反无人机

1. **选题**

关于无人机识别有两种理解方式，一是使用无人机平台上搭载的传感器，借助无人机视角完成地面有价值目标的检测（比如车辆、人流、甚至是一些测绘工作）。二是识别低空目标，即对无人机进行管理和反无人机。

无人机具有体积小、成本低、机动性强等优势，而优势也可能造成危害。近来国内外多次发生无人机非法入侵事件，不仅对公民的个人隐私与生命财产安全造成了严重危害，而且对机场、军事基地、大型集会现场、核电站、政府机部门驻地等敏感区域的安防造成了极大威胁。反无人机算法可以自动、持续地保护重要区域的安全，探测和监管无人机攻击。因此我们决定研究反无人机算法。

先明白反无人机的具体定义，反无人机指在野外发现、探测、识别和跟踪无人机（UAV）目标，同时根据RGB或热红外（IR）视频估计目标的跟踪状态。接着寻找反无人机的研究现状作为参考，北方电子设备研究所在研究反无人机项目，旨在推动无人机在野外发现、检测和跟踪的前沿发展。此外有研究组提出了一个新的数据集、评估指标，以定义高质量的在真实动态场景中捕获无人机的基准。

1. **算法**

经过资料查阅我们发现了多条“Yolo”“小目标检测”相关的词条，这说明该题目并不是罕见的操作。我们还找到了数篇基于Yolo的anti uav相关论文，在这些文章中虽然提到了自己的新算法对比yolo有着优势，但同时也可观察到Yolo本身其实是适应该题目的。因此决定使用Yolov11算法。

**（一）Yolov5**

Yolov5的输入端主要由Mosaic图像增强、自适应锚框计算以及自适应图片缩放组成。Mosaic图像增强通过组合多个不同的图像来生成新的训练图像，使模型同时处理多个不同环境的对象的上下文关系，提高网络鲁棒性。ATSS自适应锚框计算方法根据交并比IoU，自适应地选择正样本与负样本，而不需要手动设置，从而提高检测精度。自适应图片缩放可以自适应地缩放输入图像的尺寸，以适应不同尺度目标的检测。

Yolov5的Backbone层主要由Focus结构以及CSP结构组成。Focus结构用于特征提取，它按输入特征图中的像素位置进行空间切片，在通道维度上拼接后进行卷积操作，从而使特征融合，增强对全局结构的感知能力。CSP结构将输入特征图分成两部分，一部分经过一个子网络进行卷积处理，提取出相对较少的高层次特征，另一部分则直接进行下一层的处理，然后将两部分特征图拼接起来，使高层次抽象特征与低层次细节特征结合，作为下一层的输入。

Backbone中的最后一层为SPPF层。SPP结构将输入的不同大小的特征图分别进行不同尺度的池化操作，并将不同尺度的池化结果拼接起来作为SPP结构的输出，不同大小的池化核获得局部到全局的信息，从而增强模型对不同尺度目标的特征感知能力。SPPF结构通过串联多个小核池化等效替代大核池化，提升计算速度。

Yolov5的Neck层使用了FPN结构和PAN结构。输入图像经过卷积神经网络，得到一系列特征图，每个特征图对应网络的一层。FPN结构对于较深的特征图进行上采样操作，使其与较浅的特征图尺寸相同后进行融合，采用加法操作，再进行卷积进一步融合信息。PAN结构对于较浅的特征图进行下采样操作，使其与较深的特征图尺寸相同后进行拼接，再进行卷积进一步融合。两结构融合了多尺度特征，提高了目标检测的准确性。

Yolov5的Head层输出多尺度特征图，每个图对应不同尺度的预测框，每个预测框包含置信度、类别概率、边界框位置信息。Head层从输入端得出的锚框开始根据特征图不断调整得出预测框。其采用IoU loss函数确定预测框是否正确，它是一种用于衡量预测框与真实框之间重叠程度的指标。在调整过程中一个物体可能被多个预测框检测出来，为了避免对同一个物体进行多次检测，提高检测的精度和效率，需要进行NMS非极大值抑制，即对重复的预测框进行过滤。**（二）Yolov8与Yolov11**

对比Yolov5，Yolov8的Backbone将C3模块替换成了C2f模块，实现了进一步轻量化，同时沿用Yolov5中的SPPF模块。Head部分将耦合头换成了目前主流的解耦头结构，将检测和分类头分离。

对比Yolov8，Yolov11中的Backbone将C2f模块变为C3K2模块。Backbone中的最后一层（SPPF层）后增加了C2PSA模块。Head解耦头中的分类检测头增加了DWConv。

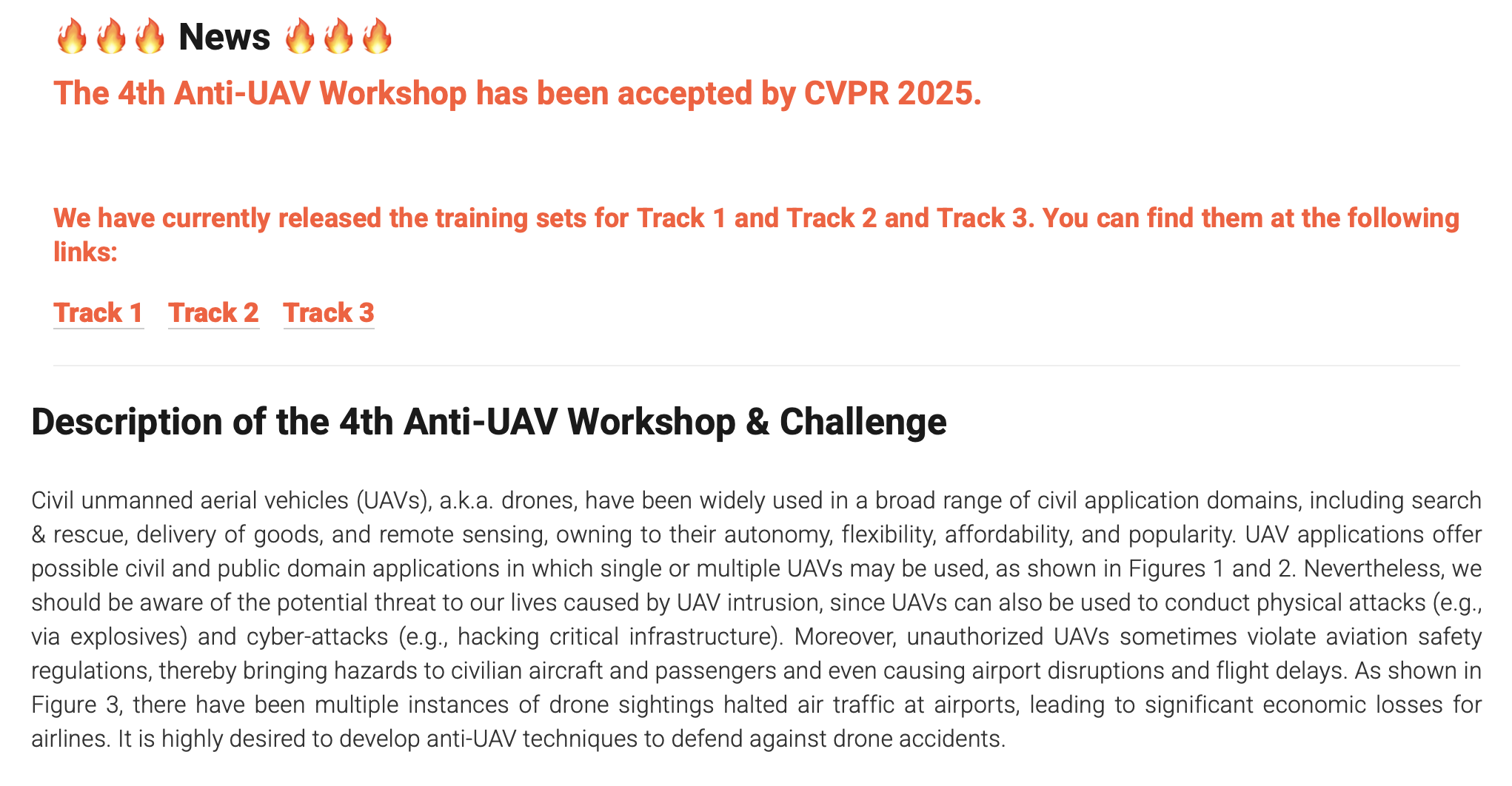
C3模块即上述CSP结构。C3k允许设置不同的卷积核大小，以便更灵活地应对各种不同的特征提取需求。C2f通过减少卷积层的数量和采用更高效的特征合并策略来提高速度。C3K2结合了C2f的速度优势和C3k的灵活性，当C3k参数设置为True时使用C3k层，否则使用标准的瓶颈层，与 C2f 类似。

C2PSA用于增强特征提取，结合了CSP结构和PSA注意力机制。特征图被分为两部分，一部分直接传递，另一部分通过PSA注意力模块处理，最终拼接融合。PSA模块通过引入不同大小的卷积核来提取多尺度的空间信息，同时结合 Squeeze-and-Excitation 模块对特征通道进行加权，从而增强网络对不同尺度目标的注意力聚焦。

DWConv（深度卷积）是一种高效卷积操作，主要用于减少计算复杂度和参数量。在标准卷积操作中，对于一个输入张量，卷积核与输入张量在每个位置都会计算与所有输入通道的点积，每个输出通道是所有输入通道的加权求和，导致计算量比较大。深度卷积将输入的每个通道单独处理，即每个通道都有自己的卷积核进行卷积，不与其他通道进行交互。

1. **数据集与部署**

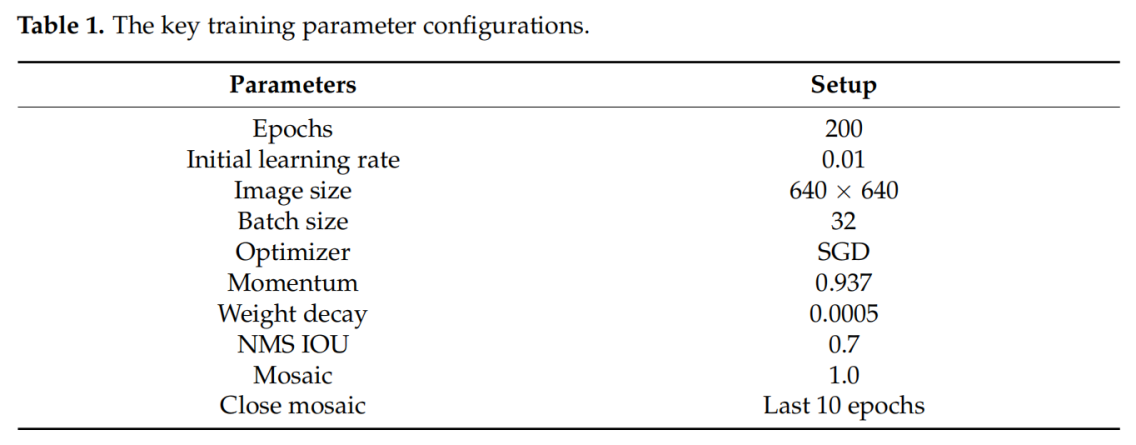
我们选择了一个无人机挑战赛项目的公开数据集，作为训练用数据集。其中已经配有预分割的验证集、约20000条训练样本、且格式为RGBD图像。我们使用conda环境下的modelscope开源平台下载了该数据集，但发现数据集本身仍存在以下问题：文件夹内层数超过一层；数据集格式不是coco格式的JSON脚本，缺少了长宽信息；本身并不是yolo格式，需要转换成yolo格式。据此，我们编写了3个python脚本以解决上述问题：addwh.py(用于补全长宽信息)、coco2yolo.py(用于将coco格式转换为yolo格式)、flatten\_img.py（将多层嵌套的文件平铺到一个单一目录中）。



挑战赛的公开数据集

经过python脚本的处理后，我们终于获得了一个可以用于训练的数据集。接下来考虑部署的问题。我们所使用的设备就是一台普通的装载3070ti的笔记本电脑，按照教程安装nvidia docker utils以及docker pull yolo11的最新镜像。因受学校断电限制，我们选择使用内网穿透，远程控制放在家中的笔记本进行训练。

关于超参数的参考与设置，在我们查询了相关资料并综合我们的算力条件后，我们发现：在batch size=32的情况下，约每20分钟可以训练一个epochs；综合参考论文，训练出能达到效果的模型大约需要训练200个epochs。



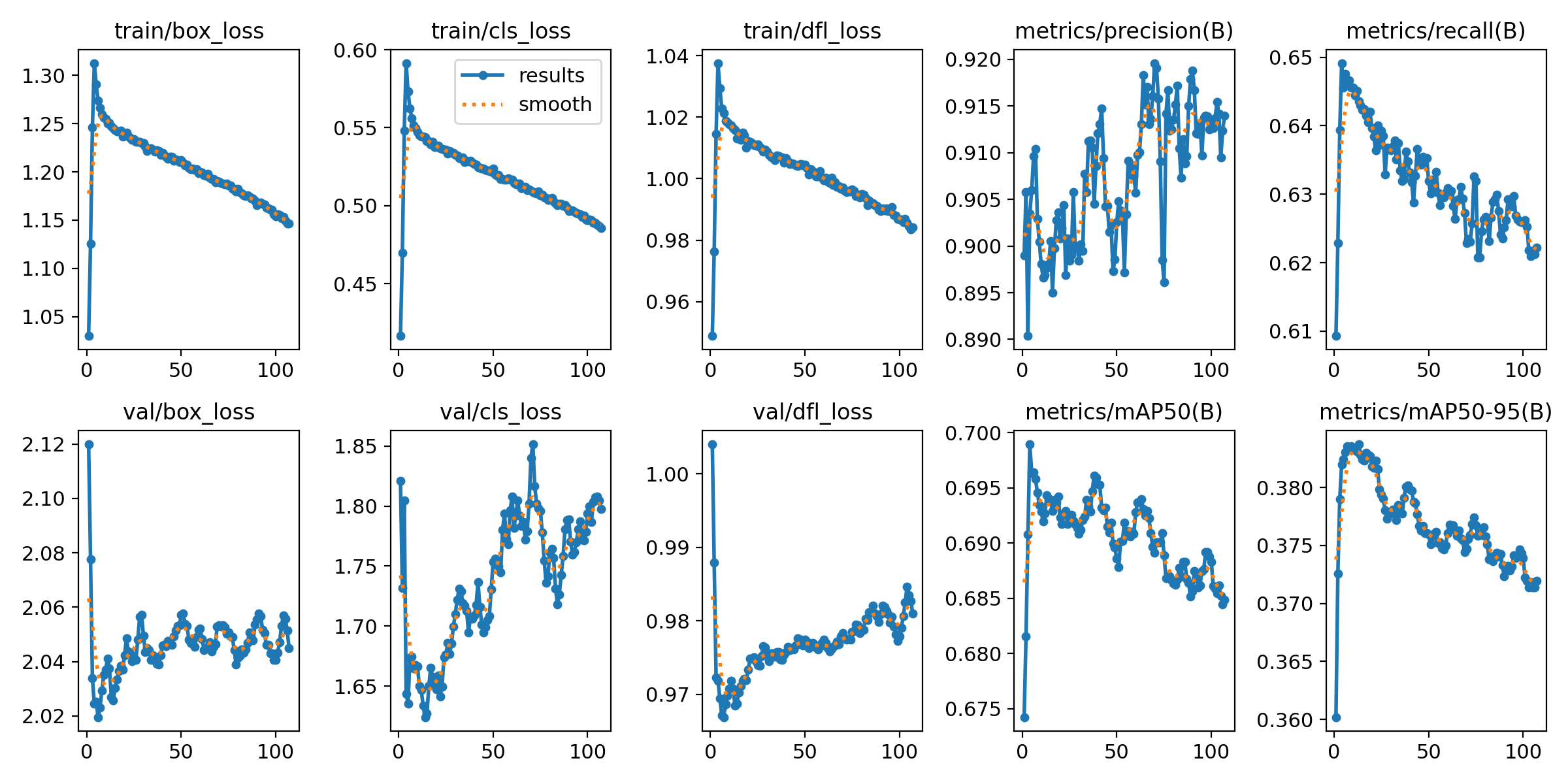
参考

到这里关于数据集合部署的准备部分已完成，等待足够长的时间直到模型训练完成即可。

1. **成果分析**

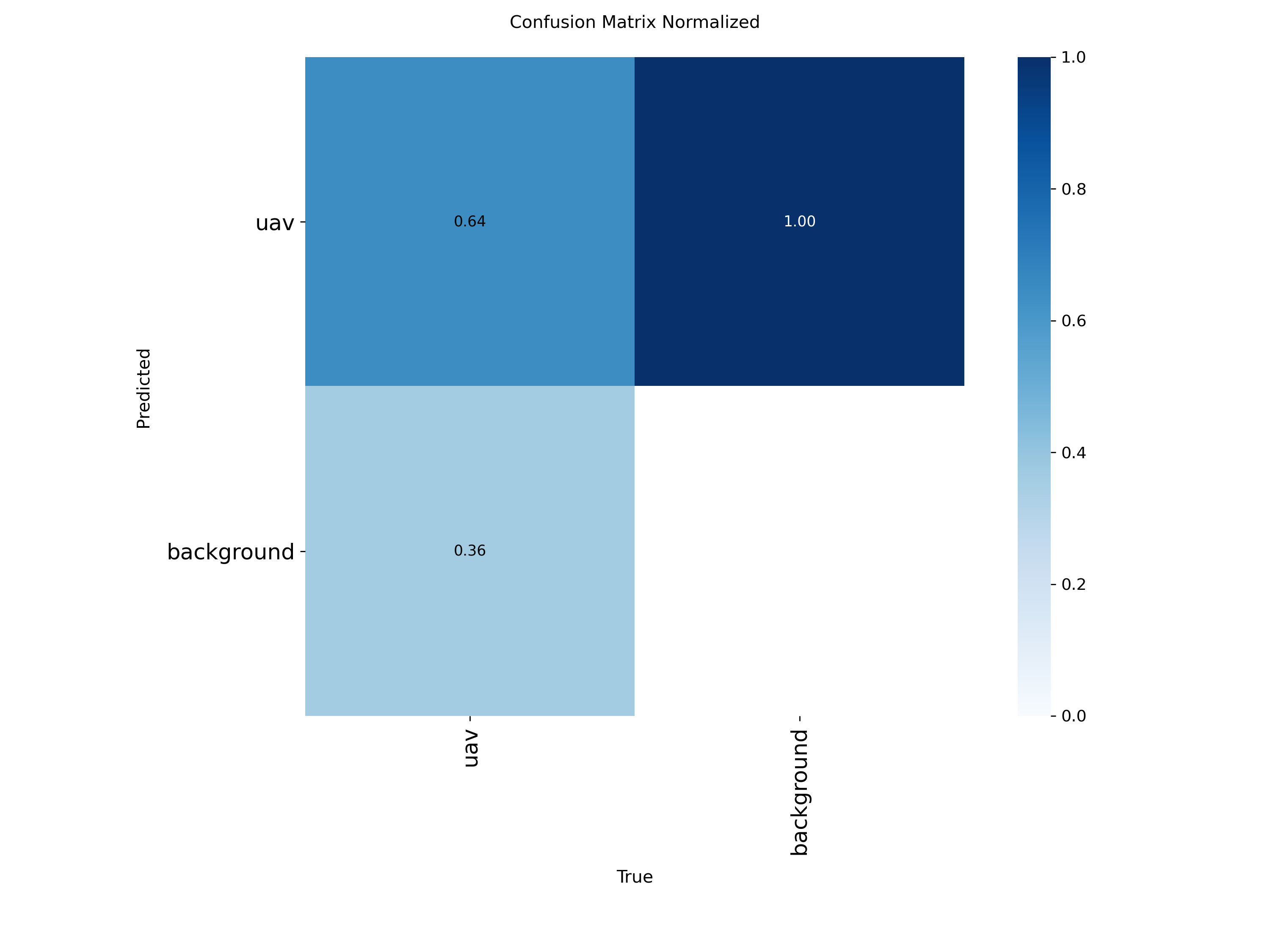
从结果来说，我们的效果并不尽如人意。

我们在训练时出现了一些失误：我们错误地认为可以先训练100轮的模型再训练100轮，和直接训练200轮的模型可以得到一致的效果，很显然这个想法过分天真了。



反应了训练效果的图表

可以看到训练损失随着训练轮次的增加逐渐减小，精度和召回率随着训练轮次的增加不断提高，说明模型的准确性和完整性都在增强。在验证集上，这些损失同样呈现出逐渐缩小的趋势，表明模型在没有见过的数据集上也有较为良好的泛化能力。



混淆矩阵

从这个混淆矩阵的结果来看，对于低空无人机目标的识别，模型预测正确的比例为64%；对于背景的识别，模型预测正确的比例为100%。可以看到我们的模型预测低空无人机目标的效果有很大的改进空间。