



湖州师范学院

## 2023 届毕业设计(论文)

课题名称: 基于 YOLOv7 模型的光学遥感图像  
典型地物检测技术的研究

课题名称(英文): Typical Ground Object  
Detection Technology System for Optical Rem-  
ote Sensing Images Based on YOLOv7 Model

学生姓名: 屈熠丽 学号: 2019082206

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 20 日

教务处制表

# 基于 YOLOv7 模型的光学遥感图像典型地物检测技术的研究

**摘要：**随着现代遥感技术的发展和遥感卫星数量的增加，遥感科技进入了高空间、高光谱、高时间分辨率时代，产生了大量高质量的遥感图像。光学遥感图像具有内容直观，纹理清晰、分辨率较高等优点，非常适合用于地物检测。近几年，伴随着人工智能和计算机视觉技术的发展，一系列基于深度学习的物体检测和识别方法被提出。本课题从对光学遥感图像信息有效利用的实际需求出发，使用深度学习目标检测技术进行研究。首先研究基于 YOLOv7 模型目标检测算法，该算法的优势是检测准确率高，速度快。然后在 Pytorch 框架下搭建目标检测模型，并在在 RSOD 数据集上对检测算法进行验证。实验结果表明，该模型对光学遥感图像具有很高的检测效率，针对飞机、油桶、立交桥、操场这四种地物，其检测率分别达到了 96.96%、97.63%、95.65%、96.70%。为了测试模型在有图像噪声下的检测精度，本论文组建了图片相同但噪声大小不同的两个数据集，共进行了三次实验。根据实验结果的对比和分析，表明加入图片噪声后将会在一定程度上降低目标的检测效率。

**关键词：**光学遥感图像；深度学习；目标检测；YOLO

# Typical Ground Object Detection Technology System for Optical Remote Sensing Images Based on YOLOv7 Model

**Abstract:** With the development of modern remote sensing technology and the increase in the number of remote sensing satellites, remote sensing technology has entered the era of high space, high spectrum and high temporal resolution, producing a large number of high-quality remote sensing images. Optical remote sensing images have the advantages of intuitive content, clear texture and high resolution, which are very suitable for ground object detection. In recent years, with the development of artificial intelligence and computer vision technology, a series of object detection and recognition methods based on deep learning have been proposed. Starting from the actual needs of effective use of optical remote sensing image information, this project uses deep learning target detection technology to conduct research. Firstly, the object detection algorithm based on YOLOv7 model is studied. The advantage of this algorithm is high detection accuracy and fast speed. Then, the object detection model is built under the Pytorch framework, and the detection algorithm is verified on the RSOD dataset. The experimental results show that the model has high detection efficiency for optical remote sensing images, and the detection rates of aircraft, oil barrels, overpasses and playgrounds reach 96.96%, 99.63%, 97.65% and 92.70%, respectively. In order to test the detection accuracy of the model under image noise, two datasets with the same picture but different noise magnitude were constructed, and a total of three experiments were carried out. According to the comparison and analysis of the experimental results, it is shown that the detection efficiency of the target will be reduced to a certain extent after adding picture noise.

**Key words:** Optical remote sensing images; Deep learning; Object detection; YOLO

# 目录

第一章 绪论 .....	1
1.1 选题背景及意义 .....	1
1.2 国内外研究现状 .....	1
1.2.1 传统目标检测算法 .....	1
1.2.2 基于深度学习的目标检测算法 .....	2
1.3 研究方案及论文组织结构 .....	2
1.4 本章小结 .....	3
第二章 相关技术概述 .....	4
2.1 光学遥感图像特性分析 .....	4
2.2 目标检测技术概述 .....	4
2.2.1 神经网络的发展 .....	5
2.2.2 YOLO 模型概述 .....	6
2.2.3 基于深度学习的 YOLOv7 算法 .....	7
2.3 目标检测评价指标 .....	9
2.4 边界框回归 .....	10
2.5 Pytorch 深度学习框架 .....	11
2.6 本章小结 .....	11
第三章 实验过程与结果分析 .....	12
3.1 数据集简介与构建 .....	12
3.1.1 数据集格式组织 .....	13
3.1.2 添加噪声 .....	13
3.2 实验环境配置 .....	14
3.3 实验流程 .....	14
3.4 实验结果对比及评估 .....	16
3.4.1 原始数据集 .....	16
3.4.2 $\sigma=0.15$ 的数据集 .....	17
3.4.3 $\sigma=0.30$ 的数据集 .....	17
3.4.4 结果对比及分析 .....	18
3.5 本章小结 .....	20
第四章 总结与展望 .....	21
参考文献 .....	22
致谢 .....	24

## 第一章 绪论

### 1.1 选题背景及意义

遥感 (Remote Sensing, RS) 技术是指在与目标物体非接触的情况下, 通过搭载在人造卫星或航空飞机等遥感平台上的传感器, 对地面对象的电磁波、光波特征进行监测, 并对目标进行感知和分析的技术<sup>[1]</sup>。遥感图像按其采集方式可划分为 SAR 合成孔径雷达图像、红外图像、多光谱图像和可见光图像等<sup>[2]</sup>。其中使用最为广泛的就是光学遥感图像, 它内容直接, 分辨率高, 信息量大, 具备丰富的空间信息、清晰的几何结构, 可以最大限度反映地形地貌, 能为人们提供大尺度的地理信息。因此, 它在城市规划、环境保护、资源管理等实际应用场景中, 可以起到重要的作用<sup>[3]</sup>。

虽然遥感图像中信息丰富, 但其中夹杂着大量无效干扰信息。要使遥感图像得到更好的应用, 就需要通过信息处理获取有用的信息。遥感图像目标检测技术就是通过算法实现信息处理过程, 能对图像中的目标对象自动分类和定位<sup>[4]</sup>。早期的传统遥感图像目标检测算法主要采用人工判读的方式, 基于手工设计的特征描述子来提取候选目标并进行验证。但由于遥感图像数据量大, 地形复杂, 且遥感图像相较于自然影像有着较大的差异等原因, 传统的判读方式对图像中各种地物目标短时间内难以判读, 实时性较差; 加之人工判读的方式受多种因素的制约, 如对判读区域的熟悉程度和判读者的经验等, 对于遥感图像数据难以实现高效运用。相比传统目标检测算法, 基于深度学习的目标检测技术不需要大量的人工参与, 通过设计合适的模型可以自动提取目标特征, 特征表达更具鲁棒性和泛化性。因此, 基于深度学习的遥感影像目标检测和分类研究在当前具有重要意义。

### 1.2 国内外研究现状

近年来, 随着高分辨率卫星遥感、高光谱遥感等遥感技术的发展, 传感器种类不断丰富, 探测能力不断提高, 产生了大量高质量的遥感图像; 同时人工智能的发展推动了机器学习的发展, 国内外众多学者相继提出了许多目标检测方法, 取得了大量研究成果。这些进步成为了遥感图像目标检测算法发展的动能。

#### 1.2.1 传统目标检测算法

传统目标检测算法主要有基于模板匹配的算法、基于传统机器学习的算法、基于先验知识的目标检测算法以及以及基于目标图像分析(Object-Based Image Analysis, OBIA) 的目标检测算法<sup>[5]</sup>等。传统目标检测框架中包含区域选择、特征提取、分类 3 个阶段。首先使用多尺度的滑动窗口对输入图像进行遍历获取候选区, 在特征提取阶段以尺度不变特征变换<sup>[6]</sup>、梯度方向直方图特征等人工特征提取方法, 提取感兴趣区域, 产生特征向量。特征向量输入第三阶段后, 支持向量机 SVM、Adaboost 等分类器会对预测区域中目标的所属类进行分类并输出检测结果。如图 1-1 所示:

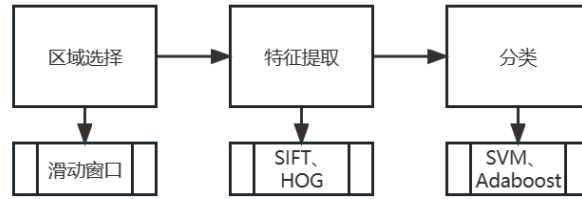


图 1-1 传统目标检测算法框架

### 1.2.2 基于深度学习的目标检测算法

近年来，在人工智能技术的推动下，国内外学者纷纷将深度学习应用于目标检测任务。深度学习是完全数据驱动的，从原始图像的像素开始，利用卷积层抽取低级特征，再利用线性或者非线性的方法组合低级特征获取高级特征，从而达到对样本数据的多层次的自动抽取，实现更精确的目标检测。

基于深度学习的目标检测算法根据是否需要候选区域提取可以分成两个类别，一是有候选区域的双阶段（two-stage）检测算法，以 R-CNN 系列为代表。这种方法借鉴了传统目标检测方法，精度高但是牺牲了算法速度，难以满足实时检测需求；二是基于回归的单阶段（one-stage）检测算法，这种方式摒弃了候选框生成阶段，直接完成目标分类和回归的操作，提高了算法的实时性。单阶段检测算法主要有 YOLO 系列、SSD 系列、RetinaNet 等。

区域卷积神经网络模型（Region-CNN，R-CNN）<sup>[7]</sup>是将深度学习算法应用于到目标检测领域的经典之作，将卷积神经网络应用于特征提取，大幅提升了检测率。Fast R-CNN<sup>[8]</sup>在 R-CNN 的基础上增加了感兴趣区域池化（ROI pooling），能减少参数量，提取出固定尺寸的特征，再由全连接层完成分类和回归操作。2015 年的 Faster R-CNN<sup>[9]</sup>模型提出区域提议生成网络（RPN），利用 Anchor 机制将区域生成和 R-CNN 结合起来，进一步提高了检测速度。

为提高检测实时性，2016 年，Joseph Redmon 等人提出了一种单阶段（one-stage）的目网络结构 YOLO（You Only Look Once）<sup>[10]</sup>。它可以对输入图像通过网络直接回归出目标位置和类别信息，舍弃了候选框生成阶段，直接对目标进行分类和回归操作，提高了目标检测算法的实时检测速度。2016 年，Liu W 等人提出了一种同样基于回归的单阶段目标检测算法 SSD<sup>[11]</sup>，综合了 Faster RCNN 与 YOLO 的思想，仅使用一个卷积神经网络就完成特征提取、目标分类和回归，有多种尺度的检测框并能利用多层次特征信息，提升了速度和精度。近年来，YOLO 系列网络发展迅速，在目标检测领域中较有代表性。

### 1.3 研究方案及论文组织结构

本研究基于 YOLOv7 的光学遥感图像典型地物检测技术，使用深度学习框架 Pytorch 搭建了 YOLOv7 算法的模型结构和损失函数，通过 ROSD 数据集对算法模型的性能进行验证，得到最终训练模型对飞机、操场、立交桥、油桶四类目标的检测效果，最后通过比较平均检测精度和召回率等指标对算法模型的优缺点和适用性进行了分析。课题组织结构具体如下：

第一章 绪论：对本文的研究背景进行阐述，并对目前相关领域的研究现状进行了总结，介绍了本文的研究方案和基本框架等。

第二章 相关技术概述：分析了遥感图像的特性，介绍目标检测技术相关理论知识，对 YOLO 模型进行了总体介绍，以及详细阐述了本文主要使用的 YOLOv7 模型基本结构和检测流程。

第三章 实验过程与结果：使用 YOLOv7 模型搭建深度学习框架进行具体实验，建立不同图像噪声的数据集，根据模型在三个数据集上的训练效果，分析比较精确率和召回率，总结并分析实验结果。

第五章 总结与展望：对论文的研究进行了总结，指出了其中存在的问题，并对今后的工作进行了展望。

根据论文组织结构画出组织结构图如图 1-2 所示。

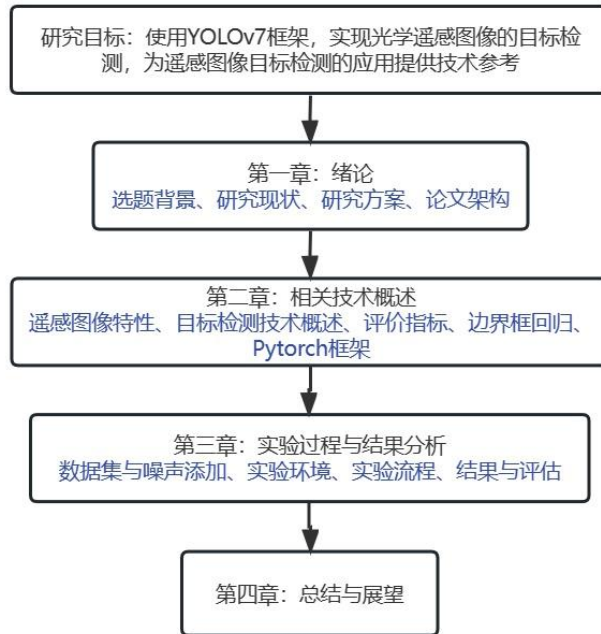


图 1-2 研究方案

#### 1.4 本章小结

本章介绍了本课题研究的背景和意义，强调遥感信息检测及利用的重要性，并介绍了传统的目标检测方法和基于深度学习的目标检测算法的国内外发展现状，分析了当前的发展趋势。最后对本课题的研究内容和研究方法进行了简单的介绍，对接下来关于深度学习遥感图像检测的研究具有重要作用。

## 第二章 相关技术概述

### 2.1 光学遥感图像特性分析

在目标检测领域，常用的遥感图像根据其获取的传感器大致可分为三种：合成孔径雷达（SAR）遥感图像、红外遥感图像、可见光遥感图像。如图 2-1 所示，不同传感器的成像存在区别<sup>[12]</sup>。

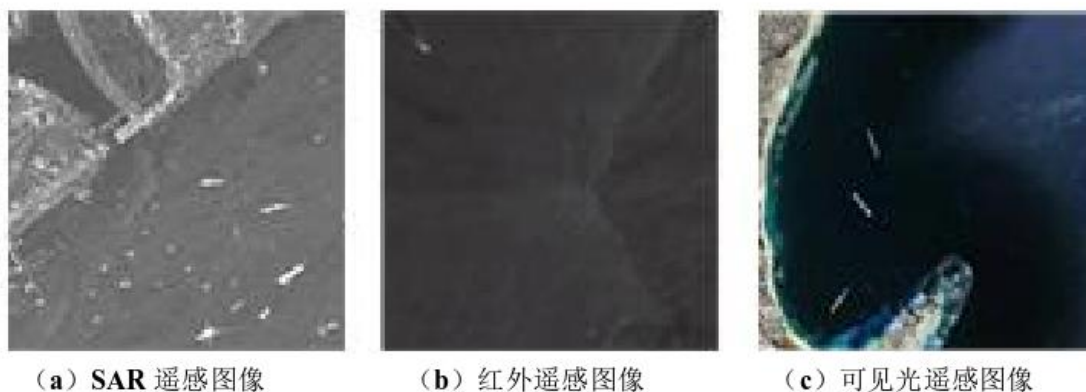


图 2-1 不同传感器下海面成像

光学图像是采用光学摄影系统获取的以感光胶片为介质的图像，光学遥感图像一般由可见光及部分红外探测器采集，往往含有多个波段的信息。相较于主动成像数据，虽然光学遥感图像易受气候影响，难以提供全天时、全天候的遥感图像，但仍具有优势，如可呈现地物本质特征、可以实现像素级分类判别等。自 2010 年我国“高分专项工程”启动并稳步推进，我国自产高分遥感卫星的发展势头迅猛，在资源管理、城市规划、灾害防备等领域发挥着重要作用<sup>[13]</sup>。

在光学遥感图像目标检测领域，光学遥感图像与自然场景下的图像差异较大，且存在大量复杂的自然和人造背景，如图 2-2 所示。此外，某些光学遥感图像中存在一些尺度小但密度大的目标群，如停车场的汽车、港口的轮船等，如图 2-3 所示。这些原因导致难以预设检测网络的锚框，从而造成目标被漏检的结果，这为深度学习算法的精度提出了更高的要求。

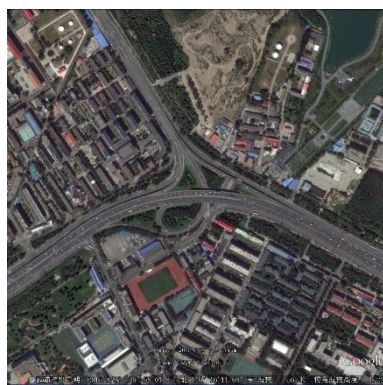


图 2-2 复杂背景下的立交桥



图 2-3 停车场车辆图像

### 2.2 目标检测技术概述

目标检测技术是计算机视觉领域中的研究热点，用于检测数字图像和视频中特定类别的语义对象，其在视频安防、自动驾驶、交通监控、无人机场景分析和机器人视觉等领域有广阔的应用前景。在这



一节将会对目标检测技术的内容和发展进行概述。

### 2.2.1 神经网络的发展

深度学习模型可以看作是具有深度结构的神经网络。深度学习主要是通过搭建人工神经网络（Artificial Neural Network, ANN）<sup>[14]</sup>来进行知识的学习。1943 年的 MAP（McCulloch and Pitts）人工神经元网络提出，最初的目的是模拟人的大脑系统去解决一般的学习问题。神经网络的输入数据通常较为复杂，规模大，维度高。反向传播算法<sup>[15]</sup>提出之后，神经网络算法逐渐流行起来。但是由于训练数据有限、过度拟合、计算能力不足以及与其他机器学习工具相比性能的不足等缺点，神经网络算法研究一度趋于冷淡。自 2006 年语音识别技术<sup>[16]</sup>的突破重新使深度学习进入发展的热潮。

深度学习的重新流行依赖于大数据、GPU 以及模型三个因素。为了得到高性能的深度学习模型，在模型训练时通常需要上万甚至十几万的有效标注数据。深度学习模型的参数量巨大，传统的 CPU 难以胜任，随着硬件的发展，GPU 以及 CUDA 计算库专注于数据并行计算，成为模型快速训练的强有力的工具。在大数据与 GPU 的支撑下，网络结构和训练策略设计方面都产生了重大进步，催生出 AlexNet<sup>[17]</sup>、VGGNet<sup>[18]</sup>、ResNet<sup>[19]</sup>等一系列优秀的深度学习模型，并且显著提高了在学习的精度及速度。

卷积神经网络（CNN）<sup>[20]</sup>是深度学习最具代表性的模型。尽管当前的目标检测算法具有各自的特点，但是一般都是先使用 CNN 对输入图像进行处理，并从中抽取出特征，从而得到更深层次的特征图，之后通过多种算法来实现区域生成和损失计算。卷积层，池化层和全连接层 CNN 的三种主要类型的层。

- (1) 卷积层：对数据用卷积核参数来进行特征抽取，再利用矩阵点乘与求和得到结果，卷积运算过程如图 2-4 所示，公式表示为  $y = \omega x + b$ 。输入  $1 \times 3 \times 3$  的特征图(x)，卷积核( $\omega$ )大小为  $2 \times 2$ ，偏置(b)为 1，按照从左到右从上到下的顺序依次将对应矩阵元素相乘相加，得到卷积结果(y)。

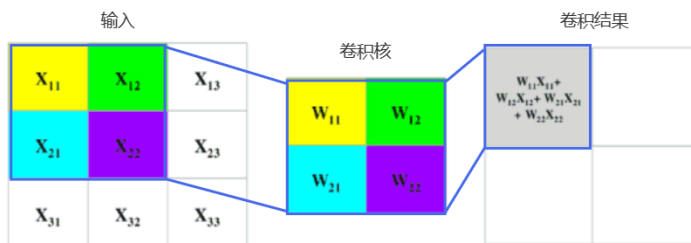


图 2-4 卷积计算的基本过程

- (2) 池化层：卷积层之间增加池化层（Pooling），减少了参数量，计算速度能够得到提升，同时能一定程度上防止过拟合。常用的方式有最大池化、平均池化、全局平均池化，全局最大池化。如图 2-5 是一个最大池化的过程，对  $4 \times 4$  的输入分割成 4 块区域，每块区域取其最大元素值，减少了数据量。



图 2-5 最大池化过程示例

- (3) 全连接层：全连接层通常放在输出层之前，由权值和偏差以及神经元组成，每一个节点连接上下层的所有节点，输入输出延展为一维向量，将卷积抽象出的特征图进一步映射到特定维度的标签空间。

除了上述三种层级，可以根据不同的任务，添加用来提高神经网络表达能力的激活函数（Activation Function）和用来度量模型预测效果的损失函数（Loss Function）等。

与传统方法相比，CNN 具有以下优势：(1)通过分层多级结构<sup>[21]</sup>可以学习到从像素到高级语义特征的多级信息；(2)网络结构更深，表达能力成倍增加；(3)其架构为多任务学习提供了可能。

如图 2-6 所示展示了目标检测的发展历程，其中标记框内为单阶段目标检测算法。可以看出图像分类算法贯穿着目标检测算法的始终，两阶段算法在前期占据主导，后期单阶段目标检测算法蓬勃发展，这是因为图像分类算法和两阶段目标检测算法中回归分析方法对于单阶段目标检测算法的发展起到了促进作用。

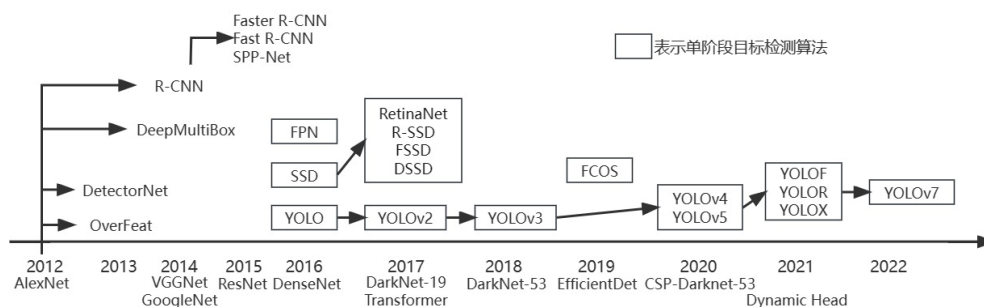


图 2-6 目标检测算法的发展

## 2.2.2 YOLO 模型概述

YOLO 由 Joseph Redmon 在 2016 年提出，核心思想是输入整张图片之后，利用分而治之的思想，对图片进行网格划分，每个网格负责预测中心位置落在该网格的目标物体，然后直接在输出层产生预测结果，包括边界框的检测位置及其所属的类别。与同期的 SSD 算法相比，在 YOLOv3 时期与 SSD 算法性能不相上下，但是随后提出的在性能上已经远远超过 SSD 算法<sup>[22]</sup>。

YOLOv1 使用单一卷积层和全连接层，直接检测整幅图像，检测速度可以达到实时检测的要求，但是由于检测精度低难以胜任小目标检测任务。YOLOv2<sup>[23]</sup>的主干网络模型改进为 Darknet-19，增加两倍特征图的通道数，使用全局平均池化层进行结果预测，并将目标分类与检测数据集结合起来进行联合训练。相比 YOLOv1 模型，YOLOv2 主要在计算量上减少了约 33%。

YOLOv3<sup>[24]</sup>算法骨干网络采用了 Darknet-53 结构，加深网络层、使用残差网络模块、使用金字塔池化，并使用适用于多标签多分类的 logistic 层，多尺度方面采用 Multi-Scales 策略，实现了在不影响速度的情况下，大幅度提高了检测精度，同时在保持图像分辨率不变的前提下，显著提高了对小目标的检测效果。

YOLOv4<sup>[25]</sup>的输入端使用了 Mosaic 数据增强，如图 2-7 所示，四张图片缩放后裁剪拼接，调整检测框的坐标位置，处理图像边缘检测框，可以增强模型对样本目标不均衡问题的鲁棒性。此外，主干网络升级为 CSP-Darknet-53，主要解决了模型卷积过程中出现的重复问题。激活函数具有更加平滑的梯度为 Mish 函数，保持模型准确度。



图 2-7 Mosaic 数据增强

Glenn 等提出了 YOLOv5，在 YOLOv4 的基础上，输入端的图片处理增加自适应锚框计算和自适

应图片缩放,提高了算法速度。在骨干网络中引入 Focus 结构,减少模型参数量,提高模型利用率。并将 CSP 结构运用到了 Neck 网络中,加强网络特征融合能力。

2021 年的 YOLOX<sup>[26]</sup>使用解耦的检测模型,模型采用平行分支分别检测目标位置、所属类别和交并比,并使用了新的 simOTA 策略区分正负样本的候选。

目前,YOLO 系列原作者也发布了 YOLOv7<sup>[27]</sup>算法。YOLOv7 主要以 YOLOv5 和 YOLOv4 为基础,提出了更加高效的网络结构,扩展了效率层网络,对模型进行缩放,并且对模型重参数化。训练方式进行了改进,用辅助监督模块进行深度监督,引入了引导头和辅助头,达到了在不影响推理的时间的同时,通过增加训练成本提升精度的效果。引入了动态标签分配策略,采用 YOLOv5 的跨网格搜索以及 YOLOX 的匹配策略。与之前的 YOLO 版本相比,YOLOv7 在速度和精度上超越了以往的 YOLO 系列,因此本文选择 YOLOv7 模型进行目标检测的实验。

### 2.2.3 基于深度学习的 YOLOv7 算法

YOLOv7 模型是由 Chien-Yao Wang 等人在 2022 年 7 月提出的一种新型网络架构模型,并且提出了适用于边缘 GPU、常规 GPU、云 GPU 的多个细化模型。在 5~60FPS 的范围内,速度和精度都超越了目前主流检测器。其网络结构主要由输入端 (Input)、主干特征提取网络(Backbone)、特征加强网络 (Head)、预测端(Prediction)构成,原始 YOLOv7 算法结构如图 2-8 所示。

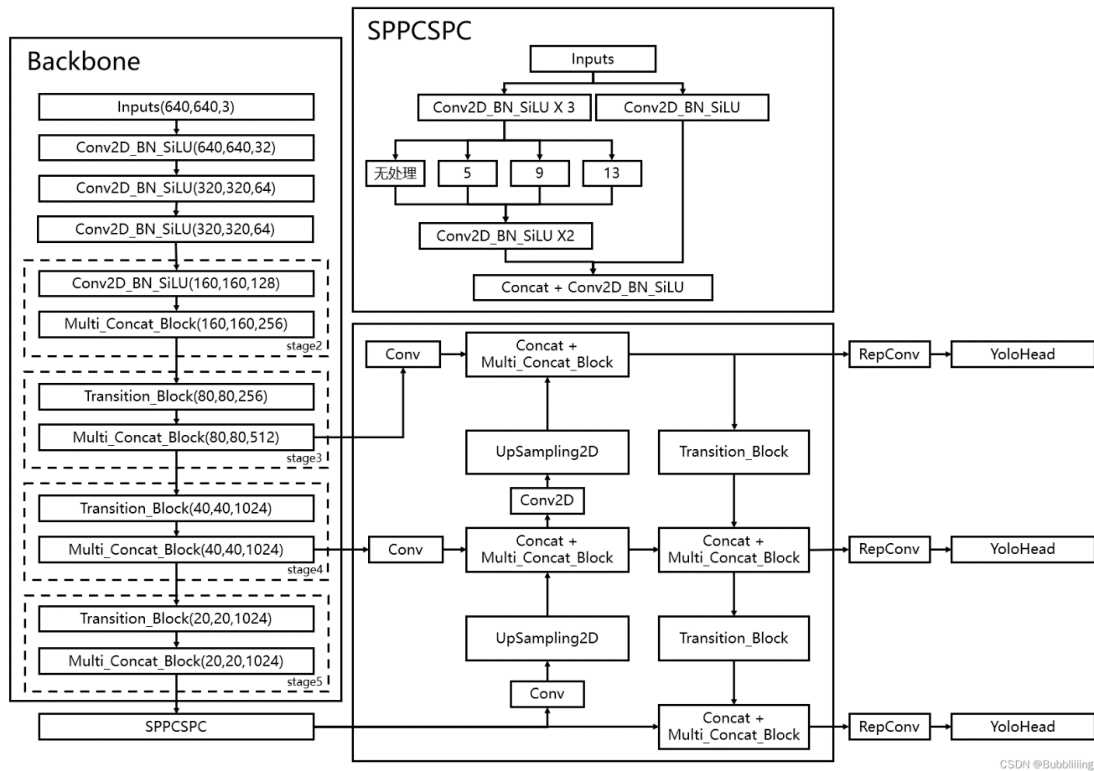


图 2-8 YOLOv7 网络结构图

输入端的作用是将输入图像缩放至指定的尺寸,并且实现数据增强。主干特征提取网络加入了 ELAN 卷积层、MPConv 卷积层、Conat 层。在 Head 模块,主要由路径聚合特征金字塔网络(Path Aggregation Feature Pyramid Network, PAFPN)结构<sup>[28]</sup>组成,采用上采样 (Upsample)、E-ELAN 卷积层、SPPCSPC 模块以及 Concat 缩放结构的方式,实现不同层次特征的高效融合。在预测阶段,通过 REP 模型重参数化结果调整特征通道数,最后经过卷积预测置信度、类别和锚框,取得预测结果。

接下来对 YOLOv7 版本的部分重点结构加以介绍。

- (1) 高效聚合网络 ELAN 与扩展的 E-ELAN: ELAN 和 E-ELAN 网络结构实际上是用多模块堆叠的方式整合特征, 用来加强特征提取。ELAN 模块包含多层由 Conv+BN+SiLU 构成的 CBS 层, 可以加深网络深度, 提升网络精度, 并使用残差网络缓解梯度消失问题。E-ELAN 网络通过扩张、变换、增加特征基数(Cardinality), 可以增强对特征的提取能力, 改进参数的使用和计算效率, 在不破坏原始梯度路径的情况下, 能提高网络的学习能力并在推理阶段节省大量的计算资源。
- (2) 模型重参数化: YOLOv7 提出了计划的重参数化卷积的概念, 将重参数化卷积<sup>[29]</sup>运用到了残差模块或用于拼接的模块, 对每个  $3 \times 3$  卷积, 在训练时构造  $1 \times 1$  卷积分支和并行的恒等连接, 但在训练完成之后的推理阶段, 重参数化成一个相对简单的结构。利用参数转换解耦训练和推理结构, 在训练过程中丰富了训练资源, 使网络能够获取更多信息, 不但能提高检测精度, 还能提升网络推理速度。
- (3) 辅助头检测及动态标签分配策略: YOLOv7 将深度监督 (Deep Supervision) 运用到了辅助检测头模块, 给中间部分增加辅助检测模块, 参与损失函数的计算, 并反向传播协助前面的卷积层更新参数。YOLOv7 采用的是软标签 (Soft Label) 分配方式, Head 部分产生的预测框和标注框一起传到分配器 (Assigner) 后才会得到目标值, 产生软标签, 并且将目标值和预测值一起传递到损失函数, 如图 2-9 所示:

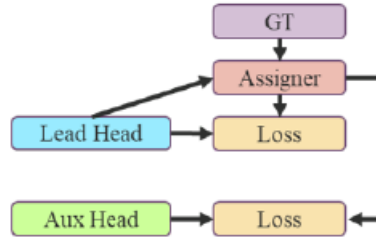


图 2-9 软标签分配策略

- (4) 损失函数配置: 在网络中目标定位依赖于边界框回归模块, 而 IoU 损失函数的功能就是使预测框接近正确目标, 进而提升目标框的定位效果。YOLOv7 模型的损失函数具体如公式 (1) 所示:

$$L_{object,loss} = L_{loc,Loss} + L_{conf,Loss} + L_{class,Loss} \quad (2-1)$$

其中,  $L_{loc,Loss}$  表示定位损失;  $L_{conf,Loss}$  表示置信度损失,  $L_{class,Loss}$  表示分类损失。置信度损失和分类损失采用 BCEWithLogitsLoss 函数进行计算, 定位损失采用 CIoU 进行计算, 计算公式如 (2-4) 所示。公式中  $b$  为预测框,  $b_{gt}$  为真实框,  $c$  为两个框的最小包围区域对角线距离。  $\alpha$  为平衡参数,  $v$  用来衡量长宽比一致性。CIoU 损失函数将边界框回归的重叠面积、中心点距离、纵横比全都考虑进去了, 提高了检测准确率。

$$L_{CIoU} = 1 - I_{IoU} + \frac{P^2(b, b_{gt})}{c^2} + \alpha v \quad (2-2)$$

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

$$\alpha = \frac{v}{(1 - I_{IoU}) + v} \quad (2-4)$$

### 2.3 目标检测评价指标

对于一个检测器，需要制定一定的规则来评价其好坏。由于目标检测模型的输出是非结构化的，输出目标的大小、位置、数量等信息是未知的，所以对于某个目标来讲，可以从预测框与真实框的贴合程度来判断检测的质量。通常使用交并比（Intersection-over-Union, IoU）来量化贴合程度。IoU 的计算方式如图 2-10 所示。

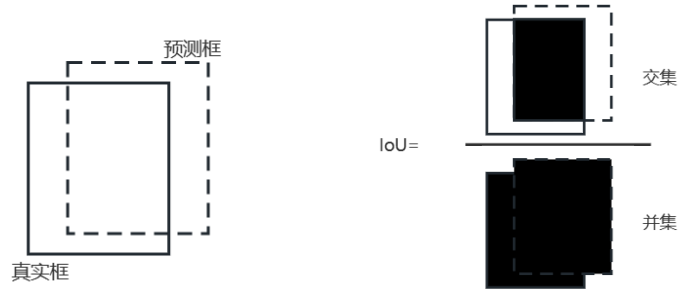


图 2-10 交并比的计算过程

这一概念源于数学，用来描述两集合 A 和 B 间关系。IoU 值的计算方式为集合交集的元素个数比上并集中的元素个数。具体公式如（5）所示。

$$IoU = \frac{A \cap B}{A \cup B} \quad (2-5)$$

IoU 的取值区间是[0,1]，IoU 值越大，表示两个框重合程度越高。通常会选取一个阈值来确定预测框是否正确，一般设为 0.5，当两个框的 IoU 大于阈值时，认为其是一个有效的检测。

对于一个检测器来说，通常用平均精度 mAP(mean Average Precision)来评价模型的性能。AP 指某个类别的精度，mAP 则是多个类别的平均精度。通过 IoU 值和阈值的比较，可以把预测框分为四种：正确检测框 TP，误检框 FP，漏检框 FN 和正确背景。在检测中通过计算可以得到模型的召回率（Recall, R）以及准确率（Precision, P），公式如式（6）和式（7）所示

$$R = \frac{TP}{TP + FN} \quad (2-6)$$

$$P = \frac{TP}{TP + FP} \quad (2-7)$$

对于每次遍历生成的 P 和 R 可以绘制成 P-R 曲线，如图 2-11。

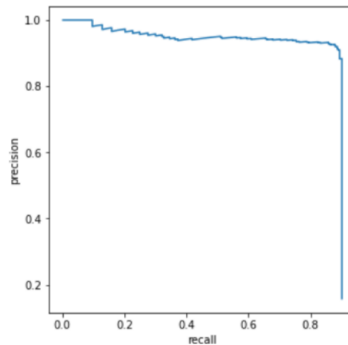


图 2-11 目标检测的 P-R 曲线

但由于评价模型是一个非线性的结构，如果直接取曲线上的点，可能会遇到召回率高但准确率



低，或者准确率高但召回率低的情况，因此引入了 AP 的概念。AP 的计算公式如式（8）所示：

$$AP = \int_0^1 P dR \quad (2-8)$$

可以看出 AP 代表了曲线的面积，能对不同召回率下的准确率综合考量。每个类别的 AP 值相互独立，在多目标检测任务中，将每个类别的 AP 求平均得到 mAP 能更好的评价模型的检测精度。

## 2.4 边界框回归

目标检测需要解决的问题就是预测图像或视频中目标物体的定位和类别。目标检测算法使用边界框回归模块进行目标的定位，而 IoU 损失函数的作用就是计算预测框与目标物体之间的差距，使得预测框能够接近真实框，提高目标框定位的正确性。此时，就需要利用边界框回归（Bounding-Box Regression）<sup>[7]</sup>来预测目标物体的检测框，确定预测框内是否包含真实物体。检测的窗口一般使用四维向量(x,y,w,h)来记录中心点坐标和宽高。如图 2-12，红色框 P 是初始候选框（Proposal），绿色框 G 代表真实标注框（Ground Truth）。边界框回归就是要利用某种映射关系，使候选框 P 能接近真实框 G，获得接近真实框 G 的回归窗口  $\hat{G}$ 。

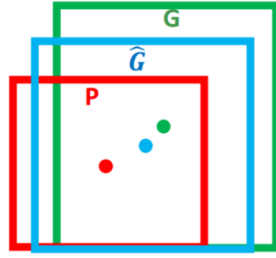


图 2-12 边界框回归

也就是说，对于给定的  $P=(P_x, P_y, P_w, P_h)$ ，需要寻找一映射关系  $f$ ，满足公式（9-10）的关系：

$$f(P_x, P_y, P_w, P_h) = (\hat{G}_x, \hat{G}_y, \hat{G}_w, \hat{G}_h) \approx (G_x, G_y, G_w, G_h) \quad (2-9)$$

$$(\hat{G}_x, \hat{G}_y, \hat{G}_w, \hat{G}_h) \approx (G_x, G_y, G_w, G_h) \quad (2-10)$$

窗口 P 向窗口  $\hat{G}$  的转换，经过了平移和尺度放缩的过程：

（1）先做平移变换：

$$\begin{aligned} \hat{G}_x &= P_w d_x(P) + P_x \\ \hat{G}_y &= P_h d_y(P) + P_y \end{aligned} \quad (2-11)$$

（2）再做尺度缩放：

$$\begin{aligned} \hat{G}_w &= P_w \exp(d_w(P)) \\ \hat{G}_h &= P_h \exp(d_h(P)) \end{aligned} \quad (2-12)$$

根据公式可以发现，边界框回归就是  $d_x(P), d_y(P), d_w(P), d_h(P)$  的四个变换，算法中最重要的即是这四个映射。当 P 和 G 较接近时，也就是 IoU 较大时，可以将这种映射视为线性回归问题。线性回归满足关系  $Y \approx WX$ ，即对于给定的输入特征向量 X，学习一组参数 W，使其线性回归后的值跟真实值 Y 非常接近。随着网络训练的逐渐收敛，这种线性关系也能一直得到满足。

## 2.5 Pytorch 深度学习框架

PyTorch 的前身是诞生于 2002 年的 Torch 框架，并且由于在计算科学领域，Python 由于其简单易用与完整的生态具有优势。于是在此基础上 Facebook 团重构了 Torch 张量上的所有模块，增加自动求导的理念，实现了高效的动态图框架。除此之外还有几种主流的深度学习框架，比如谷歌大脑团队在 2015 年发布 TensorFlow 开源框架，由 DMLC 组织创建的 MXNet 框架，以及在多个框架之上建立的 Keras 框架等。

相比其它主流框架，PyTorch 具有其优势：（1）代码风格较为简洁，包含自动求导机制；（2）用户体验较好，接口易用，更容易上手；（3）模型速度快（4）发展趋势迅猛。

## 2.6 本章小结

本章对研究将要涉及到的基本原理进行了简述。通过对光学遥感图像的介绍，引入了实验中处理的图像类型。通过对深度学习基本理论的介绍展示了目标检测研究所运用的方法和手段。接着对 YOLO 模型已经 YOLOv7 模型的概况进行了简单的描述，为接下来搭建模型进行仿真实验提供理论基础。

### 第三章 实验过程与结果分析

#### 3.1 数据集简介与构建

RSOD-Datatest 由武汉大学于 2015 年发布，它是一个开放数据集，用于遥感图像目标检测，主要数据来源是 Google Earth 等平台，数据集包括飞机、油箱、操场和立交桥四类对象，如图 3-1 数据集样例图所示。数据集中共计有 976 张遥感图像，图像大小在  $512 \times 512$ - $1083 \times 923$  像素之间。图像中的实例对象具体分布如表 3-1 所示。



图 3-1 数据集样例图

图像中四种类别的实例对象具体分布如表 3-1 所示。

表 3-1 数据集详细信息

	aircraft	oiltank	overpass	playground
图像数量	446	165	176	189
目标数量	4993	1586	180	191

RSOD 数据集已经将图片中实例目标的标注信息划分好，存放于 Annotations\xml\文件夹下的 xml 后缀文件中，文件中存储了图片的名称、图片的宽高以及目标类别名称和目标框左上角坐标和右下角的坐标信息。具体 xml 标签文件的内容如图 3-2 所示，在模型训练时，将会读取 object 的信息，即目标类别名称是 aircraft，坐标为 (363,36) 和 (96,468)。

```
<annotation>
  <source>
    <image>optic rs image</image>
    <annotation>Lmars RSDS2016</annotation>
    <flickrid>0</flickrid>
    <database>Lmars Detection Dataset of RS</database>
  </source>
  <object>
    <bndbox>
      <xmin>363</xmin>
      <ymin>36</ymin>
      <ymax>96</ymax>
      <xmax>468</xmax>
    </bndbox>
    <difficult>0</difficult>
    <pose>Left</pose>
    <name>aircraft</name>
    <truncated>1</truncated>
  </object>
</annotation>
```

图 3-2 标签文件信息



### 3.1.1 数据集格式组织

为了试验方便，需要将原始数据集中转换为 PASCAL VOC 文件格式进行处理，在 Annotations 文件夹存放 xml 标签文件，ImageSets\Main 文件夹下存放记录训练集、验证集、测试集划分的 txt 文件，JPEGImages 文件夹存放用于训练的图片文件。但是由于原始数据集 playground 类别中的图片有 189 张，而标注只有 149 张，因此对缺少标注文件的 40 张图片需要进行手动标注，产生对应的 xml 标签。标注工具使用 LabelImg，标注框的可视化界面如图 3-3 所示。

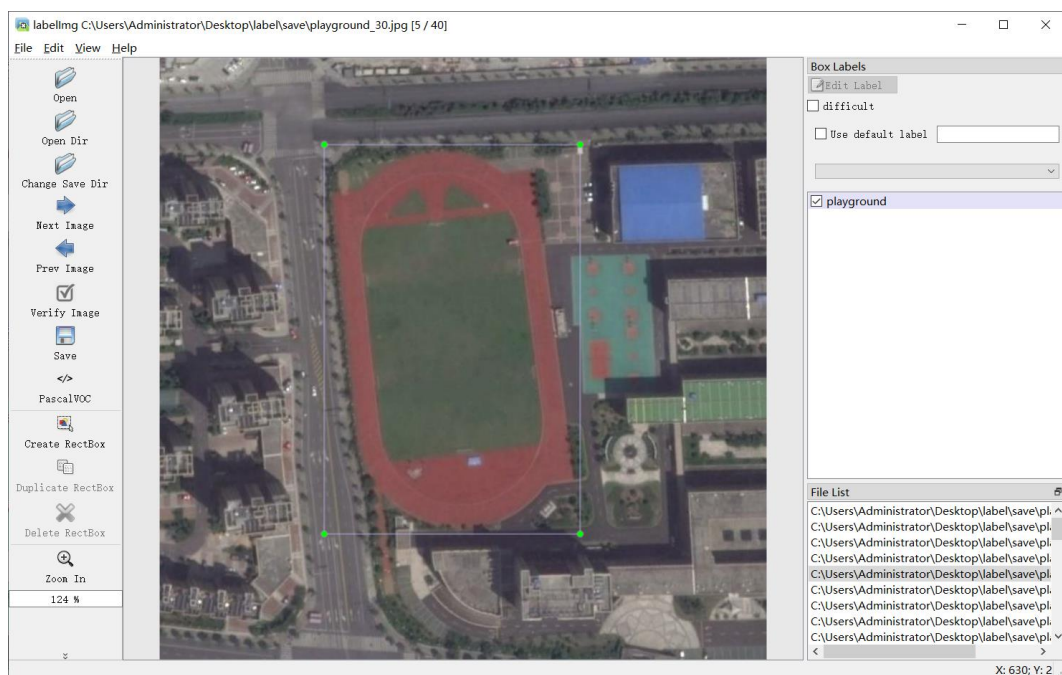


图 3-3 LabelImg 标注界面

### 3.1.2 添加噪声

为了比较模型在不同噪声条件下的检测性能，本论文对原数据集中的每张图片添加高斯噪声，同样按照 VOC 格式组建两个新的数据集。使用 Python 的 opencv 和 numpy 库编写代码，设置添加的高斯噪声的标准差 ( $\sigma$ ) 分别为 0.15 和 0.30。高斯噪声是由随机变量服从高斯分布产生的，标准差是描述高斯分布的一个重要参数。它与评价图片质量的常用指标信噪比 (SNR) 之间存在如公式 (13) 的关系。S 表示图片的信号强度， $\sigma$  表示标准差值。SNR 表示图像信号与噪声信号的比值。SNR 越大，说明图像中信号占据了更大的比例，图像的质量也就越好。因此，在图像信号强度 S 不变的情况下，标准差值  $\sigma$  越小，SNR 越高，图像的质量也就越好。

$$SNR = 10 \log \left( \frac{S}{\sigma^2} \right) \quad (3-1)$$

添加高斯噪声之后的效果如图 3-4 (a)-(c)所示，标准差越大，说明高斯噪声的强度越大，对图像的影响也越明显，也就是说， $\sigma = 0.30$  的图像质量低于  $\sigma = 0.15$  的图像质量。



图 3-4 添加噪声前后的图像效果对比

将变化后的图片和对应的 xml 文件批量重命名，使用 LabelImg 查看标注框是否正确。检查后将变化后的图片及标注文件对应是否正确，分别保存，建立初始数据集、 $\sigma = 0.15$  及  $\sigma = 0.30$  的三个数据集，用于模型的训练。

在训练时，按照 9:1 的比例，对数据集图片进行划分，得到训练集 790 张，验证集 88 张，测试集 98 张。

### 3.2 实验环境配置

YOLOv7 算法需要强大的计算力作为支撑。具体实验的相关配置如表 3-2 所示：

表 3-2 实验环境

名称	配置
CPU	12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50GHz
GPU	RTX 3080(10GB)*1
加速环境	CUDA 11.0
操作系统	ubuntu18.04
深度学习框架	Pytorch1.7.0
语言	Python3.8

### 3.3 实验流程

在准备完数据集和搭建好 Pytorch 框架下的 YOLOv7 模型后，开始进行 YOLOv7 光学遥感图像目标检测的具体实验，实验步骤如下：

(1) 配置实验环境。配置 Anaconda+pytorch+CUDA 环境。为保证训练效果和质量，提高实验配置。本课题选择在 AutoDL 平台租用云 GPU，并且在 VSCode 上使用 SSH 远程登录。

(2) 进行标签处理。新建 cls\_classess.txt 文件，键入需要区分的类别，分别是 aircraft、oiltank、overpass、playground。运行标签处理程序，将数据集划分为训练集 train.txt、验证集 val.txt 和测试集 test.txt，并且在根目录下产生记录训练集、验证集图片路径和标签的 txt 文件。

(3) 修改 train.py 的参数。设置使用 CUDA 进行训练，不使用单机多卡模式和混合精度训练。修改类别路径为上一步建立的 cls\_classess.txt 文件路径。为加快网络的收敛速度，为算法模型加载了预训

练权重 yolov7\_weights.pth。使用 Mosaic 数据增强、MixUp 数据增强，设置前 70% 个 epoch 开启数据增强。

(4) 模型优化方法采用随机梯度下降法 (SGD)，并设置权值衰减参数 (weight\_decay) 为  $5e-4$ 。调整并优化网络训练参数，模型配置参数如表 3-3 所示。

表 3-3 模型配置参数

参数名	参数设置
weights	yolov7_weights.pth
eposhes	300
workers	8
batch-size	8
max learningrate	0.01
confidence	0.5

(5) 将学习率下降方式设置为 cos，即余弦退火学习率下降方式。将网络的预测结果传入至损失函数中计算损失，实验训练过程中，每进行 10 次迭代将会进行评估，权值保存在 logs 文件夹下。

(6) 将训练后的模型运用于测试集上，计算不同目标上的召回率精确率等指标。使用训练时保存下来的权重文件，对图片进行预测，查看图片预测结果。

(7) 将上述步骤重复两次，训练添加噪声后的数据集，获取召回率精确率等指标，并与初始数据集的结果对比分析。

根据实验步骤设计实验流程如图 3-5 所示。

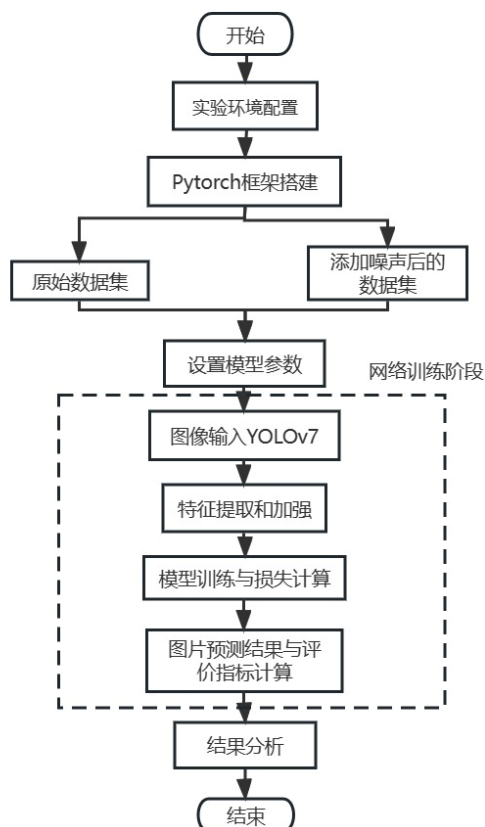


图 3-5 实验的流程

### 3.4 实验结果对比及评估

根据添加的噪声，在同样的实验环境下进行三次实验。对三次实验分别获取在模型在 aircraft、oiltank、overpass、playground 四种类别的图像预测结果，为了评估模型的检测效果，计算三次实验的效率指标。本文采用召回率（Recall）和平均精确度（AP）对实验结果进行评价，两者的定义在 2.3 节中已有阐述。

#### 3.4.1 原始数据集

原始数据集上的检测结果如图 3-6 (a)-(d)所示。

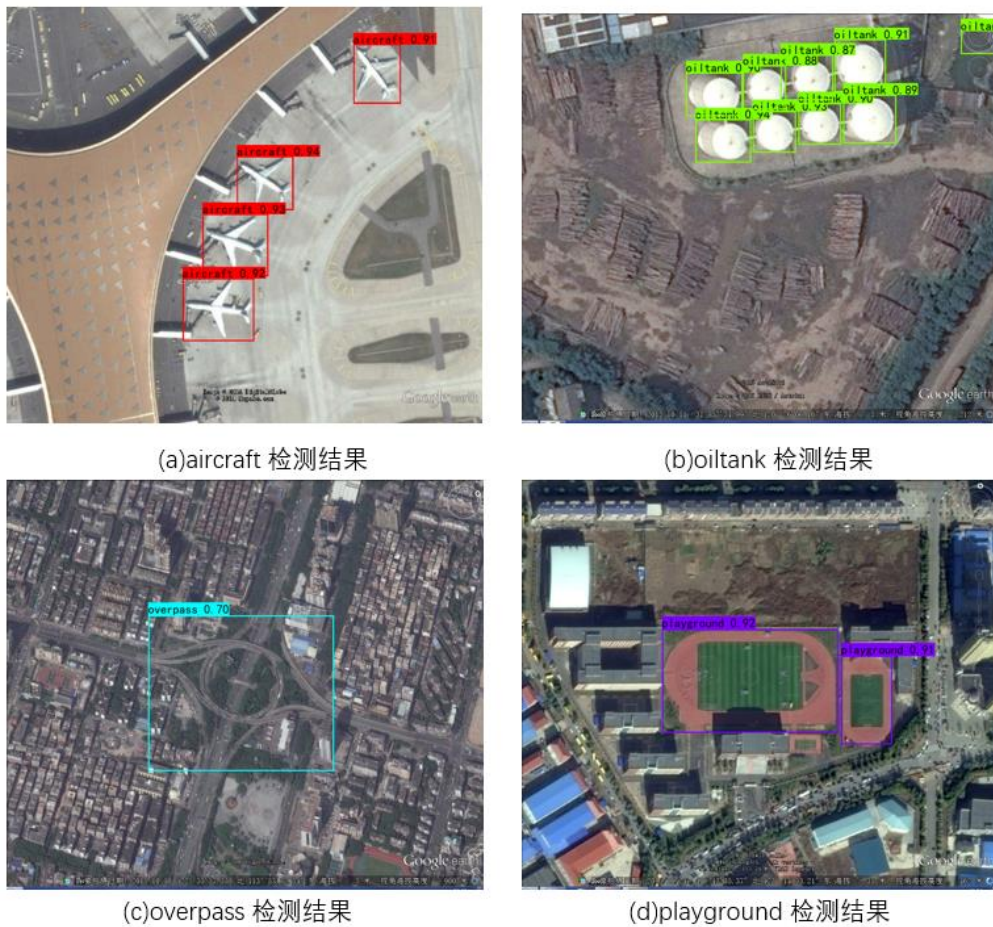


图 3-6 原数据集目标预测效果

取得检测评价指标如表 3-4 所示：

表 3-4 原数据集检测结果评价指标

	AP/%	Recall/%
aircraft	96.96	89.91
oiltank	97.63	100.00
overpass	95.65	86.67
playground	96.70	98.70
mAP/%	96.73	



### 3.4.2 $\sigma=0.15$ 的数据集

$\sigma=0.15$  的数据集上的检测结果如图 3-7 (a)-(d)所示。

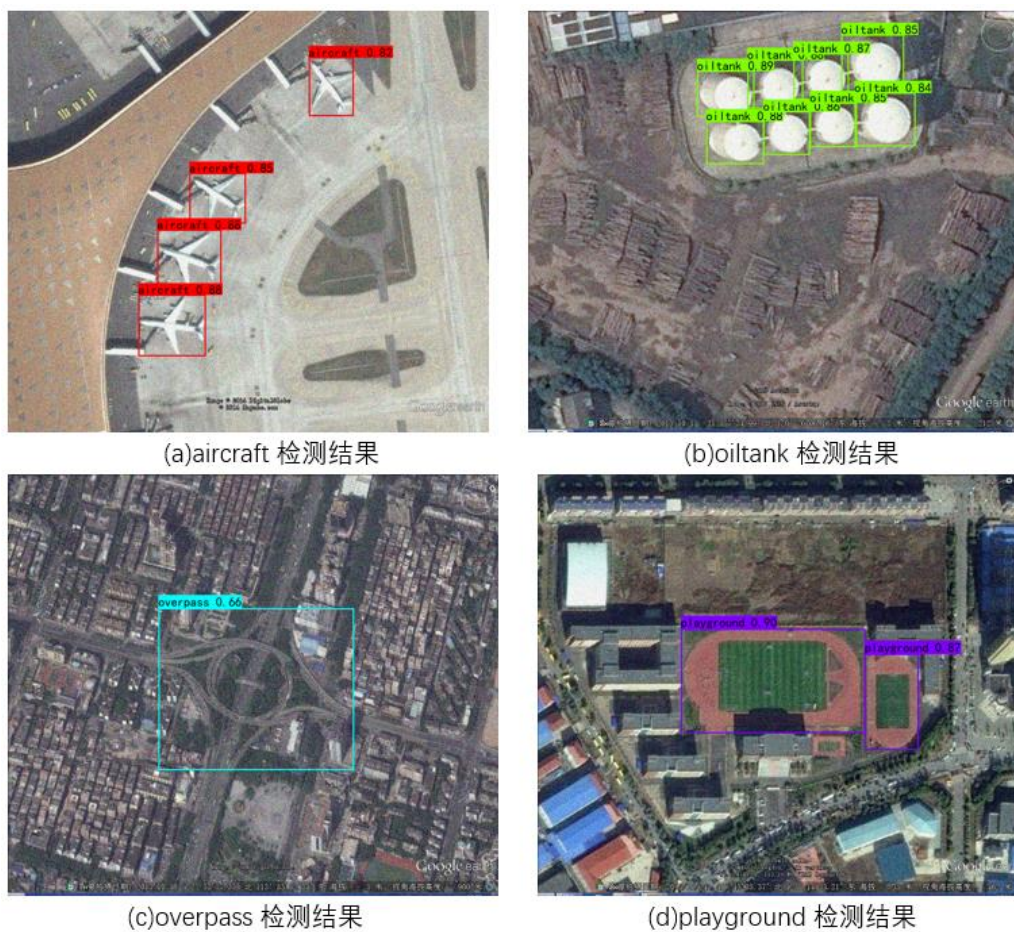


图 3-7  $\sigma=0.15$  的数据集目标预测效果

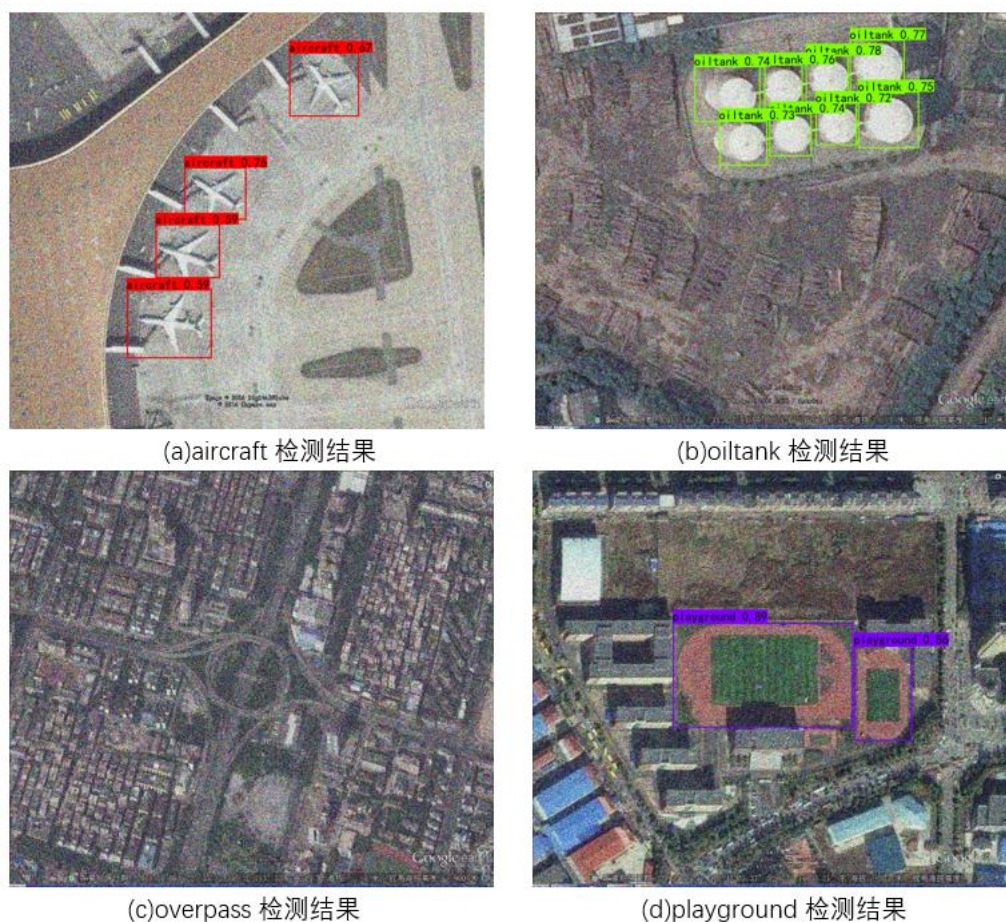
取得检测评价指标如表 3-5 所示：

表 3-5  $\sigma=0.15$  的数据集在不同目标上的性能

	AP/%	Recall/%
aircraft	94.17	84.15
oiltank	96.19	96.93
overpass	94.74	89.38
playground	95.02	93.33
mAP/%	95.03	

### 3.4.3 $\sigma=0.30$ 的数据集

$\sigma=0.30$  的数据集上的检测结果如图 3-8 (a)-(d)所示。

图 3-8  $\sigma=0.30$  的数据集目标预测效果

取得检测评价指标如表 3-6 所示:

表 3-6  $\sigma=0.30$  的数据集在不同目标上的性能

	AP/%	Recall/%
aircraft	91.78	86.86
oiltank	94.66	96.02
overpass	90.02	83.33
playground	92.06	95.65
mAP/%	92.13	

#### 3.4.4 结果对比及分析

根据原始数据集预测结果可以看到, 经过训练后的模型在 aircraft、oiltank、overpass、playground 四种类别上都取得了较好的检测效果, 平均精确度分别达到了 96.96%、97.63%、95.65%、96.70%, mAP 达到了 96.73%, 模型具有较高的检测性能。并且四种类别的预测框的置信度处于较高水平, 说明模型对于目标物体的检测能力较强, 模型的准确性和鲁棒性都比较好。但是在图 3-6 (b) oiltank 的预测结果中, 图片右上角存在对非目标物体的检测框, 出现了误检的情况, 说明模型还不能很好地区分目标物体与背景之间的特征差异, 在目标物体的背景信息的特征提取上还可以加以改进。

在  $\sigma=0.15$  的数据集上, aircraft、oiltank、overpass、playground 四类目标的平均精确度分别是 94.17%、96.19%、94.74%、95.02%, 相较原始数据集有不同程度的下降, mAP 相对于原始数据集下降了 1.7%, 预测框置信度同样降低, 说明对图片添加噪声会影响模型的检测性能, 使得模型检测性能降低。



在  $\sigma=0.30$  的数据集上, aircraft、oiltank、overpass、playground 四类目标的平均精确度分别是 91.78%、94.66%、90.02%、92.06%, 模型的 mAP 相对  $\sigma=0.15$  的数据集又降低了 2.9%, 但是总体性能还是处于较好的水平。在图 3-8 (c) overpass 的预测结果中, 由于检测框置信度小于设置好的置信度阈值(confidence)0.5, 算法将检测框划分为了误检框, 已经无法在输出图片上显示出预测框, 出现了漏检的情况。若将预测时的置信度阈值设置为 0.3, 可以看到算法将 overpass 目标重新预测出来了。如图 3-9(a)-(b) 所示是在不同置信度阈值下模型对同一目标的预测情况, 可以看到此时目标的置信度是 0.48。因此在检测噪声高、图像质量低的图片时, 可以降低置信度阈值, 从而减少目标的漏检。



图 3-9 不同置信度阈值下对同一目标的预测情况

为了验证 YOLOv7 算法的优越性, 本文又使用双阶段检测算法 Faster R-CNN、单阶段检测算法 SSD、YOLOv5 做比较。对比试验在原数据集上进行, 比较平均精度 mAP, 实验的结果如表 3-7 所示。从对比结果可以看出, YOLOv7 模型的各个类别的精度值均有所提升。虽然 YOLOv7 算法和其他算法相比, 并不能保证各个类别的精度值均是最高, 但是这不影响总体的平均精确度高于其他算法, 相较于 Faster R-CNN、SSD、YOLOv5, YOLOv7 模型的 mAP 值分别提高了 16.25%、20.37%、13.09%。YOLOv7 算法可以在保证较高精确度的同时达到较高的速度, 在遥感图像目标检测上具有一定的优越性。

表 3-7 不同检测算法对比实验结果

Model	AP/%				mAP/%
	aircraft	oiltank	overpass	playground	
Faster R-CNN	63.16	84.13	76.87	97.75	80.48
SSD	52.13	96.64	56.67	100.00	76.36
YOLOv5	89.75	79.48	71.32	94.00	83.64
YOLOv7	96.96	97.63	95.65	96.70	96.73

此外, 通过进一步测试可以发现, 在包含多目标的复合场景下, 也会出现漏检的情况。如图 3-10 (a) 所示, 图片含有 aircraft 和 oiltank 两种目标, 部分 oiltank 目标未被算法检测出来。但若修改权值文件, 便可以减少漏检, 如图 3-10 (b) 所示。说明在周边环境复杂、目标较小的情况下检测, 检测精度不够高。



图 3-10 多目标情况的预测结果

对于模型存在误检、漏检情况，无法对目标物体进行精确检测的问题，可以考虑以下原因：(1) 模型的训练数据不足。ROSD 数据集的图片数量太少，训练数据不够丰富不能很好地通过训练区分目标与背景之间的差异，导致模型鲁棒性低；(2) 遥感图像自身特性。图像成像环境复杂，背景信息丰富，且遥感图像中目标的方向多变，导致在自然场景下水平框检测不太适用于遥感图像。在后续的研究中可以从对这两方面进行改进和优化。

### 3.5 本章小结

本章以 YOLOv7 网络为基础，对光学遥感图像目标检测进行了具体的实验探究。首先对实验使用带 ROSD 数据集的来源内容和构建进行了介绍，使用添加不同大小高斯噪声的方式组织了两个新的数据集。接着使用 Pytorch 框架对模型进行了搭建，将数据集分别输入搭建好的网络中，设置训练方案参数进行模型训练，进行了三次实验。最后展示三次实验的预测效果，使用包含单一目标以及复合目标的检测图片对模型检测效果进行对比，并通过对比精确率和召回率两个评价指标对模型的检测性能进行分析，得到了 YOLOv7 在光学遥感图像上能取得较高的检测性能，但在加入了噪声的图片上性能会降低，模型仍存在改进空间的结论。



## 第四章 总结与展望

本课题从遥感图像目标检测需求出发，在对当前国内外目标检测技术发展现状进行深入分析的基础上，确定研究内容和实验目标。接着对研究所涉及到的光学遥感图像特性、目标检测技术基本原理以及 YOLO 模型的发展进行了介绍，最后利用 Pytorch 框架对 YOLOv7 模型进行搭建并进行实验训练，取得实验结果并加以分析：

- (1) 选取并划分 RSOD 数据集，对训练集图片进行预处理操作，使得其满足实验的需要。
- (2) 创建实验所需要的硬件设备和环境配置。
- (3) 根据 YOLOv7 算法的基本结构，使用 Pytorch 框架搭建网络模型。
- (4) 开始模型训练，训练阶段结束后使用保存下来的训练权重，选取测试集图片对每个模型训练的结果进行检测，分析不同对象在模型中的预测效果。
- (5) 将训练集传入网络，通过损失函数的计算获取平均检测精度和召回率评价指标。对训练模型的性能进行评估和分析，得到了 YOLOv7 在光学遥感图像上检测性能较好，但在加入了噪声的图片上性能会降低，模型仍存在改进的空间结论。

本文将 YOLOv7 模型运用于光学遥感目标检测，与其他检测算法相比有较好的精度，具有一定的优势。但本文仍存在不足之处，在之后的研究中，可以从以下几个方面加以改进：

- (1) YOLOv7 网络中大量使用常规卷积，导致检测速度在一定程度上的降低。之后可以使用更加适合遥感图像特征的卷积替换普通卷积，在保证精确率仍然较高的同时提高检测的速度。
- (2) 检测精度提升。一是在本文的实验中，仅对水平目标框做模型训练，但实际上许多目标处于倾斜状态，若能增加对倾斜目标框的训练，将能减少无效信息的干扰，使模型的学习效率更高；二是在检测时容易出现漏检和误检，多尺度目标上容易忽略显著的目标信息，可以使用多尺度特征图、改进损失函数等方式，提高多特征提取能力，提高模型的检测能力和适应性。
- (3) 数据增强和数据集扩充。为了使目标检测模型能够适用于图片质量差、噪声高的图像，可以选择旋转、缩放、改变亮度等图片处理方式进行数据增强，扩充原有的数据集，使模型学习到目标和背景的特征，增加模型对不同角度、尺度、光线条件等变化的鲁棒性，提高模型的泛化能力。

## 参考文献

- [1] 罗小波. 遥感图像智能分类及其应用[M]. 电子工业出版社, 2011.
- [2] 季顺平,秦昆,单杰.遥感:从图像获取到智能服务[J].科学世界,2016(12):88-91.
- [3] 童庆禧,孟庆岩,杨杭.遥感技术发展历程与未来展望[J].城市与减灾,2018,No.123(06):2-11.
- [4] NIE Guangtao, HUANG Hua. A survey on object detection technology in optical remote sensing images[J]. Acta Automatica Sinica, 2021, 47 (8): 1749-1768.
- [5] Zou Z, Shi Z, Guo Y, et al. Object detection in 20 years: A survey[J]. arXiv preprint arXiv:1905.05055, 2019.
- [6] Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision 60,2004,91–110.
- [7] 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. Lake City, USA, 2014: 580-587.
- [8] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE international Conference on Computer Vision. Santiago, Chile, 2015:1440-1448.
- [9] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-Time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 1137 - 1149.
- [10] 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.
- [11] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector[C]. European conference on computer vision. Springer, Cham, 2016: 21-37.
- [12] 陈鑫. 基于可见光遥感图像的典型目标自动检测技术研究[D].中国科学院大学(中国科学院长春光学精密机械与物理研究所),2022.DOI:10.27522/d.cnki.gkcg.2022.000162.
- [13] 赵坚,孟令杰,王琦,郭丁,申志强. 我国高分辨率对地观测系统建设与发展[J].卫星应用,2022(11):8-13.
- [14] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [15] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533–536.
- [16] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504–507.
- [17] Krizhevsky A, Sutskever I, Hinton GE. classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84–90.
- [18] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [19] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition[J]. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778.
- [20] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
- [21] Kavukcuoglu K, Sermanet P, Boureau YL, et al. Learning convolutional feature hierarchies for visual recognition. Proceedings of the 23rd International Conference on Neural Information Processing Systems[J]. Vancouver, Canada. 2010. 1090–1098.
- [22] 耿创,宋品德,曹立佳.YOLO 算法在目标检测中的研究进展[J].兵器装备工程学报,2022,43(09):162-

173.

[23] Joseph Redmon, Ali Farhadi. YOLO9000: Better, Faster, Stronger[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017,7263-7271.

[24] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[25] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection[EB/OL]. arXiv preprint arXiv:2004.10934, 2020.

[26] Zheng Ge, Songtao Liu, Feng Wang, et al. YOLOX: Exceeding YOLO Series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.

[27] Chien-Yao Wang, Alexey Bochkovskiy, Hong-Yuan Mark Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[EB/OL]. arXiv:2207.02696, 2022.

[28] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, et al. Feature pyramid networks for object detection[J]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, 21-26.

[29] Xie J, Miao Q, Liu R, et al. Attention adjacency matrix based graph convolutional networks for skeleton-based action recognition[J]. Neu-rocomputing, 2021, 440: 230-239.

## 致谢

经过这段时间的研究和学习，我终于完成了我的本科毕业论文。在这个过程中，从对陌生知识的学习到最终完成论文撰写，其间遇到了种种挫折，但最后收获成果那一刻，那种喜悦的心情是什么语言都描述不出来的。

在此，我真诚地感谢我的指导老师高强老师，他带领我深入接触了计算机视觉和人工智能，让我能够有幸探索这一热门领域的发展前沿。高老师平易近人，有着严谨的治学态度。在论文撰写过程中，高老师总是悉心指导，对我的问题不厌其烦，给予了我很大的帮助，论文的圆满完成离不开老师的谆谆教诲。其次，也要感谢本科四年传授知识的老师们，是他们的循循善诱，才让我学到了扎实的专业本领，对最后的毕业论文的学习奠定下了坚实的知识基础。再次，我也要对这个过程中对我提供帮助和鼓舞的同学表示感谢，在一同学习的过程中，他们刻苦的学习态度、坚毅的性格不断感染着我、激励着我。最后，感谢师院四年的悉心栽培，让我成为有能力报答社会的有志青年。

至此，四年的师院本科求学生涯将要落下帷幕。四年光阴转瞬即逝，新的生活画卷正向我徐徐展开，我将怀揣着对过往的感悟和收获，踏上通往未来的路途。