

一种基于树莓派部署的改进 YOLOv5 植物生长状态检测算法

崔强 周艳 肖洒 赖盛鑫

(联通(上海)产业互联网有限公司 上海市 200050)

摘要:植物的生长过程往往需要一定的时间,在这段时间内,它会演变成不同的阶段,从而对生长环境的要求不同。目前,确定植物生长状态的方法主要是基于人工先验经验和培育者的目视检查。这种传统方法耗时费力,效率低下且准确率有限。学术界已经提出了许多计算机视觉算法来诊断苹果叶病,但大多数算法只能在高性能 GPU 上运行,这限制了大部分算法在该领域的应用。在本文中,我们优化了 yolov5 网络结构及其卷积模块,并提出了一种新的 yolov5 网络。而这种网络具有更简洁、更轻量的结构,使其可以在树莓派上实时检测植物生长状态。我们构建了一个包含 5323 张图片的植物生长数据集,并将植物的生长状态分为五类。为了优化模型性能,我们采用了以下方法: 1. 使用 Ghost 模块替换了普通卷积,从而显著减少参数数量。2. 通过 CBAM 模块引入 attention 注意力模块来提升模型的特征提取性能。3. 修改了 mosaic 增强部分,将其改为 3x3 样式,以增强数据的多样性。实验结果表明,本文模型的最高 mAP 为 88.3%,最小模型大小为 11.67MB,在 GPU 服务器上的每秒帧数 (FPS) 最高为 85.6。在树莓派 4B 上测试达到了 11.7FPS,比标准的 yolov5 提高了 4.1% 的性能。

关键词:植物生长状态检测; YOLOv5; 注意力机制; 轻量级部署

植物是大自然基本资源,是食物、药物的主要来源。植物在地球上担任着至关重要的角色,它们是地球生态系统的基础,提供氧气和食物,维持着地球上所有生命的生存。植物也可以被用于制药、香料、染料、建筑材料和燃料等领域。虽然植物的生命周期和繁殖方式与动物有很大不同,但是它们同样具有适应环境、生长和进化的能力,是生命在地球上最为多样和美丽的表现之一。作为世界上有名的农业大国,提高农作物生产是一项非常具有挑战性的工作,因此,对植物的研究和识别对人类生活与研究有着重要性和一定的指导意义。

随着社会发展,时代的脚步不断前行,植物的生长和发育的重要性比重越来越大。然而,确认植物的生长状况是一项耗费时间和具有挑战性的任务,需要大量的人工投入和专业技能。传统的确定植物生长状态的方法依赖于培育者的目视检查^[1],这种方法依靠大量富有经验的劳务工作者,判断标准相对来说比较主观,速度慢,成本高,效率低,主观性强,准确率低,时效性差等缺点,非常容易出现误差。

近年来,计算机视觉技术在自动化植物生长分析和

提高植物生长监测的效率和准确性方面展现了巨大的潜力^[2]。人工智能的兴起与实时视频流摄像机监控的结合,加上计算硬件的迭代更新,使得通过视觉对植物状态检测具有高可行性^[3]。基于视觉的检测模型中,YOLO 系列与 SSD 已经非常成熟^[4-8],在很多场景都有标准的落地方案。因此,应用深度学习于植物检测,能够显著地降低工作负担并缩短识别所需时间^[9]。

计算机硬件与软件技术的发展,使农业与计算机工程技术的联系更加紧密。借助计算机视觉、机器学习、深度学习算法等创新工具,在植物检测方面有广阔的前途。然而,经典的机器学习大多依赖于手工制作的低级视觉特征证,这些特征是根据工程经验设计的。当捕获场景相对复杂时,这通常会导致性能不尽如人意。基于 CNN 的模型通过端到端的方式来自动学习低级判别特征和模型参数,使得非专家也更容易处理基于计算机视觉的农作物检测任务^[10]。然而,由于 CNN 网络自身的特点,参数量大和计算开销大,从而限制了大多数 CNN 模型只能在具有 GPU 加速的高性能服务器上实现。这样的特性限制了 CNN 在植物检测领域的应用^[11]。

●基金项目:国家自然科学基金项目(No. 61502220)-融合目标感知与对比度的图像和视频显著性检测技术研究(青年科学基金项目)。

已经提出了各种计算机视觉算法用于植物生长监测和病害诊断。YOLOv5 是一种流行的目标检测算法^[12], 已广泛用于各种应用, 包括植物生长监测。然而, YOLOv5 的高计算复杂性限制了其在树莓派等低功耗设备中的应用。为了使 YOLOv5 模型更加实用并适合部署在端侧设备进行实时检测, 在学术界、工业界已经提出了许多轻量化的方案, 但随之而来的是精度的降低。因此, 为了提高轻量级模型的精度, 可以在网络结构中加入注意力机制, 并以此作为输入, 在特征层的每一层分配不同的权重, 提取更广阔的全局特征, 从而提高检测效果。此外, 注意机制的特点也决定了其对模型的推理速度影响很小, 且不需要分配额外的大空间^[13]。

为了迎接这个挑战, 在本文中, 我们提出了一种改进版 YOLOv5 网络, 而这种网络的网络结构更简单、计算参数更少, 是一种轻量级的网络, 并可以在树莓派上实时检测植物的生长状态。我们引入了多种新技术, 如 Ghost 模块、注意力模块, 来优化模型的性能。我们还构建了一个全面的植物生长图像数据集, 并将植物的生长状态分为五类, 以便于模型的训练和评估。实验结果表明, 我们提出的方法有潜力改变植物生长监测, 并提供一种更为高效和准确的评估植物生长状态的方法。我们的模型轻量且实时的特点也使其能够在低功耗设备中应用, 大大扩展了其潜在的应用场景。

1 相关工作

1.1 数据集准备与预处理

本文使用的数据集图片是通过实时摄像机在植物培养的大棚中拍摄视频录像, 然后按照固定时序间隔截帧获得的。因此, 在架设摄像机时, 要求光线充足, 可以完整的覆盖植物的特征。使用摄像机采取实时视频流的原因有三点。

(1) 数据集的样本数量必须充足, 且包含了植物生长的整个周期, 获取到的视频源需要通过严格的筛选来确定数据是否符合要求。而且在数据打标的时候, 数据集的图片质量非常重要。

(2) 本文研究的算法最终面向的是落地应用, 因此摄像头监控植物生长态势与获取数据源, 与实际应用场景一致。

(3) 边缘端设备树莓派 4B 版本, 支持通过 USB 接口直连摄像机, 符合预期和整体业务设计。

为了让训练样本更符合实际情况, 本文拍摄的图片基于多种情况下不同角度对植物进行拍摄, 并对训练集的样本进行了数据增强处理, 尽可能的消除了无用照片的面积, 减少干扰物对模型的影响。使用开源标注工具对图片进行标注。共计图片总数为 5323 张。

1.2 注意力机制

注意力是一种认知过程, 它选择性的作用于相关信息而忽略深度神经网络中的其他信息。最早起源于自然语言处理, 如今已广泛的用于各种领域。Squeeze-and-Excitation(SE)在 ImageNet2017 数据集上获得了冠军^[14], 而卷积注意力模块(CBAM)结合了通道注意力和空间注意力, 更容易的与其他模块架构相结合^[15]。多个实验表明, 加载了 CBAM 模块之后, 分类和检测的性能有了显著的提高^[16]。

2 模型介绍

2.1 YOLOv5 网络概述

在 2D 目标检测与分类中, YOLO(You Only Look Once)是一种具有出色准确性和监测速度的目标检测算法。YOLOv5 利用深度和宽度的倍数来缩放网络的深度和每层中卷积核的通道数量。YOLOv5 是一种常用的目标检测模型, 其整体框架包括输入端、Backbone、Neck、Prediction 四部分。在输入端, 采用 Mosaic 数据增强和 padding 技术, 网络输入图片大小默认为 640×640 像素。Backbone 部分包含 Focus 层和 CSP 结构^[17]。Neck 层主要将物体特征进行融合, 输出端则对应的是损失函数与非极大值抑制。这些设计和结构的综合应用使得 YOLOv5 在目标检测任务中表现较好。通过这样的设计, YOLOv5 可以有效地提高模型的鲁棒性和准确性, 在各种场景下都能够良好地检测目标物体。因此, YOLOv5 被广泛应用于各种领域, 如自动驾驶、智能安防等。

2.2 Ghost 模块

使用嵌入式设备实现实时植物状态检测是一个严峻

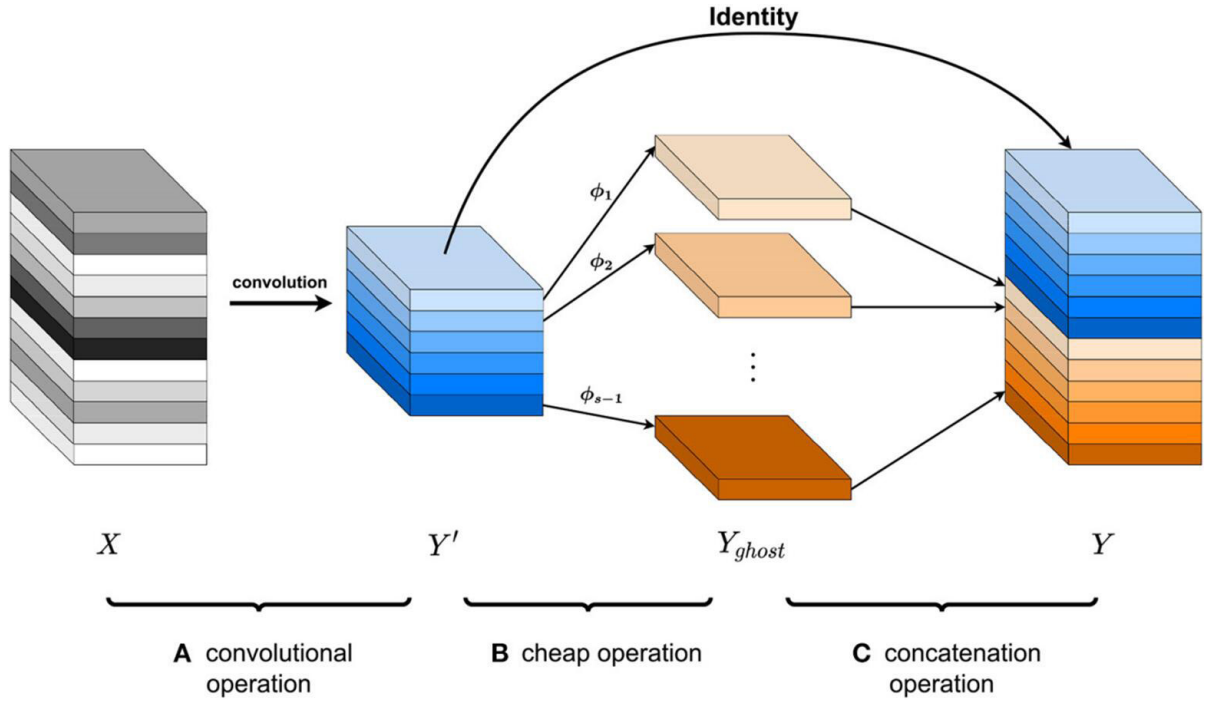


图 1: Ghost 模块运行的三个部分

的挑战,本文在模型优化中,使用 Ghost 模块来替换卷积,从而减少模型参数。如图 1 所示,首先通过卷积运算从 n 个通道的输入图生成 m 个通道的固有特征图,其次,通过 cheap operation 生成 $s-1$ 个通道的 ghost 特征图,最后将 Y 和 Y_{ghost} 拼接起来,此时输出特征图的尺寸变为 $n(n=m \cdot s)$ 。正如文献^[18]所提出的观点一样, Ghost 模块够生成更多特征图,但它所需要的计算量却非常的少。 Ghost 模块的前向传播过程可以分为三个部分。在第一部分中,他通过普通的卷积过滤器生成了一些内在特征图。然后,通过对每个内在特征进行一系列的廉价操作来生成幽灵特征图。最后,将第一部分中获得的固有特征图和第二部分中的幽灵特征图拼接起来,作为输出。

普通卷积层的操作可以表述为:

$$Y=X*f+b \quad (1)$$

在公式 (1) 中, $*$ 代表卷积操作, b 代表偏置项, $X \in R^{c \times h \times w}$ 代表输入的特征图,其中 c 代表输入通道数, h 和 w 分别代表输入特征图的高和宽, $Y \in R^{h' \times w' \times n}$ 代表 n 通道的输出特征图, $f \in R^{c \times k \times k \times n}$ 代表卷积层中的卷积滤波器。需要注意的是 h' 和 w' 分别代表输出特征图中的高和宽, $k \times k$ 是卷积核的尺寸,所以,在通过卷积操作计算输出的特征图尺寸的时候,会产生大量的 FLOPs 计算。如果输入图片的通道数和滤波器数 c 量 n 都很大的时候,

大量的 Flops 计算会耗尽移动设备的内存和计算资源。

Ghost 模块的计算过程可以表示为:

$$Y'=X*f' \quad (2)$$

$$Y_{ghost}=\phi_j(Y_i'), j \in [1, s-1] \quad (3)$$

其中, $Y' \in R^{h' \times w' \times m}$ 表示获得的输出特征图是通过卷积操作获得的,其中使用的滤波器为 $f' \in R^{c \times k \times k \times m}$,输入的特征图为 $X \in R^{c \times h \times w}$ 。为了简化整个推理过程,省略了偏置项。因此,可以通过公式 (2) 和 $h' \cdot w' \cdot k \cdot k \cdot c \cdot m$ 来计算 FLOPs。其中 Y_i' 表示第 i 张特征图, ϕ_j 代表第 j 个线性操作产生的第 j 个 ghost 特征图,这也意味着 Y_i' 产生了 $s-1$ 个 ghost 特征图。在公式 (3) 中,可以计算出 $Y_{ghost} \in R^{h' \times w' \times [m \cdot (s-1)]}$ 。最终,通过计算, $n(n=m \cdot (s-1)+m=m \cdot s)$ 的输出特征图通过下面公式计算得出:

$$Y=Y_{ghost}+Y' \quad (4)$$

在公式 (4) 中, $+$ 代表的是拼接操作,具体 ghost 模块展示如图 1 所示。

用 Ghost 模块代替普通的卷积模块可以简化卷积操作,提升模型速度^[18]。值得注意的是, Ghost 模块的计算参数比普通的卷积层更少。因此在本文中, Ghost Bottleneck 模块和 G3Ghost_X 模块是基于 Ghost 模块构建的。如图 2 所示。在骨干网中,本文使用步长为 2 的

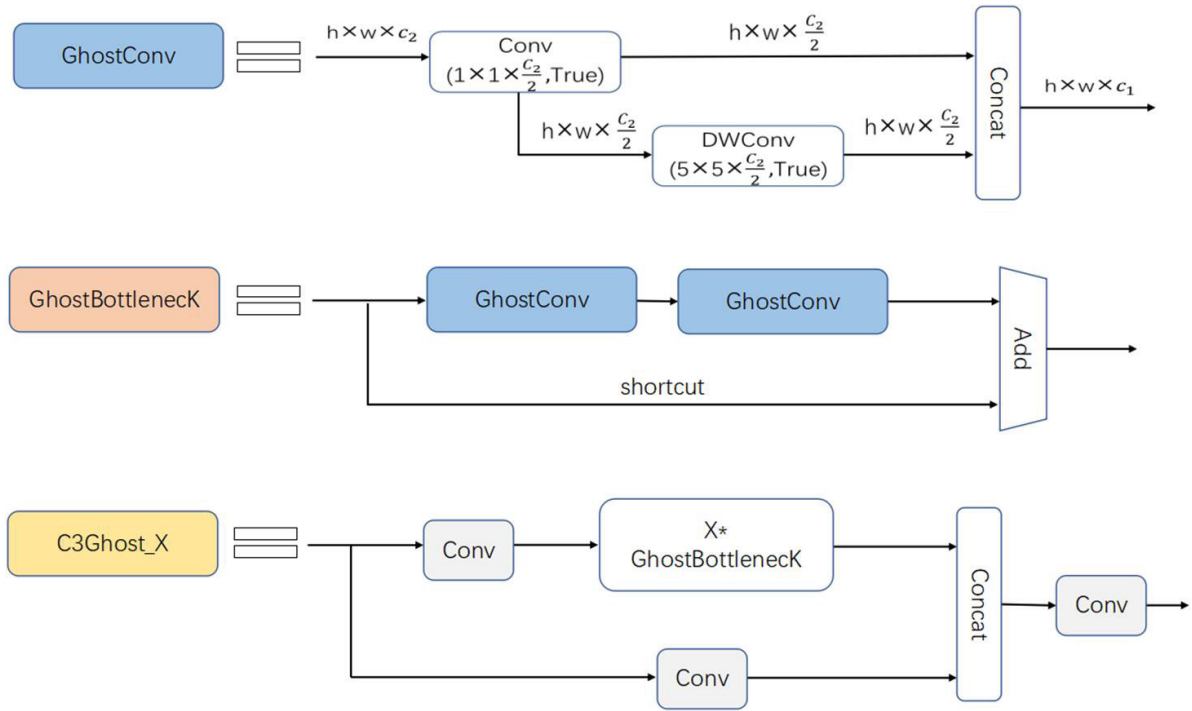


图 2: 添加 Ghost 的模块示意图

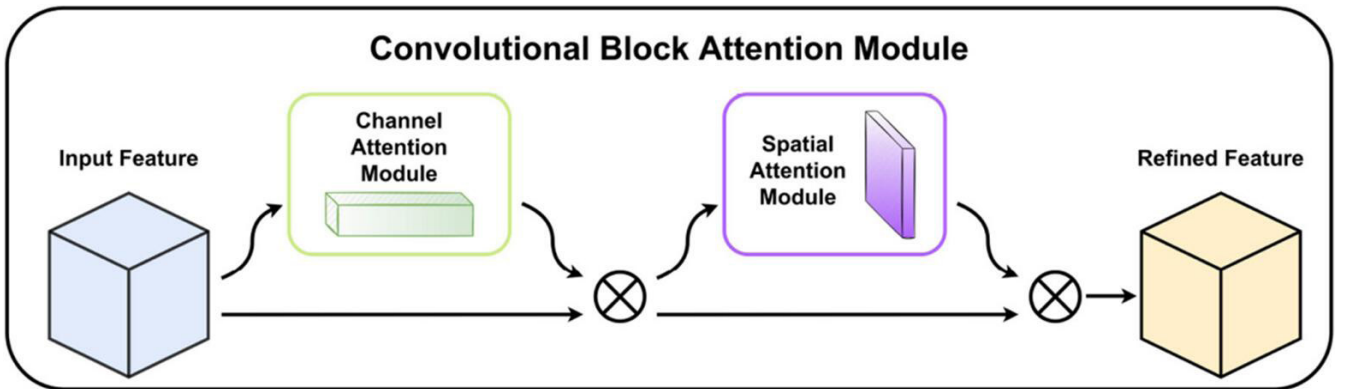


图 3: CBAM 卷积注意力模块结构

Ghost 模块，下采样并使用 C3Ghost_X 模块提取图像特征，而不是普通的卷积和 C3_X 模块。使用 MNIST 数据集进行检测，GHOST 模块提高了训练的收敛速度和检测速度，而检测精度几乎没有任何损失。

图 2 描述了 Ghost 模块、GhostBottleneck 和 C3Ghost_X 模块的结构。在 GhostConv 模块中，将输入为 $h \times w \times c_1$ 的数据，使用 $1 \times 1 \times c_2/2$ 的卷积模块，生成对应的特征图，再送入 $5 \times 5 \times c_2/2$ 的 DWConv 模块，得到对应的特征图，之后将两个特征图拼接起来，以 $h \times w \times c_2$ 的尺寸输出。“True”代表启用激活功能。GhostBottleneck 是一个残差块，而 C3Ghost_X 对应的是 C3_X 模块。

2.3 卷积注意力模块 (CBAM)

CBAM 是一种前馈 CNN 注意力模块，它在 2018 年提出，被证明是一种简单而又非常有效的方法^[19]。该模块包含两个子模块：通道注意力模块和空间注意力模块，这两个模块顺序执行。当输入特征图时，CBAM 将在通道和空间维度上生成注意力图，这些注意力图将与输入特征图相乘，生成细化的特征图。更重要的是，CBAM 是一个轻量级的模型，可以实现端到端的训练，并且可以被训练成为一个可以嵌入到任何一个卷积神经网络当中，CBAM 的基础架构图如图 3 所示。

为了将注意力机制引入到网络中，本文进行了大量

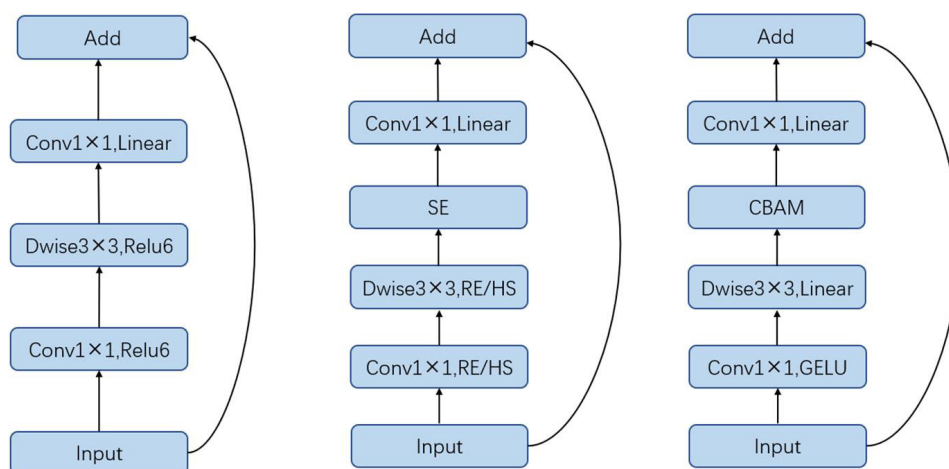


图 4: 三种不同的 Neck 块比较

的深入研究。MobileNetV2^[20] 中通过引入倒置残差和线性 Neck 层提高了模型的性能。与经典的残差网络相比，MobileNetV2 中的倒置残差利用 1×1 的卷积层扩展输入特征的通道，设定扩展率为 6，然后中间扩展层使用 3×3 深度卷积来获取非线性。最后，使用 1×1 的线性卷积层来降维。而在 MobileNetV3 中，在通道注意力提取后的扩展层附加了 Squeeze-and-Excitation (SE) 模块^[21,22]。

受到 MobileNetV2 和 MobileNetV3 的启发，本文保留了倒置残差线性瓶颈模型的一般结构具体结构，然后引入了 CBAM，增加了提取空间信息的能力，同时，MobileNet 中使用的第一个 1×1 卷积集中的 ReLU 或者 hswish^[23] 图如图 4 所示。

如图 4 所示，A 表示 MobileNetV2 块。B 表示 MobileNetV3 块，在不同的网络深度使用不同的激活函数。C 表示引入了 CBAM 的块，使用 GELU 和 Linear。

3 实验

3.1 实验相关配置

试验在高性能的服务器上进行训练，使用了 Nvidia RTX 4070 显卡。12GB 显存。操作系统为 Ubuntu21.04 (64 位)。Python 版本为 3.9，Pytorch 版本为 1.13.1。部署的树莓派版本为 4B。

为了优化网络参数，本文使用随机梯度下降 (SGD) 进行训练。初始学习率 (lr0) 设置为 0.01。

3.2 Mosaic 增强优化

一种常见的方法来增强神经网络模型的鲁棒性和泛化能力是使用 Mosaic 数据增强。该方法将多个不同的图像随机拼接在一起，形成一个新的合成图像，然后通过对合成图像进行裁剪和缩放等操作，从而产生多个不同的训练样本^[23,24]。具体来说，Mosaic 数据增强的步骤如下：

- (1) 从训练集中随机选取 4 张不同的图像，并将它们随机拼接在一起形成一张新的合成图像。
- (2) 随机生成一个裁剪区域，并对裁剪区域进行裁剪，得到一张新的图像。
- (3) 对新的图像进行缩放操作，得到不同大小的训练样本。
- (4) 将缩放后的训练样本输入神经网络进行训练。

Mosaic 数据增强不仅仅只是一种常用的图像数据增强方法，它还具有许多其他优点。在深度学习中，训练集的多样性和数量对于提升模型的鲁棒性和泛化能力至关重要，而 Mosaic 数据增强正是可以通过将多张不同的图像随机拼接在一起来实现这个目的。此外，Mosaic 数据增强还可以帮助模型学习到更多的特征，从而提高模型的分类准确率和检测精度。通过对合成图像进行裁剪和缩放等操作，Mosaic 数据增强可以生成多个不同的训练样本，从而让模型在训练过程中接触到更多的变化，从而使其更好地适应各种实际场景。因此，Mosaic 数据增强在深度学习中被广泛应用，并且已经成为了提高模型性能的有效工具之一。然而，Mosaic 数据增强也存在一些缺点。首先，由于 Mosaic 数据增强是通过随机拼

表 1: 模型消融实验

模型	mAP(%)	AP50(%)	Parameters	GFLOPs	Weight(MB)
YOLOv5s	86.1	92.6	7,701,633	16.4	14.2
YOLOv5s-ghost	86.9	91.1	3,702,991,	8.9	7.59
YOLOv5s-ghost-CBAM	88.3	93.4	6,261,157	11.2	11.67



图 5: Mosaic 原理示意图

接多张不同的图像来生成新的训练样本, 因此会导致生成的合成图像存在一定的噪声和不连续性, 可能会影响模型的训练效果。其次, Mosaic 数据增强需要较高的计算资源和时间, 因为需要随机选取图像、拼接图像、裁剪图像和缩放图像等操作。最后, Mosaic 数据增强可能会导致过拟合问题, 因为生成的训练样本与原始训练样本存在一定的重叠和相似性, 可能会让模型过于依赖训练数据, 而无法泛化到新的测试数据。

基于本文的模型需求和数据特点, 修改原有的 Mosaic 输入为 3×3 共 9 张图片, 经过测试, 模型的准确率得到了提升。如图 5 所示。

3.3 模型消融实验

表 1 中对比了 YOLOv5s, YOLOv5s-ghost, YOLOv5s-ghost-CBAM 的优劣性, 证明了本文优化后的模型具备优势。

3.4 树莓派 4B 部署

将本文的算法部署在端侧树莓派服务器上, 可以成功运行, 单张图片推理时间 11.7FPS, 满足实际场景需求。

4 结论

为了实现在树莓派上部署植物生长状态模型, 本文提出了一种改良后的轻量级 YOLOv5 模型。首先, 我们根据实时视频流挑选符合实际情况的场景, 通过视频流

截帧, 挑选训练素材。构建了一个符合业务场景的数据集, 共计 5323 张图片。其实, 为了轻量化模型, 引入了 Ghost 模块。为了提取全局语义信息, 引入了 CBAM 模块。为了增强数据集的表现能力, 优化了 Mosaic 方法。使本文优化后的 YOLOv5 模型具有精度高, 推理速度快的特点。并且部署在树莓派 4B 上成功运行, 证明了本文方法的可行性, 具有更一定的推广性和应用前景。

参考文献

- [1] Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):2999-3007.
- [2] 张慧春, 张萌, 边黎明, 等. 基于 YOLO v5 的植物叶绿素含量估测与可视化技术 [J]. 农业机械学报, 2022(004): 053.
- [3] Panbo He, Chunxue Wu, Rami Yared, Yuanhao Ma, "HRM: An Intelligent Helmet Recognition Model in Complex Scenes", Advances in Mathematical Physics, vol. 2022, Article ID 1352775, 10 pages, 2022.
- [4] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[C]// Computer Vision & Pattern Recognition. IEEE, 2016.
- [5] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017: 6517-6525.
- [6] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. arXiv e-prints, 2018.
- [7] Bochkovskiy A, Wang C Y, Liao H. YOLOv4:

- Optimal Speed and Accuracy of Object Detection[J]. 2020.
- [8] Berg A C, Fu C Y, Szegedy C, et al. SSD: Single Shot MultiBox Detector:, 10.1007/978-3-319-46448-0_2[P]. 2015.
- [9] Pathan M, Patel N, Yagnik H, et al. Artificial cognition for applications in smart agriculture: A comprehensive review[J]. Artificial Intelligence in Agriculture, 2020, 4.
- [10] Ge Z, Liu S, Wang F, et al. YOLOX: Exceeding YOLO Series in 2021[J]. arXiv e-prints, 2021.
- [11] Agarwal M, Gupta S K, Biswas K K. Development of Efficient CNN model for Tomato crop disease identification[J]. Sustainable Computing: Informatics and Systems, 2020, 28.
- [12] Mekhalafi M L, Nicolo C, Bazi Y, et al. Contrasting YOLOv5, Transformer, and EfficientDet Detectors for Crop Circle Detection in Desert[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19.
- [13] Wang P, Niu T, Mao Y, et al. Identification of Apple Leaf Diseases by Improved Deep Convolutional Neural Networks With an Attention Mechanism[J]. Frontiers in Plant Science, 2021.
- [14] Jie H, Li S, Gang S, et al. Squeeze-and-Excitation Networks[C]// IEEE. IEEE, 2017.
- [15] Strachan J P, Williams R S, Shafiee A, et al. ISAAC: A Convolutional Neural Network Accelerator with In-Situ Analog Arithmetic in Crossbars[J]. Computer architecture news, 2016.
- [16] Gao Y, Wu W, Nan H, et al. Deep Reinforcement Learning based Task Scheduling in Mobile Blockchain for IoT Applications[C]// ICC 2020-2020 IEEE International Conference on Communications (ICC). IEEE, 2020.
- [17] 吴睿, 毕晓君. 基于改进YOLOv5 算法的珊瑚礁底栖生物识别方法 [J]. 哈尔滨工程大学学报, 2022(004): 043.
- [18] Han K, Wang Y, Tian Q, et al. GhostNet: More Features From Cheap Operations[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.
- [19] Zhang Z, Wang M. Convolutional Neural Network with Convolutional Block Attention Module for Finger Vein Recognition[J]. arXiv e-prints, 2022.
- [20] Sandler M, Howard A, Zhu M, et al. MobileNetV2: Inverted Residuals and Linear Bottlenecks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.
- [21] Howard A, Sandler M, Chu G, et al. Searching for MobileNetV3[J]. 2019.
- [22] Jie H, Li S, Gang S, et al. Squeeze-and-Excitation Networks[C]// IEEE. IEEE, 2017.
- [23] 戴长江, 马远浩. 基于端到端框架的食品图像检测方法 [J]. 电子技术与软件工程, 2021(9): 4.
- [24] 肖洒, 贺盼博, 崔强. 一种复杂背景下的智能手势识别模型 [J]. 电子技术与软件工程, 2022(014): 000.

作者简介

崔强 (1993-), 男, 联通 (上海) 产业互联网有限公司员工。研究方向为数据处理。

周艳 (1991-), 女, 联通 (上海) 产业互联网有限公司员工。研究方向为科技创新。

肖洒 (1992-), 男, 联通 (上海) 产业互联网有限公司员工。研究方向为计算机视觉。

赖盛鑫 (1991-), 男, 联通 (上海) 产业互联网有限公司员工。研究方向为人工智能。