

**本科毕业设计（论文）**

题 目 基于深度学习的输液瓶瓶体位置检测算法的设计与实现

学院名称 计算机科学与技术学院

专业名称 计算机科学与技术

年级班级 计实验16-1

学生姓名 杜 鑫

指导教师 芦碧波（校内）

企业导师 许 思

2020年6月

摘要

伴随5G技术的不断发展与成熟，智慧医疗领域将会迎来更广阔的的发展空间。输液作为我国临床医疗治疗的常用手段之一，我国的病人对输液的需求往往很大，实现输液的整个流程需要由专业的人员来实施，而我国在岗护士人数不能充分满足医疗需求，护士缺口比较大，每位护士要同时负责多名病人，医护人员工作负担繁重。输液的过程往往相对漫长，在此过程中若出现意外没有及时发现并进行处理，容易出现跑针回血等情况，严重的甚至会造成空气栓塞以及产生其他并发症，危及患者生命，从而对医患关系造成不好的影响。出于安全性考虑，在传统的护理办法中，需要医护人员或患者陪护人员不断关注，每隔一定时间查看输液进度。然而这在一定程度上增加了人工的消耗，降低了社会工作效率。

科技的发展与进步给人们提供了很大的便利，从很大程度上往好的方向改变了人们的生活，但是，人类的生活节奏也不再像从前那样悠闲，而是变得愈来愈快，因此要求社会整体的效率也要加快。针对上述问题，已经有研究者从不同的方面来考虑解决问题，但是几乎每一种技术方法都其局限性，不能完全结解决。计算能力的快速提升，数据体量的爆炸式增长，计算机视觉的飞速发展，都使得人们可以借助计算机来辅助医护人员和患者进行治疗。

本文采用深度学习的方法，对在实际场景下采集来的数据进行手工标注，建立数据集，并在构建的数据集基础上使用PaddlePaddle深度学习框架，结合YOLO v3算法训练模型，通过实验证明，本文使用的方案能快速实现输液瓶瓶体位置的检测，为智能输液检测的实现打下坚实基础，进一步促进智慧医疗的发展。

**关键词：** 智慧医疗；深度学习；目标检测；YOLO v3

ABSTRACT

With the continuous development and maturity of 5G technology, the field of intelligent medical treatment will usher in a broader space for development. As one of the common means of clinical medical treatment in China, infusion is often required by patients in China, and the whole process of infusion needs to be implemented by professional personnel. However, the number of on-duty nurses in China cannot fully meet the medical needs, and the gap of nurses is relatively large. Each nurse has to be responsible for several patients at the same time, and the medical staff has a heavy workload. The process of infusion is usually relatively long. In this process, if accidents are not found and handled in time, it is easy to cause bleeding needle and other situation, which may even cause air embolism and other complications, endangering the lives of patients and thus adversely affecting the doctor-patient relationship. For the sake of safety, in the traditional nursing methods, medical staff or patient attendants are required to pay constant attention and check the infusion progress at regular intervals. However, to some extent, it increases the consumption of labor and reduces the efficiency of social work.

The development and progress of science and technology have provided people with great convenience and changed people's life to a great extent. However, the pace of human life is no longer as leisurely as before, but becoming faster and faster. Therefore, the efficiency of the whole society is required to be accelerated. In view of the above problems, some researchers have considered and solved the problems from different aspects, but almost every technical method has its limitations and cannot completely solve the problems. The rapid improvement of computing power, the explosive growth of data volume and the rapid development of computer vision have enabled people to use computers to assist medical staff and patients in their treatment.

In this paper, the method of deep learning is adopted to manually mark the data collected in the actual scene and establish the data set. Based on the built data set, the deep learning framework called PaddlePaddle is used in combination with the YOLO v3 algorithm to train the model. Through experiments, it is proved that the scheme used in this paper can quickly realize the detection of infusion bottle position, which lays a solid foundation for the realization of intelligent infusion detection and further promotes the development of intelligent medical treatment.

**Key words:** Smart medicine; Deep learning; Target detection; YOLO v3

**目录**

[1.前言 1](#_Toc22832)

[1.1研究背景与意义 1](#_Toc3129)

[1.2国内外研究现状 1](#_Toc5148)

[1.3本文主要研究内容 3](#_Toc32554)

[1.4本文组织结构 3](#_Toc17611)

[2.深度学习与卷积神经网络 5](#_Toc29369)

[2.1深度学习 5](#_Toc29426)

[2.2计算机视觉技术 7](#_Toc235)

[2.2.1图像分类 8](#_Toc25855)

[2.2.2目标检测 8](#_Toc8439)

[2.2.3图像分割 9](#_Toc15844)

[2.3卷积神经网络 10](#_Toc6221)

[2.3.1卷积神经网络结构 10](#_Toc19057)

[2.3.2卷积神经网络模型 14](#_Toc12569)

[3.基于深度学习的目标检测算法 21](#_Toc596)

[3.1两阶段检测算法 21](#_Toc18874)

[3.1.1 R-CNN 21](#_Toc8754)

[3.1.2 Fast R-CNN 22](#_Toc11516)

[3.1.3 Faster RCNN 23](#_Toc12155)

[3.2单价段检测算法 23](#_Toc24856)

[3.2.1 SSD 23](#_Toc31716)

[3.2.2 YOLO 24](#_Toc19401)

[4.利用卷积神经网络对输液瓶瓶体位置检测 27](#_Toc11694)

[4.1输液瓶数据集 27](#_Toc15352)

[4.1.1输液瓶数据集获取 27](#_Toc15135)

[4.1.2数据增强 28](#_Toc11721)

[4.1.3数据标注 30](#_Toc31455)

[4.2深度学习框架与环境搭建 32](#_Toc1040)

[4.2.1深度学习框架 32](#_Toc14355)

[4.2.2环境搭建 33](#_Toc2682)

[4.3模型训练 34](#_Toc28774)

[5. 算法评价与实验结果分析 37](#_Toc28139)

[5.1算法评价 37](#_Toc21358)

[5.2实验结果分析 38](#_Toc1250)

[6.总结与展望 43](#_Toc5634)

[致谢 45](#_Toc20447)

[参考文献 47](#_Toc384)

**1.前言**

## **1.1研究背景与意义**

在大数据时代的背景之下，5G技术的发展浪潮席卷着各行各业，为其增添了新的发展动力，当然，也包括医疗行业。在世界各地，医疗领域的发展一直是人们关注的重点，医疗关系到国计民生，与每个人的生活都息息相关。我国人口数量众多，随经济的飞速发展，老百姓的生活质量越来越好，对各方面的服务需求不断地增多，与此同时对公共医疗的服务需求量也随之增大，伴随着社会人口老龄化的速度日益加剧，我国现有的医疗资源在原本就不够充裕的基础上变得更加紧张。目前，我国大多数住院病人都需要进行静脉输液治疗，这是最简单有效的治疗方式之一，也是最常使用的护理操作。在我国，每年有大量住院病人需要进行静脉输液治疗或护理，而输液的独特程序成为普通人无法轻松进行的阻碍，需要有专业的人员来进行操作。在我国的医疗体系中，护士扮演着十分重要的角色，是治疗过程中不可缺少的辅助者，但是我国的在职护士人数相较于实际需求数量来说还有较大缺口，在一定程度上导致每位护士工作量多，工作压力大。出于治疗安全性和治疗效果的考虑，输液治疗所花费的时间一般都较长，在这期间需医护人员较频繁地前往病人病房观测输液情况，或者病人寻找陪护人员进行看护，及时注意情况，但是这在一定程度上增加了人工的消耗，从而降低了社会工作效率，在一定程度上减缓了经济效益的提高速度。在2020年初爆发的新型冠状病毒肺炎疫情中，我国的住院病人人数激增，医护人员每天都在进行超负荷工作，也不方便陪护人员进行陪护。

本课题由此入手，使用深度学习和图像处理等相关技术，突破传统的方式，快速检测图像中输液瓶瓶体的位置，为实现更快速更便捷的智能输液监测打下良好基础，减少了医院的人工消耗和患者的陪护需求，进一步促进了我国在智慧医疗领域的发展，为社会提供了更智能更便捷的医疗服务。

## **1.2国内外研究现状**

经历了近四百年的曲折发展，静脉输液技术[1]从诞生逐渐变得成熟，到现在具有完整的体系，被广泛的应用到临床治疗与护理中，成为最直接有效的方法之一。传统的输液监测依靠人工，具有安全隐患，容易造成医疗事故，为减少事故的发生，为患者提供更安全的就医环境，许多研究人员将目光投到智能输液监测上。目前主要的智能输液监测分为以下几类：电容式输液监测、光电式输液监测、重力式输液监测。

Ogawa[2]等人研究了一种基于电容传感器的输液监测装置。对称放置两个电极在输液瓶瓶壁上，将其位置设置在想要监测报警的液面水平处，如下图1-1所示。当患者进行输液治疗时，若液面下降超过报警线位置，此时输液瓶内变为空气或者真空，与充有药水时检测到的相对介电常数相比，会发生明显变化，提示报警。然而不同药水对电容器的影响程度不同，且这种方式只能在输液即将完成时才有帮助，无法对输液进行中发生的情况进行监测，且输液装置的设计需要根据输液瓶容量进行变换，操作复杂、成本高、精度差，因此无法应用到临床治疗中。

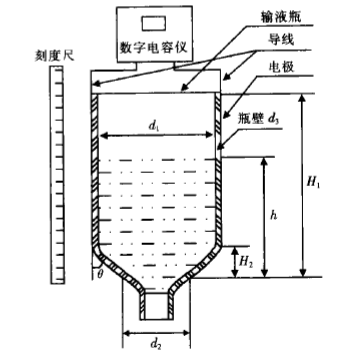


图1-1 电容式输液液位测定装置图

Cui[3]等人研究的光电式输液监测装置，依据液体对红外光具有散射作用，在输液管外侧安装光电对管，当液体通过时，光电信号强度会减弱，进而引起接收端的电脉冲变化。茂菲氏滴管内有液体和无液体时，光敏三极管的感光量有所不同，随之会引起电压、电流等参数的变化，进而传递出输液管内液体变化的信息。但是，日光或者灯光会对设备有所干扰，从而影响结果的准确性。

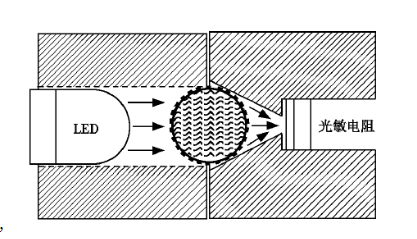


图1-2 光感式液位检测原理图

周涛[4]等人提出的重力式输液监测，其实现是利用重力传感器，将整个输液设备挂在上面，检测输液过程中的重量变化，并将其转化为电信号输出，传送到信息接收终端，从而能获取输液速度变化和液量剩余情况等信息。然而重力传感器获取的是包括药水和输液瓶一同的重量，而输液瓶或者输液袋的规格材质不同，不同药水的密度也不尽相同，由此对输液监测带来的误差还无法解决。

深度学习在医疗领域已有不少应用，如医学影像诊断、电子医疗档案等，并且，随着大数据时代的到来，以及5G技术的发展，利用深度学习技术解决智能输液监测问题，已不再是遥不可及的事。通过深度学习技术，来代替人工观察，这在一定程度上减轻医护人员的工作压力，减少病人的陪护成本。

## **1.3本文主要研究内容**

本文的主要研究内容是基于深度学习的输液瓶瓶体位置检测，目的是为智能输液监测的实现打下坚实基础。通过人工方式，在真实场景下采集输液瓶数据集并通过数据增强方式来扩充数据集，然后在构建的数据集的基础上选择合适的神经网络模型，进行训练，根据训练得到的模型的效果进而修改参数，进行优化。最后通过实验对选择的模型进行测试，证明本实验方案具有可行性，能够较准确的预测输液瓶瓶体所在位置。

针对本文研究的问题，其未来发展可以产生实际的商用价值，潜在用户为有输液治疗或护理需要的医院或诊所。利用深度学习技术与计算机视觉技术，在本文研究的问题的基础上进一步完善智能输液监测，相比于其他输液监测系统，能有效控制成本，从而扩大经济效益。

## **1.4本文组织结构**

本文主要分为五个章节，具体内容如下：

第一章为引言，主要介绍了当今环境下智能输液监测研究的背景与意义，国内外的研究现状，以及本文主要的研究内容。

第二章介绍实验所需要了解的基本概念和基础理论，介绍深度学习有关的一些概念，着重介绍计算机视觉方向的应用与卷积神经网络模型。

第三章主要介绍了目前主流的一些目标检测算法，并通过对比选择出本实验所用算法。

第四章的内容是介绍实验过程，从数据集的收集与建立到进行实验所需的环境配置，以及模型的训练及结果分析。

第五章对本实验进行总结，并提出针对本实验在现阶段所实现的功能进行上升级的合理建议。

**2.深度学习与卷积神经网络**

## **2.1深度学习**

二十世纪五十年代，“人工智能”(Artificial Intelligence)这一概念第一次出现在公众的视野，在此之后，人工智能出现的频率越来越高，直到今天，人工智能之热席卷全球，越来越多的研究者投身到人工智能领域的研究当中，并将其从实验室中带出，应用到我们的日常生活中。人工智能就是研究人类的智能活动的规律，来构建出能让机器代替人类完成以前只能由人类完成的工作的系统，通俗点的理解就是研究如何使计算机来控制其他机器仿照人类进行某种行为的理论技术。截止到目前，人工智能还不能像真正的人一样具备独立思考和推理的能力，现阶段，我们还处于弱人工智能(Narrow AI)时代，但是，弱人工智能也能代替人类从事某些特定的工作，甚至工作效率还要超过人类。例如，Facebook的人脸识别、医疗影像的诊断等。

机器学习是人工智能研究发展到一定阶段的必然产物，是人工智能的一个子集，是一门涉及概率论、统计学、逼近论、凹凸分析、算法复杂度理论等的多领域交叉学科[5]，它将研究聚焦在计算方面，通过计算的方式在计算机上从数据中训练产生“模型”，并利用重新组织已有的知识结构，一点点提高系统的性能。在近几十年中，关于数据的各方面问题，如收集保存、迁移处理等，人类都有了很好的解决办法，并且，数据规模呈爆炸式扩张，这为机器学习的发展提供了肥沃的生长土壤。机器学习的实现可概括为训练和预测两步，从一定数量的样本中，学习模型输出于输入的关系，再基于训练得到的关系，输入新的需要预测的数据，计算其对应的输出即为预测结果。

如果将机器学习技术看做是人工智能背后的推动力量，那么机器学习的关键技术必然少不了深度学习技术，它们三者之间的关系如下图2-1所示。如今，深度学习模型可以用来解决多数机器学习任务，尤其是在语音、计算机视觉、智能推荐和自然语言处理等方面表现突出，取得的成果相较于之前有了很大的进步。



图2-1 人工智能、机器学习和深度学习三者关系示意图

深度学习的概念源于人工神经网络的研究，含多个隐藏层的多层感知器就是一种深度学习结构[6]。深度学习通过组合低层特征形成更加抽象的高层来表示属性类别或特征，即学习[样本数据](https://baike.baidu.com/item/%E6%A0%B7%E6%9C%AC%E6%95%B0%E6%8D%AE/12726279" \t "https://baike.baidu.com/item/%E6%B7%B1%E5%BA%A6%E5%AD%A6%E4%B9%A0/_blank)的内在规律和表示层次来发现数据的分布式特征表示[7]。比较经典的深度学习模型有卷积神经网络模型(Convolutional Neural Network)，如图2-2所示，深度信任网络模型(Deep Belief Network)，如图2-3所示、堆栈自编码网络(Stacked Auto-encoder Network)模型，如图2-4所示。

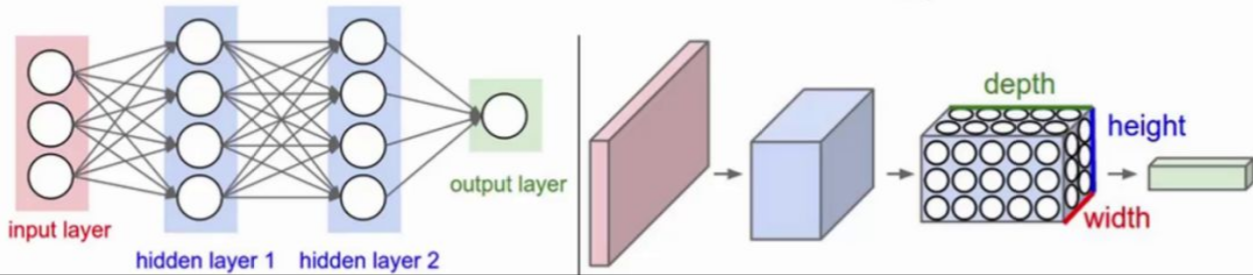


图2-2 CNN模型图

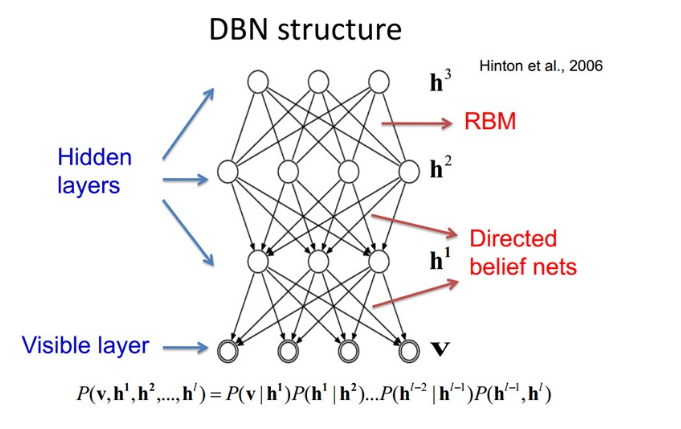


图2-3 DBN模型图

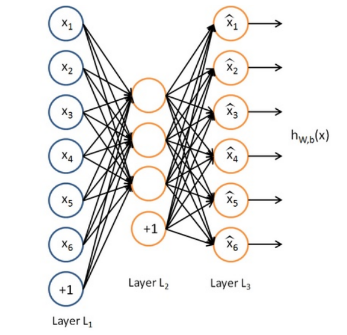


图2-4 SAE模型图

## **2.2计算机视觉技术**

随着生物的出现并不断进化发展，历经数亿年，人类的视觉系统到现在已经具备了极其高的复杂度和强大的功能，人脑中的神经元数目达到千亿数量，这些神经元相互连通，构成复杂的网络结构。物体经过光的漫反射后，反射的色光通过眼睛的晶状体折射后映射到视网膜上，再经过上面的神经，一步步传导给大脑，然后神经系统在处理后得到并反馈给人类所需要的信息[8]，如图2-5所示。

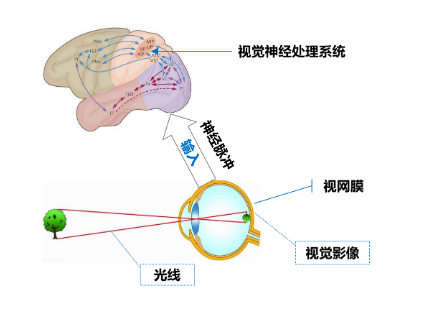


图2-5 人类视觉感知图

有实验表明，人类获得信息的途径有多种，但有超过80%的信息都是来自视觉，计算机视觉在人工智能里可以类比于人类的眼睛，是在感知层上最为重要的核心技术之一。仿造人类的结构，将摄像头看做是人类的眼睛，计算机模拟人类的大脑，摄像头获取的图片信息可以通过一定的处理过程转变为计算机能够识别的数字信息，人类思考的过程则用计算机的程序和算法来代替，转换来的数字信息经过一定的程序和算法做出相应的识别与判断。

总而言之，计算机视觉的发展依据就是希望尽可能的模拟人类的视觉，从获取到的图像中提取数据和信息，然后根据提取到的东西进行分析检测等，真正去“识别”和“理解”这些图像并进行反馈。为了使计算机能够像人一样看懂整个世界，研究者们尝试着从不同的角度来解决这一问题，由此，计算机视觉到现今为止有了不同的发展，主要分为图像分类、目标检测、图像分割等。

2.2.1图像分类

图像分类是根据图像的语义信息对不同类别图像进行区分[9]，计算机视觉的很多问题都是在实现它的基础上解决的，例如图像分割、目标检测、人脸识别、物体跟踪等。图像分类的关键是从已经分好类的集合中，选择一个标签给图像，实现的过程是通过对一大批已经标注好的数据进行特征提取与特征选择，得到特征表示，连接一个分类器，将样本标签进行训练，即可得到输出概率，输出概率最大的那一个即为预测的类别结果，如图2-6所示。目前，在实际的工业生产和社会生活中，已将出现了许多图像分类的身影，例如医学影像的分类与识别诊断、安检闸口的人脸识别、手机相册的图像检索与自动分类等。

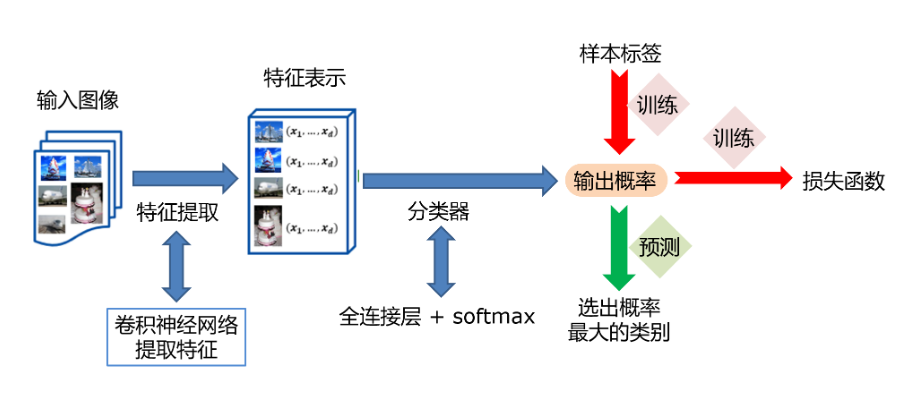


图2-6 图像分类流程

### 2.2.2目标检测

目标检测是在图像分类的基础上实现的，首先在图像中框出目标物体所在的位置，然后判断框中物体所属的类别并表示出来，如图2-7所示。传统的目标检测是利用滑动窗口的技术来进行的，就是将设置好的窗口在图片上依次移动，直至遍历整张图片，然后判断是否有在某个位置上的窗口包含有要检测的物体，但是，这个看似简单并且容易理解的过程，实现起来却有很大的难度，要完成的计算量非常庞大，而在当时的计算资源很难满足要求，在速度上也不尽人意，从而导致检测的效率很低，而基于深度学习的目标检测，突破了传统目标检测的部分局限性，性能提升显著。目标检测在计算机视觉领域也是一个非常重要的问题，是当前研究的热门方向，它的应用范围十分广泛，涉及交通场景识别、智能视频监控、工业质检、自动驾驶等多种领域，通过利用深度学习技术的计算机视对减少人力资源的消耗，提高社会的劳动生产率具有重要的现实意义。

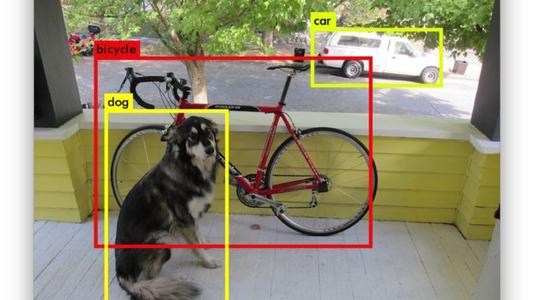


图2-7 目标检测示意图

### 2.2.3图像分割

图像分割是图像按照特定的要求将具有相同特殊性的区域标注显示出来，目前，像素级别的语义分割是常见的分割任务之一，还有一种就是实例分割。像素级别的语义分割，对图像中的每个像素都划分出对应的类别，即实现像素级别的分类[10]；而类的具体对象，即为实例，实例分割比语义分割还要再做进一步的处理，在像素级别分类后的基础之上再区别开不同的实例个体。如下图所示，在进行语义分割后得到的结果中，绿色的是背景，青色的是人，蓝色都是羊，红色的是狗，而在进行实例分割后，同样是羊这一类别，不同的羊只也被分割表示出来。

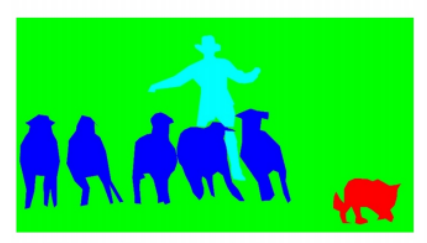


图2-8 图像语义分割示意图



图2-9 图像实例分割示意图

## **2.3卷积神经网络**

卷积神经网络(CNN)是一类包含[卷积](https://baike.baidu.com/item/%E5%8D%B7%E7%A7%AF/9411006" \t "https://baike.baidu.com/item/%E5%8D%B7%E7%A7%AF%E7%A5%9E%E7%BB%8F%E7%BD%91%E7%BB%9C/_blank)计算且具有深度结构的[前馈神经网络](https://baike.baidu.com/item/%E5%89%8D%E9%A6%88%E7%A5%9E%E7%BB%8F%E7%BD%91%E7%BB%9C/7580523" \t "https://baike.baidu.com/item/%E5%8D%B7%E7%A7%AF%E7%A5%9E%E7%BB%8F%E7%BD%91%E7%BB%9C/_blank)(FNN)[11]，是计算机视觉技术最经典的模型结构之一。二十世纪八十年代起，学者们开始对卷积神经网络进行研究，由Alexander Waibel等提出的时间延迟网络(TDNN)是第一个卷积神经网络，主要应用于语音识别问题，在此之后，越来越多的神经网络模型被提出，并被用于解决不同领域的不同问题。

### 2.3.1卷积神经网络结构

（1）卷积层

卷积是数学分析中的一种积分变化的方法，在图像处理中采用的是卷积的离散形式。在卷积神经网络中，卷积层的实现方式实际上是数学中定义的互相关 (cross-correlation)运算，卷积计算过程如下图所示。

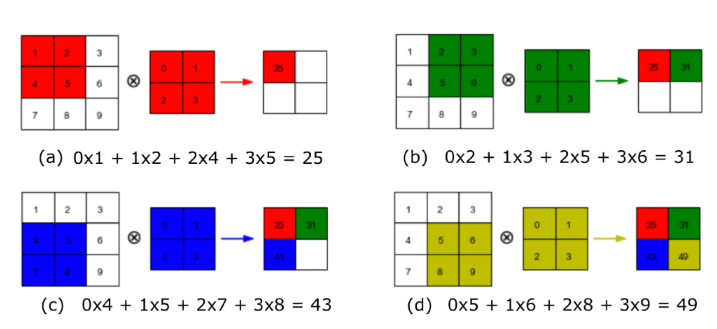


图2-10 卷积计算过程图

其中，卷积核是关键，它也被称为滤波器。若一个卷积核的高为，宽为，那么被称为卷积，如5×5卷积，表示的就是卷积核的高为5，宽也为5。卷积核计算过程的数学表达式如下：

 （2-1）

上式中的代表的是输入的图像，指的是卷积核的相关参数，表示的是卷积后输出的特征图，这些格式都是二维数组，而是对图像遍历卷积核参数与求和。

在卷积计算过程中，通常会在高度或者宽度的两侧采取等量填充，即高度方向两头各填行，宽度方向两端各填列，从而保证在进行卷积后，图片的大小不发生改变，在进行卷积核为的卷积操作后，输出图片大小的计算公式如下：

 （2-2）

 （2-3）

卷积核在图片上每次滑动的像素点个数为步幅，图2-10表示的就是步幅为1的情况，图2-11则表示的是步幅为2的情况，即卷积核在图片上移动时，每次移动大小为2个像素点。

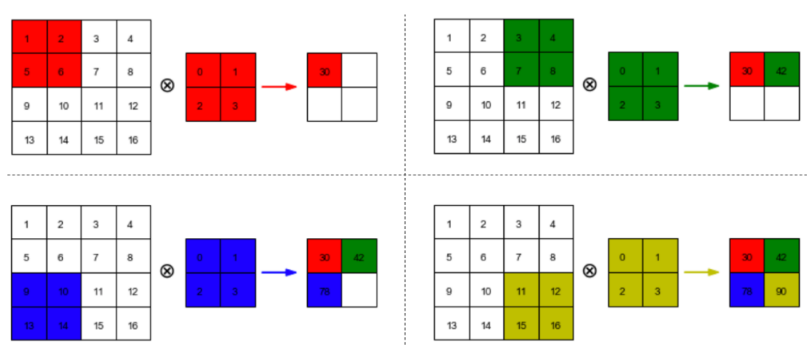


图2-11 步幅为2的卷积过程

假设宽度方向的步幅设置为，高度方向的设置为，再进行卷积计算，输出的特征图的大小为：

 （2-4）

 （2-5）

输入图片中每个大小为的区域上的像素点的数值与对应的卷积核的每个元素先相乘，再进行相加，最后得到输出特征图上每个元素的值，所以，改变输入图像区域中的任一数值大小，最终输出的特征图都极有可能不一样，这个区域叫做输出特征图上对应点的感受野。因此，在经过多层卷积后，输出特征图的一个像素点就可以反应输入图像很大范围的变化。

1. 池化层

池化层是使用某一位置的相邻输出的总体统计特征代替网络在该位置的输出[12]，主要是用来压缩图像大小的，如下图2-12所示。其好处是当输入数据做出少量平移时，经过池化函数后的大多数输出还能保持不变。例如，当想要识别一张图片是否是自行车的时候，只需要知道自行车的前边有一个轱辘，后边也有一个轱辘，它们通过车身相连接，不需要精确的知道车轱辘的位置，在这种情况下，约化提取某块区域的总体特征就非常有意义了。池化会使得到的特征图的尺寸变得更小，在之后连接的是全连接层的情况下减少神经元数量的效果十分明显，从而达到节约存储空间的效果，同时还能够提高计算效率。

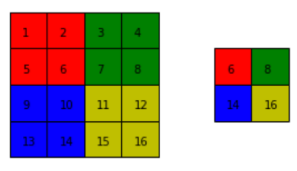


图2-12 池化示意图

在卷积神经网络中，假设使用的池化窗口大小为2×2，步幅为2，不进行填充，则输出的特征图大小为：

 （2-6）

 （2-7）

在经过上述池化方式后，输出的特征图的大小变为原来的四分之一，即高和宽都变为原来的一半，但是通道数不会发生改变。

1. 激活函数

神经网络中的每个神经元节点接受上一层神经元的输出值作为本神经元的输入值，并将输入值传递给下一层，输入层神经元节点会将输入属性值直接传递给下一层[13]。在多层神经网络中，上层节点的输出和下层节点的输入之间具有一个函数关系，这个函数称为激活函数[14]。

在不使用激活函数的情况下，每层节点的输出都是上一层输入的线性函数即使神经网络的层数再多，输出与输入之间的关系都是线性的，则网络逼近的能力就变得相当有限，但是在引入了激活函数后，相当于引入非线性因素，从而改变这种神经元之间的输出与输入的关系，使得神经网络可以任意逼近任何关系的函数，进而增强深度神经网络的表达能力。在神经网络中，常用的激活函数有一下三种：

Sigmoid函数，将输入的连续实数值变换成0-1之间的输出，可以有效缩小输出范围，但是在输入靠近0的范围内输出变化过于灵敏，在进行梯度反向传递时容易导致梯度爆炸和梯度消失。其函数解析式为：

 （2-8）

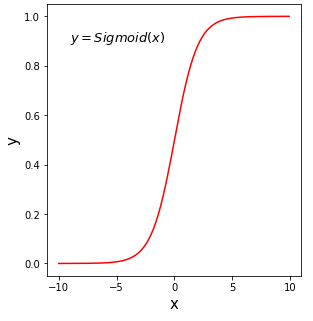


图2-13 Sigmoid函数图

ReLU函数，将函数分为两段，在x<0时，函数值为0，在时，输出为输入值x。其函数解析式为：

 （2-9）

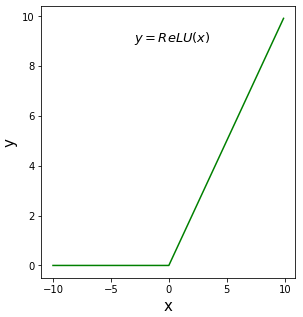


图2-14 ReLU函数图

1. 卷积神经网络反向传播

损失函数是模型优化的目标，模型的预测值与真实值存在一定误差，损失函数就是用来估量它们之间的误差，是非负值函数，通常用来表示，损失函数越小，模型的鲁棒性越好，它用于在众多的参数取值中，识别最理想的取值。在卷积神经网络反向传播中，常用的损失函数有均方误差损失函数、交叉熵损失函数等。损失函数的计算在训练过程的代码中，每一轮模型训练的过程都基本相同，分为以下几步：先根据输入数据正向计算预测输出；再根据预测值和真实值计算损失；最后根据损失反向传播梯度并更新参数。

### 2.3.2卷积神经网络模型

（1）LeNet

LeNet是最早的卷积神经网络之一。1998年，Yan LeCun第一次将LeNet卷积神经网络应用到图像分类上，并在手写数字识别任务中取得了巨大成功[15]。LeNet-5通过连续使用卷积和池化层的组合提取图像特征，其架构如下图2-15 所示。

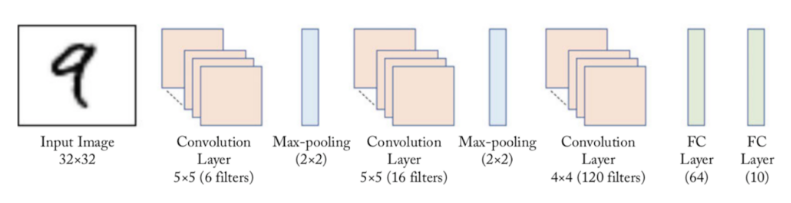


图2-15 LeNet-5模型网络结构示意图

在LeNet中的第一轮卷积和池化里，卷积的目的是提取图像中所包含的特征模式，卷积核大小为5×5，在经过卷积后，图像的大小从32×32 减小到28×28，再经过2×2的池化层，输出特征图对空间位置的敏感性有所降低，此时图像大小变为14×14；在第二轮的卷积中，卷积核的大小仍是5×5，在经过卷积操作后图像尺寸变成10×10，经过2×2的池化后又变成5×5；在第三轮卷积中，将经过第3次卷积提取到的特征图输入到全连接层；第一个全连接层的输出神经元的个数是64，第二个全连接层的输出神经元个数是分类标签的类别数，对于手写数字识别其大小是10；在这些操作之后使用Softmax激活函数就可以计算出每一种类别的预测概率。

（2）AlexNet

LeNet虽然在手写数字识别的数据集上表现出很优秀的成果，但是，将其应用到更大的数据集上，取得结果却不尽人意。在此之后的十几年中，卷积神经网络在计算机视觉领域并没有取得很好的成果，原因主要有两方面：一是神经网络的计算比较复杂，对当时计算机的算力来说，训练神经网络所花费的时间非常长；另一方面，当时还没有专门针对神经网络做算法和训练技巧的优化，神经网络的收敛性是件非常困难的事情。  
 但是，科学技术一直在不断地进步，计算机的软硬件资源也在不断地改进更新，计算能力变得愈来愈强大，特别是GPU的并行计算，这一切都使得人们更容易进行复杂的神经网络计算。除此之外，人类社会所产生的数据也越来越多，网络上的数据量迅速增长，为神经网络模型的训练提供了丰富的资源。同时也有越来越多的研究人员开始专门针对神经网络做算法和模型的优化，Alex Krizhevsky等人提出的AlexNet以很大优势获得了2012年ImageNet比赛的冠军。

与LeNet相比，AlexNet具有更深的网络结构，包含5层卷积和3层全连接[16]，如图2-16所示。AlexNet使用数据增多随机改变训练样本，扩大数据集，并且使用Dropout来有效抑制过拟合，使用ReLU激活函数减少梯度消失现象。

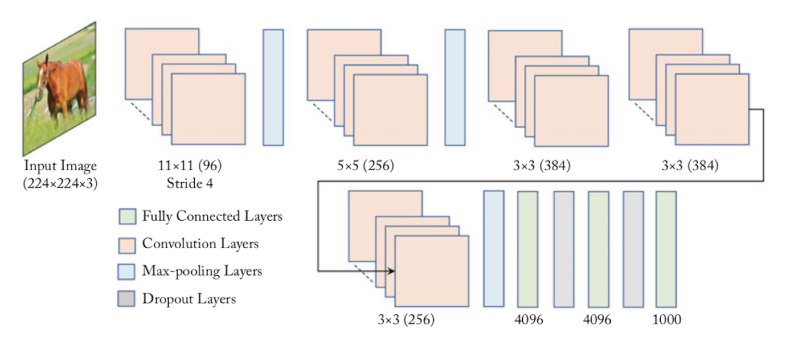


图2-16 AlexNet模型网络结构示意图

（3）VGG

2014年由Simonyan和Zisserman提出的VGG，是当前最流行的卷积神经网络模型之一。虽然AlexNet模型通过构造多层网络，取得了比较好的效果，但是它并没有明确的给出深度神经网络的设计方向，而VGG使用多个3×3的卷积核，与池化层一同来构造深度卷积神经网络，并且取得了较好的成果。VGG模型的结构相对比较简单，实际应用效果不错，深受研究者的欢迎，而且它的网络结构的设计方法也是一种突破与创新，为深度神经网络的构建提供了方向。VGG-16有13层卷积和3层全连接层，如图2-17所示。

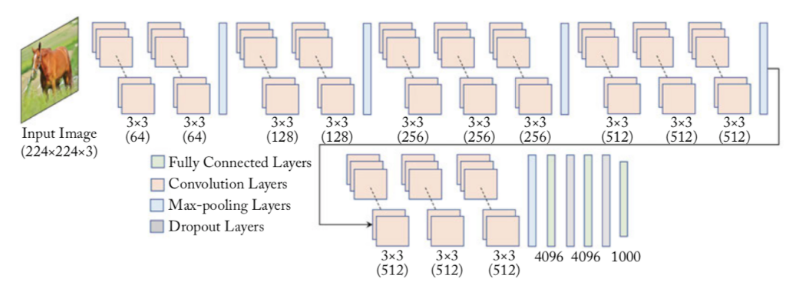


图2-17 VGG模型网络结构示意图

VGG使用3×3的卷积层和池化层来提取特征，并在网络的最后面使用三层全连接层，将最后一层全连接层的输出作为分类的预测。在VGG中，每层卷积都使用的激活函数都是ReLU函数，并在完成卷积和池化操作后的全连接层之后，添加Dropout来防止出现过拟合现象。使用小的卷积核能够有效地减少参数的个数，使得训练和测试变得更加有效[17]，比如使用两层3×3卷积层，可以得到感受野为5的特征图，而比使用5×5的卷积层需要更少的参数。卷积核的大小较小时，可以堆叠更多的卷积层，从而加深卷积神经网络的深度，更有利于图像分类任务的完成。VGG模型的成功，验证了增加网络深度有助于更好地学习图像中的特征模式。

（4）GoogLeNet

2014年ImageNet比赛的冠军是GoogLeNet，它的主要特点是网络不仅有深度，还在横向上具有“宽度”[18]。不同的图像有不同的尺寸大小，想要选择合适的卷积核大小来满足所有图像信息的尺寸以提取特征，这是一件很困难的事情。一般来说，卷积核较大的适合用在图像信息分布空间范围比较大的情况下，对于图像信息分布范围较小的来说，使用较小的卷积核来提取特征的效果比较好。而GoogLeNet提出了一种被称为Inception模块的方案就可以很好地解决上述问题，如图2-18所示。

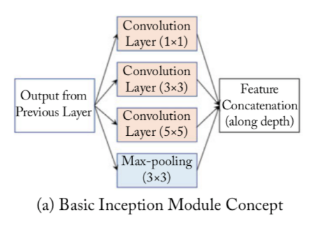


图2-18 Inception模块结构示意图

Inception模块的设计思想是分别用大小为1×1、3×3、5×5的卷积核对输入的图像进行卷积操作，还加一个3×3的最大池化，然后将这四个操作的输出整合拼接在一起，最后经过不同大小的卷积核所提取出来的特征会同时被包含在输出的特征图中。

Inception模块采用多通路的设计形式[19]，不同的支路做卷积操作时使用的卷积核大小不同，每个支路的输出通道数加在一起构成最后的输出特征图的通道数，但是，这有一个十分显著的缺点，即如果有多个Inception模块进行串联操作，输出通道数会变得很大，最后得到的模型的参数规模也会随之增大许多。因此，为了减小参数的规模，需要对原来的结构进行一定程度的改进，可增加三个1×1的卷积层，分别在3x3和5x5的卷积层之前的位置与3x3的最大池化层之后的位置，如图2-19所示。

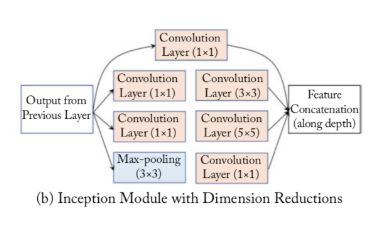


图2-19 改进后Inception模块结构示意图

GoogLeNet的结构[20]如图2-20所示，主要由5个模块来构成卷积部分，为了减小输出的尺寸，在每个模块后面进行3 ×3最大池化操作。

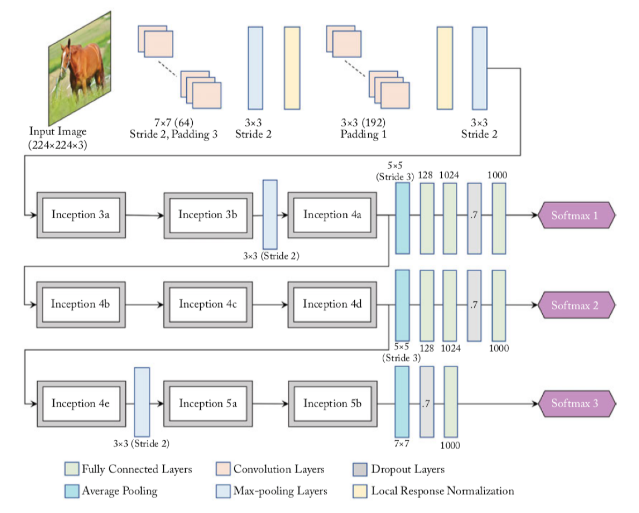


图2-20 GoogLeNet模型网络结构示意图

其中，第一模块使用一个64通道的7 × 7卷积层；第二个模块先进行一个64通道的卷积核大小为1 × 1的卷积操作，然后将通道数增大三倍后进行卷积核大小为3 × 3的卷积操作；在第三模块串联2个完整的Inception块；第四模块串联了5个Inception块；第五模块串联了2 个Inception块；在第五个模块之后，紧跟一个7 × 7的全局平均池化层，然后使用Dropout来抑制过拟合，最后跟一个全连接层，其输出个数为设置的标签的类别数。

（5）ResNet

随着深度学习技术的不断深入发展，神经网络模型也在不断发生着变化，最明显的是模型层数变得越来越多以及结构也越来越复杂，但是，一味的加深网络的层数，并不一定就能得到更好的效果，有时还会随着网络的层数的增加，训练误差不降反升。为了解决这个问题，Kaiming He等人提出了残差网络ResNet[21]。其中，输入信息通过跨层连接，能更快的向前传播数据，或者向后传播梯度，如图2-21所示。

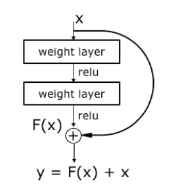


图2-21 残差块结构示意图

ResNet使用残差单元连接将比较原始的输入信息隔层直接添加到后面的网络层，这一层的网络只用学习前一层输出的残差即可，这有助于缓解梯度消失的问题，如图2-22所示。

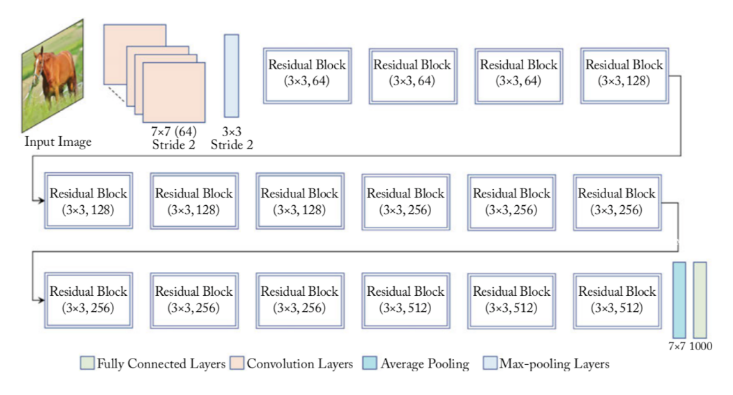


图2-22 ResNet-50模型网络结构示意图

**3.基于深度学习的目标检测算法**

目标检测问题是计算机视觉领域的一大重点。目前，主流的目标检测算法根据有无候选框生成分为两大类：两阶段检测算法和单阶段检测算法。

## **3.1两阶段检测算法**

两阶段检测算法在第一阶段先产生候选区域，在此基础上再对候选区域进行分类，最后调整位置得到检测的结果。典型的算法有：R-CNN、SPPNet、Fast R-CNN、Faster RCNN等。

### 3.1.1 R-CNN

传统的目标检测算法多数都是在图像识别的基础之上进行的。一般情况下，可以在图片上使用穷举法或者利用滑动窗口选出被检测物体可能出现的位置，即区域框，然后对选出的这些区域进行特征提取并对其进行图像识别与分类，在得到所有分类成功的区域后，通过极大值抑制来选出最合适的框，最后输出结果。仿照传统的目标检测办法，R-CNN也是利用提取框，对每个框进行特征提取、图像分类、非极大值抑制[22]这四个步骤进行目标检测，但是做了一定程度的改进：经典的目标检测算法使用滑窗法遍历整张图片，一次判断所有可能的区域，而R-CNN先缩小范围，提取出可能是目标物体的候选区域，再在这些候选区域上进行特征提取与判断，从而减少了许多计算量，提高了模型的效率；除此之外，R-CNN还将传统的特征(如SIFT、HOG特征等)换成了深度卷积网络提取特征。

R-CNN是一种基于区域的卷积神经网络算法，它在卷积神经网络上应用区域推荐的策略，形成自底向上的目标定位模型[23]，它的框架如下图3-1所示。

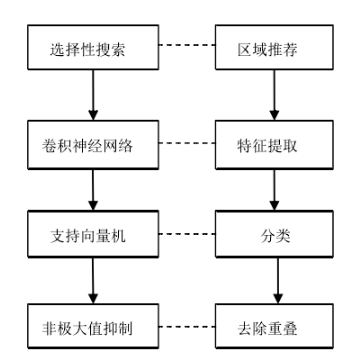


图3-1 R-CNN方法框架

其中，区域推荐常见的方法有selective search和edge boxes。假如有一张图片，经过选择性搜索算法处理后，能产生1000~2000个形状大小不同的候选边框，但是这些候选框之间可能有重叠关系或者包含关系；再使用深度卷积神经对每个候选区域进行特征提取，利用图像中的特征信息，能够在减少候选框的同时保持较高的召回率并且增强模型效果；然后将特征传递给每一种类别的支持向量机分类器，进行分类判断；最后一步进行优化，调整候选框位置，得到最终结果，如下图3-2所示。

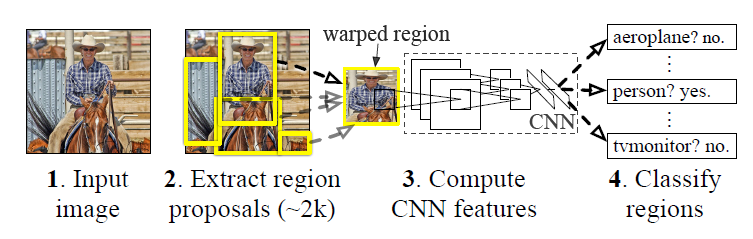


图3-2 R-CNN实现流程

### 3.1.2 Fast R-CNN

基于R-CNN，Fast R-CNN做了一些改进，但是它们的基本框架没有发生很大变化。之所以提出Fast R-CNN，主要是为解决R-CNN训练步骤繁琐，训练速度慢等问题。

Fast R-CNN的流程结构如下图3-3所示。网络有两个输入：图像和对应的候选区域。其中候选区域的产生利用的是selective search算法，流程图中没有具体体现出来。Faster R-CNN在实现过程中对每种类别都会训练一个对应的回归器，且只对非背景的候选区域进行回归，在训练时，回归与网络一同训练，原来的支持SVM分类器也被softmax取代。

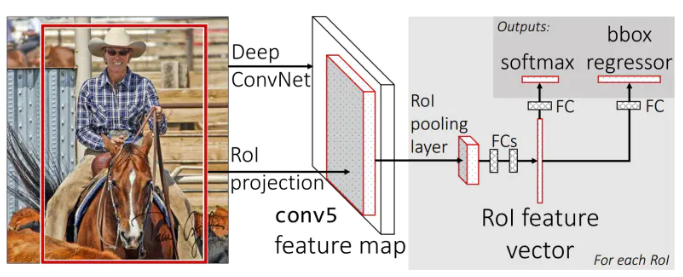


图3-3 Fast R-CNN结构图

### 3.1.3 Faster RCNN

在2016年，Shaoqing Ren等人在Fast R-CNN的基础上继续做出了创新，提出了Faster R-CNN模型[24]。相比于Fast R-CNN，Faster R-CNN不再用selective search算法来产生候选区域，而是用一种新的区域候选网络（RCN）。Faster R-CNN的结构是，首先将图片放入一个全卷积网络，在进行特征提取后送往RNP网络，产生候选区域框，再将这些候选区域框与之前得到的输出特征图一一映射，之后将其送入ROI Pooling生成固定特征图，在此基础上再提取特征，最后进行分类与回归。所以，Faster R-CNN也可以理解为RPN与Fast R-CNN的结合。Faster R-CNN的结构图如图3-4所示。

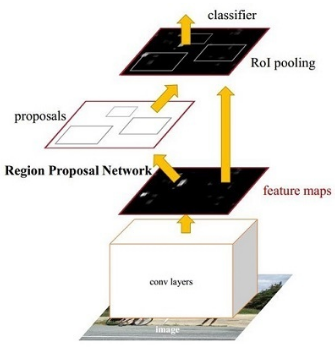


图3-4 Faster R-CNN结构图

## **3.2单价段检测算法**

单阶段目标检测算法与双阶段的不同，不用分步来进行检测，它检测目标物体的种类位置是与产生候选区域在同一阶段完成的。典型的算法有：SSD、YOLO系列等。

### 3.2.1 SSD

对于Faster R-CNN，先通过CNN得到候选框，然后进行分类和回归，而SSD可以一步到位。SSD采用CNN来直接进行检测，提取不同尺度的特征图来做检测[25]，CNN网络一般前面的特征图比较大，到后面通过步幅为2的卷积层或者池化层逐渐降低特征图的大小，其中得到的尺寸较大的特征图适合检测空间范围分布较小的物体，而空间分布范围较广的物体可用较靠后得到的特征图来检测。SSD使用VGG16作为基础模型，并在此基础上增加了多个卷积层，以此来获取更多的特征图来提高检测的精度和效率，如图3-5所示。

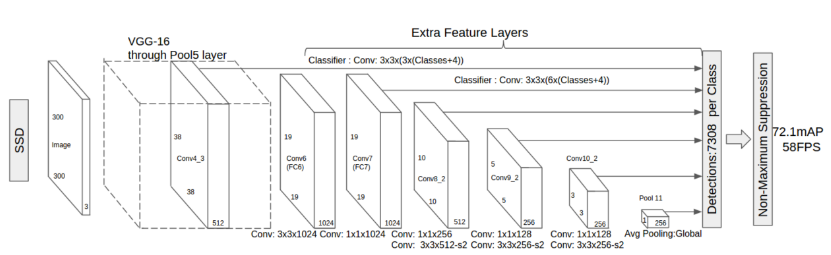
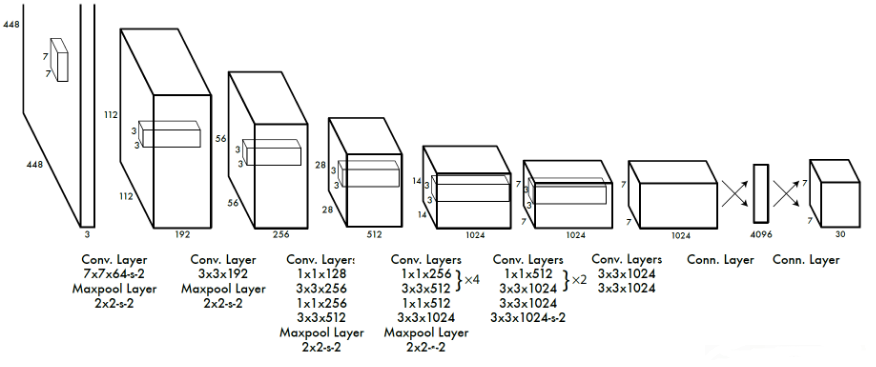


图3-5 SSD网络结构图

### 3.2.2 YOLO

YOLO算法是目前应用比较广泛的单目标检测算法，到现在版本已经迭代更新过多次。YOLO算法采用一个单独的CNN模型实现端到端的目标检测，首先将输入图片resize到448x448，然后送入CNN网络，最后处理网络预测结果得到检测的目标[26]。网络结构参考GooLeNet模型[27]，一共有24个卷积层，4个池化层以及2个全连接层，其网络结构如下图3-6所示。

图3-6 YOLO网络结构图

其中卷积层主要使用的是大小为1 × 1的卷积核，进行通道还原，然后使用3× 3的卷积核对输入图形进行卷积，在卷积层和全连接层，采用的激活函数不再是ReLU函数而是Leaky ReLU。YOLO算法采用的损失函数为均方误差函数，计算公式表示如下：

 （3-1）

第一项是边界框中心坐标的误差项，指的是第i个单元格存在目标，且该单元格中的第j个边界框负责预测该目标[28]；第二项是边界框的高与宽的误差项；第三项是包含目标的边界框的置信度误差项；第四项也是置信度误差，不过是不包含目标的；第五项则是包含目标的单元格的分类误差。

**4.利用卷积神经网络对输液瓶瓶体位置检测**

## **4.1输液瓶数据集**

### 4.1.1输液瓶数据集获取

深度学习技术极大的促进了人工智能的发展，在这些年的发展也极为快速，是近期的热门研究方向。随着数据规模的大幅度增长以及计算力的快速提升，许多以前只能想象的东西逐渐成为现实。但是，要将深度学习很好的应用到现实生活中，首先需要大量的样本数据，在此基础上进行训练，得到的模型才具有更好的泛化能力，预测结果才会更准确。因此，收集大量的输液瓶数据就很重要了。

网络上公开的输液瓶数据并不多，因此，需要人为的去创造数据集。在本文中，主要通过人工去获取数据。本实验所采集的数据，已经过医院相关人员的同意，并且在经过患者与患者家属的许可后才会进行采集，且保证数据只用于实验，不会泄露医院及患者隐私，数据不会外传。为提高工作效率，采集数据的过程中以视频拍摄为主，在后期通过Pr软件，将采集到的视频按帧批量转成图像，如图4-1所示，在转换时，预设选项选择不带Alpha通道的，以供后期进行筛选标记。



图4-1 数据处理

在本实验中，共拍摄17段视频，拍摄高度与输液瓶水平位置大概持平，拍摄距离大约在0.5米与3米之间，拍摄环境分别为病房和输液大厅两种，其中病房环境又分为普通病房内以及病房走廊，参与训练的数据集中输液瓶液体颜色也分为两种：一种为无色透明液体，另一种是黄色半透液体。具体数据类别如下图所示：



图4-2 采集数据类别

其中，(a)与(f)拍摄地点为病房走廊，(c)与(d)拍摄环境为病房内部，(b)和(e)的拍摄地点则为输液大厅，并且，输液瓶瓶内液体颜色的不同与瓶体上的标签并不影响输液瓶瓶体位置的检测。

### 4.1.2数据增强

在训练卷积神经网络模型时，往往需要很大的样本数据量，而在有些情况下，已有的数据集规模不够大，无法训练出好的模型，从而影响预测效果。因此，可采用数据增强的办法，对训练的样本图片进行随机变换，扩大样本数据集，减少过拟合现象的发生，从而得到泛化能力更强大的网络，增强适应性。本文利用numpy和PIL对输液瓶数据集进行图像增广操作。

（1）翻转。图像翻转是一种目前被广泛使用的数据增强策略，与旋转不同，这种方法得到的图像与原图像是对称的。本文使用数组切片的方式来完成，相当于将图片最后一行挪到第一行，倒数第二行挪到第二行，……，第一行挪到倒数第一行。

实现代码为：

image = Image.open('./images/0001.png')

image = np.array(image)

#垂直方向翻转

Image1 = image[::-1, :, :]

#水平方向翻转

Image2 = image[:, ::-1, :]

（2）裁剪。从图像中选择一部分，将选择的部分裁剪出来，生成新的图像。

实现代码为：

image = Image.open('./images/0001.png')

image = np.array(image)

H, W = image.shape[0], image.shape[1]

#高度方向裁剪

H1 = H // 6

H2 = H

Image3 = image[H1:H2, :, :]

#宽度方向裁剪

W1 = W//6

W2 = W//6 \* 5

Image4 = image[:, W1:W2, :]

#两方向同时裁剪

image5 = image[H1:H2, W1:W2, :]

（3）调整亮度。图像是由像素点构成的矩阵，每个像素点都有具体的数值，改变亮度，会影响每个像素点值的变化，在后期做卷积时也是新的图像。

实现代码为：

image = Image.open('./images/0001.png')

image = np.array(image)

#亮度变暗

Image6 = image\*0.5

#亮度变亮

Image6 = image\*2.0

#图片的像素值必须在0和255间，×2后要进行数值裁剪

Image7 = np.clip(image7, a\_min=None, a\_max=255.)

（4）图像压缩。将图像化为像素点构成的矩阵后，隔行或者列进行采样，得到的图片大小会发生变化，在后面还要对缩小的边界进行填充。

实现代码为：

image = Image.open('./images/0001.png')

image = np.array(image)

#纵向压缩

Image8 = image[::2, :, :]

#横向压缩

Image9 = image[:, ::2, :]

#等比压缩

Image10 = image[::2, ::2, :]

### 4.1.3数据标注

数据标注对于实现人工智能有着很重要的意义，标记过程需要借助标记工具，如LabelImg、Vatic、Sloth等；常见的数据标注类型有：图像标注、文本标注、视频标注、语音标注等；而标注的基本形式有标注画框、3D画框、文本转录、图像打点、目标物体轮廓线等。在进行目标检测的模型设计之前，首先要对原始的数据集进行处理，在上面进行人工标注，目的是让机器知道所要检测的目标对象，然后根据这来进行学习，最后实现预测。在进行标注前，需要对图像数据进行预处理，筛选掉没用的数据或者质量较低的数据，并将所有图像整理为PNG格式，将图像尺寸调整为512\*512。

由于本文所使用的数据集是自己创建的，所以还需要进行手工标注。在本实验中使用的数据标注工具为LabelImg，它可以在原始图像中框出要检测的物体的位置并设置相应的标签，最后对每一张图片生成相应的xml文件，里面包括标注后的图像的基本信息。

在使用LabelImg为训练样本和测试样本进行标注时，首先打开要标注的数据集的文件夹，需注意存放数据集的文件路径不要有中文；在左侧创建所需的标签名，以供标注时使用；在图像中用鼠标框出要检测的物体范围，注意范围尽量将物体全部包括，但也不要留出太多剩余空间，标注框尽可能贴紧物体边界，如下图4-3所示。



图4-3 数据标注

将标注好的图像保存下来，保存后得到的文件为后缀名是.xml的文本，内容如下图4-4所示，其中bndbox 里的xmin、ymin和xmax、ymax分别为标准框的左上角和右下角坐标。



图4-4 数据标注文本

## **4.2深度学习框架与环境搭建**

近些年，越来越多的领域中出现了深度学习的身影，如医疗、交通、安防、社交等，面对着各种各样的应用场景，使用深度学习框架能够帮助人们节省大量重复的简单工作，设计者可以将主要精力投入到实际业务场景和模型自身设计中。使用深度学习框架来进行模型构建，屏蔽了底层繁琐的设计与实现，能够节省设计者在此方面的大量精力，让使用者只需关注模型的逻辑结构即可，并且简化计算，降低深度学习的难度门槛；另一方面可以不用再麻烦的部署适配环境，提高了可移植性，可选择将代码部署到CPU、GPU或者移动端上，提高模型训练的效率。

### 4.2.1深度学习框架

目前，常见的深度学习框架主要有TensorFlow、Caffe、Pytorch、Keras、PaddlePaddle等。

TensorFlow是目前使用最为广泛的深度学习框架之一，它是基于计算图实现自动微分系统[29]，使用数据流图进行数值计算，其数据流图中的节点代表数学运算，边表示这些节点之间的数据交互。TensorFlow编程接口支持Python、C、C++、Java、Go、R等语言，不过就目前来讲，Python语言是用户最多的选择。

Caffe是另一个面向图像处理领域的、比较流行的深度学习框架，训练速度快、可读性好，并且它的模块化设计提高了可扩展性。Caffe把网络中的数据依次抽象成Blob,各网络层抽象成Layer，每一层之间可以实现前向传播和反向传播，多层连接在一起抽象成Net，而网络模型的求解方法抽象成Slover。Caffe完全用C++语言来实现，具有较好的移植性。

PyTorch的前身是Torch，可以用来建立深度神经网络和执行张量计算。Torch是一个基于Lua的框架，而PyTorch则运行在Python上。PyTorch是一个Python包，它提供张量计算，张量是多维数组，就像numpy的ndarray一样，它也可以在GPU上运行。PyTorch使用动态计算图，PyTorch的Autograd软件包从张量生成计算图，并自动计算梯度。

Keras由纯Python编写而成并基于Tensorflow、Theano以及CNTK后端，在这三者上也可以运行，号称10行代码即可搭建神经网络，操作相对简单，配置环境也不复杂，大大降低了建模的难度。

作为唯一一款国产的深度学习框架，飞桨(PaddlePaddle)集合了许多功能组件，部署使用灵活方便，对初学者来说安装简单，容易上手，也能够满足模型开发、训练和部署的全流程需求。综合上述的几种深度学习框架情况，本文最终采用的深度学习框架是百度出品的深度学习平台飞桨(PaddlePaddle)，主要原因如下：

1. PaddlePaddle具有丰富的工具组件， 开放 PaddleHub、PARL、AutoDL Design、VisualDL、EDL 和PGL 一系列深度学习工具组件。
2. 分为动态图和静态图两种变成方式，能根据需求选择更容易调试的命令式编程或者训练速度更快的声明式编程。
3. 支持Python和MATLAB接口，可多机多卡并行，支持CPU和GPU，甚至可将自己的代码移植到百度的AI Studio平台上运行。
4. 占用内存小，速度快，目前已应用到很多实际项目中，便于后期实现完整的智能输液监测并将其应用到实际生产中。

### 4.2.2环境搭建

工欲善其事，必先利其器。在编写自己的项目之前，需要安装配置好所需的一些工具，在开始本实验前，需要搭建好环境，以便接下来的模型训练和预测。本机系统是64位Windows10，使用Anaconda来进行PaddlePaddle的安装，开发工具使用的是Pycharm。

首先安装Anaconda。Anaconda是跨平台的，选择Windows版本，根据需要，选择64位Python3.7版本下载安装。在安装过程中，勾选添加到环境变量这一选项可以减少许多麻烦。为了加快训练速度，本机安装GPU版的PaddlePaddle，在此之前，要先安装CUDA10和cudnn7.3+，需注意CUDA版本要和cudnn版本配套。在安装完PaddlePaddle后，测试一下安装是否成功，如下图所示。

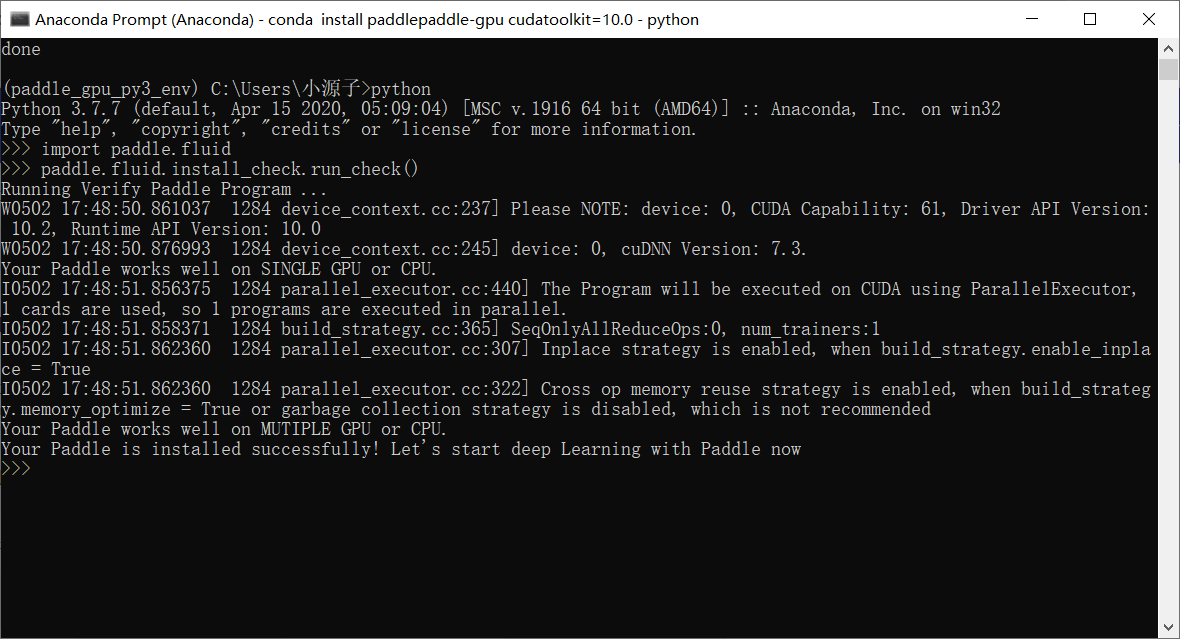


图4-5 PaddlePaddle安装

使用Pycharm开发时，在创建新项目后，要选择系统的Python环境，在这时选择外部Python环境，选择Conda Environment里python的路径，如图4-6所示。

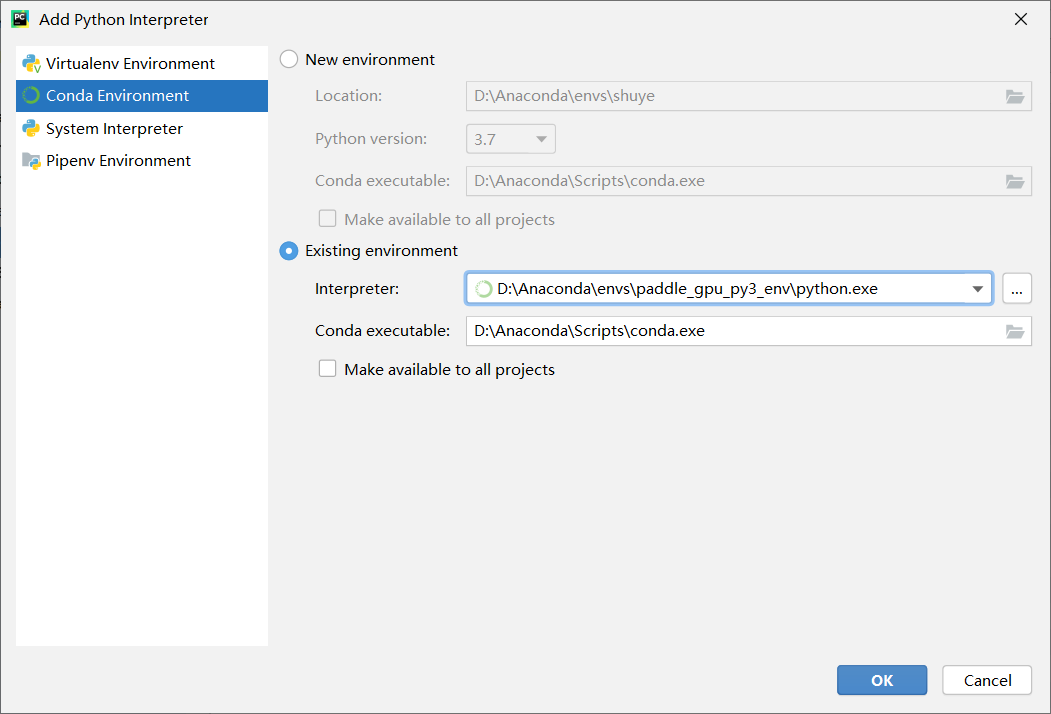


图4-6 Pycharm环境配置

## **4.3模型训练**

在本实验中，训练的模型的设置为：没有启用tiny版本；训练轮数设置是110轮；每批次训练的图片数量为10，因为对完整的YOLO V3来说，每一批的训练样本不能太多，否则内存会不够；为了增加训练速度，选择使用GPU进行训练；优化器算法选用的是随机梯度下降法，初次学习率设置为0.002；为了加快训练和预测速度，将图片尺寸从原来的512\*512压缩为384\*384。

代码参数如下：

train\_parameters = {

    "data\_dir": "data",

    "train\_list": "train\_list.txt",

    "eval\_list": "eval\_list.txt",

    "class\_dim": -1,

    "label\_dict": {},

    "num\_dict": {},

    "image\_count": -1,

    "continue\_train": True,

    "pretrained": False,

    "pretrained\_model\_dir": "./pretrained-model",

    "save\_model\_dir": "./yolo-model",

    "model\_prefix": "yolo-v3",

    "freeze\_dir": "freeze\_model",

    "use\_tiny": False,

    "max\_box\_num": 20,

    "num\_epochs": 110,

"train\_batch\_size": 10,

"use\_gpu": True,

    "yolo\_cfg": {

        "input\_size": [3, 384, 384],

"anchors": [9, 12, 15, 28, 30,21,28,56, 57, 42, 54, 110, 107, 83, 144, 183, 344, 301],

        "anchor\_mask": [[6, 7, 8], [3, 4, 5], [0, 1, 2]]

    },

    "ignore\_thresh": 0.7,

    "mean\_rgb": [127.5, 127.5, 127.5],

    "mode": "train",

    "multi\_data\_reader\_count": 4,

    "apply\_distort": True,

    "nms\_top\_k": 300,

    "nms\_pos\_k": 300,

    "valid\_thresh": 0.01,

    "nms\_thresh": 0.45,

    "image\_distort\_strategy": {

        "expand\_prob": 0.5,

        "expand\_max\_ratio": 4,

        "hue\_prob": 0.5,

        "hue\_delta": 18,

        "contrast\_prob": 0.5,

        "contrast\_delta": 0.5,

        "saturation\_prob": 0.5,

        "saturation\_delta": 0.5,

        "brightness\_prob": 0.5,

        "brightness\_delta": 0.125

    },

    "sgd\_strategy": {

        "learning\_rate": 0.002,

        "lr\_epochs": [30, 50, 65],

        "lr\_decay": [1, 0.5, 0.25, 0.1]

    },

    "early\_stop": {

        "sample\_frequency": 50,

        "successive\_limit": 3,

        "min\_loss": 2.5,

        "min\_curr\_map": 0.84

    }

}

1. **算法评价与实验结果分析**

## **5.1算法评价**

本实验所使用的算法是YOLO系列的第三个版本，相较于其他算法，YOLO v3泛化能力更强，训练速度也更加快，如下图所示。

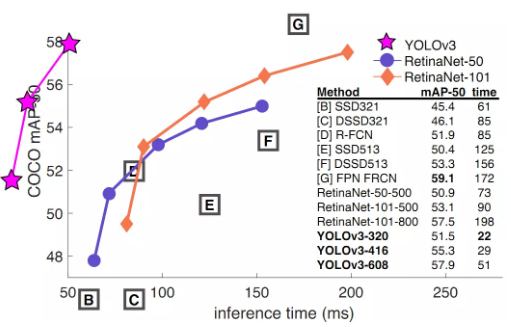


图5-1 YOLOv3与其它模型的性能对比

YOLO v3将按照规则将输入的图像分为小方格，假设一张图片分为A×A个，在每个格子上生成B个预测框，预测框的大小可手动设置，也可以自动计算生成，每个候选框预测内容包括框的坐标大小信息（中心点坐标与框的宽高）、置信度和C个类别的概率，因此YOLO v3输出层的通道数为A×A×B×（5+C）。YOLO v3的loss函数值由位置误差、置信度误差和分类误差三部分组成。YOLOv3 的网络结构由基础特征提取网络、multi-scale特征融合层和输出层组成[30]，如图5-2所示。

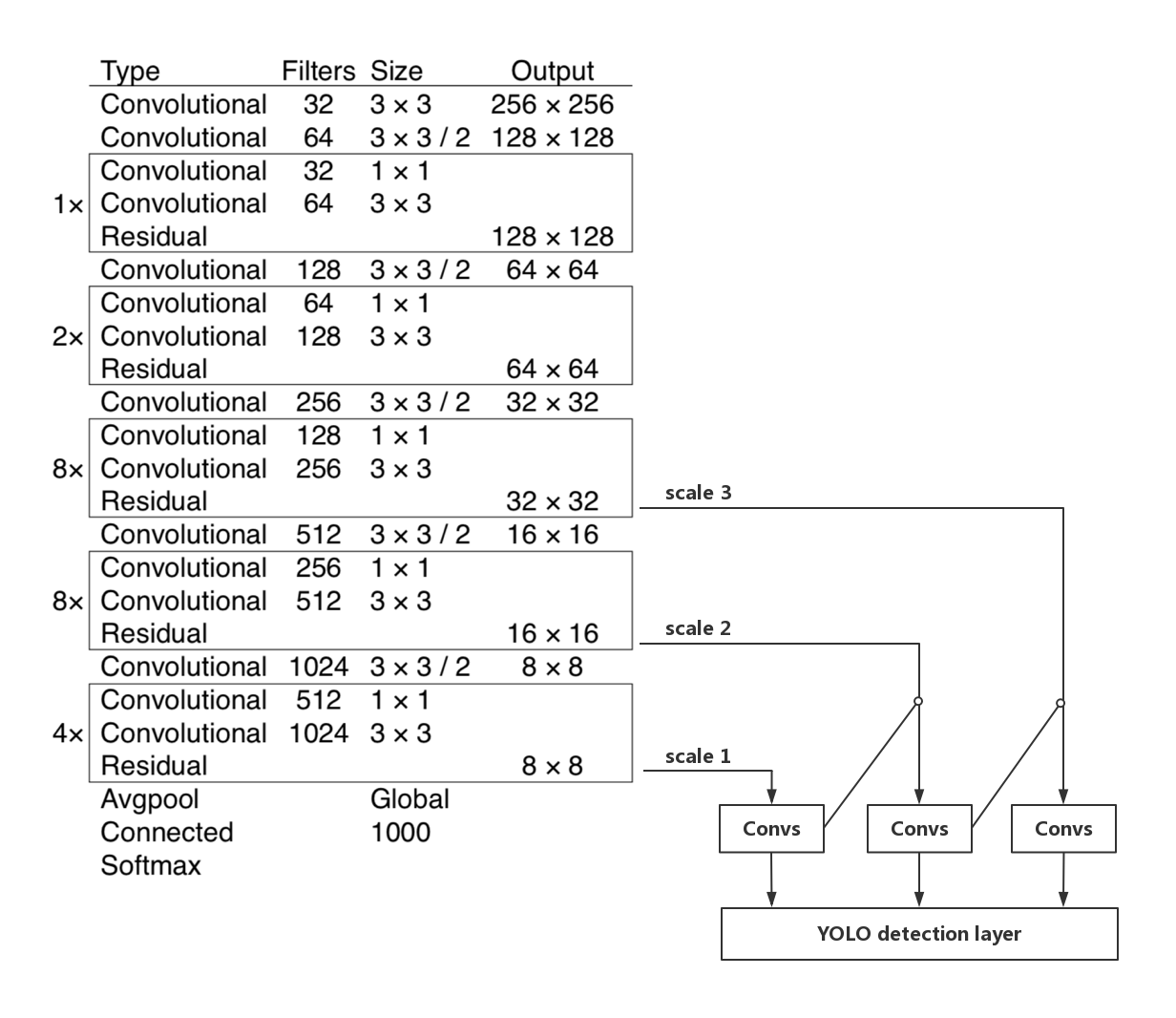


图5-2 YOLO v3网络结构

YOLO v3的骨干网络使用的是 [DarkNet53](https://arxiv.org/abs/1612.08242" \t "https://aistudio.baidu.com/aistudio/projectdetail/_blank)，DarkNet53 基本采用了全卷积网络，没有池化层和全连接层，使用步长为2的卷积层代替，以此来减小输出图像尺寸，为了避免网络层数过深时发生梯度消失现象，[DarkNet53](https://arxiv.org/abs/1612.08242" \t "https://aistudio.baidu.com/aistudio/projectdetail/_blank)还添加了残差单元。与之前的YOLO算法相比，YOLO v3同时使用尺度不同的特征图来进行目标检测，有效的解决了小目标识别不敏感问题，13×13尺度的用来检测较大的目标，26×26尺度的可以用来检测空间分布范围比较适中的目标，而52×52的则用来检测较小的目标。

## **5.2实验结果分析**

通过前几节的介绍，证明了本实验所设计的算法对于输液瓶瓶体位置检测的可行性以及具有较强的鲁棒性。最终，本实验训练出来的YOLO v3神经网络模型，在被检测图片中有较少数目标物体，以及目标物体尺寸相对于整张图像尺寸占比较大时，能够快速准确的检测出输液瓶瓶体位置，效果如下图所示。

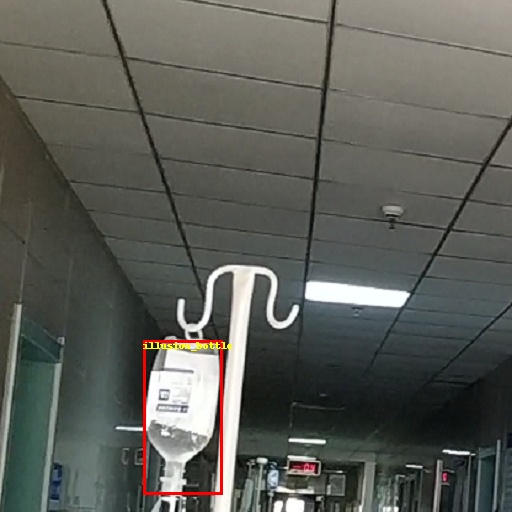


图5-3 输液瓶位置检测效果图1



图5-4 输液瓶位置检测效果图2



图5-5 输液瓶位置检测效果图3

在模型的训练过程中，将数据集按8:2比例划分为训练集与测试集，每个batch的大小为10，学习率为0.002，在经过一个batch的训练后，对权值进行优化，并且根据预测框与真实框的相差位置，计算出平均loss值，如下图5-6所示。使用随机梯度下降法，loss值在训练之初下降明显，随后loss值下降幅度变缓，并逐渐趋于稳定，在经历110轮的训练后，loss值趋于2.0左右，准确率在94%左右。

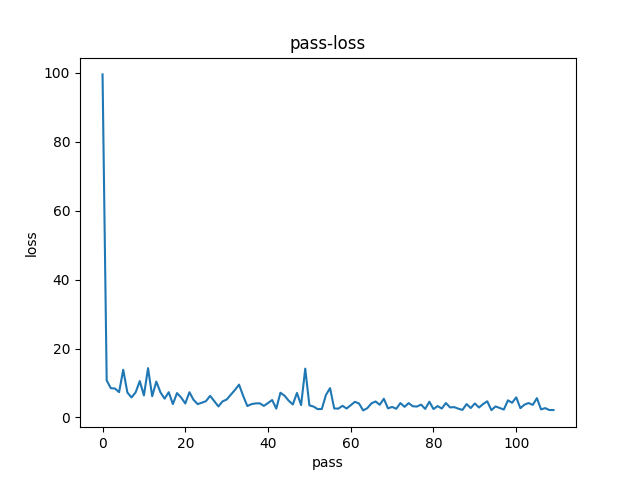


图5-6 学习率为0.002的loss幅值变化图

SGD算法每计算一个样本，数值就迭代更新一次，在样本数量特别多的情况下，为提高效率可以只选用其中的某些样本进行迭代，同样可以得到最优解。因为不需要计算全部样本，使用随机梯度下降法训练的速度会快很多，特别适合用在规模比较大的数据集上。但是，由于是样本是随机抽取的，得到的梯度肯定是有误差的，只是保持在整体大方向上是正确的，每一次迭代的结果与所抽的样本有关，所以绘制出的函数图可能会有明显的锯齿状，即梯度含有较大噪声，因此不能准确的反映真实梯度情况。

改变参数，将学习率设置为0.001，在进行模型的训练，计算出的平均loss值变化如下图5-7所示。通过与上图对比，可以明显发现改变学习率后，模型的效果好了很多，梯度噪声减少许多，loss值也有所降低，准确率升高到98%左右。

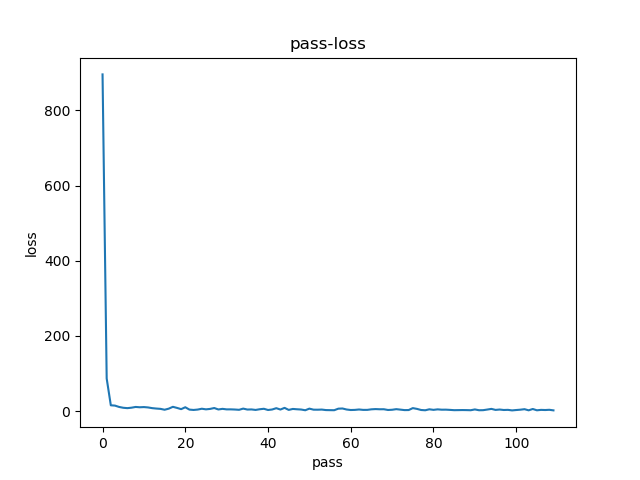


图5-7 学习率为0.001的loss幅值变化图

但是，本实验还有需要改进的地方，例如针对同一张图片中有较多需检测的目标物体，且物体不太清晰时，训练出来的模型不能很好的进行准确预测，如下图5-8所示。

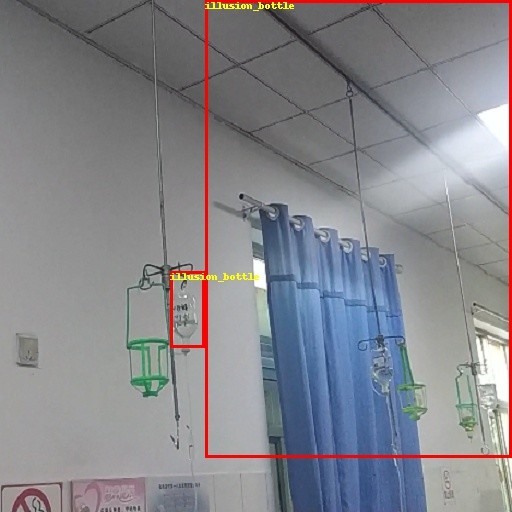


图5-8 输液瓶位置检测效果图4

**6.总结与展望**

随着科技的不断发展，人工智能被越来越多的应用到各行各业之中，医疗作为与我们生活息息相关的一个领域，人工智能在此方面还有很大的发展潜力。本文首先介绍了智能输液监测的意义与在国内外的发展状况，由此引出本文的主题，即基于深度学习的输液瓶瓶体位置检测，为实现基于深度学习的智能输液监测打下坚实基础，并且还大致介绍了目前比较典型的一些目标检测算法与深度学习的一些框架，结合查阅的相关文献资料，进行比较选择，然后设计本文的实验方案，搭建环境，训练自己的神经网络模型，最后成功实现输液瓶瓶体位置的检测。但是，由于数据集需要人工采集以及硬件资源有限，目前训练出的模型还不是很完美。

目前，基于深度学习的智能输液监测只实现了前期的功能，还不是一个完整的系统，且适应场景有较大局限性，需要再增大数据集的规模。之后，在此基础上，还要再进一步设计与实现，来完成完整的智能输液监测系统，减轻医护工作人员的工作负担，减少病人的陪护需求与成本，为我国的智慧医疗事业发展添砖加瓦。

致谢

2020年是一个非同寻常的年份，在年初发生蔓延的疫情，给今年带来了一个不一样的毕业季。我们不能像往常一样，在寒假假期结束后返回校园进行常规的教育教学活动，只能宅在家中，通过网络联系师友。在这个特殊的时期，大学生活即将匆忙的画上句号，为此深感遗憾。回想起这四年里所遇见的人和经历的和事，都感觉如在眼前。在论文即将完成之际，我要感谢在此过程中给予过我支持与帮助的人。

感谢我的毕设指导老师芦碧波老师，从一开始芦老师就很关心学生情况，在我开始进行毕设选题到最终完成毕业论文的整个过程中，芦老师一直悉心指导，认真负责。受全国疫情影响，不能面对面指导此次毕设，但是芦老师每周都会在线上给我们开会，解决所遇到的困难。由于我的毕设题目涉及医疗，需前往医院实地采集数据，芦老师还再三叮嘱要注意个人安全。芦老师深厚的专业素养，严谨的治学态度以及关切的提醒问候，都另我十分敬佩与感谢。

感谢驻马店市第一人民医院的护士许思及相关医护工作人员，为我介绍输液相关背景知识，允许并辅助我进行数据采集工作，同时还要感谢同意我进行拍摄的患者及患者家属，谢谢他们的理解支持与配合，为我顺利完成本次毕设创造了一个好的开端。

感谢河南理工大学与计算机学院，为我提供了良好的教育平台，让我能够在这里接受到更多更深的知识；感谢计算机学院的每一位领导、任课老师与辅导员，感谢他们四年如一日的付出与孜孜不倦的教导，为我完成毕设打下了坚实基础；感谢我的室友以及同学们，感谢他们在我遇到困难时伸出的援助之手，给予我很多关心与爱护，陪着我一同成长。

最后要感谢我的家人，一直在背后默默的奉献，鼓励并支持我求学，如果没有他们，就没有今天的我。

参考文献

1. 封宇飞,裴艺芳,倪倩.静脉输液技术发展沿革[J].临床药物治疗杂志,2014,12(06):11-15.
2. Ogawa H, Maki H, Tsukamoto, et al. A new drip infusion solution monitoring system with a free-flow detection function [C]. // Engineering in Medicine & Biology Society. 2010 : 1214-1217.
3. Cui Jianwen, Zhang Han. Design of Intravenous Infusion Monitoring and Alarm system Based on Wireless Communication Technology[C]. // IEEE International Conference Mechatronics & Automation. 2011.
4. 周涛.医院无线输液系统的设计应用[J].无线互联科技,2016(18):64-65.
5. 陈春霞.浅析机器学习的发展与应用[J].信息系统工程,2017(08):99-100.
6. 韦坚,刘爱娟,唐剑文.基于深度学习神经网络技术的数字电视监测平台告警模型的研究[J].有线电视技术,2017(07):78-82.
7. 巴桂.基于卷积神经网络的图像分类算法[J].电脑与信息技术,2020,28(01):1-3.
8. 孔维芳. 小鼠视网膜片层化及神经干细胞增殖与分化的研究[D].河南大学,2012.
9. 王永超. 基于深度学习的车道线检测[D].燕山大学,2018.
10. 李新叶,宋维.基于深度学习的图像语义分割研究进展[J].科学技术与工程,2019,19(33):21-27.
11. 王雪丽.基于卷积神经网络的双目测距研究[J].电子质量,2019(10):1-3+7.
12. 周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017,40(06):1229-1251.
13. 霍启明. 基于深度学习的个性化草图检索系统的研究与实现[D].北京邮电大学,2019.
14. 周双双. 基于学习策略的地形合成技术研究[D].华东师范大学,2018.
15. Y. LeCun,B. Boser,J. S. Denker,D. Henderson,R. E. Howard,W. Hubbard,L. D. Jackel. Backpropagation Applied to Handwritten Zip Code Recognition[J]. Neural Computation,1989,1(4).
16. 王泽瑜. 室内场景智能识别系统的设计与实现[D].辽宁大学,2018.
17. 田东. 基于卷积神经网络的深度学习算法研究与实现[D].上海交通大学,2017.
18. Szegedy C , Liu W , Jia Y , et al. Going Deeper with Convolutions[J]. arXiv preprint arXiv:1409.4842, 2014.
19. 罗聪. 基于深度学习的自然场景文本检测与定位方法研究[D].华南理工大学,2018.
20. Hua Cui,Gege Yuan,Ni Liu,Mingyuan Xu,Huansheng Song. Convolutional neural network for recognizing highway traffic congestion[J]. Journal of Intelligent Transportation Systems,2020,24(3).
21. Dawei Dai,Liping Yu,Hui Wei. Parameters Sharing in Residual Neural Networks[J]. Neural Processing Letters,2020,51(3).
22. 马增妍.基于卷积神经网络和上下文模型的目标检测[J].中国安全防范技术与应用,2018(06):28-33.
23. 李光曜. 基于深度学习的遥感图像目标检测方法研究[D].石家庄铁道大学,2019.
24. Ren Shaoqing,He Kaiming,Girshick Ross,Sun Jian. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE transactions on pattern analysis and machine intelligence,2017,39(6).
25. 单倩文,郑新波,何小海,滕奇志,吴晓红.基于改进多尺度特征图的目标快速检测与识别算法[J].激光与光电子学进展,2019,56(02):55-62.
26. 李名波.基于机器学习的目标检测算法综述[J].计算机产品与流通,2019(06):154-155.
27. 杨梓豪. 基于区域卷积神经网络的物体识别算法研究[D].北京邮电大学,2018.
28. 霍沫霖,林国强,高昆仑.人工智能在能源服务中的应用[J].供用电,2019,36(01):54-60.
29. 朱汇龙,刘晓燕,刘瑶.基于贝叶斯新型深度学习超参数优化的研究[J].数据通信,2019(02):35-38+46.
30. 陈宏彩,任亚恒,郝存明,程煜,张效玮.基于YOLOv3的医药玻璃瓶缺陷检测方法[J].包装工程,2020,41(07):241-246.