哈爾濱工業大學

人工智能数学基础
大项目结题报告

题　　目	基于 YOLOv8 的空对地目标检测
学　　院	计算机科学与技术
专　　业	人工智能
学　　号	2022113416 2022112018 2022112047
学　　生	刘子康 张文浩 张子鉴
任 课 教 师	刘绍辉

哈尔滨工业大学计算机科学与技术学院

2024.3

基于 YOLOv8 的空对地目标检测

一、 项目内容梳理及成员分工

1.1 项目应用前景

配备摄像头的无人机凭借自身成本低、灵活性高、操作简单、体积小等优点，已经被广泛应用于军事、农业、航空摄影、快速交付和监视等领域或场景。目标检测技术使得无人机具有更强的环境感知和识别对象能力，也因此取得了显著成果。当前无人机对采集视频的处理，主要通过数据链传输到地面站，这对数据链带宽要求很高，存在容易被干扰、同时传输视频路数有限、通信时延较长等缺点。

而考虑无人机机载嵌入式设备，在任务中直接对采集的视频进行处理的方法，可实现目标检测并将相关信息传输给地面指挥单元，该方法对传输带宽要求不高，可大大增加指控无人机数量，实现无人机集群侦察，同时降低通信时延，缩短从目标到作出反应的时间。相较于雷达、无线电等探测技术的应用，基于深度学习的视觉方法具备低成本、高效率等优点。因此，提高识别精度与提升训练、推理速度对无人机目标检测的研究具有重要意义。

图 1-1 无人机识别地面目标

1.2 项目研究内容及关键技术

本项目的研究内容是利用 VOLOv8 模型形成空对地目标检测系统。YOLO(You Only Look Once)，是一个实时目标检测模型和算法，用来在一张图篇

中找到某些特定的物体，目标检测不仅要求识别这些物体的种类，同时要求标出这些物体的位置。YOLOv8 是 YOLOv5 的下一个重大更新版本，它建立在以前 YOLO 版本的成功基础上，并引入了新的功能和改进，以进一步提升性能和灵活性，支持图像分类、物体检测和实例分割任务。

项目主要实现三个任务：

（1）构建和处理数据集：配置 NWPU VHR-10 数据集，利用数据增强技术，如旋转、缩放、裁剪等，扩充数据集规模，提高模型的泛化能力；

（2）模型训练与评估：配置 YOLOv8 模型，对数据集进行训练，得到适用于无人机空对地目标检测的模型，并对模型的性能进行评估；

（3）模型预测：在不同场景下对算法进行测试，分析算法在不同光照、遮挡、角度等条件下的表现。

1.3 小组成员分工

刘子康：整合 PPT 及结题报告，介绍项目，搜集数据集

张文浩：调试和运行代码，改进模型，记录实验结果

张子鉴：查阅文献，搜集资料，撰写结题报告

二、 空对地目标检测研究现状和发展趋势

2.1 现阶段问题及发展趋势

现阶段的空对地目标检测研究，由于无人机在对地目标检测中存在目标较小、目标尺度差异较大与场景复杂等问题，采用传统检测的手工特征方法难以适应多样化的数据，且经典的滑动窗口检测虽然不需要预处理即可应用于不同的检测任务，但是处理速度很慢，难以满足实时性检测的需求。

而基于深度学习的目标检测方法可以通过大量的数据集训练，提取到目标的特征表示，具有强大的特征学习和表达能力，且算法实时性较强，可应用于无人机机载实时目标检测，如 R-CNN、YOLO 等目标检测算法在以往实验中都体现了较好的效果。

2.2 研究现状及最新进展

Wenyuan Xu 提出团队了一种基于增强型 YOLOv8 模型的无人机图像小目标检测算法，称为 YOLOv8-MPEB[2]。模型用轻量级的 MobileNetV3 骨干网络替代了交叉阶段部分暗网 53（CSPDarknet53），从而降低了模型参数和计算复杂度，同时也提高了推理速度，并且精心设计了专门的小目标检测层，以优化多尺度目标的特征提取，在卷积到特征（C2f）模块中集成了高效多尺度关注（EMA）机制，旨在加强重要特征的提取，抑制多余特征。最后，在 "颈部 "部分使用了

双向特征金字塔网络（BiFPN），以改善尺度变化和复杂场景造成的检测误差，从而增强模型的泛化能力。本研究通过进行消融实验，并将实验结果与其他算法进行比较，以证实所提算法的有效性，尤其是在检测性能方面。

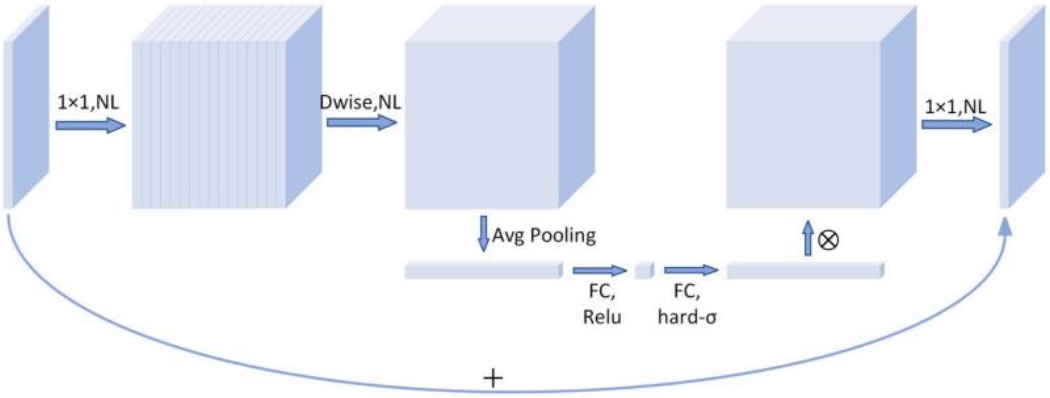

图 2-1 MobileNetV3 骨干网络

三、 VOLOv8 模型介绍，算法思想及其实现细节

3.1 YOLO 最新发展和应用

3.1.1 YOLO 算法简介

YOLO 是一种 one-stage 目标检测算法，即仅需要“看”一次就可以识别出图片中物体的 class 类别和边界框。而 YOLOv8 是 Ultralytics 公司最新推出的 YOLO 系列目标检测算法，可以用于图像分类、物体检测和实例分割等任务。

根据官方描述，YOLOv8 是一个 SOTA 模型，它建立在 YOLO 系列历史版本的基础上，并引入了新的功能和改进点，以进一步提升性能和灵活性，使其成为实现目标检测、图像分割、姿态估计等任务的最佳选择。其具体创新点包括一个新的骨干网络、一个新的 Ancher-Free 检测头和一个新的损失函数，可在 CPU 到 GPU 的多种硬件平台上运行。

此外，YOLOv8 还有一个特点就是可扩展性，Ultralytics 没有直接将开源库命名为 YOLOv8，而是直接使用"ultralytics"，将其定位为算法框架，而非某一个特定算法。这也使得 YOLOv8 开源库不仅仅能够用于 YOLO 系列模型，而且能够支持非 YOLO 模型以及分类分割、姿态估计等各类任务。

YOLOv8 的网络结构如图 2-1 所示，主要可分为 Input 输入端、Backbone 骨干神经网络、Neck 混合特征网络层和 Head 预测层网络共 4 个部分。

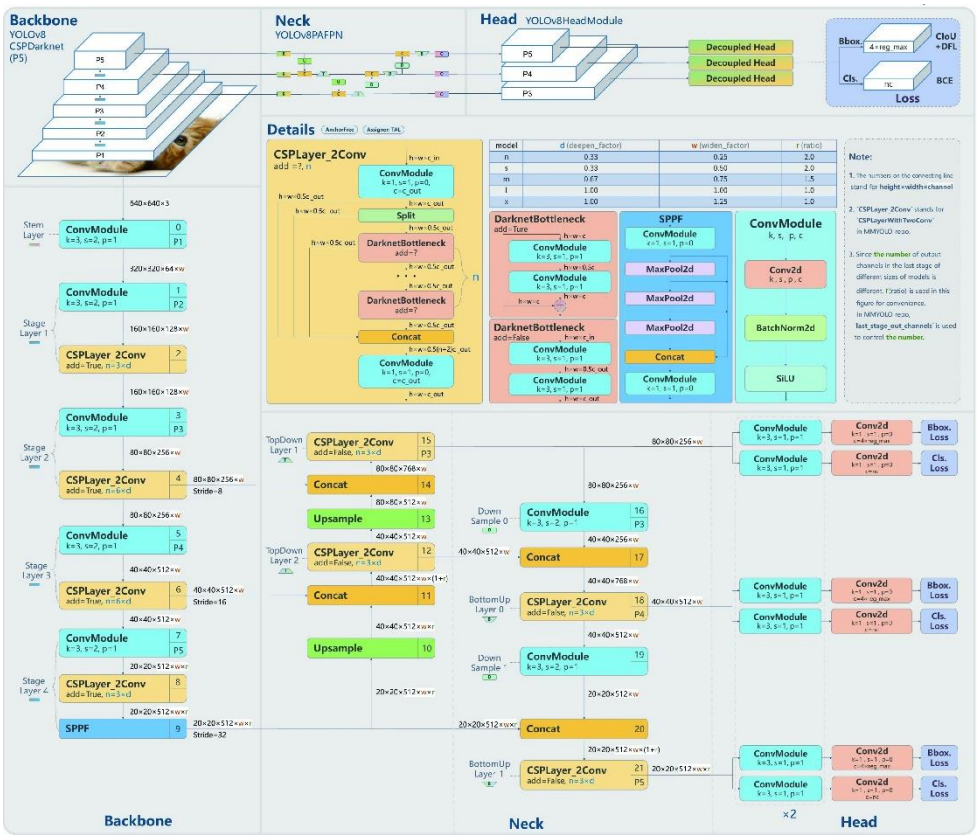

图 3-1　YOLOv8 网络结构

总而言之，YOLOv8 是 YOLO 系列模型的“最新王者”，各种指标全面超越现有对象检测与实例分割模型，借鉴了 YOLOv5、YOLOv6、YOLOX 等模型的设计优点，在全面提升改进 YOLOv5 模型结构的基础上实现，同时保持了 YOLOv5 工程化简洁易用的优势。

3.1.2 YOLOv8 模型最新应用

2023 年 10 月，福建理工大学高良彭团队基于 YOLOv8n 网络模型提出了一种新的交通标志检测算法—Faster-YOLOv8[3]。该模型在 Neck 部分采用 C2f-Faster 模块（C2f 和 FasterNet 的高效融合）来优化 YOLOv8n 网络结构，降低模型参数量及模型大小；引入 EMA 注意力机制，并应用于模型主干网络，实现了更好的多尺度感知和空间感知，增强了模型的特征提取能力；通过添加小目标检测层，有效地结合了不同尺度特征信息，保留更多的细节信息，从而提高了对小目标的检测能力；采用 SioU 作为边界损失函数提高检测精度。

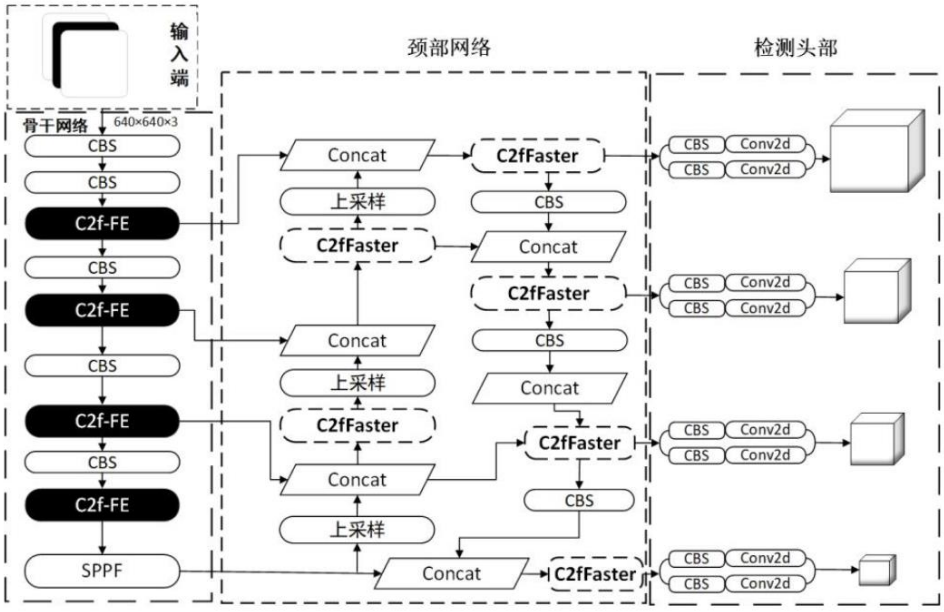

图 3-2　Faster-YOLOv8 网络结构

而江祥奎团队为了改善在动态场景下同步定位与地图绘制(Simultaneous Localization And Mapping, SLAM)算法定位精度低的问题，提出一种基于轻量化 YOLOv8n 的动态视觉 SLAM 算法[4]：利用加权双向特征金字塔网络(Bidirectional Feature Pyramid Network, BiFPN)对 YOLOv8n 模型进行轻量化改进，减少其参数量；在 SLAM 算法中引入轻量化 YOLOv8n 模型，并结合稀疏光流法组成目标检测线程，以去除动态特征点，利用经过筛选的特征点进行特征匹配和位姿估计。

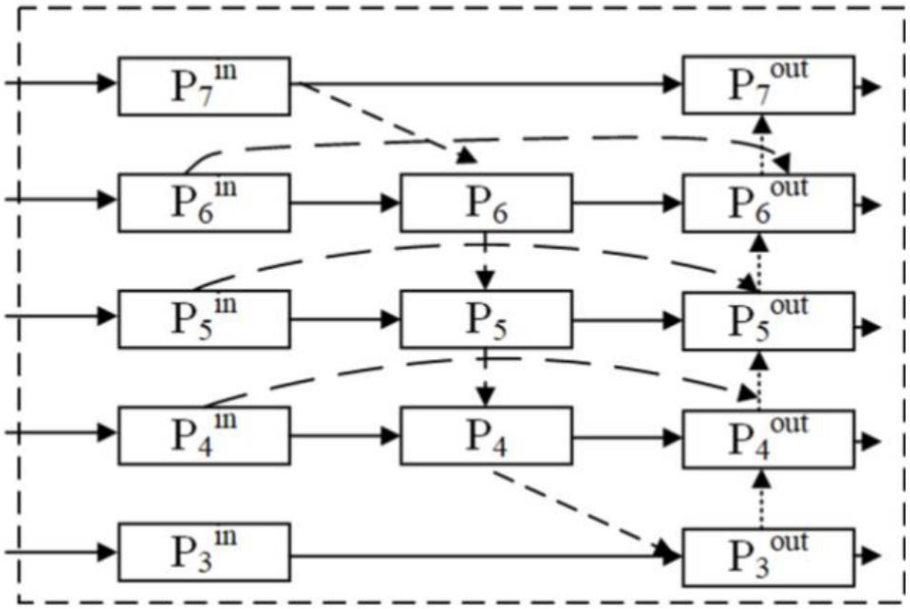

图 3-3 轻量化 BiFPN 网络结构

3.2 YOLOv8 算法细节

3.2.1 评价指标

记目标的结果为正例(true)和负例(false)，而预测的结果为正例(positive)和负例(negative)。则准确率(查准率，precision)是真实目标占网络预测目标总数的比例，表示该网络的分类准确率，而召回率(查全率，recall)是网络成功预测的真目标数与实际真目标数的比值，相对应的公式[5]为：

$$precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

$$recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

此外，目标的交集/并集比 IoU (intersection over union)常用来衡量物体检测结果与真实值（真实的物体边界）匹配的好坏。一般当 IoU 大于 0.5 时可作为真目标，若记真实区域为 A，预测区域为 B，则 IoU 可表示为：

$$IoU = \frac{A \cap B}{A \cup B}$$

在 IoU 的基础上增加中心点距离，而 CIoU Loss 则又向其中添加了相对比例，用于对预测框和真实框不一致的结果进行惩罚，公式如下：

$$CIoU = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v$$

式中 b 和 b^{gt} 分别是预测框和真实框的中心点，p 表示两个矩形框的欧式距离，c 表示两个矩形框的闭包区域的对角线的距离；而 v 用来衡量预测框和真实框两个矩形框相对比例的一致性，α 是权重系数。这样处理后，不仅考虑了两个框之间的距离，还考虑了矩形框的相对比例，更是解决了梯度为 0 和梯度爆炸的情况。

3.2.2 损失函数改进

由于正负样本有不平衡问题，可能会导致评价指标的误判，如几乎全是正例的模型中采用查准率(accuracy)作为评价指标会导致生成的模型无法预测反例。解决这个问题的常见思想是调整类别权重：即算法实现过程中，对于分类中不同样本数量的类别分别赋予不同的权重。

因而 YOLOv8 使用变焦损失函数(varifocal loss function, VFL Loss)作为分类损失函数，方程式如下：

$$VFL(p,q) = \begin{cases} -q(qlg(p) + (1-q)\lg(1-p)) & q > 0 \\ -\alpha p^{\gamma} \lg(1-p) & q = 0 \end{cases}$$

主要改进是提出了非对称的加权操作，其中 p 表示前景的预测概率，q 是标签，正样本时 q 为预测结果和真实标注框的 IoU,而负样本时候 q 值为 0。该损失函数其他符号中由于 p 表示前景的预测概率，故 1-p 和 p 的系数次方分别是前景类和背景类可以减少简单样例的损失贡献，即相对增加误分类样例的重要性。

回归损失为 CIoU 损失函数+分布焦点损失(distribution focal loss, DFL)，记 y

是真实标签，靠近真实标签 y 的两个标签分别是第 i 次和 i+1 次的预测。则有

$$S_i = \frac{y_{i+1} - y}{y_{i+1} - y_i}, S_{i+1} = \frac{y - y_i}{y_{i+1} - y_i}$$

故 DFL 损失函数表示如下：

$$DFL(S_i, S_{i+1}) = -((y_{i+1} - y)\log(S_i) + (y - y_i)\log(S_{i+1}))$$

四、 基于 RSOD-数据集的实验设置及结果分析

4.1 数据集来源与介绍

此次项目使用的数据集为 RSOD-数据集 RSIA-LIESMARS-WHU/RSOD-Dataset-：遥感影像中目标检测的开放数据集 (github.com)[6]。RSOD 是一个开放的目标检测数据集，用于遥感图像中的目标检测，包含飞机，油箱，运动场和立交桥四种地面目标，以 PASCAL VOC 数据集的格式进行标注，也可用于无人机视角下的目标检测。

数据集包括 4 个文件夹，每个文件夹各包含一种对象：

（1）飞机数据集，包含 446 幅图像中的 4993 架飞机

（2）操场数据集，包含 189 副图像中的 191 个操场

（3）立交桥数据集，包含 176 副图像中的 180 座立交桥

（4）油箱数据集，包含 165 副图像中的 1586 个油箱

4.2 实验设置及实现步骤

4.2.1 数据集构建与处理

下载飞机、操场、油罐数据集，并将三个数据集中的图片和注释 annotations 分别合并在文件夹中。由于 annotations 中为.xml 文件，不符合 YOLOv8 的标签格式，需要清洗转换成.txt 格式，该格式要求每行五个数，第一个数是标签，其他四个数是框住物体的框的四个坐标。清洗结束后，划分数据集，以 7：2：1 的比例划分为训练集，验证集，测试集。

```python
def split_img(img_path, label_path, split_list):
    try:
        Data = 'D:/airtoground/dataset'
        # Data是你要将要创建的文件夹路径（路径一定是相对于你当前的这个脚本而言的）
        # os.mkdir(Data)

        train_img_dir = Data + '/images/train'
        val_img_dir = Data + '/images/val'
        test_img_dir = Data + '/images/test'

        train_label_dir = Data + '/labels/train'
        val_label_dir = Data + '/labels/val'
        test_label_dir = Data + '/labels/test'

        # 创建文件夹
        os.makedirs(train_img_dir)
        os.makedirs(train_label_dir)
        os.makedirs(val_img_dir)
        os.makedirs(val_label_dir)
        os.makedirs(test_img_dir)
        os.makedirs(test_label_dir)
```

(a) 代码块 1

```python
    except:
        print('文件目录已存在')

    train, val, test = split_list
    all_img = os.listdir(img_path)
    all_img_path = [os.path.join(img_path, img) for img in all_img]
    # all_label = os.listdir(label_path)
    # all_label_path = [os.path.join(label_path, label) for label in all_label]
    train_img = random.sample(all_img_path, int(train * len(all_img_path)))
    train_img_copy = [os.path.join(train_img_dir, img.split('\\')[-1]) for img in train_img]
    train_label = [toLabelPath(img, label_path) for img in train_img]
    train_label_copy = [os.path.join(train_label_dir, label.split('\\')[-1]) for label in train_label]
    for i in tqdm(range(len(train_img)), desc='train ', ncols=80, unit='img'):
        _copy(train_img[i], train_img_dir)
        _copy(train_label[i], train_label_dir)
        all_img_path.remove(train_img[i])
    val_img = random.sample(all_img_path, int(val / (val + test) * len(all_img_path)))
    val_label = [toLabelPath(img, label_path) for img in val_img]
    for i in tqdm(range(len(val_img)), desc='val ', ncols=80, unit='img'):
        _copy(val_img[i], val_img_dir)
        _copy(val_label[i], val_label_dir)
        all_img_path.remove(val_img[i])
    test_img = all_img_path
    test_label = [toLabelPath(img, label_path) for img in test_img]
    for i in tqdm(range(len(test_img)), desc='test ', ncols=80, unit='img'):
        _copy(test_img[i], test_img_dir)
        _copy(test_label[i], test_label_dir)
```

(b) 代码块 2

图 4-1 数据集划分

使用上述代码进行划分，划分好后整理文件，结构如下图所示：

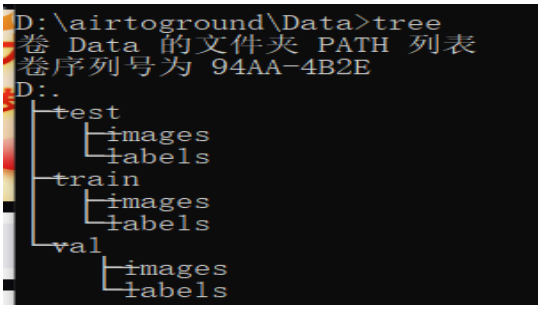

图 4-2 数据集文件结构

4.2.2 模型配置及调用

首先编写 yaml 文件，如下图所示：

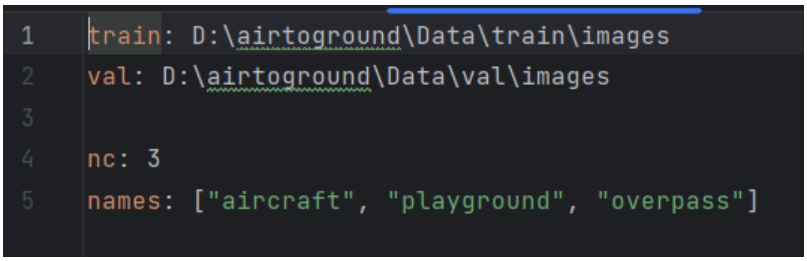

图 4-3 yml 文件

然后进行模型训练，进入 conda 的 YOLOv8 终端，cd 进入 airtoground 项目中，单卡训练输入命令：

YOLO task=detect mode=train model=YOLOv8n.pt data=data/airtoground.yaml batch=32 epochs=100 imgsz=640 workers=16 device=cpu

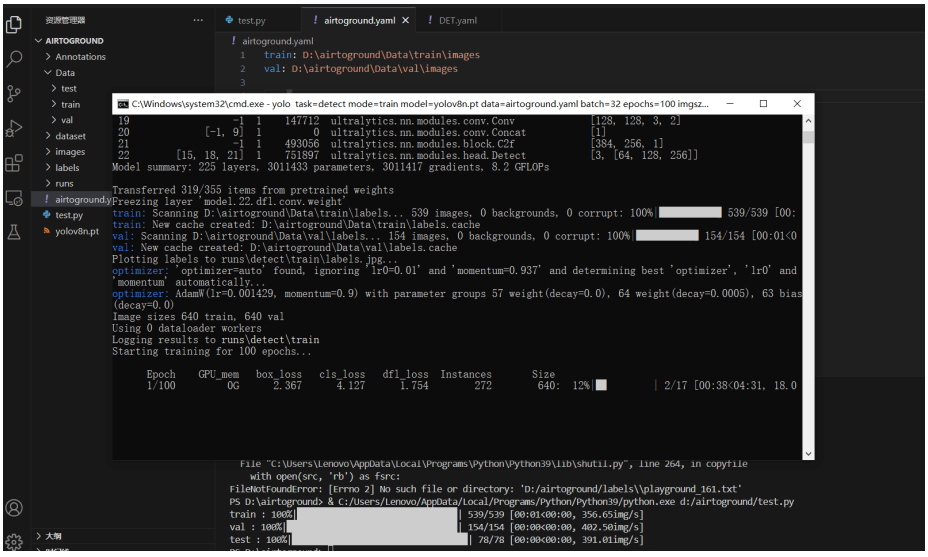

图 4-4 YOLOv8 模型训练

4.3 实验结果分析

等待训练结束，出现 best.pt 后分析结果。

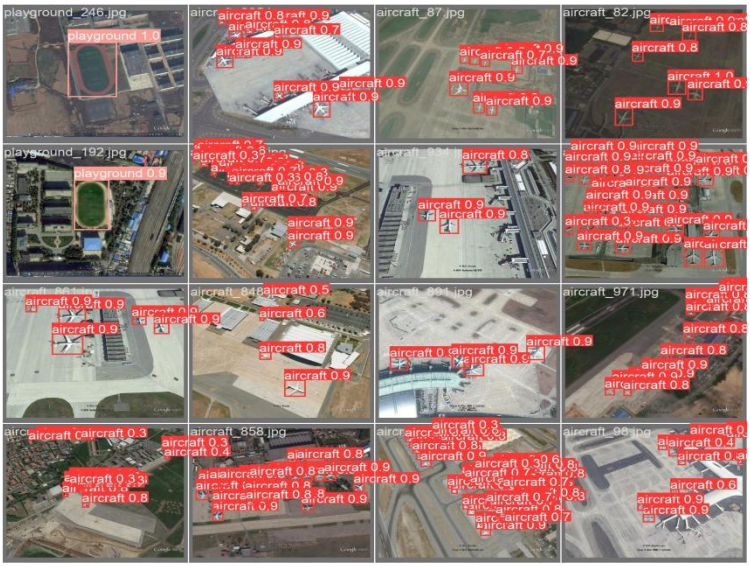

图 4-5 训练及预测结果

模型预测准确率的混淆矩阵图如下所示：

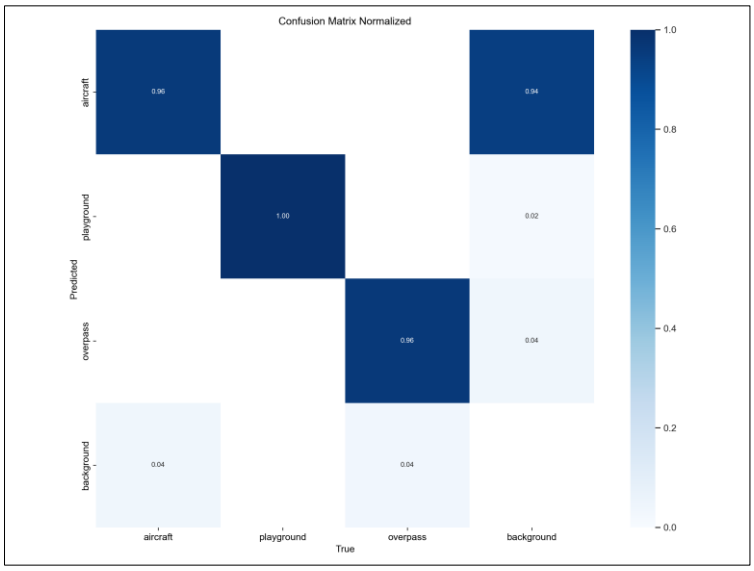

图 4-6 预测准确率混淆矩阵

F1 分数（F1-score）是分类问题的一个衡量指标，是精确率和召回率的调和平均函数，介于 0-1 之间。一些多分类问题的机器学习竞赛，常常将 F1-score 作为最终测评的方法。

对于某个分类，综合了 Precision 和 Recall 的一个判断指标，F1-Score 的值是从 0 到 1 的，1 是最好，0 是最差：

$$F_1\ \ Score\frac{2 * Precision * Recall}{Precision + Recall}$$

简而言之就是想同时控制 recall 和 precision 来评价模型的好坏。

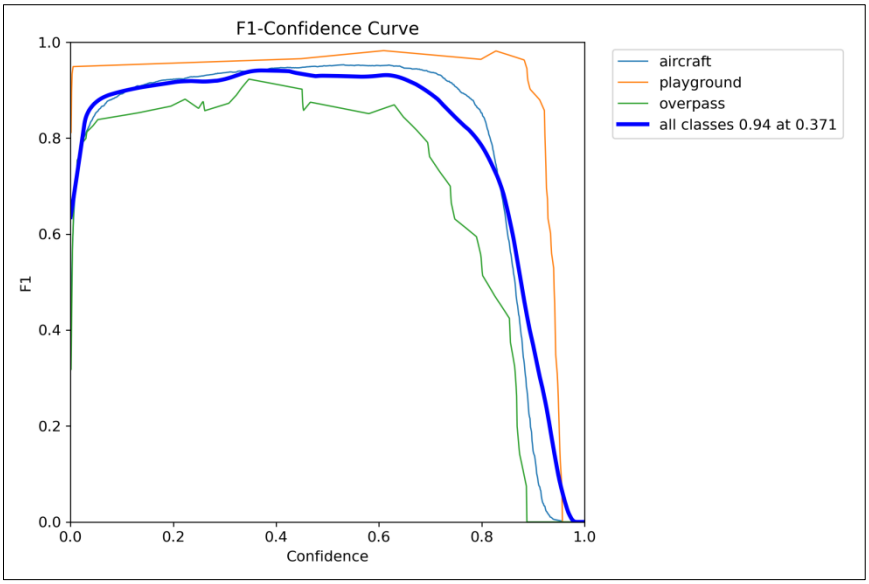

图 4-7 F1 分数-置信度关系曲线

标注所得 Labels 如下图所示，各部分含义：（1，1）表示每个类别的数据量，（1，2）真实标注的 bounding_box，（2，1）真实标注的中心点坐标，（2，2）真实标注的矩阵宽高。

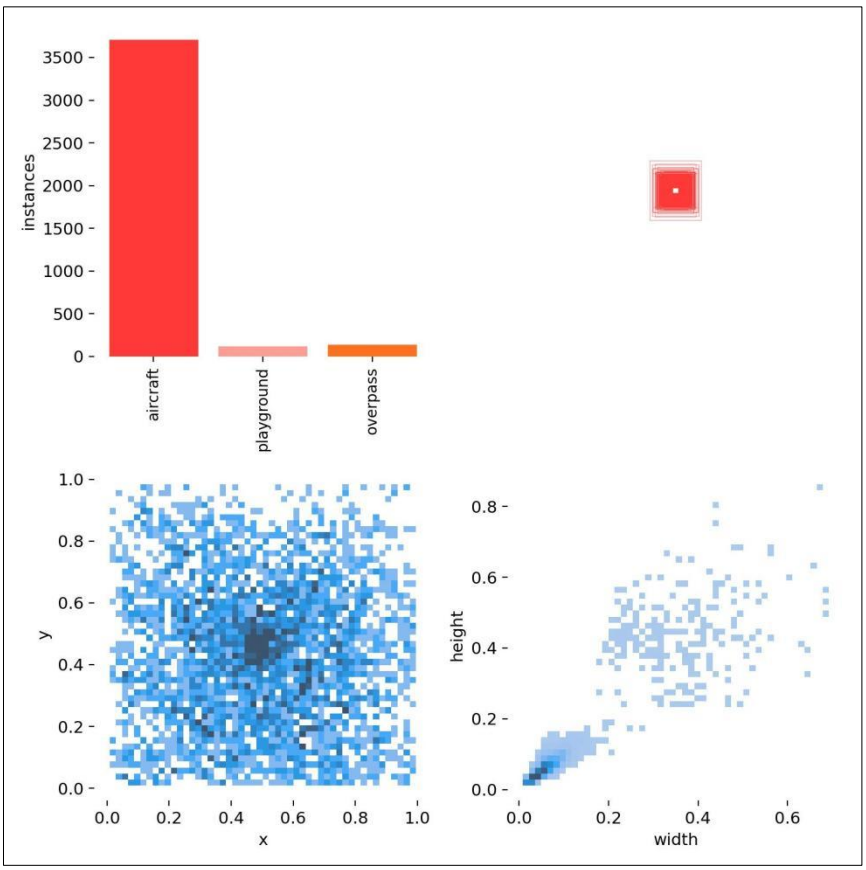

图 4-8 真实标注标签

下图为表示准确率与置信度的关系图线，横坐标为置信度，可以看出置信度越高，准确率越高。

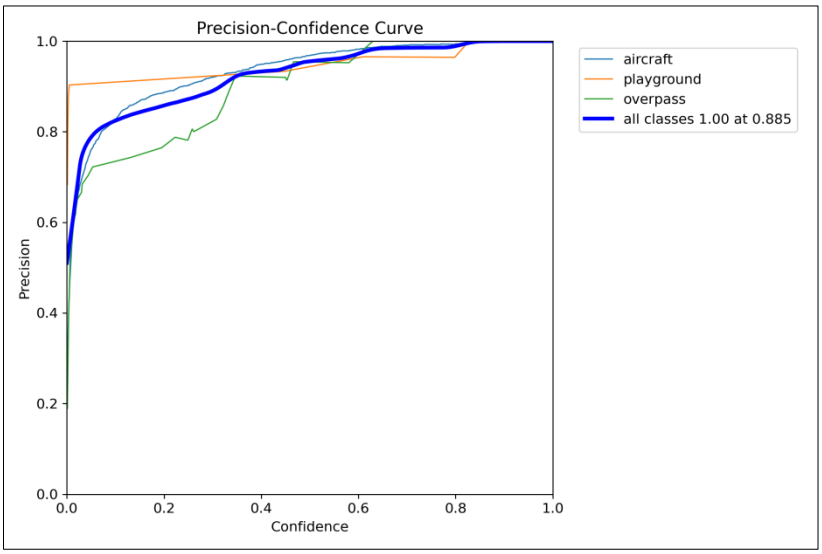

图 4-9 准确率-置信度关系曲线

PR 曲线中的 P 代表的是 precision（精准率），R 代表的是 recall（召回率），其代表的是精准率与召回率的关系，一般情况下，将 recall 设置为横坐标，precision 设置为纵坐标。PR 曲线下围成的面积即 AP，所有类别 AP 平均值即 Map。

平衡点（BEP）是 P=R 时的取值（斜率为 1），F1 值越大，我们可以认为该学习器的性能较好。

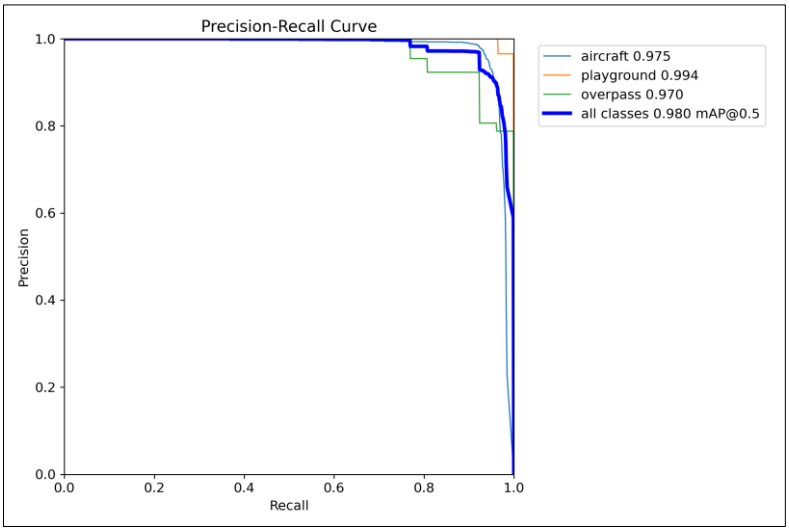

图 4-10 准确率-召回率关系曲线

不同结果均值如下图所示，（1，1），（2，1）：YOLOV5 使用 GIOU Loss 作为 bounding box 的损失函数，该图分别表示训练时和验证时 GlOU 损失函数的均值，越小方框越准；（1，2），（2，2）：推测为目标检测 loss 均值，越小目标越准；（2，4），（2，5）：表示在不同 IoU 阈值时计算每一类中所有图片的 AP 然后所有

类别求取均值。mAP50-95 表示从 0.5 到 0.95 以 0.05 的步长上的平均 mAP。

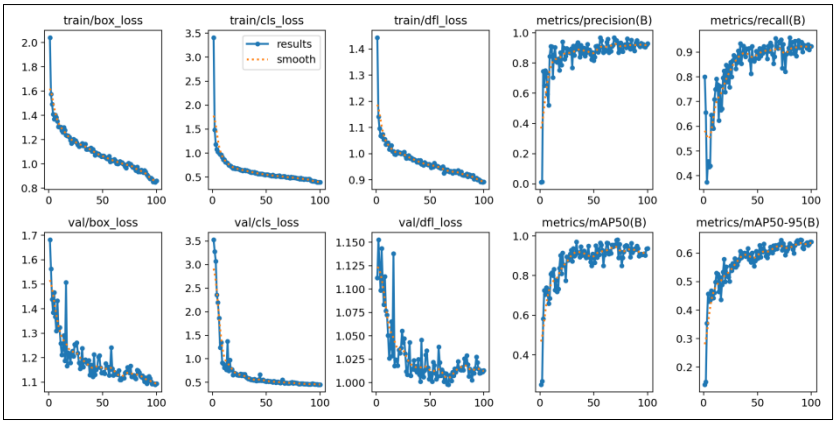

图 4-11 各类结果均值

最后构建可视化测试界面，使用 pyqt5 进行窗口实现，可以选择模型，检测图片、视频以及打开摄像头实时监测

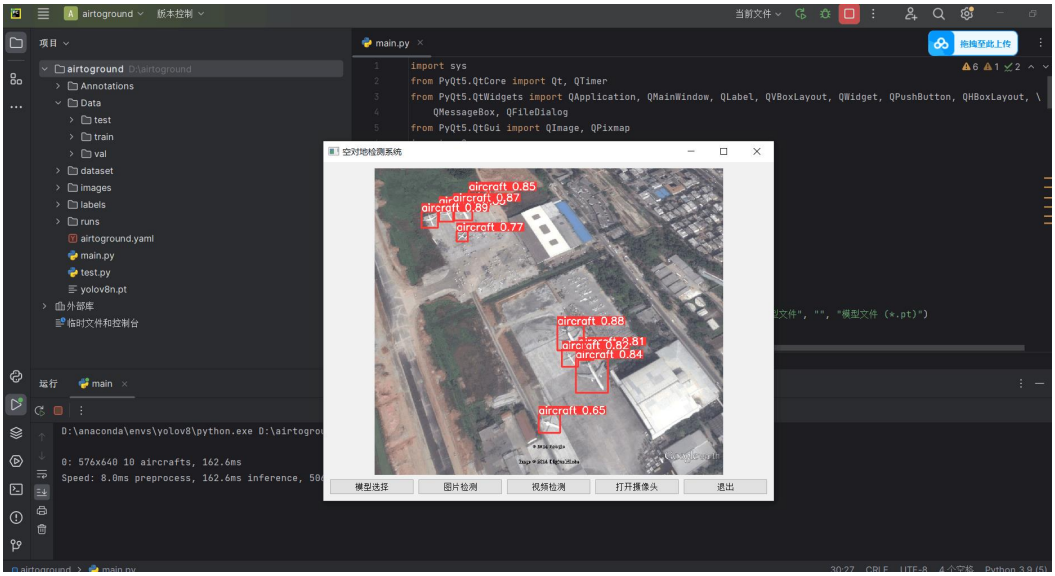

图 4-11 pyqt5 工具窗口实现

图 4-12 可视化测试界面

经检验，视频以及打开摄像头实时监测功能也有良好效果。

五、 项目结论和算法改进

5.1 项目主要思想及结论

本项目使用搜集到的数据集对 VOLOv8 模型进行训练和预测，实现无人机视角下对地面物体和建筑设施等目标的识别和检测，并在此基础上进行算法改进。针对空对地目标检测的特性，从网络结构、损失函数、训练策略等方面对 YOLOv8 算法进行优化和改进，提升检测精度和速度。同时，结合 pyqt 等工具，进一步实现视频的实时监测。

经实验验证，改进后的 YOLOv8 模型可以很好地应用于无人机空对地目标检测问题，识别效率高、成本低、操作简便。在模型具体应用时，仅使用 CPU 进行训练时速度很慢，可以借助 GPU 提高训练速度；通过调整 epoch 的值，有助于解决欠拟合与过拟合的问题。

5.2 算法改进的基本思想

YOLOv8 模型存在参数量过大的问题，可以对其进行轻量化改进，如高良鹏团队在 Neck 部分采用 C2f-Faster 模块（C2f 和 FasterNet 的高效融合）来优化 YOLOv8n 网络结构，降低模型参数量及模型大小，这是一个可行的方法，能够解决模型体积大、不宜部署等问题。

在特征提取上，还可以在模型主干网络上引入注意力机制，实现更好的多尺度感知和空间感知，增强模型的特征提取能力。对于不易检测的小目标，可以二外设计检测层，结合不同尺度的特征信息、保留更多细节，提高对小目标的检测能力。

六、 参考文献

[1]Ultralytics:UltralyticsYOLOv8.[EB/OL].[2023-3-26].https://github.com/ultralytics/ultralytics.
[2]Xu W ,Cui C ,Ji Y , et al.YOLOv8-MPEB small target detection algorithm based on UAV images[J].Heliyon,2024,10(8):e29501-.
[3]高良鹏,赵博文,简文良.基于 Faster-YOLOv8 网络模型的车载交通标志检测算法研究[J].重庆交通大学学报(自然科学版):1-9.
[4]江祥奎,杨刚,杜遥遥.基于轻量化 YOLOv8n 的动态视觉 SLAM 算法[J].西安邮电大学学报:1-8.
[5]刘瑞锦,何章鸣.基于 YOLOv8 的卫星遥感图像快速目标检测方法[J].空间控制技术与应用,2023,49(05):89-97.
[6]Y. Long、Y. Gong、Z. Xiao 和 Q. Liu，“基于卷积神经网络的遥感图像中的精确目标定位”，载于 IEEE Transactions on Geoscience and Remote Sensing，第 55 卷，第 5 期，第 2486-2498 页，2017 年 5 月。doi：10.1109/TGRS.2016.2645610