大家好，我是第二组的汇报人陈思成。我们汇报的主题为好跌yo——基于yolov8的跌倒检测模型性能优化。我们汇报内容分为以下三个部分。

下面进行第一个部分，项目背景。

我们的项目是一个基于yolov8深度学习网络的检测人类跌倒的计算机视觉目标检测模型。如左图所示，目标检测是找出图像或视频中人们感兴趣的物体，并同时检测出他们的位置和大小。我们使用的yolo V8是one stage算法，也就是省略掉专门生成候选框的模块来直接生成边界框的模型算法，这也是yolo系列网络的名称来源：you only look once。相比需要专门生成候选框模块的two stage算法损失了一些精度，但大幅提升了模型的推理速度。目前yolo V8在目标检测领域中属于最新且最稳定的算法，因此我们决定以其为基础，通过改变数据集，数据增强，调整超参数等方式，在一定时间内尽可能提升模型的性能。

接下来介绍一下目标检测的评价指标。所有样本可以被分为如左上图所示的四个类别，其中我们关注的是TP，即模型检测到正确的正例。围绕该需求可以引申出两个较为关键的指标：精确率和召回率。前者反映了被分类器判定正例中真正的正例比重，与误判有关;后者反映了被正确判定的正例占总检测目标的比重，与漏判有关。

根据以上两个指标可以绘制PR曲线，曲线下方面积即检测器在各个召回率下的平均值，称为AP，而mAP则是多分类AP的平均值。在我们的跌倒检测中只有一个类别，因此mAP=AP。mAP常取的两个指标为mAP50和mAP50-95，分别对应IoU，即模型的预测框和真实边界框的交并比阈值，50指的是二者交并比大于50%即视为检测到的正样本；同理，50-95指的是步长为5%的交并比阈值下mAP的平均值。我们的模型可靠度主要以以上提到的四个指标为参考依据。

接下来是项目介绍。右图为yolov8的结构图，因为相对比较复杂，在此不做展开介绍。为了方便多方同步，我们决定采用AutoDL云服务器作为模型部署设备，并使用github仓库来保证项目的稳定实时更新。我们租用的设备配置和模型的运行环境如左方所列。在数据方面，我们的数据渠道主要来源于各大图片网站收集的真实图片和利用Stable Diffusion，即AI生成的图片。详细地说，我们共制作了三个数据集。v1版本体量较小，被用于作为前期模型运行稳定性的测试样本集；v2版本在样本类别和数量上相比v1获得了大幅提升；v3版本修正了v2的部分类别标注问题，并引入了部分AI生成的样本。总体来说，我们的数据集基于样本的种类、质量和数量的提升而不断迭代优化。我们的标注工具使用的是Yolo自带的YoloLabel 1.2.1

在整个项目中，我们的工作重心在于模型的性能优化。因为模型的网络结构已经集成了大部分目前最新的工作，已经具备较为出色的表现，所以我们没有针对其结构进行修改，而是通过调整超参数配合升级数据集的方式来优化模型性能。我们主要调整的超参数如图所示，可以被归类为模型本身的迭代参数和数据增强设置两大类。通过无数次的尝试，我们最终大体确定了目前最适合该项目背景的超参数。

右图是模型优化前后各项指标随迭代次数的变化。可见优化前模型性能变化趋势很难收敛，在经过优化后表现明显趋于平缓，证明我们的工作聊有成效。

右图是我们训练的模型表现结果的可视化，左图为不同参数量下模型的性能表现汇总。可见在同一实验环境下，并非参数量越大的模型表现越好，我们推测可能有以下几个原因：小模型在当前训练环境下对于测试集达到了过拟合，即其有限的参数恰好只针对测试集的相似样本进行了拟合，从而相对获得了更好的性能表现。但这只说明该模型在特定环境具有较好的性能，而不能说明其具有良好的通用性能；第二个可能性是模型的训练次数较少。因为时间有限，我们只能把每个模型的训练轮次限制在100轮，这可能导致参数量较大模型还没有达到其最佳性能；最后一个可能是我们的超参数设置不当，因为不同参数量大小的模型可能有不同的最佳超参数设置，但我们没有足够的时间把每个模型的超参数都设置到最佳。

最后进入项目总结。本次项目我们虽然已经取得了阶段性成果，但是因为时间原因还有很多地方可以优化。经总结在未来我们可以基于模型从以下几个角度入手，从更加深入和层次去优化模型性能。

同时，研究表明模型表现中数据质量占很大一部分比重，因此还可以从数据的角度入手，从多个方面通过提升数据质量的方法来提升模型性能。

如果之后我们的成果要部署在嵌入式设备上，则需要做出一些相应的改动，可以从以下几个方面入手，最终目标是使其在嵌入式设备上在精度和速度上取得最佳平衡点。

我们在本次项目中学到了以下五个方面。首先是性能度量。每一种任务都有其特有的评判标准，但这些标准又是建立在共有的性质上，比如准确率，错误率等。只有明晰任务的评价标准的意义，才知道如何根据得到的数据来认识当下的不足和思考改进的方向；其次是默认的基准模型。确实基准很重要，一个高质量的基准可以更好的让自己通过对比来发现不足和优势，从而有针对性的提出优化目标；再次是收集数据。数据并不是越多越好，过多的数据可能造成模型不能很好的拟合从而导致性能的下降，因此选择合适的数据量很重要。除此之外，数据的质量一定要得到保证，否则可能会导致模型性能随着训练而下降；接下来是超参数。超参数是最能直接影响模型性能的数据，因此选取合适的超参数是很重要的。可以参考前人的经验，结合自己在训练中观察到的结果来建立自己对于超参数的选取准则；最后是调试策略。将结果可视化是最好的调试辅助，因为其能帮助我们快速直观的判断模型的性能表现，从而进一步做出相应的应对策略。

以上是我们本次答辩的全部内容，请各位批评与指正！