# 毕业设计（论文）开题报告

|  |
| --- |
| 1. 课题背景   建筑业是我国国民经济的重要产业，也是国民经济的重要物质生产部门。近些年来，随着我国经济的快速增长，建筑产业也一直保持着高速增长的姿态。建筑业作为劳动密集型产业，包含了大量的从业者，且建筑产品的生产现场存在大量的危险源以及复杂的环境，建筑产业的生产安全受到全社会的高度关注。根据国务院安全生产委员会办公室在2018年通报的上半年全国建筑业安全生产形式中指出：上半年全国建筑业共发生生产安全事故1732起，死亡1752人，同比分别上升7.8%和1.4%，事故总量已连续9年排在工矿商贸事故第一位，事故起数和死亡人数自2016起连续“双上升”。  建筑业复杂的生产环境和严峻的安全生产形式对自身健康发展产生了极为不良的影响，因此在建筑生产过程中，保证生产从业者对安全防护设备的正确佩戴显得尤为重要。  信息技术产业作为21世纪蓬勃发展的新兴产业，对生产生活的各个方面都产生了极大的影响。人工智能作为信息技术的一个分支，近些年来热度持续上升，各种利用人工智能技术的产品层出不穷。因此，可以利用人工智能技术来规范建筑产业从业者的安全防护设备的佩戴。  利用人工智能技术可以实现实时对象检测。目前，实时对象检测的应用面非常的广泛，比如自动驾驶汽车、人脸识别、智能机器人等等。而在上述检测建筑产业从业人员的安全防护设备的场景中，实时对象检测也有很大的发挥空间。  “You Only Look Once”或“YOLO”是一种对象检测算法，由Redmon等人在2016年发表的论文You Only Look Once:Unified, Real-Time Object Detection中提出。YOLO算法实现了实时对象检测。YOLO将对象检测重新定义为一个回归问题。它将单个卷积神经网络(CNN)应用于整个图像，将图像分成网格，并预测每个网格的类概率和边界框。YOLO算法处理图像的基本步骤是：调整输入图像的大小，在图像上运行单个卷积网络，根据模型的置信度对检测结果进行阈值化。这一模型具有很多优点，第一，YOLO的速度非常快，因为检测框被定义成一个回归问题，不需要复杂的管道。作者在论文中提到，他们在实验时可以达到在不到25毫秒的延迟时间内实时处理流视频，平均精度是其它实时系统的两倍以上。第二，YOLO在对图像进行预测时会有全局性的推理。与滑动窗口和region proposal技术不同的是，YOLO在训练和测试时可以看到整个图像，并能很好的利用上下文信息，因此YOLO能很好的避免背景错误。第三，YOLO具有高度的泛化性。YOLO在学习时可以得到高度泛化的特征，因此应用在新领域和意外输入时具有很好的稳定性。  YOLO算法会将输入的图片划分为s\*s个区域，每个区域负责检测中心落入其中的目标。每个区域会预测B个边界框（bounding box）以及边界框的置信度（confidence score）。所谓置信度其实包含两个方面，一是这个边界框含有目标的可能性大小，二是这个边界框的准确度。前者记为Pr(object)，当该边界框是背景时（即不包含目标），此时Pr(object)=0。而当该边界框包含目标时，Pr(object)=1。边界框的准确度可以用预测框与实际框（ground truth）的IOU（intersection over union，交并比）来表征，记为IOU。因此置信度可以定义为Pr(object)\*IOU。边界框的大小与位置可以用4个值来表征：（x,y,h,w），其中(x,y)是边界框的中心坐标，而和是边界框的宽与高。还有一点要注意，中心坐标的预测值(x,y)是相对于每个区域左上角坐标点的偏移值，并且单位是相对于区域大小的。而边界框的w和h预测值是相对于整个图片的宽与高的比例，这样理论上4个元素的大小应该在[0,1]范围。这样，每个边界框的预测值实际上包含5个元素：(x,y,w,h,c)，其中前4个表征边界框的大小与位置，而最后一个值是置信度。因此，每个区域需要预测B\*5+C（C为数据集中的类别），因为图片被划分为s\*s个区域，所以整张图片的预测结果是s\*s\*(B\*5+C)大小的张量。而在原论文中，作者取s=7，B=2，C=20，因为作者使用的PASCAL VOC数据集中有20个类，因此，最终预测结果是7\*7\*30大小的张量。  YOLO模型被实现为一个卷积神经网络，网络的初始卷积层从图像中提取特征，而完全连通层预测输出概率和坐标。YOLO的网络架构基于GoogLeNet模型的启发，拥有24个卷积层和2个全连接层。  YOLO的发展经过了几个阶段，第一个阶段，即上文提到的在2016年发表的论文中提出的，也被称为YOLOv1，这是YOLO的第一个版本。尽管YOLOv1实现了很多，但是仍有许多不足，比如对相互之间靠的很近的目标以及很小的目标检测能力偏弱，还有就是由于避免了一些背景错误导致检测的准确率偏低。因此他们改进了算法，提出了YOLO9000，也就是YOLOv2，相对于YOLOv1最大的改变就是使用了anchor box。增加了anchor准确率下降了，但是召回率上升不少。在这之后就是本设计使用的YOLOv3, YOLOv3使用了darknet-53的前面的52层（没有全连接层），YOLOv3的网络是一个全卷积网络，使用了大量的残差的跳层连接，并且为了降低池化带来的梯度负面效果，作者直接摒弃了POOLing，用conv的stride来实现降采样。在这个网络结构中，使用的是步长为2的卷积来进行降采样。  本设计是基于YOLOv3的安全帽实时目标检测的研究与实现，需要解决以下问题，首先是数据集的整理，需要采集相关图片，并将数据集进行标注，在这里使用的工具是LabelImg，该工具需要手动框选目标进行标注，并生成xml格式的标注文件。接下来就是处理数据集，使用python语言编写程序将数据格式化成训练框架需要的类型。在这之后就是训练，这里使用的darknet框架，训练好之后会得到权重文件，这是YOLO进行目标检测的基础。在编写代码之前需要对环境进行配置，首先需要安装Visual Studio 2017编译环境，然后安装和机器配置显卡版本相符合的CUDA和cuDNN，之后安装vcpkg包管理器，并使用vcpkg集成程序所需依赖。如此一来，准备工作就已经做好了，接下来就是正式的代码编写。这里使用的开发工具是Visual Studio Code，使用的语言是python，使用到的技术有flask框架和Vue框架。  本设计应该具有的功能有：接受输入的视频流，将视频流的每一帧处理成图片，使用YOLOv3算法对图片进行检测，返回处理结果到前端页面上，这是后端应该完成的工作。前端需要将输入的视频流显示在页面上，并根据用户选择显示处理后的图像。在前端应该给用户提供是否打开AI显示的按钮，根据用户的需要启动算法，以避免算法一直运行浪费大量的计算机资源。  本设计使用了YOLOv3算法进行实时目标检测，应比使用同类其他算法在速度和准确率方面有明显提高。  参考文献  (1)Redmon J, Divvala S, Girshick R. You only look once: unified, real-time object detection [C]. Proceedings of CVPR 2015, 2015: 779-788.  (2)Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger [C]. IEEE Conference on Computer Vision and Pattern Recogition. 2017: 6517-6525.  (3)Redmon J, Farhadi A. YOLOv3: An Incremental Improvement [C]. IEEE Conference on Computer Vision and Pattern Recogition. 2018.  (4)Lawal Mubashiru Olarewaju. Tomato detection based on modified YOLOv3 framework [J]. Scientific Reports, Volume 11, Issue 1. 2021.  (5)刘智嘉, 汪璇, 赵金博, 夏寅辉, 高旭辉. 基于YOLO算法的红外图像目标检测的改进方法 [J]. 激光与红外, 2020.  (6)王兵, 乐红霞, 李文璟, 张孟涵. 改进 YOLO 轻量化网络的口罩检测算法 [J]. 计算机工程与应用, 2021.  (7)Mohamed Loey a, Gunasekaran Manogaran, Mohamed Hamed N. Taha, Nour Eldeen M.Khalifa. Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection [J]. Sustainable Cities and Society, 2021.  (8)单美静, 秦龙飞, 张会兵. L-YOLO：适用于车载边缘计算的实时交通标识检测模型 [J]. 计算机科学, 2021.  (9)Weijun Chen, Hongbo Huang, Shuai Peng, Changsheng Zhou, Cuiping Zhang. YOLO-face: a real-time face detector [J]. The Visual Computer, 2021.  (10)刘素行, 吴媛, 张军军. 基于YOLOv3的交通场景目标检测方法 [J]. 国外电子测量技术, 2021. |
| 二、毕业设计（论文）研究方案  本设计是基于YOLOv3的安全帽实时目标检测的研究与实现，需要解决以下问题，首先是数据集的整理，需要采集相关图片，并将数据集进行标注，在这里使用的工具是LabelImg，该工具需要手动框选目标进行标注，并生成xml格式的标注文件。接下来就是处理数据集，使用python语言编写程序将数据格式化成训练框架需要的类型。在这之后就是训练，这里使用的darknet框架，训练好之后会得到权重文件，这是YOLO进行目标检测的基础。darknet框架是YOLO的作者自己编写的一个深度学习框架，在YOLOv3中使用的是darknet-53。在编写代码之前需要对环境进行配置，首先需要安装Visual Studio 2017编译环境，然后安装和机器配置显卡版本相符合的CUDA和cuDNN。之后安装vcpkg包管理器，并使用vcpkg集成程序所需依赖。  如此一来，准备工作就已经做好了，接下来就是正式的代码编写。这里使用的开发工具是Visual Studio Code，后端程序使用的语言是python，使用到的技术有Flask框架。Flask 是一个微型的 python 开发的 Web 框架，基于Werkzeug WSGI工具箱和Jinja2 模板引擎。 Flask使用BSD授权。 Flask也被称为“microframework”，因为它使用简单的核心，用extension增加其他功能。Flask没有默认使用的数据库、窗体验证工具。然而，Flask保留了扩增的弹性，可以用Flask-extension加入这些功能：ORM、窗体验证工具、文件上传、各种开放式身份验证技术。前端页面使用HTML语言进行编写，使用到的技术有Vue框架和Bootstrap框架。Vue.js是一套构建用户界面的渐进式框架。Vue只关注视图层，采用自底向上增量开发的设计，目标是通过尽可能简单的API实现响应的数据绑定和组合的视图组件。Bootstrap 是一个用于快速开发 Web 应用程序和网站的前端框架。Bootstrap 是基于 HTML、CSS、JavaScript的。 |
| 三、毕业设计（论文）预期成果及创新  本设计是基于YOLOv3的安全帽实时目标检测的研究与实现，需要解决以下问题，首先是数据集的整理，需要采集相关图片，并将数据集进行标注，在这里使用的工具是LabelImg，该工具需要手动框选目标进行标注，并生成xml格式的标注文件。接下来就是处理数据集，使用python语言编写程序将数据格式化成训练框架需要的类型。在这之后就是训练，这里使用的darknet框架，训练好之后会得到权重文件，这是YOLO进行目标检测的基础。在编写代码之前需要对环境进行配置，首先需要安装Visual Studio 2017编译环境，然后安装和机器配置显卡版本相符合的CUDA和cuDNN，之后安装vcpkg包管理器，并使用vcpkg集成程序所需依赖。如此一来，准备工作就已经做好了，接下来就是正式的代码编写。这里使用的开发工具是Visual Studio Code，使用的语言是python，使用到的技术有Flask框架和Vue框架。  本设计应该具有的功能有：接受输入的视频流，将视频流的每一帧处理成图片，使用YOLOv3算法对图片进行检测，返回处理结果到前端页面上，这是后端应该完成的工作。前端需要将输入的视频流显示在页面上，并根据用户选择显示处理后的图像。在前端应该给用户提供是否打开AI显示的按钮，根据用户的需要启动算法，以避免算法一直运行浪费大量的计算机资源。  本设计使用了YOLOv3算法进行实时目标检测，应比使用同类其他算法在速度和准确率方面有明显提高。 |
| 指导教师审核意见：  指导教师（签名） 2021 年 1 月 9 日 |
| 教研室审核意见：  教研室主任（签名） 2021 年 1 月 9 日 |

注：此表中的一、二、三项，由学生在教师的指导下填写。