

学号：2015019051

# 北京化工大学

## 毕业设计（论文）

论文题目：危化工场景下火苗识别方法研究

学院名称：国际教育学院

专业：机械设计制造及其自动化

学生姓名：寇露丹

导师姓名：胡标

## 诚信声明

本人声明：

本人所呈交的毕业设计（论文），是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

作者签名： \_\_\_\_\_ 日期： \_\_\_\_\_

## 毕业设计(论文)任务书

设计(论文)题目: 危化工情境下火苗识别方法研究

学院: 国际教育学院 专业: 机械设计制造及其自动化 班级: 国机 1502

学生: 寇露丹 指导教师: 胡标副教授 专业负责人: 马秀清副教授

### 1. 设计(论文)的主要任务及目标

熟悉和掌握深度学习网络理论基础和识别流程,对不同网络效果进行调研,查阅相关文献和代码,调研危险化工过程的火情情况;搭建深度神经网络识别火情,并给出火情种类。

### 2. 设计(论文)的主要内容

- (1) 危险化工环境容易出现火情,研究火情种类。
- (2) 设计一套基于深度学习的识别算法识别火情。

### 3. 设计(论文)的主要要求

- (1) 完成 5000 汉字的与设计内容有关的英文资料的翻译。
- (2) 阅读相关文献,完成毕业毕业论文。

### 4. 主要参考文献

- [1] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. arXiv, 2018.
- [2] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. arXiv preprint arXiv:1506.01497, 2015.
- [3] Zhen Z , Minjuan W , Yukun S , et al. A convolutional neural network-based flame detection method in video sequence[J]. Signal, Image and Video Processing, 2018.

### 5. 进度安排

	设计(论文)各阶段名称	起止日期
1	收集相关资料、阅读相关文献、查阅相关资料。	2019.01.23~2019.02.25
2	书写开题报告、外文翻译、任务书等	2019.02.26~2019.03.04
3	训练网络、编写程序、修改程序、得出相关结论	2019.03.05~2019.04.30
4	根据相关资料和结论书写论文。	2019.05.01~2019.05.15
5	修改论文、检查格式、制作 PPT、熟悉答辩流程。	2019.05.16~2019.05.31

## 危化工场景下火苗识别方法研究

### 摘 要

如今,深度学习已经在图像分类和计算机视觉任务的其他方面有最先进的表现。它们在火苗探测系统中的应用将最大提高检测精度从而减少火灾,减少生态损失。本文旨在探究可以高效完成检测效果的神经网络识别方法,从而可以实现搭载于机器人的平台等,而实现实时检测效果。由于真实数据集较难采集、数量较少,本文使用虚拟数据集进行预训练,训练网络为 YOLOV3 和 faster RCNN,进行效果比对。为了实现实时的效果,使用 YOLOV3 进行进一步训练。为了达到真实场景的应用效果,使用真实数据集进行迁移学习训练,从而达到了使用较少训练数据、训练资源达到识别效果。最后,使用 PyTorch 框架自主实现 YOLOV3 网络,将实现的网络效果与原网络进行效果比对。

**关键词:** 深度学习, 图像检测, 火苗检测

# **REARCH ON FIREARM IDENTIFICATION METHOD IN DANGEROUS CHEMICAL ENVIRONMENT**

## **ABSTRACT**

Today, deep learning has the most advanced performance in image classification and other aspects of computer vision tasks. Their use in fire detection systems will maximize detection accuracy, thereby reducing fire and reducing ecological losses. The purpose of this paper is to explore a neural network recognition method that can efficiently perform detection effects, thereby realizing a real-time detection effect by implementing a platform mounted on a robot. Since the real data set is difficult to collect and the number is small, this paper uses the virtual data set for pre-training, and the training network is YOLOV3 and faster RCNN for comparison. In order to achieve real-time effects, use YOLOV3 for further training. In order to achieve the application effect of the real scene, the real data set is used for the migration learning training, thereby achieving the recognition effect by using less training data and training resources. Finally, the PYL Torch framework is used to implement the YOLOV3 network autonomously, and the network effect achieved is compared with the original network.

**KEY WORDS:** deep learning, image detection, flame detection

## 目 录

前言.....	(1)
第 1 章 概论.....	(2)
第 1.1 节 火苗识别方法的发展概况.....	(2)
第 1.2 节 火苗识别方法的发展及研究现状.....	(2)
1.2.1 火苗识别方法的展展.....	(2)
1.2.2 火苗识别方法的研究现状 .....	(3)
第 2 章 火苗识别的不同深度学习方法实验.....	(5)
第 2.1 节 火苗识别数据集及实验环境介绍.....	(5)
第 2.2 节 YOLOV3 方法.....	(5)
第 2.3 节 faster RCNN 方法.....	(8)
第 2.4 节 迁移学习.....	(11)
第 2.5 节 基于 PyTorch 的 YOLOV3 方法.....	(14)
第 2.6 节 训练过程与结果对比.....	(15)
2.6.1 训练过程.....	(15)
2.6.2 结果对比.....	(15)
结论.....	(19)
参考文献.....	(20)
致谢.....	(22)

## 前 言

最近，图像识别已经引入了各种传感器不同的应用，如火警检测<sup>[1]</sup>，车辆障碍物检测，可视化内部人体诊断<sup>[2]-[4]</sup>，动物和船舶监测和监视<sup>[5]</sup>。在这些应用中，检测得到了研究人员的重视。使用智能监控系统，如道路交通事故，火灾，医疗突发事件等各种事故都可以在早期阶段被检测到，并通过结果可以自主报警<sup>[6]-[7]</sup>。火灾是可能在很短的时间内对生命造成重大损害的事件<sup>[8]</sup>。此类灾难包括人为错误或系统故障导致严重的人命损失和其他损害<sup>[9]</sup>。在欧洲火灾每年造成 10 000 平方公里的植被被毁坏；在北美和俄罗斯，损害高达 10 万平方公里。2013 年 6 月，火灾造成 19 名消防员死亡并破坏了美国亚利桑那州的 100 所房屋。同样，2013 年 8 月另一个在加利福尼亚的森林大火毁坏了一片土地面积 1042 平方公里，造成亏损 1.273 亿美元<sup>[10]</sup>。根据年度灾难报告<sup>[11]</sup>，受火灾受影响达 494000 人，导致 31 亿美元的损失。在 2015 年。为了避免这种灾难，在早期阶段利用智能监控探测火灾这很重要。

人们的生活离不开化工品。化工厂为人们的生活带来便捷的同时，也为我们的生活带来了危险性，比如一些人为或者化学制品反应而造成的燃烧反应。火灾发生时，如果不能及时扑灭火苗，就会造成巨大的人力、物力、财力及环境破坏的损失。使用机器人平台搭载识别系统，可及时监控工厂的情况，并阻止已产生的危险进一步的扩大，可减少不必要的损失。

## 第 1 章 概论

### 第 1.1 节 火苗识别的发展概况

火苗识别大致有定两大类方法：1) 传统火警和 2) 图像检测火灾探测。传统的火灾报警系统基于传感器的使用。物体燃烧时会产生热量、强光、微小烟雾颗粒等，以及伴随着会产生红外线、紫外线等不可见光。传统检测系统会对现场的烟雾浓度、热量、各种不可见光的强度等因素与设定的阈值做对比，从而判断是否发生火灾并进行警报。根据此原理，可以将他们分温度、光强、烟雾、气体、及复合检测类型。这些传感器不具备定位功能，需要人为参与，在拉响警报的情况下确认火灾的情况，从而确定火的位置。此外，这样的系统不能通常提供有关大小和燃烧的程度等具体信息。

如今，随着图像检测的大力发展，因为图像检测具有较高的准确性，可以提供准确信息，越来越多的研究员投身到此领域，对火焰的特征进行研究。

### 第 1.2 节 火苗识别方法的发展及研究现状

#### 1.2.1 火苗识别方法的发展

在早期没有大力推广视频设备的时候，研究人员是通过燃烧产生的烟雾粒子、热量和燃烧时发出的强光来检测是否有火焰产生。检测火焰使用的传感器通过对温度，光强和烟雾颗粒的浓度等进行监测，当人工设定的阈值小于监测结果时，即已经有火灾发生，预警装置立即启动。但这些监测设备都有些致命的缺陷，如传统的温度、光强装置只能监测着火点部分区域，否则无法高效探测到火苗的发生，而对于很大一部分情况下，人们无法对哪个地点发生火灾产生预先且准确的判断。其次，这些传感监测装置的探测结果非常容易受所处位置的影响，如在天气异常的情况下、场地开阔时，这些情况以及一些人为原因都会导致监测结果出现差错。并且，火焰检测装备由于经常安放在易损害电子元件的环境下，设备会产生老化的情况，从而会严重影响传感元件的敏感度，出现误判情况。随着视频设备的普及和科学的发展进步，传统的火焰检测设备已经越来越不能满足人们的需求，越来越多的研究人员开始探究基于视频设备的研究方法，极大的提高了结果的准确性。



### 1.2.2 火苗识别方法的研究现状

如今,许多研究人员开始探究基于图像识别系统的方法研究<sup>[12]-[15]</sup>;这些都有其优点人为干扰少,响应速度快,价格合理,并且有更大的监控范围。此外,这样的系统可以确认火灾并且不需要确定火灾的位置,可以提供有关火灾的详细信息,包括它的位置,大小,程度等。尽管有这些优点,这些系统仍然存在一些问题,例如复杂性观察中的场景,不规则的照明,而且基于传统图像识别的检测框架质量较低,并没有较高的鲁棒性;研究人员已经做了几次努力解决兼顾颜色和运动特征这些方面。

陈等人<sup>[16]</sup>研究了火灾的动态行为使用 RGB 和 HSI 颜色模型并提出了一个规则的辅助火灾探测方法,它用于检测火灾的不规则特性。他们的方法是基于帧到帧的差异,因此不能区分火灾和火色移动区域。马尔巴赫等人<sup>[17]</sup>使用了研究 YUV 颜色空间运动信息将像素分类为火灾和非火灾成分。Töreyin 等<sup>[18]</sup>使用时间和空间小波确定火灾和非火灾区域的分析。他们的方法使用许多启发式阈值,这极大地限制了它的真实的差异。汉和李比较用于隧道火灾探测的彩色信息帧;此方法仅适用于静态火灾,因为它有大量参数。Çelik 和 Demirel<sup>[19]</sup>探索了 YCbCr 色彩空间并提出了一种像素分类检测火焰,为此,他们提出了新的分离规则色度和亮度分量的方法。然而,他们的方法无法从远距离检测到火灾在小规模,这在早期火灾检测中很重要。除了这些基于色彩空间的技术,Borges Izquierdo<sup>[20]</sup>利用了低级特征包括颜色,偏斜度和粗糙度用于火灾识别的贝叶斯分类器。福贾等人<sup>[22]</sup>提出了基于 YUC 颜色模型的<sup>[21]</sup>的修改版本,获得了比 RGB 版本更好的结果。另一种基于颜色的类似方法信息和 SVM 分类器在中给出。这个方法可以处理 20 帧/秒;但是,它无法检测到大火或小尺寸的火灾,这些都可能发生在火灾中现实世界的监控录像中。通常基于颜色的方法由于阴影和阴影的变化而产生更多的错误警报亮度,并经常错误分类穿红色衣服的人或红色车辆。Mueller 等试图解决这个问题通过分析火灾形状和变化的问题刚性物体的运动。他们的算法可以区分在刚性移动物体和火焰之间,基于特征从光流和物理中提取的矢量火的行为。DiLascio 等结合颜色和用于检测火灾的运动信息视频。Dimitropoulos 等使用时空特征基于纹理分析,然后是 SVM 分类器将视频帧的候选区域分类为火和无火。此方法严重依赖于用过的参数;例如,小尺寸的块会增加错误率警报,而较大的块会降低其灵敏度。同样的,时间窗口对此的表现也至关重要系统;较小的值会降低检测精度较大的值会增加计算复杂性。这些依赖性极大地影响了这种方法的可行性在实际监控系统中

实施。最近，福贾等人提出了一种实时火灾探测算法基于颜色，形状和运动特征，在 multiexpert 系统中结合起来。这种方法的准确性高于其他方法;但是，误报的数量仍然很高，火灾探测的准确性也可以进一步改进。对现有文献的调查显示计算上昂贵的方法具有更好的准确性，更简单的方法会影响准确性和速率误报。因此，需要找到更好的权衡这些指标之间的几个实际应用场景兴趣，现有计算成本高昂方法不合适。

## 第 2 章 火苗识别的不同深度学习方法实验

### 第 2.1 节 火苗识别数据集及实验环境介绍

本节主要针对危化工场景下的火苗识别实验中所采用的数据集和实验环境进行简要介绍。

本实验的机器配置为 GPU:GTX 1080ti, 操作系统: UBUNTU16.04, 使用的深度学习框架为 Darknet, Caffe, Keras 和 Pytorch。编程语言主要是 Python。

本实验的数据集分为两部分, 第一部分视频资源是基于 UNITY 软件的仿真化工场景图片, 分辨率为  $960 \times 544$ , 格式为 MP4 格式。第二部分是真实数据集, 实验视频集均来自于 Bilkent 大学火焰数据集库, 视频分辨率为  $320 \times 240$ , 格式为 MP4 格式。

由于本实验是基于小型火焰, 且化工场景下的火情场景较难拍摄, 所以使用仿真软件制作的数据, 仿真软件可制造大量图像数据集, 再用此数据集进行预训练。为了达到在真实场景使用的效果, 使用真实场景(如森林小型火灾)等进行训练。因其数据集数量小, 很难在在深度学习的场景下达到拟合状态, 故使用虚拟场景下训练好的网络进行迁移学习, 可大大节省时间和计算资源, 从而达到理想效果。

虚拟场景数据集包含 4513 张图片, 考虑到火焰的多样性, 数据集包含 3 个类别的火焰, 平均每个类别的图片数量为 1500 张, 测试集、训练集比例为 7:3。真实场景下数据集包含 1600 张, 有两个类别的火焰, 测试局训练集划分为 7:3。数据集格式为 VOC2007。

### 第 2.2 节 YOLOV3 方法

YOLO 网络共有 3 个版本, 基础的 YOLO, 改进后的 YOLOV2 和 YOLOV3。其中, YOLOV3 的结果较基础的 YOLO 网络有了极大的提升。YOLO 是基于 GoogleNet 的基础网络进行进一步改进, 它并没有使用 inception 结构。它的基础网络共有 24 个卷积层, 后面连接两个全连接层, 并且它使用了  $1 \times 1$  的卷积核。

在论文中, 作者训练网络时先用 ImageNet 预训练了前 20 个卷积层, 训练好分类网络之后, 作者将图像修改成  $448 \times 448$ , 已达到检测需要的较大分辨率的效果。另外,

YOLO 使用的激活函数为 Leaky Rectified Linear 激活函数，而没有使用常用的 ReLU 函数。在损失函数上，作者选用了均方误差作为损失函数。

如图 2-1 所示为 YOLO 结构图。

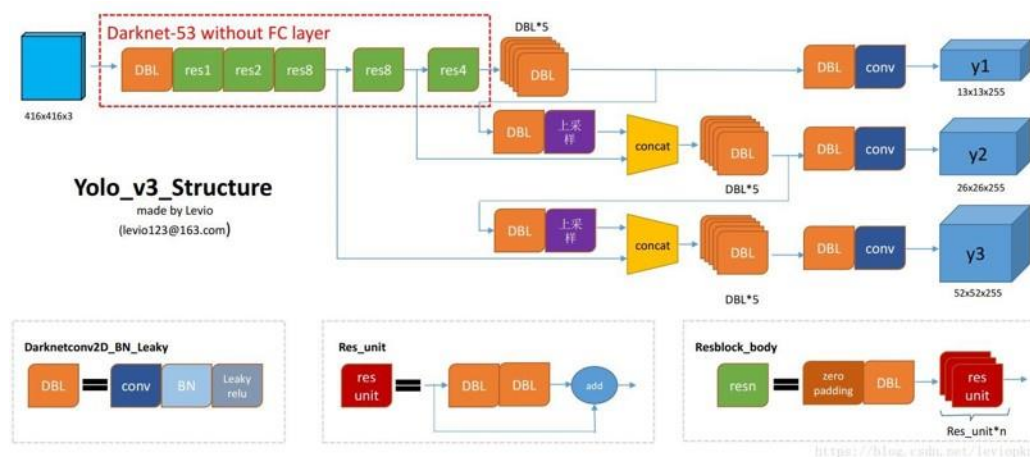


图 2-1 YOLO 结构图

此外，YOLO 网络对每一层使用了批归一化，提高了收敛速度，增加了准确度。在初始的 YOLO 中，bounding box 的坐标是通过全连接层进行预测的，而 YOLOV2 参考了 Faster R-CNN 的做法，引入了 anchor。首先将网络的池化层和全连接层去掉，使最后的卷积层获得了更高的分辨率的特征图，再缩减网络，用  $416 \times 416$  的大小进行训练。YOLO 将图片划分成  $7 \times 7$  的网格，通过引入 anchor，预测的格子的大小超过了 1000 个。增加了预测格子的数目可极大的提高准确率。Anchor box 的大小和比例是通过经验设定的，YOLO 网络通过使用 k-means 进行聚类，从而找到了 Anchor box 的大小。

对比 YOLO 的几个网络，YOLOV1 检测速度比 R-CNN 要快很多，但是在检测精度却远不及其程度，YOLOV1 在定位方面存在不够准确的问题，且无法检测出小物体。YOLOv2 针对此提出了几种改进策略来进一步提升网络的准确率和精度，从而提高 mAP，YOLOv2 在改进中遵循一个原则：保持检测速度，这也是 YOLO 模型的一大优势。YOLOv2 的优势也是他的检测速度，所以他的改进策略是保持其检测速度，进一步提高精度。

在 YOLOV1 网络中，输入图片被分成  $7 \times 7$  的格子，每个单元格预测得出两个预测框。YOLOV1 对边框的预测使用的是全连接层，而其中对边框的边长是针对对整张图片的比例进行计算。由于在图片中的物体存在不同的尺寸，形状不同，YOLOV1

在适应不同物体尺寸进行网络训练学习比较困难，因此 YOLOV1 在定位的精确性上表现较差。YOLOV2 借鉴了 RPN 中的提取先验框的方法，并且移除了其中一个池化层，而达到了所检测的特征图分辨率更高。在网络模型中，YOLOV2 放弃了  $418 \times 418$  的图像输入，而是采用  $416 \times 416$  的大小。由于 YOLOV2 下采样的步长是 32，所以，对于  $416 \times 416$  大小的图像，得到的特征图为  $13 \times 13$ ，以便得到特征图的中心位置。对于大的检测目标，其中心点往往落入图像的中心位置，所以，用一个中心点去预测相应物体的边框位置较容易，因此，YOLOV2 最终确定特征图有奇数个位置。对于 YOLOV1，每个单元格预测了两个边框，每个边框包含五个参数值(x,y,w,h,c)，x,y,为边框的位置，w,h 为边框的长宽大小，c 为边框置信度，其中包含了含有任何可能物体的概率及预测框与检测目标的重叠概率。但是每个单元格的两个边框只会预测一份分类概率，进行共享。YOLOV2 网络使用了 anchor box 之后，每个位置的预测框都会分别预测一份分类概率，这点和 SSD 相似。

YOLO 采用了多级检测，改进后的 YOLO 网络增加了自顶至下的多级预测，解决了 yolo 因图片划分太过不精细而产生的漏检测小物体的问题。YOLOV2 的检测器只有 1 个，YOLOv3 增加到了 3 个，分别为一个下采样，两个上采样，特征图分别为  $13 \times 13$ ， $26 \times 26$ ， $52 \times 52$ 。YOLOV3 的特征图最大可到  $52 \times 52$ ，而 YOLOv2 把数据的处理用到上采样上，所以只用到了  $16 \times 16$  的特征图，因此，在 YOLOV2 上容易发生小物体的漏检。且论文从单层网络预测 5 个预测框，变成了预测 3 个。

此外，YOLO 网络还对损失函数进行了改写。在 YOLOV3 网络中，将 YOLOV2 的损失函数 softmax 函数改写成了 logistic 函数。Logistic 函数可在分类时可输出多个结果，从而在最终达到同时分类多个目标的结果。由于每个点的对应的预测边框差异大且数量少，每个预测边框与 softmax 函数的结果对应为 1 比 1。故当预测的目标类别较为细小庞杂时，采用 logistic 回归函数是更有效的，比如在论文中提出的，在 Open Images Datase 数据集中，许多标签相互重叠，使用 Softmax 表示每个候选框只能对应一个类别，使得检测效果变差，无法达到预期的效果。所以，使用 logistic 回归函数可以将数据标签更好的拟合，得到正确的建模效果。

YOLOV3 的进一步改进是加深了网络。新的 YOLO 网络采用简化的残差块代替了原来  $1 \times 1$  和  $3 \times 3$  的模块，即加了一个短路 (short cut) 部分，从而达到了既加深了网络又防止网络系数爆炸。从 YOLOV2 的 19 层网路到 YOLOV3 的 53 层网络进行了上采样，卷积层的数目变多了。此外，作者还用了连续的  $1 \times 1$  和  $3 \times 3$  的过滤器

增加通道数目。 $1 \times 1$  的卷积核的作用是压缩  $3 \times 3$  的过滤器的卷积结果。此外,YOLOV 网络又新增了路由器模块,用于拼接相应层的过滤器参数。

YOLOV3 的 anchor boxes 和 YOLOV2 处理方法相同,通过 kmeans 聚类得到。YOLOV3 对每个预测框预测四个值(tx,ty,tw,th),将预测的图片划为分  $S \times S$  个网格框,根据左上角的偏移值(cx,cy),以及之前聚类得到的预测框的宽 pw 和高 ph,进行计算。其中,在训练时,几个坐标值均采用了均方误差损失。YOLOV3 通过通过每个框进行逻辑回归的打分的比较来预测这个物体是否为目标物体,及其类别。如果真实的边框与预测的边框的重合度高,且比其他的预测边框重合的要好,那么输出的回归值即为 1。如果所有重叠部分均没有达到阈值,则这个预测框就会被忽略,即损失值为 nan。改进后的 YOLO 网络既能最大限度保证识别的准确性,又可以保证识别的速度,可做到实时检测的效果。

### 第 2.3 节 faster RCNN 方法

Faster RCNN 作为一个 2015 年推出的目标检测算法如今仍有极高的价值。由它推出的网络结构层出不穷,具有强大的生命力。Faster RCNN 在其他领域也应用广泛,比如图像分割,描述,目标追踪等领域,为科学界,工业界带来了许多研究方向和新的思路。

Faster RCNN 网络首先对数据进行处理,它会缩放图像,使其分辨率在  $1000 \times 600$  以内,对相应的预测框也会进行通尺度的缩放。由于 Faster RCNN 是基于 Caffe 的 VGG16 预训练模型,其像素值需在 0-255 之内,三原色的排列为 BGR 格式,再对其三通道进行处理,使图像像素均值为 0。

一般是以下四个值用于网络训练,第一个是图像值(images),向量大小为  $3 \times H \times W$ ,BGR 三个通道,宽为 W,高为 H。第二个是预测框的值(bboxes),向量大小为  $4 \times K$ ,K 的意思是预测边框,每个预测边框的数值为每个预测边框左上角和右下角的坐标值( $Y_{min}, X_{min}, Y_{max}, X_{max}$ ),Y 和 X 表示第 Y 行,第 X 列。第三个数值是标签值,向量大小为 K,表示 K 个预测边框的标签值(对于 VOC 的取值范围是 [0-19])第四个值是缩放倍数,将原图  $H' \times X'$  缩放到人为规定的  $H \times W$ 。需要注意的是,目前大多数 Faster RCNN 的实现是基于批大小为 1 的训练。

Faster RCNN 使用提取器进行提取特征,其中,训练时提取器提取特征使用的是与训练好的模型。论文里使用的是基于 Caffe 的 VGG16 模型,并进行了修改,将前

四层卷积层的学习率设置为 0。

Faster RCNN 的一大创新点在于提出了 RPN, RPN 用于代替传统网络中的 Selective Search, 从而极大的减少了时间开销。在 RPN 中, 作者提出了一个概念, Anchor, Anchor 是一种特殊的候选框, 其大小和尺寸均确定。论文中使用了  $128 \times 128$ ,  $256 \times 512$  和  $512 \times 256$  三种尺寸的组合框, 其组合结果共有 9 种 anchor。9 种 anchor 在特征图中进行移动, 并且每个特征图的点都有对应的 9 个 anchor。特征图中 9 个 anchor 进行上下左右个方向移动, 即每个特征图都有 9 个 anchor。经过 16 倍下采样之后, 生成  $(H/16) \times (W/16) \times 9$  个 anchor。例如  $512 \times 62 \times 37$  的特征图即生成 20000 个 anchor。

不同的特征图对应的 anchor 数量不一样, 即 Anchor 的数量和特征图的尺寸有关。RPN 在提取器输出的特征图的基础上, 增加了卷积层, 并进行二分类和位置回归, 其使用的是两个  $1 \times 1$  的过滤器。进行分类的过滤器的通道数是  $9 \times 2$ , 使用交叉熵回归, 并且有  $9 \times 4$  个回归卷积核通道数。由于 RPN 是一个全卷积网络, 这样对图像的输入尺寸没有要求。

接下来 RPN 利用 Anchor Target Creator 将上述 20000 个候选框进行筛选, 选出 256 个 anchor 进行回归和分类。对于每一个目标候选框, 选择的正样本为与它重叠度最大的 anchor。再剩下的 anchor 中选择正样本, 即从中选择和任意一个目标候选框重叠度大于 0.7 的 anchor, 其数目不超过 128 个。再选择负样本, 随机选择和目标候选区域重叠度不超过 0.3 的 anchor, 其正样本数目和负样本数目总和不超过 256 个。每一个 anchor 和目标检测框由前景为 1, 背景为 0 组成。而其位置由  $(tx, ty, tw, th)$  四个位置参数组成, 其效果好于直接回归位置坐标。在 Faster RCNN 中计算回归损失用的是 smooth L1 损失, 只计算前景损失, 不计算背景损失。而分类损失用的事交叉熵损失。

训练 RPN 时, 还会给 Fast RCNN 提供 RoIs 进行训练。RPN 生成 RoIs 对每张图片利用其特征图计算其 anchor 的位置参数及正样本概率。再选取概率较大的 12000 个 anchor, 进一步使用回归修正其位置参数, 得到 RoIs。利用极大值抑制将此 12000 个 anchor 进行处理, 再选出概率最大的 2000 个 RoIs, 并将其最为返回值返回。进一步训练时, Fast RCNN 的 RoI Head 在 RPN 的与选矿中再进行进一步的分类和位置回归。因为 RoIs 给出的 2000 个候选框分别位于特征图的不同大小及位置, 所以首先利用 PTC 挑选出一部分 ROIs, 然后再使用 RoI 池化将其尺寸缩放至同一大尺寸上。RoI 池化是一种特殊的池化操作, 对于特征图的大小是  $(512 \times H/16 \times W/16)$ , 相当于把其批大小

置为 128，通道数为 512，大小为  $7 \times 7$  的特征图。

Faster RCNN 的网络结构分为如下几个部分。

第一部分是卷积层。作为基于 CNN 的目标检测网络，Faster RCNN 首先用几个基础模块（卷积层+ReLU+池化层）提取图片的特征图，并一一用于后续的提取 anchor boxes 的 RPN 层和全连接层。卷积层的特征图为 256 或者 512 张，由于其过滤器是  $3 \times 3$ ，所以卷积核的尺寸是  $3 \times 3 \times 256$  或  $3 \times 3 \times 512$ 。

第二部分是 RPN 网络。RPN 网络用于生成位置的预测，其原理是通过 softmax 判断 anchors 是属于前景还是背景，并利用边框回归修正 anchors 获得精确的预测，此网络包含前景背景的判断和对边框的修正两部分。

第三部分是 RoI 池化。此层收集输入的特征图和预测值，结合这两部分信息后提取预测的特征图，再将其值放入全连接层，从而判定目标类别。RoI 池化层主要是解决各个预测尺寸不同，而直接改变其大小会破坏其特征。

第四部分是分类。利用预测的特征图计算预测目标的类别，同时，通过边框回归获得准确的检测目标预测位置。分类是利用已获得的预测特征图来预测目标属于哪个类别，其判断是通过全连接层的 softmax 来计算，输出概率向量。同时利用边框回归获得每个目标的位置偏移，并通过再次计算获得更精确的目标检测框。

Faster RCNN 由 RCNN 和 Fast RCNN 发展而来，因其发现 selective search 导致检测速度过慢，达不到实时的效果，因此，作者采用 RPN 网络取代 SS 方法，并通过共用 fast RCNN 的分类和回归网络的提取层，减少了计算量。实验证明，Faster RCNN 的速度和准确率都得到了提高。故 RPN 是 Faster RCNN 的重点所在。也因此，有了 RPN 的存在，faster RCN 才得以在速度上慢于其他主流网络，而精度大于其他网络。

如图 1-2 为 faster RCNN 结构图。



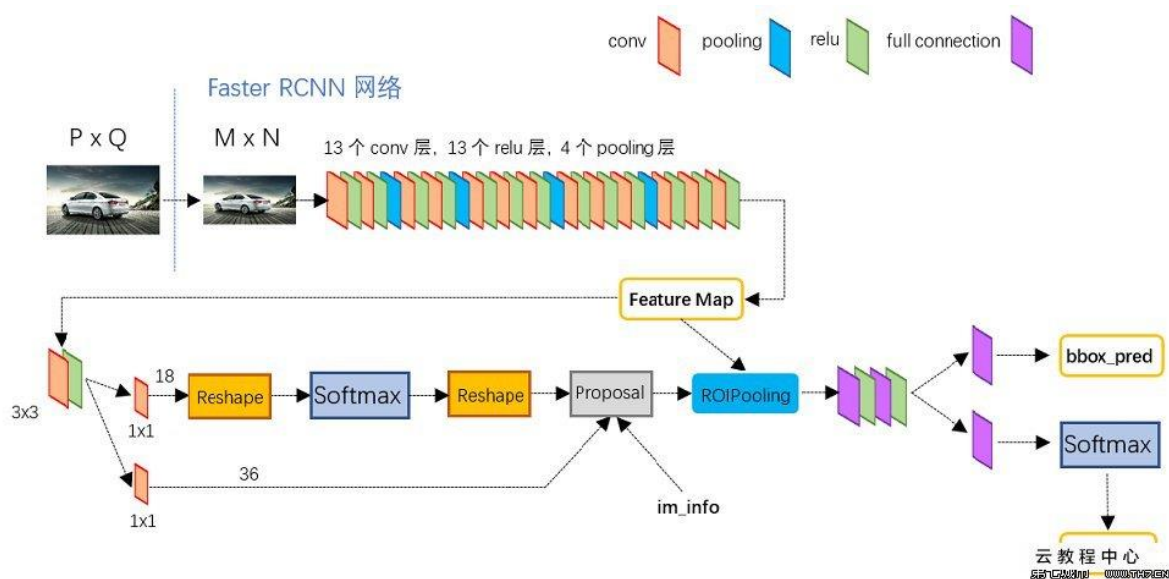


图 2.2 faster RCNN 结构图

## 第 2.4 节 迁移学习

迁移学习即为将由某个数据集已经训练好的模型参数用于辅助训练另一个任务的新的模型来进行训练。由于数据集有较大相关性，通过迁移学习我们可以将我们已经训练得到的参数迁移给新模型加速其训练速度，不需要从头开始重新训练。

在深度学习分类任务中，一般有两个基本假设：（1）新的测试样本和用于学习的样本基于同样的独立分布，即两样本出于相同类别，（2）需要大量的数据集才能训练出来一个好模型。但是实际场景下，这两个条件往往无法完全满足。首先，标签数据可能会首先，经过一段时间之后，原本可用的标签可能因为客观原因而与要标注标签产生语义上等影响，且在工业上更新图片标签需要消耗大量的人力资源。另外，标签样本十分缺乏，而且获取困难。这也就引出了研究人员关注的另一个问题，如何用少量样本数据训练出基于大量标签的理想效果的网络。迁移学习是针对已学习的知识对另一个不同但是存在共性的领域来解决新的任务的机器学习方法。它减少了机器学习中两个假设的影响，让网络通过已经学到的知识来解决目标的新任务领域的问题。如今，迁移学习的定义即为使用以前学习的知识运用到新的领域中。

对迁移学习有所了解，要先了解一下几个概念，域，源，任务以及目标：

域包括两个部分，特征空间和概率。当域有所不同时，分为两种情况，一种为空

间特征不同，另一种为空间特征相同，概率不同。任务也分为两个部分：标记空间和目标预测函数。当任务不同时，同样分为标记空间不一样，或者标记空间一样但是目标预测函数不同。源用于训练模型的域/任务，任务是要用已有的模型对自己的数据进行机器学习的源或任务进行各项操作，如分类，预测等。

迁移学习广泛应用于人类的行为中，两个不同的任务相同的因素越多，学习的迁移就越容易，反之越困难，甚至产生副作用。所以，迁移学习是具有局限性的，我们需要明白迁移学习的范围在哪了，以及其源和任务之间的可迁移性。随着科技的发展，很多研究人员对迁移学习开展的大量的研究，使用不同的技术、算法等对迁移学习展开研究。

迁移学习已在文本的分类，图像分类，识别等各个方面起到了很大作用，迁移学习的表示形式及其分类分为以下方面。

基于实例的迁移学习方法代表有 Dai 等人[3]提出的基于实例的 TrAdaBoost 迁移学习算法。当目标域中的样本被错误地分类之后，可以认为这个样本是很难分类的，因此增大这个样本的权重，在下一次的训练中这个样本所占的比重变大。如果源域中的一个样本被错误地分类了，可以认为这个样本对于目标数据是不同的，因此降低这个样本的权重，降低这个样本在分类器中所占的比重。Masashi Sugiyama 等人[4]，考虑了模型选择的重要性，在协变量变换下，根据测试和训练输入密度的比例一致，标准学习方法（如最大似然估计）不再是一致加权变量，首先分别估计训练和测试输入密度，然后通过取估计密度的比率来估计重要性。

基于特征选择的迁移学习方法和基于特征映射的迁移学习方法组成了基于特征的迁移学习方法。首先，基于特征选择的迁移学习方法通过识别目标领域和源领域的相同分布特征，再利用这些特征进行参数的迁移。Jiang 等人对于和目标领域分布相同的特征权重应该赋予更高的得分，获得更高的权值，因此，他们在领域适应任务中提出了一个两阶段的特征选择网络。基于实例的迁移学习方法表示由 Dai 等人提出的基于实例的 TrAdaBoost 迁移学习算法。在目标域中的样本被错误分类之后，可以认为样本难以分类。因此，该样品的重量增加，并且在下一次训练中该样品的比例变大。如果源域中的样本被错误分类，则可以认为样本对于目标数据是不同的。因此，该样品的重量减少，并且该样品在分级器中的比例降低。Masashi Sugiyama 等人。标准学习方法（例如最大似然估计）不再是一致的加权变量，首先分别估计训练和测试输入密度。然后通过采用估计密度的比率来估计重要性。

将每个域的数据从原始高维特征空间映射到低维特征空间的做法被称为基于特征映射的迁移学习方法。在该低维空间中，源域数据具有与目标域数据分布大致相同。以这种方式，标记的源域样本数据可用于训练低维空间中的分类器以预测目标测试数据。Pan 等人通过最小化隐式语义空间中源域数据和目标域数据的最大均值差异 (Maximun Mean Discrepancy), 从而求解得到降维后的特征空间的新的维度降低的迁移学习方法。在隐式空间中，不同的域具有相同或非常接近的数据分布，因此监督学习算法训练模型可用于预测目标域数据。Blitzer 等人 然后就在这个“轴”特征下进行训练学习，从而提出了一种结构对应学习算法(Structural Corresponding Learning, SCL)。

基于模型的迁移学习方法将从源域学习的模型应用于目标域，然后根据目标域学习新模型。来自中科院计算所的赵等人首先基于现有的标记数据，决策树用于构建鲁棒的行为识别模型，提出了 TransEMDT 方法。在 2014 年的 CVPR 上，Oquab 等作者建议可以对大型数据集（例如 ImageNet）执行预训练，然后将训练好的网络权重迁移到小数据集的 TRCNN 网络。也就是说，使用小数据集对网络进行微调，以便网络可以应用于小数据集。

基于关系的迁移学习方法利用源域来学习逻辑关系网络，然后将其应用于目标域。此方法侧重于源域和目标域之间的关系。目前，关于基于关系的迁移学习方法的研究工作很少，而且只有少数相关文章需要讨论。这些文章使用 Markov Logic Net 挖掘不同域之间的相似性。Davis 等人提出的二阶马尔可夫逻辑。

如图 1-3 为迁移学习示意图。



图 2.3 迁移学习示意图

## 第 2.5 节 基于 PyTorch 的 YOLOV3 方法

PyTorch 框架是现在在图像领域愈发流行的深度学习库。基于对 PyTorch 有一定了解，在实现 YOLOv3 时决定使用此库。

首先，先去查阅 YOLO 的三篇论文，对 YOLO 基础结构有了一定的了解之后，查阅了 Darknet 的代码。结合一些博客和对 PyTorch 的了解，进行编写程序。编写完网络结构等各个部分的代码之后，进行测试及调试。调试结束即可进行训练和测试。

因时间有限，并没有编写多尺度处理图像的代码，所以程序里硬性规定图像大小为  $416 \times 416$  大小，需要在输入网络之前对图像进行缩放。代码分为以下几个部分：

训练部分(train.py), 测试部分(test.py), 视频测试部分(video\_test.py),网络部分(Darknet.py),配置部分(util.py), 以及检测部分(detector.py)。其中, 训练部分主要是导入终端的命令, 比如迭代次数, 学习率等, 网络部分为构建了基础的网络结构, 导入了原本的配置文件(.cfg)文件, 将图像进行处理等。

## 第 2.6 节 训练过程与结果对比

### 2.6.1 训练过程

本实验分别使用了基于 Darknet 的 YOLOV3 和基于 Caffe 的 Faster RCNN 的官方源码, 以及基于 keras 框架下的 YOLOV3 开源代码和自己编写的基于 PyTorch 框架下的 YOLOV3 代码。

首先先对基于 Darknet 的 YOLOV3 网络进行训练。安装好 ubuntu 系统的环境之后下载编译 Darknet 代码, 再依照官方指南将网络进行测试。测试成功之后, 使用 Unity 软件制作虚拟图片数据集, 使用 labelME 进行标注。按照 VOC 格式制作相应的数据集, 放入 Darknet 代码, 修改配置文件, 再进行训练。15 个小时之后, 训练结束, 再使用视频文件进行测试, 导出标注好的视频文件, 作为对比文件。

其次, 对基于 Caffe 的 Faster RCNN 进行训练。下载配置 Caffe 等相应环境之后, 首先用 VOC2007 数据集进行测试, 判断是否能使用, 再使用已制作好的数据集文件进行训练。最后用视频文件进行测试导出标注好的视频文件, 作为对比文件。

下面的实验是迁移学习的实验。用基于 Keras 的 YOLOV3 网络进行训练。由于要使用真实数据集进行训练, 需要搜集相关的真实火焰数据集, 并进行截取, 标注。制作成 VOC 数据集后, 修改相关配置文件, 进行训练。

最后, 进行编写 YOLOv3 网络, 比较其性能。

### 2.6.2 结果对比

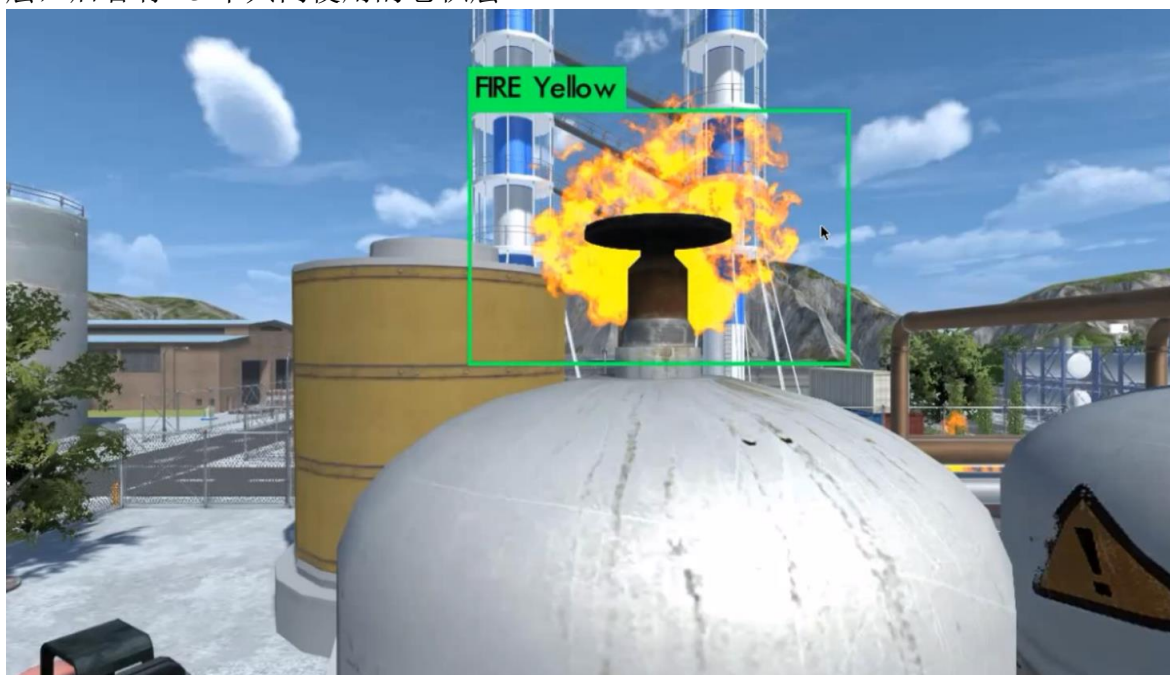
首先是 YOLOv3 网络和 faster RCNN 网络的效果对比。总体而言, YOLOV3 检测速度快于 faster RCNN, 而 faster RCNN 的精确度高于 YOLOV3。

YOLOv3 网络的速度快的的原因主要在于作者采用了和一般检测网络不同的方法。作者将整张图像分割成  $S \times S$  的区域, 再送入一个简单的神经网络, 预测出其预测边框和相对的概率。此外, 预测的概率值还会对边框进行反馈加权。

YOLOv3 模型相比以前提出的分类检测网络有许多优势，比如在测试时候，它将整张图片输入网络，预测的消息会由全局图片信息引导。另外，此模型与 RCNN 检测系统不同的一点是，RCNN 需要对一张图片进行上万次预测，而此模型只需要单个神经网络即可完成整体预测，整个过程十分简洁。

YOLOv3 的优势就是他的速度非常快，比 RCNN 快了 1000 倍，比 Faster RCNN 快了 100 倍。

Faster RCNN 网络的精度高的原因在于其创新点 RPN。RPN 网络可以输入多尺度图片，输出一系列目标预测的矩形区域，每个目标预测区域都会被打分。作者用全卷积网络来模拟这个过程，并将此结果和 Fast RCNN 网络来共享结果，并以此来计算，所以两网络共享了一系列相同建卷积层。作者依照 the Zeiler and Fergus model (ZF) 和 he Simonyan and Zisserman model (VGG-16) 的做法，前者有 5 个共同使用的卷积层，后者有 13 个共同使用的卷积层。



如图 1-4 为 YOLO 结果图。

图 2-4 YOLO 结果图

在生成目标预测区域时，我们通过在卷积层的特征图上用一个小网络进行卷积，这个特征图是由最后的共同使用的卷积层输出。此外，对于这个小网络，一个卷积特征图的  $n \times n$  的窗口在这个小网络中是非常重要的。每个滑动窗口的结果都反应了一个低维特征的分布，如 ZF 网络是 256 维的，VGG 是 512 维并需要通过 ReLU 激活函

数。这个特征被输入共用的全连接层中。作者在文中使用了特征图窗口大小是  $3 \times 3$ ，这就使得图像的感受野很大，如 ZF 网络为 171 个像素，VGG 网络为 228 个像素。因此全卷积网络通过滑动窗口的方式华东分，因此全连接层的空间位置相互共享。此网络通过  $n \times n$  的卷积层实现，后面接了两个  $1 \times 1$  的卷积层。

经两个网络各自训练后，结果与以上分析一致。YOLOV3 的识别速度远远快于 faster RCNN，其表现为 YOLOV3 的标注效果十分流畅，而 faster RCNN 会出现延迟现象。但是 faster RCNN 的识别能力更强，模型具有更高的鲁棒性，准确性高于 yolov3。最后，基于实现实时处理的效果，选择 YOLOV3 网络。

图 1-5 为 faster RCNN 准确率图。

```
Average Precisions:
+ Class '0' - AP: 0.9951456310679612
+ Class '1' - AP: 0.7346474066709153
+ Class '2' - AP: 0.8404390328863088
mAP: 0.8567440235417284
```

图 2.5 faster RCNN 准确率图

其次是虚拟数据集模型和真实数据集模型比较。在同样硬件条件下，基于 Darknet 的训练时间为 15 个小时，而使用迁移学习的训练时间仅为 2 小时，模型即收敛。且使用迁移学习可以应用于 cpu 训练上，可以节省大量计算资源。

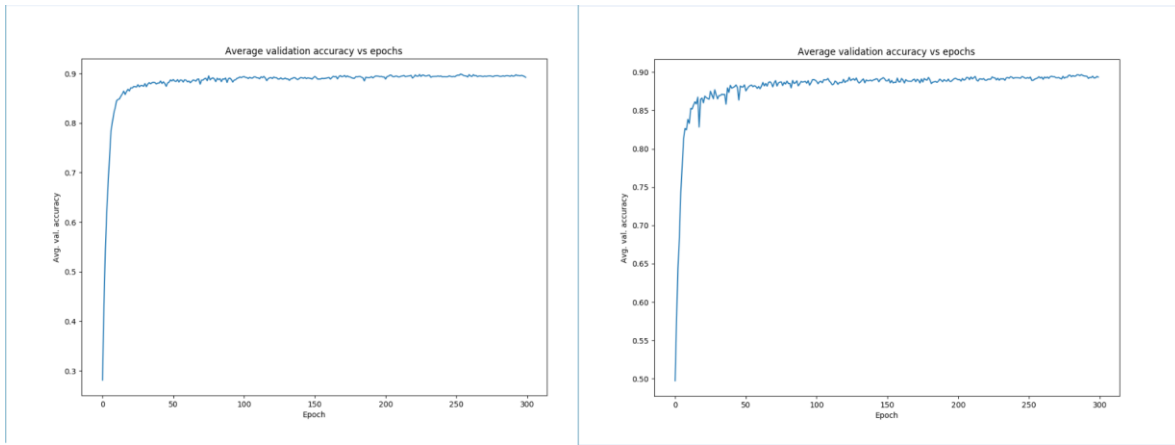
最后是基于 pytorch 的 yolov3 模型和基于 Darknet 的模型进行比较。由于真实火焰图片较少，所以选用虚拟数据集进行训练。Pytorch 框架下的网络实现了原本网络的大部分功能，但是由于搭建框架的编程语言分别为 python 和 C 语言，可看出 pytorch 框架下的 yolov3 网路速度略慢于 Darknet 框架的网络。

如图 1-6 为迁移学习结果。



图 2.6 迁移学习结果图

如图 1-7 为两框架结果对比。



(a) DARKNET 训练结果

(b) PyTORCH 训练结果

图 2.6 框架对比结果图



## 结 论

本课题主要是基于各个框架下的 YOLOV3 网络和基于 Caffe 网路下的 faster RCNN 网络进行的方法研究，在将模型进行对比之后得出以下结论：

- 1、 Faster RCNN 网络的识别精确性略优于 YOLOv3 网络，但是速度远不及 YOLOv3 网络，所以基于实时的检测系统使用 YOLOv3 是更优选择。
- 2、 使用迁移学习的 YOLOv3 网络大大节约了计算资源，数据集的数量如果可以足够多，网络训练可以达到更好的效果。
- 3、 使用 Darknet 的框架下的 YOLOv3 网络识别速度略快于 Pytorch 框架下的网络。

工作虽已完成，但是还有待改进，改进方面为如下几点：

- 1、 数据集数量不够大量，且真实场景下的数据集并不完全是基于化工场景下的数据集。
- 2、 网络还有进一步改进余地，比如在精度方面可以再进一步提高，提出针对火苗识别网络的改进网络。

## 参考文献

- [1] 相徐斌.基于视频的烟雾检测算法研究[D].杭州: 浙江大学, 2017.
- [2] 熊国良,苏兆熙,刘举平等.火焰特性识别的 Matlab 实现方法[J].计算机工程与科学,2013,35( 7) : 131—136.
- [3] 严云洋,吴茜茵,杜静,等.基于色彩和闪频特征的视频火焰检测[J].计算机科学与探索,2014,8(10) : 1271—1279.
- [4] 胡勤,陈琛,刘敏. 一种基于动态纹理的烟雾和火焰检测方法[J]. 消防科学与技术,2014.33( 6) : 667—669.
- [5] 陈磊,黄继风. 基于视频的火焰检测方法[J]. 计算机工程与设计, 2014, 35( 9) : 3143—3147.
- [6] STADLER A,WINDISCH T,DIEPOLD K.Comparison of intensity flickering features for video based flame detection algorithms[J].Fire safety journal, 2014, 66: 1—7.
- [7]王华利,邹俊忠,张见,等. 基于深度卷积神经网络的快速图像分类算法[J]. 计算机工程与应用,2017,53( 13) : 181—188.
- [8] GIRSHICK R , DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detection and Segmentation[J]. IEEE Trans Pattern Anal Mach Intell, 2015,38( 1) : 142—158.
- [9] 曹诗雨,刘跃虎,李辛昭.基于 Fast R-CNN 的车辆目标检测[J]. 中国图象图形学报, 2017, 22( 5) : 671—677.
- [10] REN S, HE K,GIRSHICK R,et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis & machine intelligence, 2017, 39( 6) : 1137.
- [11] JIA Y,SHELHAMER E, DONAHUE S, et al. Caffe: convolutional architecture for fast feature embedding[C]/ /Proceedings of the 22nd ACM International Conference on Multimedia.Orlando, Florida: ACM, 2014.
- [12] CHMELA R P, BENK R ID A. Efficiency of HSV over RGB Gaussian Mixture Model for fire detection[C]/ /Radioelektronika. Slovakia: IEEE, 2014.
- [13] RONG J Z, ZHOU D C, YAO W, et al. Fire flame detection based on GICA

and target tracking[J]. *Optics & laser technology*, 2013, 47: 283—291.

[14] FRIZZI S, KAABI R, BOUCHOUICHA M, et al. Convolutional neural network for video fire and smoke detection[C]//Industrial Electronics Society, IECON 2016—42nd Annual Conference of the IEEE. USA: IEEE, 2016.

[15] 余启明.基于背景减法和帧差法的运动目标检测算法研究[D].江西赣州:江西理工大学,2013:2-4.

[16] 严红亮,王福龙,刘志煌.结合三帧差分的 ViBe 运动检测算法[J].*计算机系统应用*, 2014, 23(11):105-110.

[17] 赵建.基于三帧差法的运动目标检测方法研究[D].西安:西安电子科技大学, 2013:2-4.

[18] 郭峰,曹其新,谢国俊,等.基于 OHTA 颜色空间的瓜果轮廓提取方法[J].*农业机械学报*, 2005, 36(11):119-122.

[19] 严云洋.色彩和轮廓特征的火焰检测 [J].*微电子学与计算机*, 2011, 28(10):137-141.

[20]周忠涛. 基于视频特征的火情监测研究[D]. 江西农业大学,2017.

[21]戴静.火焰图像快速检测方法研究及应用[D].西南科技大学,2017.

[22]严云洋,杜静,高尚兵,周静波,刘以安.融合多特征的视频火焰检测[J]. *计算机辅助设计与图形学学报*,2015,27(03):433-440.

## 致 谢

首先要感谢我的指导老师胡标老师，他在我做毕业设计的这段时间给予了我非常多的指导和帮助，他治学风格严谨，专业知识渊博，对工作要求精益求精，以及严于律己、宽以待人的崇高品格这些无不影响启迪着我，也让我掌握了基本的研究方法，树立了学习的目标，对自己的要求也不断提高。从这次毕业设计的选题到完成，每一步都包含着老师的悉心指导，在此，对胡标老师表示最衷心的感谢和最崇高的敬意！同时，还要感谢师兄师姐对我的帮助和指导，也感谢在我毕业设计期间帮助过我的老师同学们。

时光匆匆流逝，转眼已是大学时光结束之时，随着毕业设计的顺利完成，离不开老师同学们对我的帮助，在此，我要向国际教育学院的机械设计制造及其自动化专业所有老师报以最诚挚的谢意！感谢你们四年来的教诲。