segmented, and the images were labeled in accordance with the requirements of YOLOv3 algorithm, and training samples were made. At last, YOLOv3 deep learning algorithm was used to identify the ears of rice in the farmland so as to obtain accurate positioning box and some good recognition effect.

Key Words: real time field on farmland; deep learning; object recognition; Dji unmanned aerial vehicle; sample label

(责任编辑:熊文涛)