



湖州师范学院

2023 届毕业设计(论文)

课 题 名 称: 基于 YOLOv7 模型的菜品识别技术研究

课 题 名 称(英文): Research On Dish Recognition

Technology Based On YOLOv7 Model

学 生 姓 名: 钟尚 学 号: 2019082419

专 业 名 称: 计算机科学与技术

指 导 教 师: 高强 职 称: 讲师

所 在 学 院: 信息工程学院

完 成 日 期: 2023 年 3 月 27 日

教务处制表

基于 YOLOv7 模型的菜品识别技术研究

摘要：餐饮业是劳动密集型行业，分菜、收款等工序需要耗费大量人力资源。因此，研究菜品识别技术可以有效地降低餐饮企业的劳动需求，提高工作效率，具有非常重要的意义。本课题基于 YOLOv7 (You Only Look Once, YOLO) 模型对菜品识别技术进行研究。采用 UNIMIB2016 数据集进行实验验证。首先，对数据集进行标注，并对数据集进行数据增强处理。然后，在 320 和 640 两个分辨率下针对数据增强前后两个数据集进行训练，并收集实验结果。通过对比准确率、召回率以及 mAP (Mean Average Precision) 等指标，实验发现在 640 分辨率下训练的 YOLOv7 模型在测试集上效果优于 320 分辨率下的 YOLOv7 模型；同时数据增强后的模型召回率和准确率高于数据增强之前且有效避免了模型过拟合。实验表明，数据增强对模型训练的效果提升显著。此外，提高图像分辨率也是提升菜品检测效果的有效手段。

关键词：计算机视觉，深度学习，数据增强，YOLO

Research On Dish Recognition Technology Based On YOLOv7 Model

Abstract: The catering industry is a labor-intensive industry, and processes such as food distribution and payment require a large amount of human resources. Therefore, studying dish recognition technology can effectively reduce the labor demand of catering enterprises and improve work efficiency, which is of great significance. This project is based on the YOLOv7 (You Only Look Once, YOLO) model to study dish recognition technology. Experimental validation was conducted using the UNIMIB2016 dataset. Firstly, annotate the dataset and perform data augmentation processing on the dataset. Then, train the two datasets before and after data enhancement at resolutions 320 and 640, and collect experimental results. By comparing accuracy, recall, and mAP (Mean Average Precision) metrics, the experiment found that the YOLOv7 model trained at 640 resolution outperformed the YOLOv7 model trained at 320 resolution on the test set; At the same time, the recall rate and accuracy rate of the model after data enhancement are higher than those before data enhancement, and overfitting of the model is effectively avoided. Experiments have shown that data augmentation significantly improves the effectiveness of model training. In addition, improving image resolution is also an effective means to improve the effectiveness of dish detection.

Key words: Computer Vision, Deep Learning, Data Augmentation, YOLO

目 录

第一章 绪 论	1
1.1 研究背景和意义	1
1.2 国内外研究现状	1
1.2.1 人工智能	1
1.2.2 目标检测	2
1.2.3 菜品检测	3
1.3 研究内容	4
1.4 本章小结	5
第二章 YOLO 模型介绍	6
2.1 YOLO 模型介绍	6
2.2 YOLOv7 模型介绍	7
2.3 本章小结	10
第三章 实验验证	11
3.1 实验数据集	11
3.1.1 实验数据集	11
3.1.2 标注工具介绍	11
3.1.3 数据增强	12
3.2 实验环境和实验步骤	14
3.2.1 实验环境	14
3.2.2 实验步骤	15
3.3 实验结果与分析	15
3.3.1 评估指标	15
3.3.2 实验结果与分析	16
3.4 本章小结	17
第四章 总结与展望	18
4.1 总结	18
4.2 展望	18
参考文献	19
致谢	21

第一章 绪论

1.1 研究背景和意义

随着大数据和人工智能的快速发展，产生了许多与深度学习有关的研究领域和技术，例如：菜品识别、智能健康、智能零售和智能家居等。这些技术极大程度地方便了人们的生活，使得人们的生活质量显著地提高了。吃饭一直是人们十分关心的话题，但是在餐厅里，人工结算效率低下，就餐者为了付款需要花较多时间去排队，容易导致餐厅拥堵。为了解决这些问题，智能餐厅应运而生。智能餐厅能够节省人工成本，增长人效。为了实现智能餐厅，自动识别菜品是智能餐厅必备的技术之一。随着人们对生活品质和健康的要求不断提高，餐饮业已成为现代城市生活中不可或缺的一部分。对于消费者来说，了解所摄入食物的营养成分和热量等信息，对于维护健康十分重要。同时，随着餐饮业的不断发展，餐厅菜单种类繁多，而菜品的名称和配料也可能存在歧义或者翻译不准确等问题，这给消费者点餐和记录饮食信息带来了困扰。因此，开发一种能够自动识别菜品的技术就显得尤为重要和必要。菜品自动识别技术可以通过对菜品图片或视频的分析，实现对菜品的自动识别和分类，并提取相关的菜品信息，如名称、营养成分、热量等，从而让消费者能够更加方便地了解所摄入食物的情况，并更好地进行饮食管理。另外，菜品自动识别技术还可以应用于餐饮行业中的点餐系统中，实现智能点餐和菜品推荐等功能，提高餐厅的服务质量和效率。此外，在菜品溯源方面，菜品自动识别技术也可以帮助监管部门对餐饮业的食品安全进行监管和管理。因此，菜品自动识别技术的研究与应用，将对消费者的健康管理、餐饮业的服务质量和效率以及食品安全监管等方面都产生重要的意义和作用。

综上所述，研究菜品识别技术可以降低餐厅的用人成本，提升人民的生活质量，帮助人们进行饮食管理，具有较高的研究价值。

1.2 国内外研究现状

1.2.1 人工智能

人工智能（Artificial Intelligence，AI）是指让机器通过某些方法模拟人类智能行为的技术。随着计算能力、算法和数据等技术的不断提升，AI 已经成为全球科技领域的重要发展方向之一。以下是 AI 国内外发展现状的概述

国际发展现状：自 20 世纪 50 年代以来，AI 一直是全球科技研究的重点。随着计算能力的提升、互联网和移动设备的普及，AI 技术在各个领域得到了广泛应用，例如自然语言处理、计算机视觉、机器学习等。美国、欧洲、日本、韩国、加拿大等国家一直是 AI 领域的领先者，他们在科技研发、产业布局、人才培养等方面投入巨大，并取得了不少成果。近年来，AI 在医疗、金融、教育、智能家居等领域得到广泛应用，取得了许多重要进展。例如，美国的 DeepMind 公司开发出了 AlphaGo，击败了人类围棋大师^[1]；Google 翻译能够准确翻译多种语言；亚马逊的 Alexa 能够听懂人类语言并作出相应的回答。除此之外，各大科技公司也在竞相研发智能驾驶、智能医疗、智能安防、智能家居等领域的技术。

国内发展现状：中国政府将 AI 作为未来国家战略发展的核心方向之一。在国内，政府、企业和高校联合推动 AI 技术的研究和应用，AI 技术在医疗、金融、智能交通、智能家居、智能制造等领域得到广泛应用。例如，蚂蚁金服的“蚂蚁智能风控”系统可以通过机器学习实现自动化的风控；京东物流的“智能仓储”系统可以根据不同的订单分配仓库和员工等资源；云从科技的“人脸识别”技术可以实现门禁管理、支付验证、考勤签到等功能^[2]。

1.2.2 目标检测

传统目标检测方法的发展可以追溯到上世纪 80 年代,随着人工智能和模式识别技术的发展,传统目标检测方法逐渐成为研究热点。以下是传统目标检测方法的发展历程:

1980 年代:这个时期的传统目标检测方法主要采用模板匹配和特征提取的方法,如卡尔曼滤波、Haar 特征等。这些方法在一定程度上可以实现目标检测,但存在着很多限制,如对光照、旋转、尺度变换等变化比较敏感,对于复杂场景下的目标检测精度较低^[3]。1990 年代:基于特征的方法渐渐地成为了目标检测的主流。在这个时期,研究者们提出了很多特征提取算法,如尺度不变特征转换(Scale-invariant feature transform, SIFT)、加速稳健特征(Speeded-Up Robust Features, SURF)^[4]等。这些算法在一定程度上提高了目标检测的精度和鲁棒性,但仍存在一些问题,如对噪声、光照变化等干扰比较敏感。2000 年代:基于机器学习的目标检测方法开始受到关注。研究者们提出了一些基于机器学习的目标检测算法,如 AdaBoost、支持向量机(Support Vector Machine, SVM)等^[5]。这些算法利用训练集中的正负样本进行训练,并学习到分类器模型。这些方法可以处理一些复杂场景下的目标检测问题,但仍存在一些限制,如需要手工选择和提取特征,难以适应不同场景下的目标检测需求。

这些传统的目标检测方法存在一些缺点,如对光照、旋转、尺度变换等变化比较敏感,在复杂场景下的目标检测的准确率较低,而且需要手工提取特征,难以自适应地适应不同场景下的目标检测需求。因此,随着深度学习的发展,基于深度学习的目标检测方法出现了,如快速区域卷积神经网络(Faster Region-CNN, Faster R-CNN)^[6]、YOLO^[7]等已经成为目前主流的目标检测方法。

基于深度学习的目标检测是一种使用深度学习算法自动从图像或视频中检测出目标的技术。这种技术通常需要训练一个深度神经网络来识别和分类不同的目标。目前常见的深度学习模型包括卷积神经网络(Convolutional Neural Networks, CNN)、循环神经网络(Recurrent Neural Networks, RNN)和深度置信网络(Deep Belief Networks, DBN)^[8]等。

常见的基于深度学习的目标检测算法包括 Faster R-CNN、YOLO、单步多框目标检测(Single Shot MultiBox Detector, SSD)^[9]等。这些算法通常包括两个主要步骤:区域提取和分类器。首先,算法在输入图像以后,会在其中检测出一些可能包含目标的区域,这些区域被称为候选框。接着,分类器会对这些候选框进行分类,判断它们是否包含特定的目标。

近年来,基于深度学习的目标检测是计算机视觉领域中的重要研究方向之一,其发展历程大致可以分为以下几个阶段:

基于深度学习的目标检测的起步阶段(2012-2014 年):在这一阶段,主要是基于卷积神经网络的分类模型,如 AlexNet、计算机视觉组(Visual Geometry Group, VGG)^[10]等,用于图像分类任务。然而,这些模型并不能直接应用于目标检测,因为它们无法直接输出目标的位置和大小等信息。因此,在这一阶段,出现了一些基于 CNN 的区域提取方法,如选择性搜索算法(Selective Search)和候选区域算法(Edge Boxes),用于生成候选框。基于深度学习的目标检测的中期阶段(2014-2016 年):在这一阶段,出现了一些基于 CNN 的端到端的目标检测算法,如 OverFeat 和区域卷积神经网络(Region-CNN, R-CNN)等。这些算法主要通过候选框上提取特征,并使用 SVM 或全连接层来进行目标分类。这些算法虽然效果较好,但计算速度较慢,难以在实际应用中实时处理。基于深度学习的目标检测的现代阶段(2016-至今):在这一阶段,出现了一些新的目标检测算法,如 Faster R-CNN、YOLO 和 SSD 等,这些算法采用了更加高效的区域提取方法和分类器,使得目标检测速度更快,并且能够达到更好的检测精度。

总的来说,基于深度学习的目标检测技术在近年来得到了飞速的发展,在不断地创新和优化中,其应用场景也在不断地拓展。

1.2.3 菜品检测

菜品检测是指利用计算机视觉技术来检测和识别菜品的种类、数量和质量等信息。随着人们对食品安全和品质的要求不断提高,菜品检测技术逐渐得到广泛应用。以下是菜品检测国内外发展现状的概述:

(1) 国外研究现状

在国外,研究主要面向本土菜肴,例如:西餐、印度菜品及日料。与总体菜品的研究量相比,中餐菜品的研究量较小。2014年,Kagaya^[11]等人在菜品分类中采用卷积神经网络,发现颜色特征是影响菜品识别的关键因素,并且在10分类菜品数据集上进行实验,实验的正确率超过70%。同年,Kawano^[12]等人将SIFT、颜色特征和Fisher矢量编码这三种特征提取算法整合在预训练AlexNet模型中,在日本菜品数据集UEC-FOOD100中准确率可达到72.3%。2015年,Yanai^[13]等人通过提取AlexNet网络在ImageNet数据集上训练好的参数,激活相应特征后在UEC-FOOD100数据集上Top-1准确率达到78.77%,在UEC-FOOD256数据集上Top-1准确率达到67.57%,研究表明,使用预训练模型参数可以有效提高小图像数据集的识别率。2016年,Aguila^[14]等人将食物检测和语义分割进行并行组合,分别对分割和识别这两个问题进行处理。首先对图片中不含食物的部分与含食物的部分进行分割,确定其中的食物边界像素信息。然后在包含食物的部分进行检测识别,通过使用非极大值抑制算法避免无效检测。使用了上述方法的YOLOv2模型在UNIMIB2016数据集上准确率为91.1%。2017年,Aguilar^[15]等人针对不同卷积神经网络模型的特点进行研究,给在ILSVRC数据集上完成预训练的ResNet50和InceptionV3模型分配上不同组合形式的分类器,然后采用不同分类器的模型在Food-11数据集上进行实验,得到Top-1准确率均高于93.5%,并且在Food-101数据集上准确率可达86.07%。Mezgec等人开发了一款能够促进帕金森患者治疗的饮食管理系统,采用的是全新的卷积网络结构NutriNet,该网络基于AlexNet对不同分辨率的卷积层进行修改,并且使用随机梯度下降(Stochastic GradientDescent,SGD)、内斯特罗夫加速梯度(Nesterov Acceleration Gradient,NAG)、自适应梯度算法(AdaGrad)三种优化器对模型进行优化处理,提高了食物图像识别效果。2020年,Metwalli^[16]等人提出了食品图像识别算法DenseFood,这个算法基于密集连接卷积网络DenseNet架构。通过密集连接的方式加深了模型深度,采用了最大池化对特征进行下采样,并且使用ELU激活函数替换ReLU,在中餐数据集VIREO-FOOD上的Top-1准确率达到81.23%。国外学者的研究主要针对国外本地的餐饮现状,由于西餐、日料、印度菜品与中餐在颜色、数量、纹理等方面有着巨大的差异,相关成果直接在中式菜品上套用的效果并不明显,但可以充分学习并借鉴其使用的深度学习理论和相关模型,能够对中式菜品图像识别的研究产生指引作用。

(2) 国内研究现状

在国内,对中式菜品图像的研究在2012年的时候就开始了。Chen等人使用稀疏编码(Sensory Coding),解决了食物识别的特征描述符问题,并结合颜色特征和Gabor特征共同表示食物相关特征信息。在此基础上,为每一个特征匹配训练一个多标签的支持向量机分类器,并将独立的分类器同多类Adaboost算法组合。该文还为后续学者整理了第一个含有50类菜品,每类100张图像的中餐菜品数据集,在此数据集上使用该文的算法,得到了68.3%的准确率。2017年,Chen等人收集并整理了含有208类中餐菜品,一共185628张图片的数据集ChineseFoodNet,同时将投票方法与VGG、DenseNet、ResNet等5个不同卷积网络的预测结果相结合,提出了两步数据融合方法TastyNet,在ChineseFoodNet数据集上Top-1得到了81.55%的准确率^[17]。2019年,Liang^[18]等人采用多层级联结构,将特征从全局转移到了局部,并在每级网络中融合特征金字塔结构,形成了多级卷积特征金字塔的细粒度识别模型,在数据集Food-101、ChineseFoodNet中的准确率分别达到了91.4%、82.8%。2020年,Liang^[19]等人提出多视图注意力网络MVANet,利用语义特征具有更多信息的特点,将食物成分以及配方信息作为辅助信息添加到食物识别进程当中,基于此算法优化后的RestNet50模型在数据集ChineseFoodNet上提高了3.2%的Top-1准确率。

菜品识别不同于一般的目标检测任务，因为在菜品图像中，菜品造型各异、色彩缤纷、光照和拍摄角度的不同使得即使是同类菜品也会出现不同的特征。

1.3 研究内容

针对目前市场所需，本文设计一种基于 yolov7 算法的菜品检测模型。研究方案如图 1-1 所示，主要工作包含以下几个方面：

1) 采集实验数据集。本课题使用 UNIMIB2016 数据集，该数据从真实的餐厅环境中收集而来，一共有 73 种菜品，总计 1027 张照片，其中包含了一个餐盘和餐盘上不同的食物，有些食物放在餐具垫上而非碟子中。通过对 UNIMIB2016 前二十种食品进行采集，标注以及数据增强，防止过拟合并提升其通用性。

2) 研究算法模型。首先学习 YOLOv7 模型的理论，掌握相关知识，然后进行菜品识别程序的编写。通过对模型的研究，发现 YOLOv7 具有以下优势：

更高的检测精度：YOLOv7 相对于 YOLOv5 在目标检测精度上有所提升，这得益于模型结构的优化和更加充分的训练。

更快的速度：YOLOv7 在保持较高检测精度的情况下，能够比 YOLOv5 更快地完成目标检测任务，这得益于对模型的进一步优化。

更加灵活的设计：YOLOv7 对模型结构进行了灵活的设计，可以根据不同的应用场景进行个性化的调整，从而更好地满足不同用户的需求。

更加易用的工具：YOLOv7 提供了一系列易用的工具，方便用户进行模型训练、测试和部署，大大降低了使用门槛，适用于菜品检测。

3) 实验验证。分别在 320 分辨率和 640 分辨率的图像进行实验验证，并进行数据增强处理。同时进行消融实验，对比数据增强前和数据增强后的结果，找出最优模型。

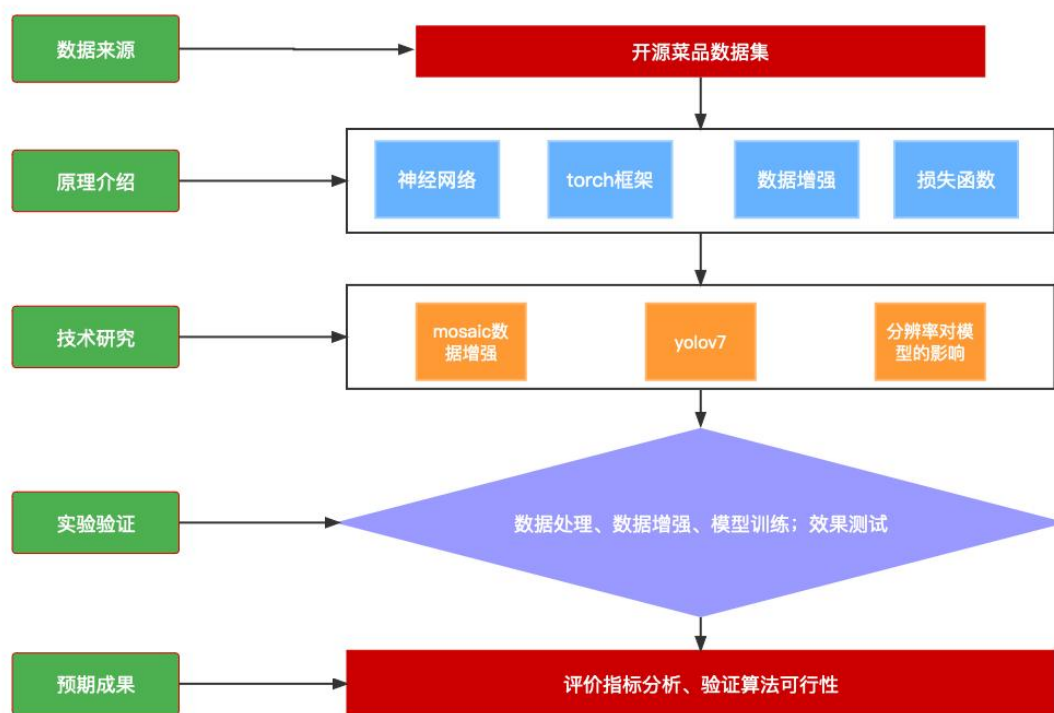


图 1-1 研究方案

1.4 本章小结

通过分析国内外深度学习与人工智能在菜品识别上的发展以及结合本课题的研究背景。检测算法是一种基于图像处理和计算机视觉的技术，通过对图像中的物体进行检测和分类，理解并识别图像中的内容。在菜品检测领域，检测算法可以自动识别菜品的种类、数量和状态等信息，可以大大提高餐饮管理的效率和准确性，同时也有助于食品安全监管等方面的工作。菜品检测技术在国内外都有一定的发展，但仍存在以下不足：数据集不足：菜品种类繁多，但公开的菜品数据集相对较少，这限制了菜品检测算法的发展。多样性不足：不同地区的菜品制作方式和口味存在较大差异，但目前的菜品检测算法往往只能适应特定地区或特定菜系的菜品。精度不够：目前的菜品检测算法在面对菜品形状复杂、摆盘不规则、颜色相近等情况时，检测精度仍有待提高。实时性不足：一些菜品检测算法需要较长的推理时间，无法满足实时检测的需求，这在一些场景下可能会带来问题。模型应用不足：尽管有许多菜品检测算法已经被提出，但在实际场景中的应用还比较少，这可能是由于模型的稳定性和可靠性还需要进一步验证。

因此，为了解决上述问题，提高餐饮管理效率，本人采用了最新的 YOLOv7 模型研究菜品识别技术。在研究过程中，本人首先进行了 YOLOv7 模型的理论学习，掌握了相关知识，其次进行了菜品识别系统的程序编写，然后对代码进行训练，最后对本次菜品识别技术的研究进行总结并思考其在未来的发展趋势。

第二章 YOLO 模型介绍

2.1 YOLO 模型介绍

YOLO 是一种基于深度学习的目标检测算法，由 Joseph Redmon 等人于 2016 年提出^[20]。与传统的目标检测算法不同，YOLO 可以在一张图像中同时检测出多个目标，而且速度相对较快。

YOLO 算法的主要思想是将目标检测问题转化为一个回归问题，即在图像中直接回归出目标的边界框和类别。该算法采用卷积神经网络进行特征提取，然后将提取到的特征与一个小型的全连接层进行连接，用于输出预测的边界框和类别信息。与传统的目标检测算法不同，YOLO 的输出包含了每个边界框的类别、置信度以及位置信息。YOLO 算法的优点在于速度快、准确度高，并且可以直接在输入图像的原始分辨率下进行检测。此外，YOLO 还可以在连续的视频帧中跟踪目标，进行实时目标跟踪，从而实现目标追踪的功能。图 2-1 和图 2-2 是 YOLO 的特征提取示例。



图 2-1 特征提取^[20]

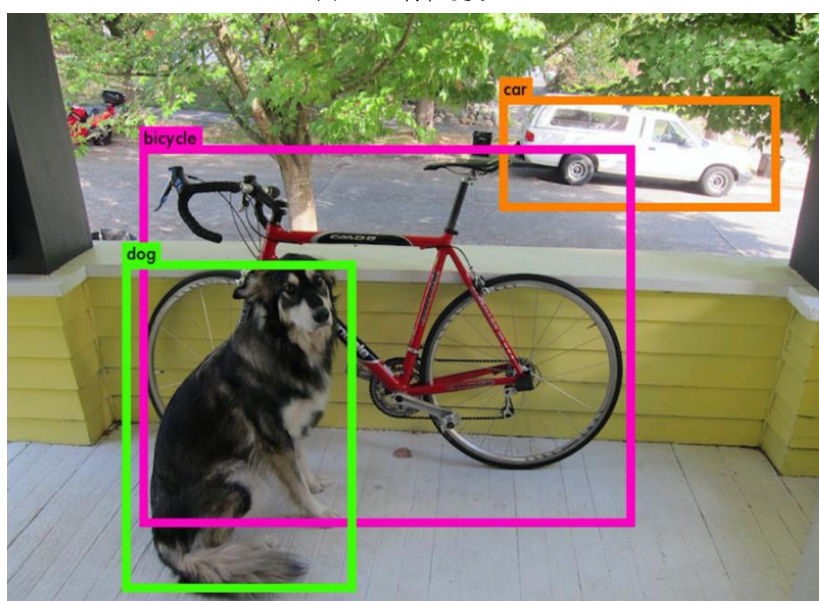


图 2-2 回归结果^[20]

不过,与其他目标检测算法相比,YOLO 在小目标检测和密集目标检测方面的表现不如其他算法,因为 YOLO 的特征图大小较大,导致对小目标的检测不够精细。此外,YOLO 对目标的定位误差较大,可能会导致目标漏检或误检。为了解决这些问题,YOLO 的后续版本进行了改进和优化,如 YOLOv2、YOLOv3 等。YOLOv2 (2017 年):第二个版本的 YOLO 在第一个版本的基础上进行了改进,采用了更深的神经网络结构和一些技巧来提高精度。此外,YOLOv2 也引入了一些新的功能,例如锚点框和多尺度预测。YOLOv3 (2018 年):第三个版本的 YOLO 在第二个版本的基础上进行了改进,引入了一些新的技术,例如多尺度训练、更好的特征提取、更精细的边界框预测等。此外,YOLOv3 还支持检测更多种类的物体,并在速度和精度方面都有了大幅提升。YOLOv4 (2020 年):第四个版本的 YOLO 是比前三个版本更为先进的版本,采用了更深的神经网络结构和一些新的技术来提高精度和速度。此外,YOLOv4 还支持更多的数据增强方法和更复杂的模型优化策略,使得它在各个指标上都超越了之前的版本。总体来说,YOLO 的发展历史可以概括为不断提高检测速度和精度的过程。每个新版本都引入了新的技术和方法,使得 YOLO 在物体检测领域一直处于领先地位^[21]。

2.2 YOLOv7 模型介绍

本文采用 2022 年 7 月 YOLOv5 原作者新发布的 YOLOv7 作为菜品检测模型。YOLOv7 主要的贡献在于:

- 1.模型重参数化:YOLOv7 将模型重参数化引入到网络架构中,重参数化这一思想最早出现于 REPVGG 中。
- 2.标签分配策略:YOLOv7 的标签分配策略采用的是 YOLOv5 的跨网格搜索,以及 YOLOX 的匹配策略。
- 3.ELAN 高效网络架构:YOLOv7 中提出的一个新的网络架构,以高效为主。
- 4.带辅助头的训练:YOLOv7 提出了辅助头的训练方法,主要目的是通过增加训练成本,提升精度,同时不影响推理的时间,因为辅助头只会在训练过程中。

其主要结构为:REP^[22]。REP 模块分为两个,一个是 train,也就是训练,一个 deploy,也就是推理。训练模块,它有三个分支。最上面的分支是 3x3 的卷积,用于特征提取。中间的分支是 1x1 的卷积,用于平滑特征。最后分支是一个 Identity,不做卷积操作,直接移过来。最后把它们相加在一起。

推理模块,包含一个 3x3 的卷积, stride(步长)为 1。是由训练模块重参数化转换而来。在训练模块中,因为第一层是一个 3x3 的卷积,第二层是一个 1x1 的卷积,最后层是一个 Identity。在模型重参数化的时候,需要把 1x1 的卷积,转换成 3x3 的卷积,把 Identity^[23]也转换成 3x3 的卷积,然后进行一个矩阵的一个加法,也就是一个矩阵融合过程。然后最后将它的权重进行相加,就得到了一个 3x3 的卷积,也就是说,这三个分支就融合成了一条线,里面只有一个 3x3 的卷积。它们的权重是三个分支的叠加结果,矩阵,也是三个分支的叠加结果。

YOLOv7 的 backbone 主要采用了 E-ELAN 网络^[24],通过控制最短最长的梯度路径,更深的网络可以有效地学习和收敛。基于 ELAN 设计的 E-ELAN 可以在不破坏原有梯度路径的情况下不断增强网络学习能力。结构如图 2-3 所示:

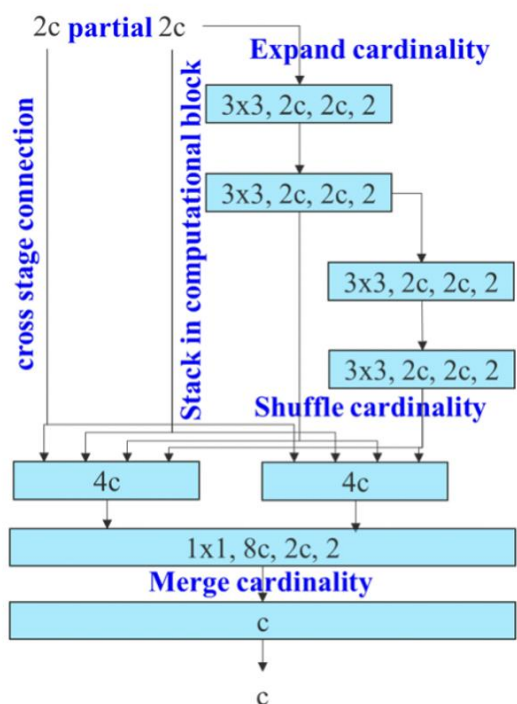


图 2 - 3 E-ELAN 网络

YOLOv7 采取的匹配策略结合了 YOLOv5^[25]的 anchor-base 和 YOLOv6 的 anchor-free。具体措施为：

1. 训练前，会基于训练集中真实框（ground truth, gt），通过 k-means 聚类算法，先验获得 9 个从小到大排列的预测框（anchor）。

2. 将每个 gt 与 9 个 anchor^[26]匹配：YOLOv5 为分别计算它与 9 种 anchor 的宽与宽的比值（较大的宽除以较小的宽，比值大于 1，下面的高同样操作）、高与高的比值，在宽比值、高比值这 2 个比值中，取最大的一个比值，若这个比值小于设定的比值阈值，这个 anchor 的预测框就被称为正样本。一个 gt 可能与几个 anchor 均能匹配上（此时最大 9 个）。所以一个 gt 可能在不同的网络层上做预测训练，大大增加了正样本的数量，当然也会出现 gt 与所有 anchor 都匹配不上的情况，这样 gt 就会被当成背景，不参与训练，说明 anchor 框尺寸设计得不好。

3. 扩充正样本。根据 gt 框的中心位置，将最近的 2 个邻域网格也作为预测网格，也即一个 gt 框可以由 3 个网格来预测；可以发现粗略估计正样本数相比前 YOLO 系列，增加了三倍（此时最大 27 个匹配）。如所图 2 - 4 示的浅黄色区域，其中实线是 YOLO 的真实网格，虚线是将一个网格四等分，在这个例子中，gt 的中心在右下虚线网格，则扩充右和下真实网格也作为正样本^[27]。

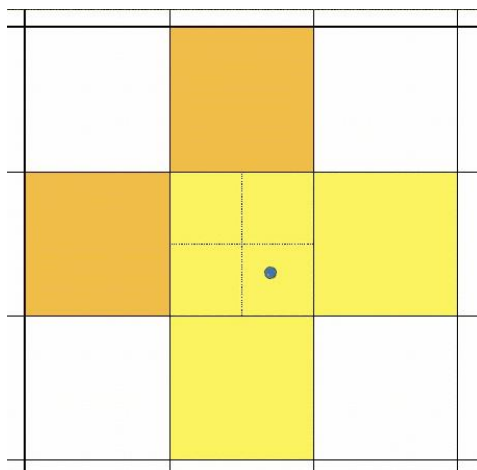


图 2 - 4 扩充正样本

4. 获取与当前 gt 有 top10 最大交并比 (Intersection over Union, IoU) 的预测结果。将这 top10 IoU 进行加和, 就为当前 gt 的 k。k 最小取 1。

IoU 计算公式如下:

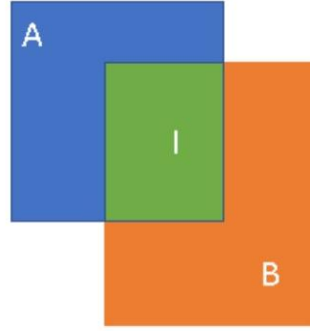


图 2-5 IoU

$$IoU = \frac{A \cap B}{A \cup B} \quad (1)$$

如图 2-5 所示, IoU 的值等于框 A、B 的交集与框 A、B 的并集的比值。

5. 根据损失函数计算每个 gt 和候选 anchor 损失。前期会加大分类损失权重, 后面降低分类损失权重, 并保留损失最小的前 k 个。损失函数计算公式如下:

$$\mathcal{L}_{CIoU} = 1 - IoU + \frac{\rho^2(p, p^{gt})}{c^2} + \alpha v \quad (2)$$

$$v = \frac{4}{\pi^2} \left(\arctan\left(\frac{w^{gt}}{h^{gt}}\right) - \arctan\left(\frac{w}{h}\right) \right)^2 \quad (3)$$

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (4)$$

其中, $\rho(p, p^{gt})$ 是指 anchor 框的中心点和 gt 框的中心点之间的欧氏距离, c 是能够同时包含 gt 框和 anchor 框的最小外接矩形的对角线长度, w 和 h 分别是框的宽和高。

6. 去掉同一个 anchor 被分配到多个 gt 的情况。此做法可以提高效率且不失准确度地检测待检测目标。

YOLOv7 的结构如图 2-6 所示。

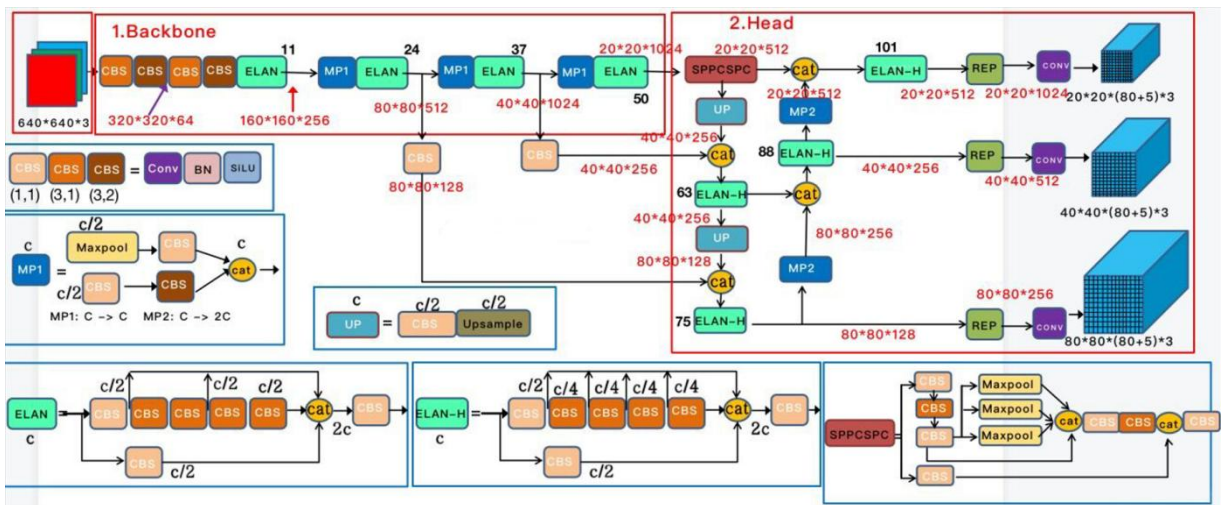


图 2-6 YOLOv7 结构

2.3 本章小结

本章通过对深度学习以及检测算法的研究，对 YOLO 发展历程以及主要版本模型的比较，同时对 YOLOv7 模型的分析，对 E-ELAN 以及 YOLOv7 正负样本分配策略深入研究后，决定采用 YOLOv7 作为菜品检测实验的神经网络算法模型，之后将对采集到的数据集进行对比实验。

第三章 实验验证

3.1 实验数据集

3.1.1 实验数据集

优秀的数据集对模型的提升效果非常显著。一个好的数据集可以提供多样化、真实、丰富的数据样本，包含不同的场景、光照、角度、尺寸、姿态等，这样的数据集能够有效地提高模型的鲁棒性和泛化能力。此外，一个好的数据集还应该准确的标注信息，标注信息应该清晰、准确、全面，包含目标的位置、大小、形状等信息，这样的数据集可以让模型更好地学习和理解目标。当然，一个好的数据集还应该足够大，数据集中的样本数量越多，模型的训练效果也就越好。而且，数据集中的样本应该具有足够的多样性，这样才能够确保模型的泛化性能。综上所述，优秀的数据集对于模型的提升效果是非常重要的，它不仅可以提高模型的准确性和鲁棒性，而且可以提高模型的泛化能力，为实际应用提供更好的支持。

本文选取 UNIMIB2016 数据集作为 YOLOv7 训练的数据集，如图 3 - 1 所示。由于 UNIMIB2016 的样本多、复杂度高，对训练出的模型具有普适性。



图 3 - 1 数据集

3.1.2 标注工具介绍

标注工具为 labelImg, labelImg 是一款开源的图像标注工具，用于为计算机视觉项目创建标注数据。它支持常见的图像格式，如 JPEG、PNG、BMP 和 GIF 等，并能够将标注数据保存为常见格式，如 PASCAL VOC、YOLO 和 TensorFlow 等。labelImg 是使用 Python 和 Qt5 开发的，它具有对用户友好的界面和快捷操作，使得标注过程更加高效和简单。标注工具可以标注矩形、多边形、线条和点等多种形状，并

且可以为每个形状添加标签和注释。此外，labellmg 还支持自定义快捷键和标签，可以针对特定的应用场景进行配置，以适应不同的需求。同时，它还支持批量标注，可以快速处理大量的图像数据。

总之，labellmg 是一款非常实用的图像标注工具，如图 3-2 所示，可以帮助开发人员快速创建高质量的训练数据，为计算机视觉项目的成功提供强有力的支持。

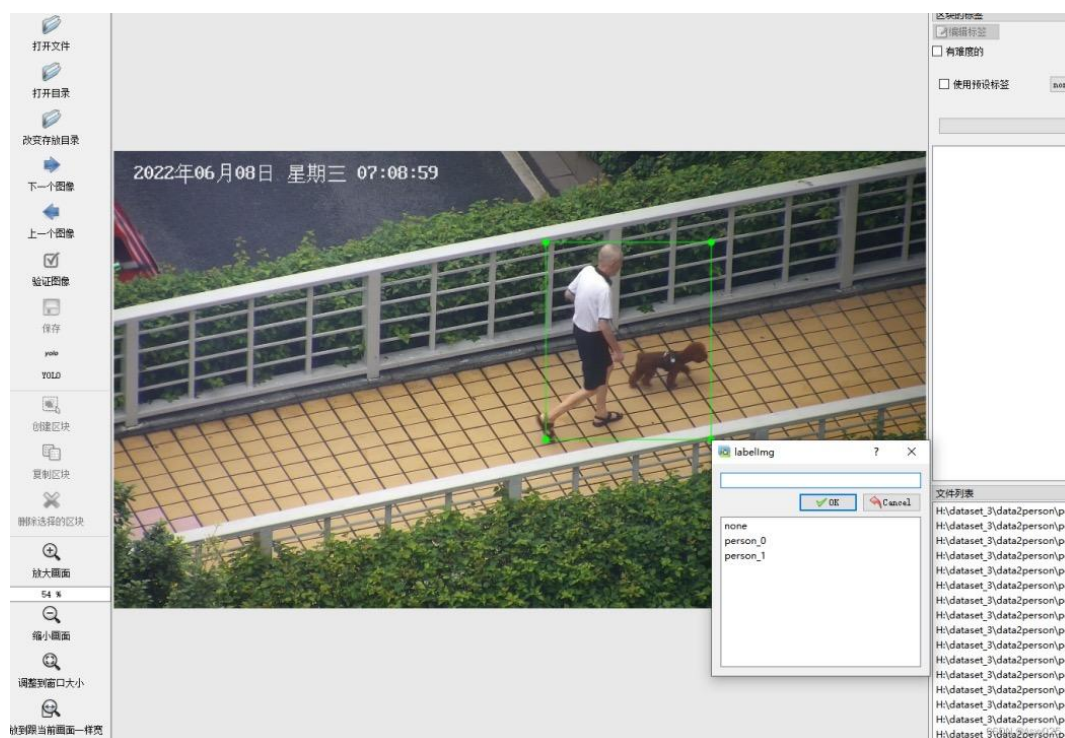


图 3-2 标注工具

3.1.3 数据增强

数据增强是一种常用的数据预处理技术，用于增加训练数据集的数量和多样性，从而提高模型的泛化能力和鲁棒性。数据增强可以通过对原始数据进行各种变换和处理，生成新的数据样本。常见的数据增强技术包括：随机裁剪：对图像进行随机裁剪，生成多个不同尺寸的图像，从而增加训练数据的多样性。翻转和旋转：对图像进行随机翻转和旋转，生成多个不同方向的图像，从而增加训练数据的多样性。色彩变换：对图像进行随机的色彩变换，如调整亮度、对比度、饱和度等，从而增加训练数据的多样性。噪声添加：向图像中添加随机噪声，如高斯噪声、椒盐噪声等，从而增加训练数据的多样性；平移和缩放：对图像进行随机平移和缩放，生成多个不同尺度和位置的图像，从而增加训练数据的多样性；通过数据增强技术，可以有效地增加训练数据的数量和多样性，从而提高模型的泛化能力和鲁棒性。

本文的数据增强部分采用 Mosaic 数据增强模块对已有样本进行数据增强。Mosaic^[28]是一种数据增强技术，用于增加训练数据集的多样性和数量。Mosaic 的基本思想是将多张不同的图片合并成一张大图，并随机裁剪出多个小图进行训练。

具体来说，如图 3-3 所示，Mosaic 将四张不同的图片拼接成一张大图，每张小图都覆盖了大图的四分之一区域，然后随机裁剪出多个小图进行训练。在裁剪过程中，大图中的不同部分可能会组合成新的图像，从而增加了训练数据的多样性。

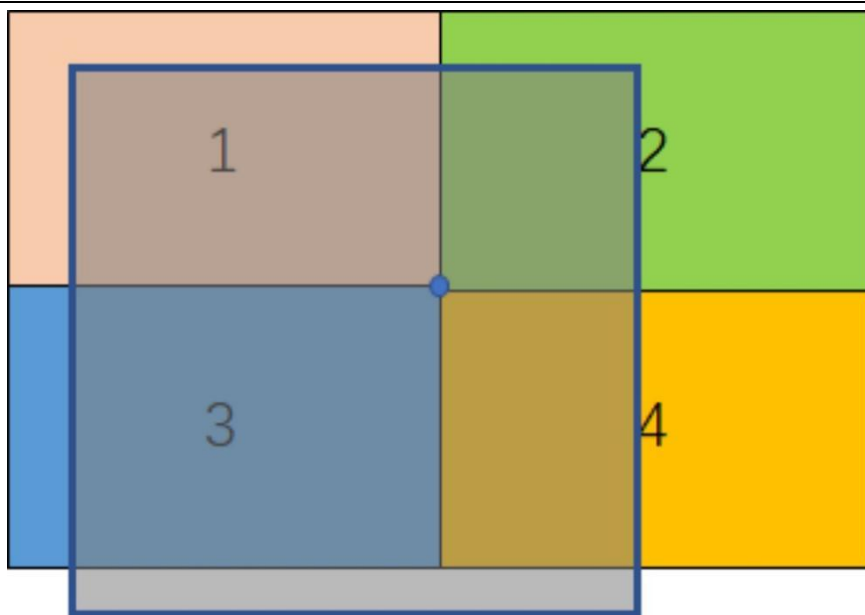


图 3 - 3 数据增强

通过 Mosaic 数据增强技术，可以有效地增加训练数据的数量和多样性，从而提高模型的泛化能力和鲁棒性。此外，Mosaic 还可以减少过拟合现象的发生，提高模型的性能。Mosaic 数据增强技术已经在多个计算机视觉领域的应用中得到了验证，包括目标检测、图像分类、语义分割^[29]等领域。图 3 - 4 既是数据增强的应用例子。



图 3 - 4 数据增强的应用

本文中 MOSAIC 数据增强的参数为 1.0，示意图如图 3 - 5 所示：



图 3 - 5 Mosaic 数据增强

3.2 实验环境和实验步骤

3.2.1 实验环境

本文将使用 Pytorch 框架搭建 yolov7 网络。PyTorch 是一个开源的机器学习框架，广泛用于开发和训练神经网络^[30]。它是由 Facebook 的人工智能研究团队开发的，现在由一个大型开发者社区维护。PyTorch 有着强大且关键功能：

1.动态计算图：PyTorch 使用动态计算图，这意味着该图是在运行时动态创建的。这允许更灵活和高效地执行复杂模型。

2.Pythonic 编程接口：PyTorch 设计为易于使用，并与 Python 编程语言很好地集成。它有一个简单直观的 API，允许快速实现模型和实验。

3.自动微分：PyTorch 支持自动微分，这意味着为图形中的每个操作自动计算梯度。这使得训练具有许多层的复杂模型变得容易。

4.GPU 加速: PyTorch 可以在 GPU 上运行计算, 以实现更快的训练和推理。它支持 CUDA 和 cuDNN 用于 GPU 加速。

5.大型生态系统: PyTorch 拥有一个庞大而活跃的开发人员社区, 他们为其开发做出贡献, 并为机器学习创建了一系列有用的库和工具。

总体而言, PyTorch 是一个强大而灵活的机器学习框架, 非常适合研究和生产应用。它的动态计算图和 Pythonic 编程界面使其易于使用和高度可定制。

与此同时还需要配置如下环境:

1.CUDA:YOLO 使用 CUDA 进行 GPU 加速。需要确保拥有兼容的 NVIDIA GPU 并安装 CUDA 工具包。

2.cuDNN^[31]: YOLO 需要 cuDNN 来加速深度神经网络训练和推理。从 NVIDIA 网站下载并安装 cuDNN。

3.OpenCV^[32]:YOLO 使用 OpenCV 进行图像和视频处理。使用包管理器安装 OpenCV 或从源代码下载并编译它。

3.2.2 实验步骤

YOLOv7 是一种目标检测算法, 采用单阶段检测器 (single-stage detector) 的方式进行物体识别。其训练过程可以分为以下几个步骤:

(1) 数据准备: 首先需要准备训练数据, 包括训练图片和标注文件。标注文件通常采用标记图形文件的 yolo 格式。在这个步骤中, 还需要将训练数据集按照一定比例分为训练集和验证集。

(2) 模型选择和配置: 选择合适的 YOLOv7 模型, 本文选择 yolov7.pt 作为预训练模型, 根据硬件环境选择合适的 batchsize 和 img_size, 本文的 img_size 有两个, 分别为 320、640。数据集的 num_classes 主要有 pizza, yogurt 等。

(3) 数据增强: 为了增加数据样本量和提高模型泛化能力, 需要对训练数据进行增强。YOLOv7 支持的数据增强方式包括旋转、翻转、随机裁剪、变形等。

(4) 模型训练: 使用训练集对模型进行训练, 然后更新模型参数。在每个 epoch 结束后, 使用验证集进行模型评估, 调整模型参数。

(5) 模型保存和测试: 训练完成后, 可以将模型保存为.pt 文件, 然后对测试集进行测试, 评估模型的性能。

本实验数据如下, 数据共包含菜品照片 1027 张, 训练集与验证集比例为 9:1, 使用数据增强中的随机裁剪、旋转、翻转与 paste in。参数设置如表 3 - 1 所示。

表 3 - 1 训练参数设置

参数名	参数设置
batchsize	8
epoch	300

3.3 实验结果与分析

3.3.1 评估指标

本文主要评估指标为准确率、召回率和 mAP。

Precision: 准确率指目标检测模型判断该图片为正样本, 该图片确实是正样本的概率。

$$P = TP/(TP + FP) \quad (5)$$

Recall: 召回率是指分类器判为正样本且真实类别也是正样本的图像数量/真实类别是正样本的图像数量，衡量的是一个分类器能把所有的正类都找出来的能力。

$$R = TP/(TP + FN) \quad (6)$$

在公式(5)和(6)中, TP(True Positive)是真正例, 是指真实值和预测值都是正样本; FP(False Positive)是假正例, 是指真实值是负样本, 预测值是正样本; FN(False Negative)是假负例, 是指真实值是正样本, 预测值是负样本。

mAP: AP 是 Average Precision, 即平均准确率。mAP 是 mean Average Precision, 即各类别 AP 的平均值, 计算出所有类别的 AP 后除以类别总数, 就是 mAP。

$$AP = \frac{1}{N} \sum_{i=1}^m (precision_i) \quad (7)$$

其中, N 是样本总数, m 是同一类别的样本个数, $precision_i$ 是第 i 个样本。

$$mAP = \frac{\sum_{i=1}^k AP_i}{k} \quad (8)$$

其中, k 是类别数, AP_i 是第 i 个类别的平均准确率。

3.3.2 实验结果与分析

实验结果如图 3 - 6 所示:



图 3 - 6 实验结果

320 与 640 分辨率下评价指标结果如表 3 - 2 所示:

表 3 - 2 模型评价指标对比

分辨率	数据增强	准确率	召回率	mAP
320	是	0.05	0.22	0.06
	否	0.698	0.43	0.043
640	是	0.915	0.17	0.146
	否	0.79	0.2	0.036

由表 3 - 2 可知：在没有数据增强的情况下，320 分辨率下的准确率为 69.8%，而 640 分辨率下的准确率为 79%，提升了 13.1%。而数据增强后 320 分辨率下的准确率只有 5%，原因是分辨率变小后，样本中很多标签都找不到了。在 640 分辨率下，数据增强前的准确率为 79%，数据增强后的准确率为 91.5%，提升了 15.8%。由此可见，数据增强后且高分辨率的图像对模型检测率的提升最有帮助。

320 和 640 分辨率下的数据增强后的召回率均低于数据增强前的召回率，是因为图片中食物分辨率占比过小，数据标注不准确所导致的。同时，通过分析 640 分辨率下的召回率可以发现，数据增强前的召回率与数据增强后的召回率相似，只相差了 0.03，而在 320 分辨率下，数据增强前的召回率几乎是数据增强后的召回率的 2 倍。由此可见，数据增强需要严格对数据进行标注，本实验由于数据分布不均匀，因此召回率均偏低。

在 320 分辨率下，数据增强前的 mAP 为 4.3%，数据增强后的 mAP 为 6%，提升了 39.5%。在 640 分辨率下，数据增强前的 mAP 为 3.6%，数据增强后的 mAP 为 14.6%，提升了 3 倍。由此可见，数据增强后 mAP 值均高于数据增强前的 mAP 值，数据增强能显著提升 mAP 值。

实验结果表明：在数据输入相同时，分辨率越大，准确率越高；而且数据的标注是实验中至关重要的一个步骤；同时，数据增强能显著提升 mAP 值。

3.4 本章小结

本章分别训练了数据增强前后在 320 和 640 两个分辨率下的菜品检测模型。并且在验证集上进行验证。通过消融实验的方式验证四个模型在相同验证集下的准确率、召回率以及 mAP。通过对比得出数据增强后的 640 分辨率下训练的菜品模型是最优解。并且通过实际推理展示菜品检测效果，准确率可达到 0.9 以上。

第四章 总结与展望

4.1 总结

本文验证得出, YOLOv7 是一种具有快速、准确的基于深度学习的目标检测模型。此模型在菜品检测方面也可以得到广泛的应用。它通过多尺度特征图, 多制度预测以及高效聚合网络等技术, 在前几代 YOLO 基础上不断更新迭代, 在检测精度和速度方面得到了很大的提升。同时 Mosaic 数据增强解决了菜品数据集少以及模型泛化能力差的问题。在菜品检测方面, YOLOv7 可以快速准确地检测出不同种类的菜品, 帮助餐厅管理者更好地管理菜品的品种和数量。本文使用的 YOLOv7 采用了高分辨训练模型, 即该网络在保持高精度的同时, 也可以实现实时检测。同时在公开数据集上进行了大量的实验, 证明了该算法的有效性和优越性。总的来说, YOLOv7 是一种非常优秀的菜品检测算法, 具有高精度、高实时性和高鲁棒性等优点, 可以应用于实际的菜品检测场景中。菜品检测在国内还未得到重视, 有很大的市场前景。目前存在的问题有: 1.国内菜品数据集几乎没有, 国外数据集也都是一些做法简单容易识别的菜品; 2.模型未部署到移动端, 不能确保部署后的检测准确率与召回率是否会大幅度降低。针对以上问题, 本文提议应发动人力大规模拍摄中国餐厅菜品, 或者通过图片网络搜集, 确定中国菜品分类, 构建规范且规模大的数据集, 然后针对本模型在移动端的部署, 进行大量实验, 得出相应结论。

4.2 展望

菜品检测是人工智能技术在餐饮领域的应用之一, 随着人们对食品质量和食品安全的要求越来越高, 菜品检测技术具有广阔的应用前景和发展空间。菜品检测未来展望的几个方面: (1) 精度和速度的提高: 随着深度学习算法的不断优化和计算机硬件性能的不提高, 菜品检测的精度和速度将得到更大的提升, 可以更准确地识别不同菜品的成分和质量, 同时在较短时间内完成大量的检测任务。(2) 应用范围的拓展: 目前, 菜品检测主要应用于餐饮行业中, 未来随着技术的发展, 可以将其应用于更广泛的领域, 如食品安全监管、烹饪教育、健康管理等。(3) 个性化需求的满足: 随着人们对食品健康和营养需求的不断提高, 菜品检测技术可以根据个人需求进行菜品的定制和匹配, 比如可以识别菜品中的卡路里、营养成分、过敏原等, 从而满足人们的个性化需求。(4) 智能化和自动化的发展: 菜品检测技术的智能化和自动化水平不断提高, 可以实现自动识别菜品成分、自动计算卡路里、自动调整菜品口味等功能, 从而更好地服务于人们的生活。总之, 菜品检测技术在未来具有广阔的发展前景和应用空间, 可以更好地满足人们对食品安全、营养和健康的需求, 同时也可以为餐饮行业的管理和服务提供更加智能化和高效化的解决方案。

参考文献

- [1] Goodfellow I, Bengio Y, Courville A. Deep learning[M]. MIT press, 2016.
- [2] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.
- [3] Huang R, Pedoeem J, Chen C. YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers[C]. 2018 IEEE international conference on big data (big data). 2018: 2503-2510.
- [4] Law H, Deng J. Cornernet: Detecting objects as paired keypoints[C]. Proceedings of the European conference on computer vision (ECCV). 2018: 734-750.
- [5] Wang X, Girshick R, Gupta A, et al. Non-local neural networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7794-7803.
- [6] 李晓磊, 杨涛, 魏斌, 等. 基于深度学习的目标检测算法研究进展[J]. 计算机科学, 2019, 46(1): 11-18.
- [7] 李凌飞, 陈卫平, 周乐, 等. 深度学习在目标检测中的研究综述[J]. 计算机技术与发展, 2018, 28(4): 1-8.
- [8] 范丽丽, 马金凤, 李卫, 等. 深度学习在目标检测中的应用研究进展[J]. 电脑知识与技术, 2018, 14(9): 178-180.
- [9] 张云涛, 谢智强, 李丽华, 等. 深度学习在目标检测中的应用研究综述[J]. 现代电子技术, 2018, 41(4): 61-64.
- [10] 胡苏珂, 陈光明, 张景涛, 等. 基于 YOLO 的移动端车辆检测算法设计[J]. 河北工业科技, 2020, 37(8): 13-17.
- [11] Kagaya H, Aizawa K, Ogawa M. Food detection and recognition using convolutional neural network[C]. Proceedings of the 22nd ACM international conference on Multimedia, 2014: 1085-1088.
- [12] Kawano Y, Yanai K. Food image recognition with deep convolutional features[C]. Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2014: 589-593.
- [13] Yanai K, Kawano Y. Food image recognition using deep convolutional network with pretraining and fine-tuning[C]. Proceedings of the 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2015: 1-6.
- [14] Aguilar E, Bolaños M, Radeva P. Food recognition using fusion of classifiers based on CNNs[C]. International Conference on Image Analysis and Processing, 2017: 213-224.
- [15] Aguilar E, Remeseiro B, Bolaños M, et al. Grab, pay, and eat: Semantic food detection for smart restaurants[J]. IEEE Transactions on Multimedia, 2018, 20(12): 3266-3275.
- [16] Metwalli A S, Shen W, Wu C Q. Food Image Recognition Based on Densely Connected Convolutional Neural Networks[C]. 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC), 2020: 027-032.
- [17] 陈雅丽. 基于深度学习的菜品图像分类方法研究[D]. 电子科技大学, 2018.
- [18] Ciocca G, Napoletano P, Schettini R. Learning CNN-based features for retrieval of food images[C]. International Conference on Image Analysis and Processing, 2017: 426-434.
- [19] Ciocca G, Micali G, Napoletano P. State recognition of food images using deep features[J]. IEEE Access, 2020, 8(1): 32003-32017.
- [20] 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. 2016: 779-788.
- [21] 王文娟, 张超, 张彦, 等. 基于 YOLO 的车辆检测算法在城市交通视频中的应用[J]. 电子技术应用, 2020, 46(8): 26-30.
- [22] 吴俊杰, 王红霞, 王建华, 等. 基于 YOLO 的室内物品识别系统设计[J]. 计算机工程, 2020, 46(11): 269-274.

- [23] 袁浩, 李晓蓉, 张世瑞, 等. 基于 YOLO 的遥感图像目标检测方法[J]. 遥感技术与应用, 2020, 35(6): 1236-1242.
- [24] 吴钊, 马永前. 基于 YOLO 的重载机动车目标检测算法研究[J]. 电子测量技术, 2020, 43(8): 124-128.
- [25] 段秀凤, 谭梅, 张志华. 深度学习在目标检测中的应用研究[J]. 现代计算机, 2017, (10): 142-144.
- [26] 刘志坚, 邢宁波, 马君, 等. 目标检测算法研究综述[J]. 软件导刊, 2018, 17(12): 25-29.
- [27] 邓青龙, 郑永生, 陈浩, 等. 目标检测算法研究进展[J]. 现代计算机, 2017, (6): 107-109.
- [28] 吴峰, 付小东, 程思梦. 目标检测技术的发展综述[J]. 电脑知识与技术, 2018, (13): 7-9.
- [29] 王恒, 李建平. 目标检测中的深度学习方法研究进展[J]. 自动化与仪器仪表, 2018, (1): 1-6.
- [30] 刘维, 王彦, 王奕, 等. 目标检测算法研究进展综述[J]. 计算机工程与设计, 2018, (12): 2851-2855.
- [31] 刘建, 张思玉, 杨云鹏, 等. 目标检测技术的研究与发展[J]. 计算机工程与应用, 2018, 54(3): 141-147.
- [32] 杨迪, 黄一川, 郑文翔, 等. 目标检测算法的研究与应用综述[J]. 信息安全, 2018, (7): 131-134.

致谢

经过一段时间的学习和研究，我终于完成了毕业设计和论文。在此期间，从查阅文献，从中学习理论知识和研究思路，然后到代码的编写和修改，最后到论文的撰写与修改，我感觉收获良多，受益匪浅。我也从中体验到了科研工作的艰辛，但是，在历经困难以后所获得的成果，是十分令人喜悦的。

首先，我要感谢高强老师。从选题、开题、编写代码到撰写论文，高老师总是关心学生的研究情况，及时给予帮助。高老师渊博的理论知识、认真的科研精神和平易近人的态度一直影响着我，促进了我的科研工作的顺利进行。在此，我要向高老师表示最诚挚的感谢。同时，我要感谢大学四年里向我授课的各位老师，感谢他们的悉心教导，让我学到了扎实的专业知识。然后，我要感谢我的父母、同学和朋友，感谢他们在平日里对我的关心和引导。最后，感谢师院的耐心栽培，让我成为了一个能够为社会发展和国家建设做出贡献的人。

四年的大学时光转瞬即逝，回望走过的求学之路，有坎坷有曲折也有春风得意。在这四年间，我不仅学到了扎实的理论知识，还学到了做人做事的方法。虽然即将要离开学校步入社会，我有很多的不舍和惋惜，但在今后的生活和工作中，我会坚持努力，乘风破浪。