**基于卷积神经网络的肺炎X光片病灶识别**

**Recognition of Pneumonia by X-ray Based on Convolutional Neural Network**

**专 业：物联网工程**

**姓 名：杨雨洁**

**指 导 教 师：杨巨成**

**申请学位级别：学 士**

**论文提交日期：2020年5月25日**

**学位授予单位：天津科技大学**

摘 要

21世纪以来,随着科技飞速发展，人工智能逐渐融入到人们的日常生活中。深度学习是人工智能领域近几年来发展的最火的方向之一，其中卷积神经网络是深度学习的一种模型，它被广泛应用到各种智能任务中，包括图像识别、自然语言处理、目标检测等方面，在医学图像处理中的功能也不容小觑，其中就包括对肺炎X光片的病灶识别。

通过计算机技术对医学影像进行辅助检测可以节省大量的人力、财力和物力，而且随着目前计算机的各个硬件性能逐渐提高、海量数据和数据挖掘技术的快速发展，深度学习的模型算法被广泛应用在计算机视觉领域，是图像处理的最优方法之一。使用卷积神经网络搭建模型算法，将肺炎医学图像作为算法的输入，输进训练好的模型，最终得出分类结果，这一过程比传统人工判断识别肺炎X光片的病灶要更快且准确率更高。

卷积神经网络（Convolutional Neural Networks，CNNs）通过反向传播过程学习训练模型，而且具有共享权重、稀疏性连接等特性，大量地减少了参数数量，对提升整个网络的计算速率具有重要作用。残差神经网络（Residual Neural Network，ResNet）属于CNNs的一种，相比于一般的CNNs，它具备一种特殊的结构—残差模块（Residual Block），这种结构能够很好地解决梯度消失或梯度爆炸现象，通过加深网络的层数来提高网络的性能。根据ResNet的层数不同，可以将其划分为ResNet18、ResNet34、ResNet50、ResNet101、ResNet152等。其中ResNet34是一种常用的残差网络模型，该模型网络结构较为简单，网络由7\*7卷积层、3\*3最大池化下采样层、一系列的残差结构以及平均池化下采样和全连接层构成，它采用批量归一化（Batch Normalization，BN）对数据进行标准化，能够加快网络的收敛速度和运算速率，加以优化后运用到肺炎CT影像的病灶识别检测上，在Bitahub服务器平台的训练准确率达到了84.4451%，对智能医疗的发展具有重要的影响。

关键词：深度学习； 卷积神经网络； 残差神经网络； 肺炎； 医学影像

**Abstract**

Since the 21st century, with the rapid development of technology, artificial intelligence has gradually integrated into people's daily lives.Deep learning is one of the hottest directions in the field of artificial intelligence in recent years. Convolutional neural network is a model of deep learning. It is widely used in various intelligent tasks, including image recognition, natural language processing,Target detection and other aspects, the function in medical image processing can’t be underestimated, which includes the identification of lesions of pneumonia X-rays.

Auxiliary detection of medical images through computer technology can save a lot of manpower, financial resources and material resources, and with the current performance of various hardware of computers and the rapid development of massive data and data mining technology, deep learning model algorithms are widely used in.The field of computer vision is one of the best methods for image processing.Use a convolutional neural network to build a model algorithm, take pneumonia medical images as input to the algorithm, enter the trained model, and finally get the classification results. This process is faster and more accurate than traditional manual judgment to identify the lesions of pneumonia X-rays.The rate is higher.

The core idea of Convolutional Neural Networks (CNNs), like deep learning, is to learn the training model through the back-propagation process, and has the characteristics of shared weights and sparse connections.The calculation rate of the entire network plays an important role.Residual Neural Network (ResNet) is a type of CNNs. Compared with general CNNs, it has a special structure-Residual Block (Residual Block), which can solve the gradient disappearance well.Or the gradient explosion phenomenon is a major breakthrough in improving the performance of the network by deepening the number of layers of the network. The maximum number of layers has exceeded 1,000 layers.According to the different layers of ResNet, it can be divided into ResNet18, ResNet34, ResNet50, ResNet101, ResNet152 and so on.Among them, ResNet34 is a commonly used residual network model. The network structure of this model is relatively simple. The network consists of a 7 \* 7 convolutional layer, a 3 \* 3 maximum pooled downsampling layer, a series of residual structures, and an average pooled downsampling.It is composed of a fully connected layer. It uses batch normalization (BN) to standardize the data, which can greatly accelerate the network convergence speed and operation rate. The training accuracy on the Bitahub server platform has reached 84.4451%, After being optimized and applied to the lesion recognition and detection of pneumonia CT images, it has an important impact on the development of intelligent medical treatment.

**Key words：**Deep Learning; Convolutional Neural Network; Residual Neural Network; Pneumonia; Medical Image

目 录

[第一章 绪论 1](#_Toc42516887)

[第一节 研究背景及生活意义 1](#_Toc42516888)

[第二节 国内外研究现状 1](#_Toc42516889)

[第三节 课题研究的主要内容 2](#_Toc42516890)

[第四节 本文结构 3](#_Toc42516891)

[第二章 涉及技术的相关介绍 4](#_Toc42516892)

[第一节 卷积神经网络 4](#_Toc42516893)

[第二节 激活函数 5](#_Toc42516894)

[第三节 数据增强 6](#_Toc42516895)

[第四节 批量标准化BN 8](#_Toc42516896)

[第五节 特征分类 10](#_Toc42516897)

[第三章 基于卷积神经网络的肺炎X光片病灶识别模型设计 13](#_Toc42516898)

[第一节 网络结构设计 13](#_Toc42516899)

[第二节 卷积神经网络层次设计 16](#_Toc42516900)

[第三节 预处理模型ResNet34 19](#_Toc42516901)

[第四节 网页前后端设计 20](#_Toc42516902)

[第四章 基于卷积神经网络的肺炎X光片病灶识别算法实现 22](#_Toc42516903)

[第一节 数据集 22](#_Toc42516904)

[第二节 医学图像预处理 23](#_Toc42516905)

[第三节 LBP算法特征提取 25](#_Toc42516906)

[第四节 模型参数初始化及反向传播 26](#_Toc42516907)

[第五节 肺炎病灶识别 28](#_Toc42516908)

[第五章 总结 30](#_Toc42516909)

[参考文献 X](#_Toc42516910)

[致 谢 X](#_Toc42516911)

# 第一章 绪论

## 第一节 研究背景及生活意义

21世纪以来，随着我国科学技术和社会经济的稳速发展，全国人民的生活水平及生活质量不断提高，从而使得人们生活的基本矛盾发生了变化，人民对美好生活的需要愈来愈高[1]。虽然近几十年以来，我国大力发展科技，经济水平大幅度增长，但是却给人民的生活带来一些问题—我们赖以生存的自然环境遭受了一定的破坏，空气中充斥着难闻的气味，肺炎患病率随之上升。

肺炎可以按照不同的分类方式进行分类，其中根据解剖类型的不同，可将其分为大叶性肺炎、支气管肺炎、间质性肺炎等。迄今为止，对肺炎X光片的研究识别主要是依靠专业的医生凭借丰富的经验作出正确的判断。但是，仅仅依靠医生的力量是不够的，因为肺炎X光片中不同的肺炎类型和正常的X光片的差别是微乎其微的，仅依靠医生将会浪费大量的时间，而且有些偏远地区的医务人员短缺，不能对患者进行及时的治疗。尤其2020年，全球爆发了新冠肺炎，仅中国确诊病例就达8万余人，促使我们使用科技的力量战胜疫情。

计算机辅助诊断（computer aided diagnosis，CAD）是一种分析计算方法，该方法通过其医学影像及其技术处理来实现，并且可以结合利用计算机的高速度、优性能来实现对医学影像的精准判断[2]。深度学习属于人工智能领域的一个研究方向，目前深度学习的发展处于鼎盛时期，在海量数据挖掘、自然语言处理、人脸识别、图像识别处理等领域都有很大的突破。卷积神经网络是深度学习相关算法之一，它能够自主地进行特征学习，该模型具有能够进行卷积计算和包含多层网络的网络结构，网络结构中包括多个卷积层和亚采样层，卷积核进行参数共享和卷积层之间的稀疏性连接，能够减少计算的规模，提高运行速度，在计算机视觉、医学影像识别方面具有重要的作用。

## 第二节 国内外研究现状

迄今为止，计算机技术在医学影像处理方面的应用在全球发展的如火如荼[1]，基于卷积神经网络的算法被认为是计算机视觉领域内发展速度最快、性能最优的算法，深受广大科研和技术人员的青睐，是现如今人工智能研究的热点。

早期，在对医学影像进行分类时，主要是人工进行特征提取，或者是利用图像处理的相关的算法模型对医学影像进行特征提取，一般情况下，这些特征都是经过定性表达的[2]。当我们使用卷积神经网络训练模型时，对用于训练的数据集有着严苛的要求。首先，数据集规模要足够大；其次，用来训练和验证的数据需要带有标签，能够进行有监督的学习。当然，很多平台上也有许多公用的数据集可以检验算法的优劣，用来训练模型的数据集的规模越大，算法的预测准确率也会随之升高。目前，常见的一些公共数据集平台有Kaggle、UCI机器学习库、VisualData等，其中大家最熟悉的机器学习公共数据集应当是ImageNet，该数据集由计算机界女神李飞飞参与创建，在计算机视觉领域具有很强的影响力。Kaggle平台为全球热爱机器学习的程序员们提供了权威的竞赛平台，该平台创建于2010年，创建地在墨尔本，安东尼·高德布卢姆（Anthony Goldbloom）作为联合创始人兼首席执行官。此外，ImageNet是一个超大的数据库，并且还具备可视化的功能，与其同名的比赛名誉国内外，被认为是深度学习革命的开端。此外，ImageNet项目每年都会举行一次大型的计算机视觉挑战赛，即ImageNet大规模视觉识别挑战赛（ILSVRC）。

由于近些年各种轻重工业的无限制的发展，导致全球空气质量不断恶化，各种呼吸系统疾病大量爆发，肺炎作为呼吸疾病的一种，在全球各地的发病率都很高。仅在中国，每年死于肺炎的患者就高达10多万人，这个数字不容小觑。对肺炎的检测，目前主要是依靠专业的放射科医生，通过对胸部X光线的研究、识别、最终作出判断，这个过程需要耗费大量的时间，而且有时也会由于人为等难以避免的原因而造成误诊[3]。近三十年来，人工智能逐渐走入寻常百姓家，人工智能的相关技术也被广泛应用在医学方面，智能医疗经常出现在人们的日常生活中。利用计算机视觉领域的相关技术处理医学影像，能够从各方面给人们带来极大的便利[26]。对肺炎胸部X光片的病灶识别，早期的方法最主要的工作是通过特征提取得到感兴趣区域，对X光片进行特征提取显得至关重要，这也是传统方法的难点所在[3]。

在深度学习方面颇有造诣的专家吴恩达教授，在前些年和其团队提出了一个用于检测肺炎的模型CheXNet，该模型是具有121层的卷积神经网络。据参与该模型研究的团队人员介绍，ChXNet模型在搭建过程中分别采用了密集连接（DenseNet）、批量归一化（Batch Normalization）两种方法进行优化，并称该模型对肺炎X光片检测准确率已经超过了专业医生。该模型的输入是胸部X光片，输出的分类结果通过sigmoid函数表示出来，sigmoid函数是一种非线性函数。最终的输出结果是患者患肺炎的概率，取值在0-1之间。

## 第三节 课题研究的主要内容

本课题研究的主要内容包括对肺炎X光片病灶的识别以及病灶数量的检测，模型的输入是肺炎X光片，输出是肺炎病灶的数量，以及使用边界框（bounding box）（一种矩形的框）将病灶的位置标注出来。

病灶识别的主要内容是目标图像检测的问题[13]，该算法首先将用于训练模型的数据集进行数据增强，采用的是最常用的数据变换方式transforms，再使用Compose方法将各个变换串在一起。然后将数据集划分为测试集和验证集，加载预训练模型ResNet34。我们都知道，从理论上推导，卷积神经网络的卷积层数与训练得到的分类结果的准确率成正比，这是因为能够提取出更加丰富且抽象的特征。但是事实并非如此，网络层数简单地叠加将会带来一个副作用，极有可能出现梯度消失或者爆炸现象[4]。因此，可以采用残差神经网络来解决这个问题，该网络模型赢得了2015年的ILSVRC比赛。

最后，为了更好地实现人机交互，本课题还使用HTML、CSS、JavaScript等相关知识制作了网页，包括登录界面和上传医学影像界面。上传图片后返回病灶数量和边界框出病灶位置的X光片。为了实现通信，该模型代码是在本地GPU是运行，虽然运行速度较慢，但最高准确率达到了84.4%。

## 第四节 本文结构

依照本课题的研究内容，可以将本文分为以下几个章节：

第一章绪论部分主要讲述了本课题的研究背景，以及研究结果能够给人民的生活带来怎样的便利；其次就是本课题在国内外的研究现状，介绍了一些值得学习和借鉴经验；最后简单的概括本文主要的文章结构以及研究内容等。

第二章主要介绍了开展本实验过程中采用的卷积神经网络的相关技术以及一些典型的技术之间的对比。

第三章主要描述了基于卷积神经网络的肺炎X光片算法模型的设计过程及其基本框架。

第四章简要介绍了本实验采用的卷积神经网络进行肺炎X光片病灶识别的实现过程。

第五章主要是对本实验的简单总结。

# 第二章 涉及技术的相关介绍

神经网络实际上就是一种仿生学产品，它在20世纪中叶就出现了，但由于当时技术不足，直到21世纪才真正受到人们的青睐，它的工作原理就像人的大脑一样，具有自主学习表征特征的能力，能够像人一样思考。简单的说，神经网络的主要功能是对目标物体进分类识别，分类识别的对象主要有图像、语言、文本等，其中语言属于密集性矩阵，文本属于稀疏性矩阵。神经网络能够解决很多计算机视觉领域的问题，例如：医学图像识别、语言文字处理、情绪分析、目标检测等问题。

## 第一节 卷积神经网络

卷积神经网络（CNNs）是深度神经网络的一种，它能够进行卷积计算，其网络结构由众多类型的层组成。其中最主要的包括输入层、卷积层、下采样层、全连接层等[5]。

在人工智能的相关领域中，神经网络被认为是一种能够模仿生物神经网络的、（尤其是人的大脑中枢神经网络），并且具有自主学习和认知功能及结构的计算模型（数学模型）[6]，一个简单的神经网络结构如下：

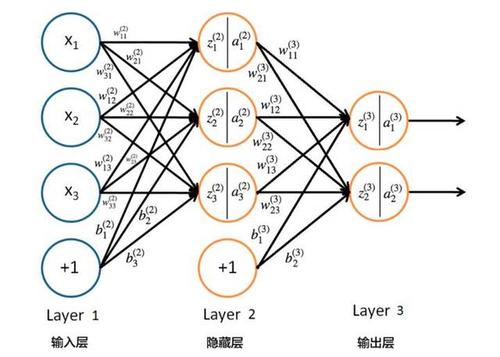


图2-1 简单的神经网络结构

由图2-1可知，此网络的网络结构极其简单，仅仅包括一个输入层、一个隐藏层以及一个输出层，这些层次相互连接构成了一个完整的神经网络，在图中我们可以观察到，各个层次均由多个神经元组成，表示输入信号，y表示输出的结果，隐藏层又可以称为中间层，它可以包含多个层次。由图2-3所示的神经网络的模型可以看出，该模型的输入层拥有三个网络节点，输出层仅有一个网络节点。

LeNet是目前最早最有名的卷积神经网络框架之一，诞生于20世纪90年代，该网络是各大神经网络框架兴起的起点[10]；另一个比较有名的框架是AlexNet，2012年Alex Krizhevsky等人发布了该网络框架，它相比于LeNet，深度和广度都提升了，并且同年该神经网络在ILSVRC（ImageNet大规模视觉识别挑战赛）上以显著优势获得胜利，该神经网络为以后CNNs的广泛应用作出了重大贡献。除此之外，GoogleNet、VGGNet、ResNet、DenseNet等也是大家耳熟能详的卷积神经网络框架。典型的卷积神经网络模型的大致框架图如图2-2所示：

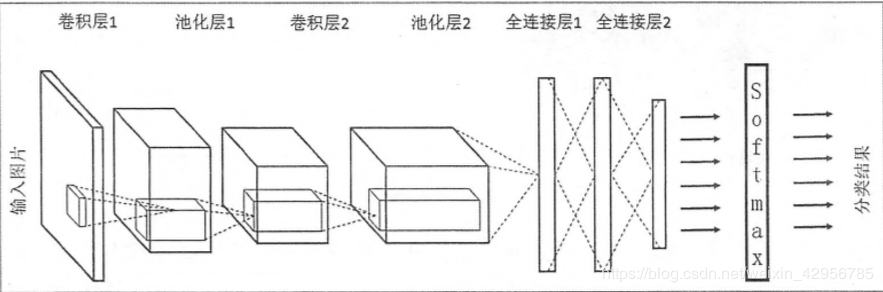


图2-2 典型的神经网络模型框架图

从图中可以看出，池化层和卷积层交互连接共同组成了CNNs的隐藏层，这也就意味着CNNs的隐藏层并不仅仅是一层网络，它可以包含很多层，层数越多，网络的深度就越深，网络结构也就越复杂，同时提取到行为特征就越抽象，得到的算法模型也就越逼近数学函数。

## 第二节 激活函数

激励函数最主要的作用是为模型提供规模化的、非线性化的能力，当前大家熟知的激活函数（Activation Function）包括ReLu函数（Rectified Linear Unit）又称线性修正单元、Sigmoid函数、阶跃函数、交叉熵损失函数、Softmax回归函数以及tanh函数等[7]，其中Sigmoid函数最大的优点是在函数区间的任意点均可导，不会出现导数不存在的现象，因此，为后面部分进行反向传播（实质上就是求导）提供便利，当然Sigmoid函数也有一定的缺点，其中最明显的是函数图像不是中心对称的，需要对数据进行中心化；tanh函数是由Sigmoid函数平移之后得到的，解决了中心对称的问题，由两个函数的图像可以看出，当数据量特别大或者特别小的时候，函数图像的变化就特别平缓，因此，当神经网络进行计算的时候，计算的速度（学习效率）会受到很大的影响，甚至会影响到整个神经网络的计算周期。据不完全统计，当前使用最多的激励函数是ReLu函数，其表达形式为：

(2‑1)

由此可以看出，ReLu函数在本质上就是取最大值。从另一方面来看，ReLu函数是一个分段函数，在每一个段上，它是线性的，当变量x小于0的时候，它的函数值都取0，也就是把所有的负数都映射为0，这也就意味着所有输入信号小于0的神经元不会被激活，仅有部分神经元在同一时间被激活，这使得网络结构变得很稀疏，能够提高整个神经网络的计算速率，专业术语称之为“单侧抑制”，也正是因为单侧抑制的存在，才使得神经网络中的每一个神经元拥有稀疏激活这种特性，这种情况在卷积神经网络（Convolutional Neural Networks，CCNs）体现的尤其明显。就理论而言，对于使用ReLu函数作为激活函数的模型，当网络模型的层数增加N层时，对应的神经元激活率会降低倍。

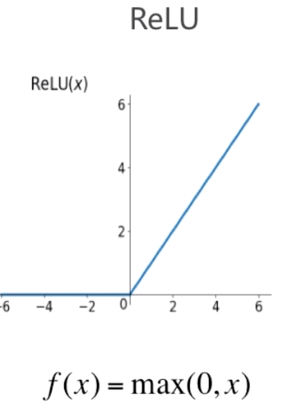


图2-3 ReLu激活函数

## 第三节 数据增强

采用数据增强（Data Argumentation）技术通常是因为数据集的规模过小，而神经网络在训练模型的时候又需要大量的数据，否则可能会导致网络不能正常工作，还容易产生过拟合现象[10]。常见的数据增强的方法有随机翻转（Flip）、裁剪（Crop）、移位（Translation）、高斯噪声（Gaussian Noise）、旋转（Rotation）、缩放（Scale）、颜色增强，除此之外，还有一些高级的数据增强技术，比如条件性生成对抗网络（Conditional GANs）、神经风格迁移等。

在图像几何变换过程中，经常会使用到插值算法，常见的几种插值算法包括近邻取样法又称最近邻插值法、双线性内插值法、三次卷积法[2]。

（a）近邻取样法

该方法是三种方法中最简单的，图像的基本单位是像素，也就意味着没有比像素更小的单位了，如果在数据增强时进行旋转或者缩放操作时，得到的新图像对应原始图像的坐标出现了小数，我们只需要简单地对它取整，使其变为整型坐标即可，一般情况下，取的是最靠近浮点坐标左上角的整数像素点。虽然该方法较为简单，算法复杂度低，系统资源的开销也比较小，但是经过该方法处理过后的图像会遭到破坏，一些重要的信息将会损失。

（b）双线性内插值法

双线性插值算法又称为二维线性插值算法，从数学的思路出发，双线性插值实质上指的就是插值函数对线性插值上的扩展[8]，只不过其中的插值函数具有两个变量，实质上就是在两个不同的方向上做了两次一维线性插值。双线性插值在处理图像时的性能要优于近邻取样法，但是由于使用的方法要比近邻取样法复杂的多，所以如果应用此方法处理图像，可能会耗费大量的时间[8]。除此之外，如果我们在处理图像上四个相邻的像素点时采用了双线性插值法，会得到与其他各个领域处相吻合的区域面，但是斜率却截然不同，与其他领域处差别很大。除此之外，双线性内插法也有一些缺点，其缺点之一：具有平滑作用，可能会使得图像的细节信息被丢弃。

（c）三次卷积法

在双线性内插值法的基础上稍作修改就可以得到三次卷积法了，因此，该算法相比于双线性插值要复杂的多[6]。在OpenCV中，有时也把三次卷积法称之为立方卷积法，它是一种三阶的内插值算法，并未出现上述两种方法的缺陷部分，并且化具有较高的精度的计算结果。当然，三次卷积法也不是完美无缺的，它也有一些缺陷，最主要的是它需要考虑的像素点很多，包括与一个浮点坐标相邻的16个像素点，这就可能导致算法的计算量很大，从而使算法整体的计算效率低。

以上三种算法在图像进行旋转、放缩、一般的线性变化或者非线性变换均可使用，但各有千秋。最邻近插值法的实现最简单，计算速度快，系统资源的开销也相对较小，但能实现的功能也是有限的，使用该算法处理后的图像质量较差。双线性内插法改善了最近邻插值法的缺陷，它在某种程度上能够保证图像的灰度值的连续性。三次卷积法对图像进行插值处理的时候，因为算法中涉及到三次卷积，因此计算量将会很大导致计算过程耗费大量的时间，但是三次卷积法在很大程度上改善了最邻近插值法和双线性内插法带来的缺陷，并且三次卷积法的算法的鲁棒性较低，同时又能得到很高的计算精度。

除了以上简单介绍的数据增强技术之外，还有移位、高斯噪声、颜色增强等数据增强的方法[9]。移位可以使神经网络观察到图像的任意位置。高斯噪声能够避免产生过拟合现象，基本原理是使高频特征失真，降低其对网络模型的影响。具体的实现方式为：使用零均值的高斯噪声使所有的频率都可以产生数据点。但是这种方法不仅能够减弱高频特征，连带着我们关心的特征同时也会受到影响，只不过这种影响可以在神经网络的学习过程中被忽略。椒盐噪声有时也被称为脉冲噪声，也可以把它看成是低版本的高斯噪声，因为它的基本原理与高斯噪声相同，只是产生的效果相对较差。颜色增强的具体实现方式有多种，包括色彩的饱和度、对比度、亮度等，在此就不再一一赘述。

常见的滤波方式有线性滤波、图像锐化滤波器（Sharpness Filter）、浮雕滤波器、均值模糊（Averaging）有时也称Box Blur、高斯模糊（Gaussian Blur）等。

其中线性滤波是最常见的滤波方式，经常用于图像处理的过程，它实现过程是最简单的，集体操作为：准备一个卷积核和等待处理的二维图像，然后再将等待处理图像中的所有像素点及其邻域点分别和卷积核的相对应元素一一作乘法，这样就完成了线性滤波过程。

图像锐化在基于深度学习的图像处理方面的应用也有很多，锐化是为了更好的描述图像的轮廓和边缘[23]，它和图像平滑的操作和作用刚好相反，主要操作是将图像的高频特征增强从而使图像中的模糊减少，灰度反差随之增强，给之后的要进行的图像识别和处理带来了极大的便利，但是图像锐化也有一定的缺点，最主要的是增加了图像的噪声，图像锐化常用的方法包括高通滤波法和微分法。

高斯模糊也需要相应的滤波矩阵来实现，高斯分布也叫正态分布，函数图像特征是越靠近中心点，函数取值越大，离中心点越远，函数值越小。由于处理的图像都是二维的，因此需要使用二维的正态分布计算出滤波矩阵（权重矩阵），计算过程中，要注意归一处理，这是因为当权重的总和大于1时，图像亮度将会变大，反之图像变暗。得到滤波矩阵之后，将每个像素点乘对应的权重值，再把所有的结果相加，就完成了高斯模糊的过程。如果图像是彩色的，则应该分别对RGB三个通道作同样的处理。

对图像实施降噪操作前，会先对图像加噪，这是因为有些图像的噪声比较小，而去掉不明显的噪声往往是很困难的。常见的噪声类型有高斯噪声、白噪声等。上述的高斯滤波法能起到图像去噪的作用，除此之外，中值滤波、P-M方程去噪、TV法去噪等是常见的一些去噪方法。

## 第四节 批量标准化BN

使用BN对数据进行正则化，目的是使一批feature map满足均值为0，方差为1的分布规律，该方法提出于2015年，由谷歌公司首次发布于《Batch Normalization：Accelerating Deep Network Training by Reducing Internal Covariate Shift》论文中，使用该方法对输入的数据进行标准化处理，能够提高网络收敛的速度并且使网络具有更优的性能，比如使其分类结果拥有更高的准确率。具体实现过程如图2-4所示：

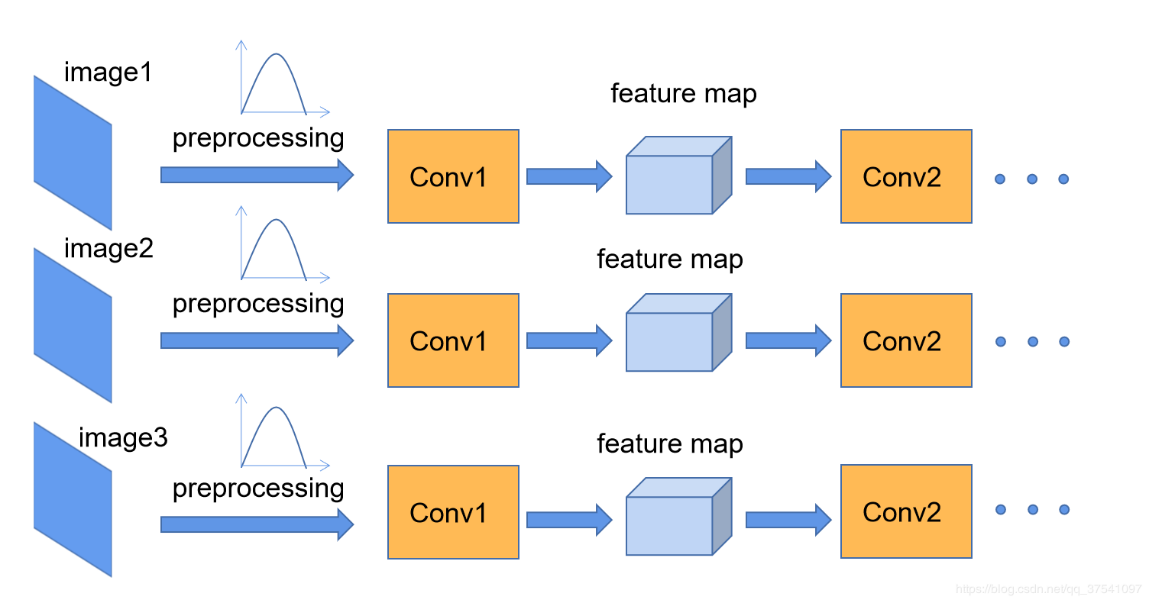


图2-4 BN实现过程

其中，Conv1的输入是一个特征矩阵，该矩阵满足某种分布规律，而Conv2的输入特征却不一定能够满足该要求，因此，当我们使用BN对数据进行标准化后，就可以使每一批（Batch）数据的Feature Map满足一定的分布规律，从而使训练样本集也满足该分布规律（期望值为0，方差为1）。当输入数据x的维度是d的时候，需要对x在d个维度分别进行标准化处理，也就是说，当我们输入的数据（图像）是彩色图像时，需要分别对其中的每一个通道（也可以称为维度，通常是RGB三个通道）进行标准化处理。BN标准化处理的公式如图2-5：

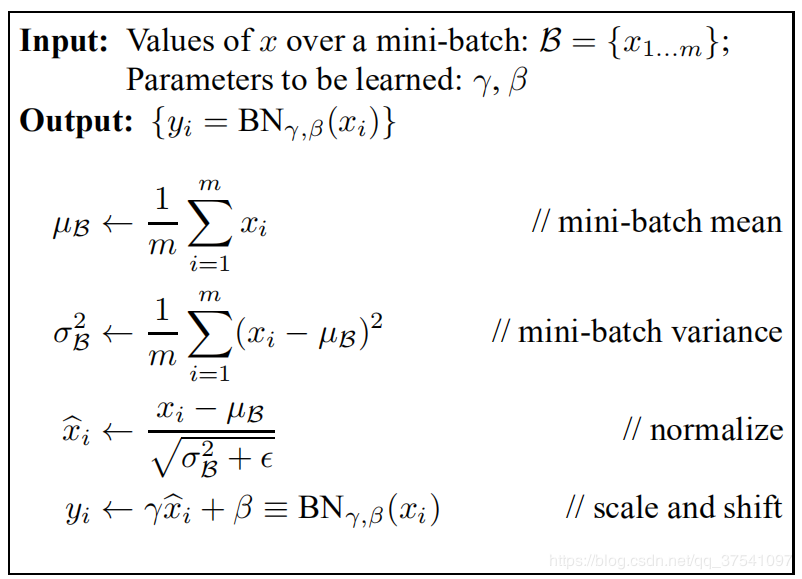


图2-5 BN标准化公式

其中表示的是均值，代表的方差，是一个非常小的数，为了防止出现分母为0的情况，γ是用来调整方差的数值，β是用来调整均值的大小。假如我们不使用γ、β来调整均值和方差的数值大小，理论上得到的应该均值为0，方差为1的数据分布规律，但很可能此时的网络模型的分类效果不一定能达到最优，因此，添加了两个学习参数γ、β在训练过程中来对网络作进一步的调整，这两个参数主要是在反向传播过程中学习得到的，均值和方差是通过对一批批的数据计算并且进行统计得到的。

使用BN进行数据的标准化处理，需要注意一些问题，首先是在训练过程中需要将training这个参数设置为True，因为训练过程中需要不断统计均值和方差，而在验证的时候training设置为false，因为验证过程使用的是历史统计的均值和方差，而不是使用当前计算得到的均值和方差的值；其次是batch\_size值的设置要适当，该值设置的越大就越接近整个数据集，但是会极大的拉低计算效率，设置的过小会导致分类结果的准确率降低；最后尽量将BN层夹在卷积层和激活层中间，网络的性能将会更好。

## 第五节 特征分类

根据机器学习中的定义，可以将数据集带标签的学习过程称为监督学习，数据集不带标签的学习过程称为非监督学习。在过去众多的传统分类方法中，一般是将整张图片用于分类，但是准确率却不高，如果我们将图像的一些特征提取出来，再基于图像的这些特征进行分类，分类的准确率将会得到很大的提高。机器学习中常见的特征分类器，主要包括两大类—回归和分类，而回归主要就是线性回归（Linear Regression），分类包含的种类就比较多了，逻辑回归（Logistics Regression），Softmax回归，SVM支持向量机等等都属于分类问题的范畴。其中的逻辑回归和Softmax回归都属于二分类问题。除此之外，还有贝叶斯分类器以及K近邻算法等等，

根据上文的叙述，我们知道图像的特征包括LBP特征、HOG特征、SIFT特征、HSV特征等，而上述特征中的HOG特征是图片的纹理特征，该特征不会提取图像的颜色信息，而颜色直方图（HSV）则和HOG特征相反，HSV将会将图像的颜色信息提取出来，而忽略了图像的纹理特征，因此，把这两种特征结合起来形成特征向量作为卷积神经网络的输入，输入到网络中进行训练，将会得到较好的分类效果。

在传统的肺炎图像识别分类的过程中，有一些比较典型的分类器，例如模板匹配、SVM支持向量机和Adaboost等特征分类器，但是目前最流行的对图像的识别结果进行分类的方式是基于卷积神经网络的。

（1）模板匹配

模板匹配（Template Matching）起源于20世纪70年代，它本质上解决的是匹配问题，主要的工作原理是通过研究一张图片中某一特定的目标位于图片中的位置，进而识别目标，图片中的特定目标就是模板，如果图片中存在特定目标，并且与模板的方向、大小尺寸都相同的，就可以通过相关算法确定目标的坐标信息。模板匹配相关的算法有相关法、误差法和二次匹配误差法等。模板匹配是模式匹配中最基本的方法之一，因其年代久远，也存在一些问题，主要表现在只能对图像平移，而不能对原始图像进行旋转、翻转等数据增强操作，否则模板匹配算法将无效，它的这种机制将会对数据集小的模型训练带来致命性的影响，实用性不强。

（2）SVM支持向量机

支持向量机（Support Vector Machine，SVM）机制是1964年被苏联的学者V.N. Vapnik等提出来的，并且在20世纪90年代末衍生出一系列的相关改进算法，得以迅速发展。SVM从数据集是否带有标签上分类属于监督学习的一种，而且由于其分类结果有两种，因此属于二分类的范畴，而肺炎X光片的病灶识别是一个多分类问题[13]，若使用该分类器对训练结果进行分类则需要在一定程度上作修改。SVM特征分类器是一种广义的线性分类器，其结构具有稀疏和稳定特性，常被用于模式识别领域的图像及文本识别。

当一个分类问题具有线性可分性时，采用最大边距超平面作为其决策边界是可取的，否则将会对分类结果造成一定的损失，该损失具体表现为部分支持向量将会位于间隔边界的内部或者落入决策错误的一端，而不是落在间隔的边界上。如果使用SVM去处理线性不可分的分类问题时，为了更直观地观察到分类结果的损失，需要采用损失函数（Loss Function）对分类结果的损失进行量化，铰链损失函数（Hinge Loss）是SVM常用的损失函数，其表达式如式2-2所示：

（2-2）

研究表明，当一种分类问题不具有线性可分性时，有可能具有非线性可分性，所谓非线性可分性即在其特征空间存在的是一个超曲面（而非超平面）可以将分类结果分为正类和负类。非线性可分问题在某些情况下可以转化为线性可分问题，主要实现方法为：通过非线性函数操作原始特征空间，将其映射到希尔伯特空间，该空间具有更高的维度，可以实现将非线性问题线性化。

（3）Adaboost特征分类器

Adaboost (Adaptive boost)算法是一种基于boost的自适应机器学习算法，通过迭代(回溯)实现。该算法的核心是从相同的数据集中选取不同的样本，训练不同的弱分类器，然后根据这些弱分类器的权重比例进行组合。首先是给每一个样本赋予权重，初始化为相同的值；其次是采用弱学习算法训练模型，计算出错误率ε；再次是利用错误率计算出弱学习算法的权重，计算公式如式2-3：

（2-3）

其中α为算法的权重，ε为错误率。最后是对样本权重数值的更新替换，一直这样迭代下去，最后得到的各个样本的弱分类器，将这些弱分类器按照不同的权重组合起来就可以得到最终的决策分类器了[5]。该算法具有检测效率高和测试精度高等特点，而且不容易出现过拟合现象，但是由于每次训练弱分类器的时候采用的是贪婪算法，因此得到的弱分类器都是局部最优的弱分类器，这也意味着最终加权得到的强分类器可能不是全局最佳的，其次最后得到的强分类器过于依赖弱分类器的选择，因此如果训练弱分类器的时间过长，将会导致整个网络的运行周期延长。

# 第三章 基于卷积神经网络的肺炎X光片病灶识别模型设计

人工神经网络有时也简称其为神经网络，它从本质上来讲就是一种数学算法模型，具有模仿生物神经系统的行为特征的功能和结构，它主要是通过调整各个节点相互连接的情况来实现信息数据处理。研究本课题搭建的卷积神经网络模型引用了典型的残差神经网络ResNet34，在ResNet34这个残差神经网络中，具有34层的网络结构[10]。

## 第一节 网络结构设计

神经网络研究的内容范围很广，与其相关的研究领域也是数不胜数，这从侧面体现了人工智能领域具有多学科技术交叉的特点，当然，这也是大多数学科共有的特点。在神经网络上应用的算法一般都是使用矩阵来做向量乘法来完成对应的计算。神经网络的算法中包含了大量的符号函数和这些函数的无限接近，和传统算法的计算方法相比，神经网络的计算方法具有很多优势，例如能够实现并行计算的功能[5]（大幅度的提高了运算速率）、容错性能强也是神经网络最大的优势之一、具有自主学习行为特征的能力等。神经元是神经网络的基本组成单元。单独的一个神经元也是一个神经网络，是最简单的神经网络，其本质是数学函数，在数学上，仅仅是一个数学表达式，当然，仅仅包含一个神经元的神经网络，其功能也是最简单的，目前的大部分的神经网络都拥有几十层甚至上百层的网络结构在一定范围内，我们称之为“阈值”，随着深度的增加，计算的速度就会越快，计算能力也逐渐增强，网络层数增加的过程也就是把要解决的问题逐渐细化的过程，随着网络层次的加深，提取到的样本特征就越抽象，也就意味着该模型越接近于数学函数。

目前，M-P模型仍然是最受欢迎的神经网络模型[6]，此外它还是经典的神经元模型，如图3-1为M-P模型的简单结构：

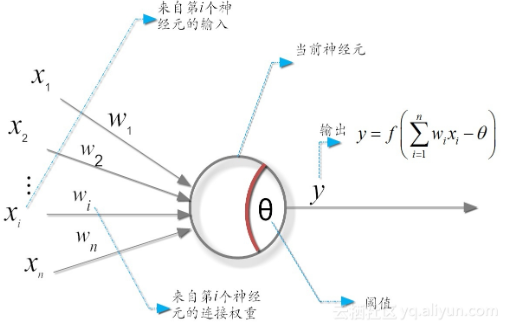


图3-1 M-P模型的结构

在M-P模型中，当前的神经元会接收来自其他的n个神经元传输过来的信息作为输入信息，这些输入信息的表达方式通常是权重的大小，神经元之间的连接通过权重值来实现。具体过程为：首先，当前的神经元接收到输入信号；第二步获取权重值；第三步用获取到的权重值以某种方式与输入信号整合起来；第四步将得到的整合后的结果与当前的神经元 “阈值”大小进行比较；最后通过相应的“激活函数”对外界输出最终结果，这一般在专业术语上称之为“感知机”。简单地来说，神经元模型就是一种能够通过模仿生物神经系统的行为特征来调整神经网络内部各个神经元之间的连接关系，以达到分布式并行处理数据的目的的数学模型，同时该模型还拥有自适应和自主学习的能力。

本课题引用的ResNet34的网络结构首先是一个卷积层，紧接着是一个池化层，随后是由一系列的残差结构堆叠而成以及进行平均池化下采样操作，最后再利用全连接层也就是该网络最后的输出层将最终的分类结果输出。

网络结构中堆叠的一系列的残差模块具有不同的形式，如图3-2所示：

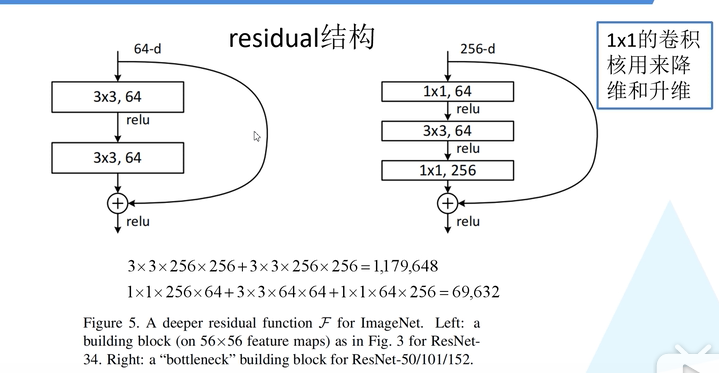


图3-2 两种不同的残差模块

其中，左边的残差模块主要是针对层数较少的残差神经网络，比如ResNet18、ResNet34等，右边的残差模块主要是应用在较深的残差神经网络中，主要包括ResNet50、ResNet101、ResNet152等，从图中表达式还可以得出使用右边的残差模块和使用左边的残差模块相比，网络模型的参数数目呈数量级式减少，使模型整体的运行效率得到明显的提升。因此，不同层数的残差神经网络选择不同的残差模块具有很重要的意义。ResNet34的网络结构如图3-3：

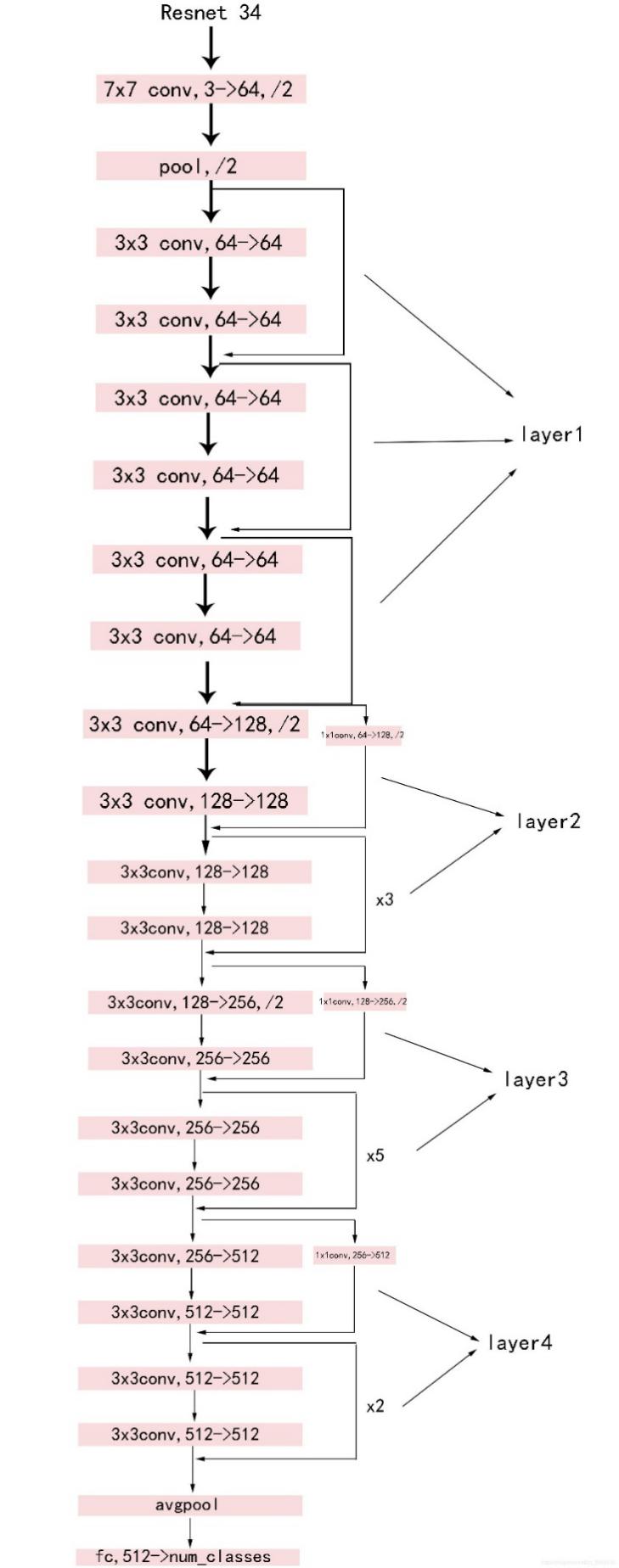


图3-3 ResNet34的网络结构图

## 第二节 卷积神经网络层次设计

与传统的神经网络相比，CNNs具备很多优点，它通过权值共享、局部感受野以及降采样（按比例缩小特征图的长和宽的过程）三种策略来降低算法模型的复杂程度，同时还可以采用pooling技术保持特征图的某种扭曲不变形，包括平移、翻转、伸缩或其它形式的高度不变性[5]。Pooling的过程实质上就是将各个小邻域的特征进行特征融合得到新的特征，Pooling操作可以使网络模型的参数大幅度减少，提高了计算效率。

（1）卷积层

在CNNs的网络结构中，最重要部分的是卷积层，在卷积层中，每一个神经元的输入只是上一层网络的其中一部分（也称作卷积核），这是与传统的全连接层最大的不同之处。通常情况下，卷积核的大小大约为3\*3或者5\*5，有时也会出现1\*1大小的卷积核。有些情况下也会将卷积层称之为内核（kernel）或者过滤器（filter）[11]。

卷积层主要功能是对输入图像进行特征提取[2]，也正是因为网络结构中存在卷积层，才使得CNNs不同于其它的深层次神经网络，卷积的主要目的是为了得到医学图像的边缘特征。此外，大量的数据计算也都在这一部分完成，进行卷积运算时，卷积层将会使用特定大小的卷积核（convolution kernel）对给定的输入图像卷积。在图像处理中，卷积是最常见的一种滤波方法，它是一种线性运算，能够实现图像的降噪以及图像的锐化等滤波功能。

卷积核常出现在图像处理相关的领域，它只关心局部特征，具有局部性，卷积核的大小决定了它局部的大小，CNNs中的卷积核与一般神经网络中的卷积核本质上没有什么不同，比如在图像处理时，都是将卷积核分别和输入图像不同的输入位置一依次做卷积运算。为了更好地理解卷积层，我们假设现在有一个2\*2\*3的普通的节点矩阵，希望能够通过卷积层的卷积操作将其变为1\*1\*5的单位矩阵，此时的卷积层的参数个数就是2\*2\*3\*5+5=65， 5代表的是偏置项的个数。公式表示为如式3-1：

（3-1）

其中，表示的是第i个偏置项的参数，表示第i个权重。使用式3-1进行卷积运算后会得到特征图，特征图的数量不止一个，卷积运算得到的数据越大，说明该部分越符合卷积核的特征[14]，由此可以通过多次训练以调整卷积核的参数，最终达到对图像进行最优化特征提取的目的。

卷积运算就是卷积层向前传播的过程，在该过程中过滤器实现了从CNNs的当前层的左下角滑动到右下角，移动的方向不是固定的，可以横向滑动或者纵向滑动，滑动结束后就可以得到一个新的二维特征图。但是由于滑动的步长（stride）大小设置的不同，因此得到的二维特征图也存在不同程度的差异，图3-4显示了同一个输入图像的情况下，不同步长得到的不同二维特征图。当输入图像的大小为5\*5像素块并将滑动步长的大小设置为1时，根据公式计算得输出图像的大小为3\*3像素，而当滑动步长变化为2时，同样的输入图像，输出图像的大小为2\*2像素，这反映了步长的大小会影响卷积运算的结果。输出图像的大小可以通过以下计算公式得出：

（3-2）

其中为输出图像的大小，为输入图像的大小，stride是滑动步长，是卷积核的大小。由公式3-2可以看出，通过该公式计算出的结果不是整数，这是因为步长不能整除输入图像的像素宽度。因此，我们可以将卷积运算分为越过边界取样和不能越过边界取样两种。当可以越过边界取样时，得到的取样面积将和输入图像的面积大小相等，而不越过边界取样得到的取样面积必定会小于输入图像的面积[2]。

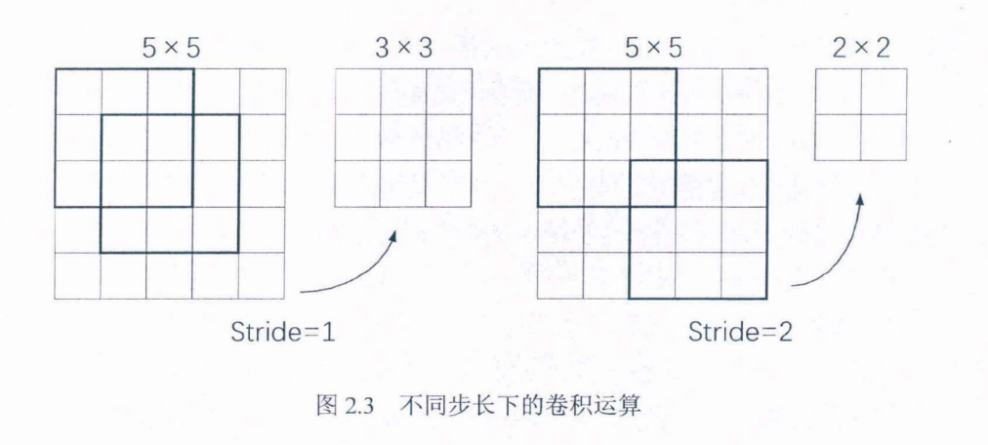


图3-4 不同步长下的二维特征图

通常卷积神经网络中还包括激活层，激活层一般都是紧跟在卷积层的下一层。当神经网络做完卷积运算后，通过引入激活函数（激励函数）来增加网络的表达能力，因为激活函数是非线性的，比线性函数拥有更强的表达能力[15]。在一般的线性网络模型中，网络节点的输出结果是输入数据的加权和，我们称这个过程为线性变换，并且各种线性变换的线性组合仍然是线性的。然而，现实中的大部分问题用线性模型是没有办法解决的，现实问题是高度抽象的，只能通过非线性模型来处理。因此，神经网络中的每一个神经元节点得出输出结果后，还要经过一个非线性的激励函数处理后才能执行输出操作。

（2）下采样层

下采样层最常用的名称为池化层（Pooling Layer），有时也称其为子采样层（Subsample Layer），池化层主要是用来将提取到的行为特征进行优化和信息过滤，本质就是从图像的大量特征中选取有用的特征。一般情况下，卷积层的后面就是池化层，但不排除是激活层，池化操作常用的方法包括最大值池化（Max-pooling）和平均值池化（Mean-pooling）两种。其中Max-pooling是通过获取领域内特征点的最大值来实现，同理Mean-pooling就是通过取领域内特征点的平均值实现具体操作。下采样层还可以通过采用降低特征图维度的方法，实现减少网络参数量，以达到降低整个网络计算量的目的。此外，还通过优化特征和降低输入图像的分辨率使得网络模型的抗噪能力增强[8]。

下采样层的具体实现方法与卷积层类似，图3-5为最大池化的实现过程图，只是下采样层的卷积核仅使用对应像素值的最大值和平均值，这与最大池化和平均池化操作相对应，而且下采样层也没有反向传播的过程。

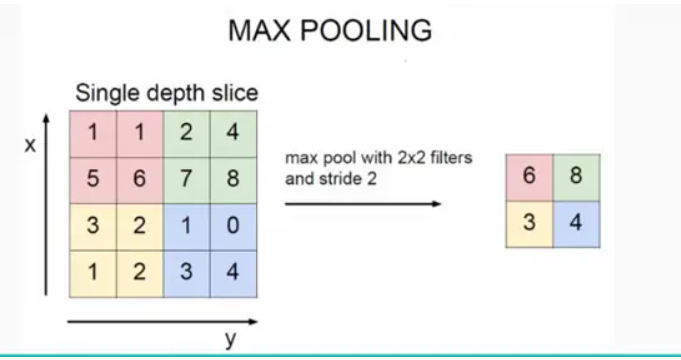


图3-5 最大池化的实现过程

采用下采样层的好处有很多，其中最主要的就是它可以减少参数数目，同时还稀疏化了网络结构，既降低了网络模型的复杂程度从而提高了计算效率，还能够防止过拟合现象的产生[16]；另一方面，加入池化层可以保证图像行为特征的扭曲和位置不变形，这种特性对图像处理来说有利的。

（3）全连接层

全连接层（fully connected layer，FC），当前层的每一个神经元与前一层的每个神经元之间均存在连接关系时，此当前层就被称为全连接层。全连接层在整个卷积神经网络中充当“分类器”的角色[8]，起到的作用是将已经学习到的“分布式表征（Distributed Representation）”映射到样本空间，实现将分类结果显式的表达出来。在模型训练过程中可以通过卷积操作来实现全连接层的功能，分布式表征是研究人工神经网络的一个重要的概念，并且具有很多优点，最主要的是信息的表达不会因为某些层的部分神经元出现故障时而导致毁灭性的损坏，其次，分布式表征也能起到防止过拟合现象发生的作用，辛顿等人在其2012年发表的一篇论文中提到了Dropout的概念，Dropout可以简单地概括为当神经网络向前传播的时候，人为地将网络中的某个神经元以特定概率p处于停止激活的状态，减少过拟合现象的发生，这是因为阻碍了部分特征检测器进行相互作用，使其对某些局部特征的依赖性降低[18]，使模型拥有更强的泛化能力。Dropout方法是在一个前馈神将网络的网络结构特别复杂而用于训练模型的数据集却特别小的基础上提出来的，该方法的提出奠定了CNNs在图像分类算法中的核心地位。

正常情况下，全连接层会出现在网络模型的最后几层,当卷积层和池化层依次完成对图像特征提取和优化后，再使用全连接层得到分类结果，最后再使用Softmax逻辑回归函数将分类结果以概率的形式输出[14]。

（4）输出层

在大多是的图像处理问题上，输出层均采用Softmax函数作为激活函数，因为Softmax函数适用于处理多分类问题[15]，Softmax函数表达式如式3-3：

（3-3）

其中表示输入的数据，表示第k层神经元的输出结果，n代表当前输出层总的神经元个数。由于采用Softmax函数输出的结果是一个概率值，故输出的取值范围在0-1.0之间，并且所有的输出值加起来和为1。但是，Softmax函数在计算机上运算时可能会产生“溢出”现象，由Softmax函数的公式我们知道，该函数需要进行指数计算，而指数计算又呈“爆炸式”增长的趋势，如果出现“溢出”情况，计算的结果将是不确定的。为了防止这种情况的发生，我们可以在Softmax函数进行指数运算时，加上某个常数C，常数C相对于指数的变化情况是可以忽略不计的，当将C设置为输入信号中的最大值时，“溢出”问题就被解决了。

## 第三节 预处理模型ResNet34

残差神经网络（Residual Network，ResNet）也是一种卷积神经网络，该网络模型由微软公司实验室的四位华人团队提出的，并且在2015年的ImageNet大规模视觉识别大赛ILSVRC上获得了巨大的胜利，拿到了两个图像识别相关领域的冠军，同时还获得MS COCO数据集中的两项冠军。

从理论上来讲，在一定的范围内，我们可以通过增加网络的深度和宽度来适

当地提高网络的性能，例如VGGNet，该网络是在AlexNet的基础上加深了网络的层数，从而使网络的性能大幅度地提高，但是对于传统的卷积神经网络，一味地加深网络会很可能会导致退化现象，深层次的网络识别的准确率将会低于浅层的网络，即当网络增加到一定的层数时，准确率达到饱和状态，此时再继续加深网络，准确率不但不会提高反而会下降[25]。这是因为产生了梯度消失或爆炸现象，可以通过归一初始化（Normalized Initialization）以及中间归一化（Intermediate Normalization）方法解决该现象，此方法可以使不到一百层的神经网络在反向传播时可以快速收敛。

残差神经网络不同于传统的深度神经网络，它可以通过只加深网络的层数来提高测试结果的准确率，这是因为它具有独特的网络结构—残差块，网络的层与层之间通过残差块跳跃来接[17]，可以防止产生梯度消失（爆炸）现象。因此，它相比于其他网络更容易优化。残差块的结构如图3-6所示：

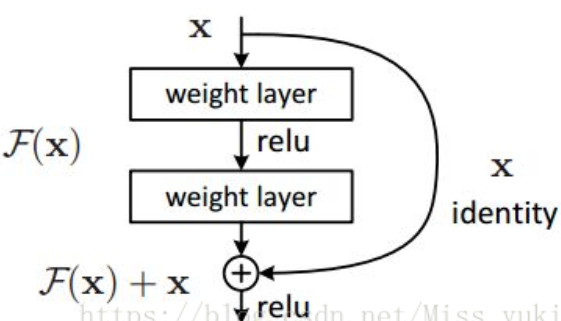


图3-6 残差块的结构图

由图3-6所示，可以发现残差块的输出包括了两个部分，一个是两个卷积层级联，另一部分仅包括输入元素，两部分相加后再通过激活函数ReLu得到最终的输出[22]，需要注意的是主分支输出特征的shape和shortcut输出特征的shape必须相同，才能实现主分支的特征矩阵和shortcut特征矩阵在相同维度上相加操作。

通过上面的介绍可以总结出ResNet的一些特点，其中最大的优点就是它能够搭建很深的神经网络，最大层数突破了1000层，而反过来看其他的网络，比如VGGNet、AlexNet等一般层数是在十几到二十几层，原因之一是层数太深会出现梯度爆炸或梯度消失现象，该现象的明显程度和层数的大小成正比[20]，单纯地加深网络使得网络模型的性能大幅度降低，得不偿失；另外一个原因是当梯度消失或者梯度爆炸现象被残差神经网络的残差块解决之后，可能还会出现退化问题（Degradation Problem），即层数深的网络仍然没有层数浅的网络的效果好。另外，残差神经网络具备其他网络模型所没有的残差模块，使优化网络的过程更加简单。

从图3-3可以看出，ResNet网络结构首先是一个7\*7的卷积层，紧接着是一个3\*3的最大池化下采样操作，然后又把一系列的残差结构分成了四层，layer1、layer2、layer3、layer4，由图可知layer1是由三个残差结构组成，最后两层是平均池化下采样和全连接层，然后再将Softmax函数概率化后的分类结果输出。

## 第四节 网页前后端设计

为了更好地与用户实现交互，本实验还使用网页编写了界面，编写语言为HTML、CSS、JavaScript，主要包括登录和上传图片功能，图片上传后能够保存到本地文件夹中，便于预测模型获取图片。

1. 登录界面

登录界面的功能设计主要是靠一个form表单实现，包括用户名、密码文本框和一个提交按钮，可以通过点击提交按钮“type=submit”将前后端连接起来，用户名和密码已经提前保存到后端，若匹配后端则将上传图片界面返回给用户。

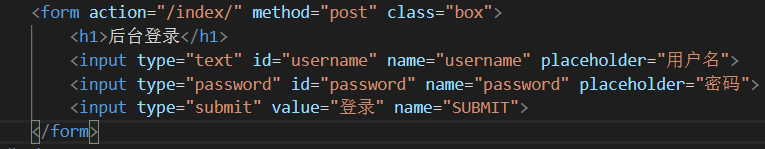


图3-7 登录功能代码实现

1. 上传图片界面

上传图片界面的主体是一个div，div内包括两个form表单，表单中的提交图片按钮的类型为“submit”，点击提交后后端将其保存在“uploads”文件夹中。整体背景样式如下图所示：

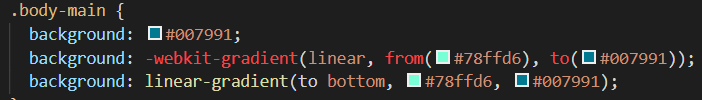


图3-8 上传图片界面背景样式

1. 后端设计

用户上传图片后，服务器端接收图片，并将图片保存在本地静态文件夹中。

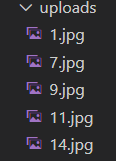
 

图3-9 后端服务器保存图片至本地文件夹

# 第四章 基于卷积神经网络的肺炎X光片病灶识别算法实现

## 第一节 数据集

本课题实验采用的数据集是从AI研习社开展的肺炎识别比赛中下载的，共计20013张训练集，按照4:1的比例将数据集划分成了训练集和验证集，6672张测试集。如图5-1所示为数据集中不同病灶数量的数据分布情况：

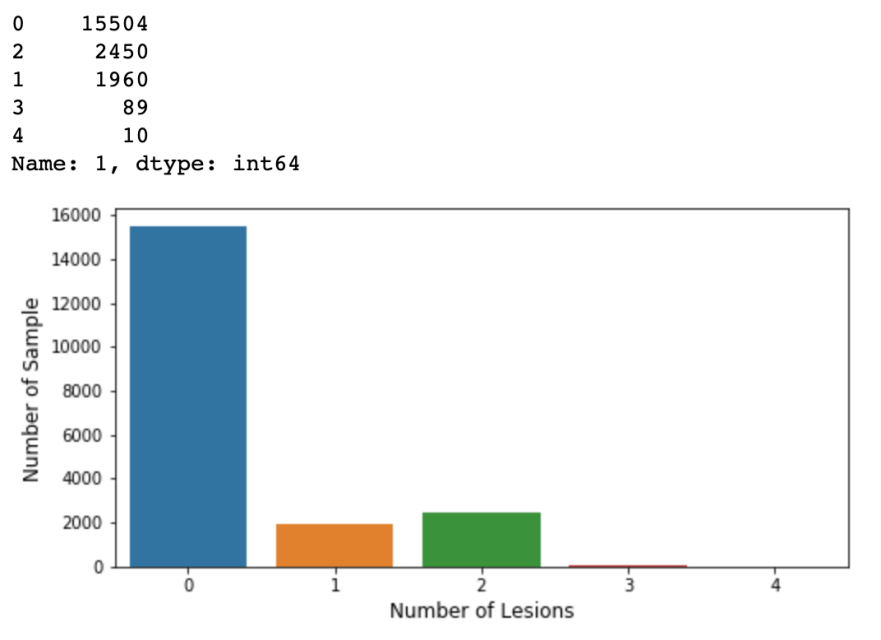


图4-1 数据分布情况

从图中可以看出，数据集的类别分布极其不均匀，这就要求我们在对数据集进行采样的时候随机选择，尽可能使样本数据体现出随机性，保证训练的模型具有很强的泛化能力。划分数据集时也要选择合适的比例，便于提高训练准确率和降低损失。

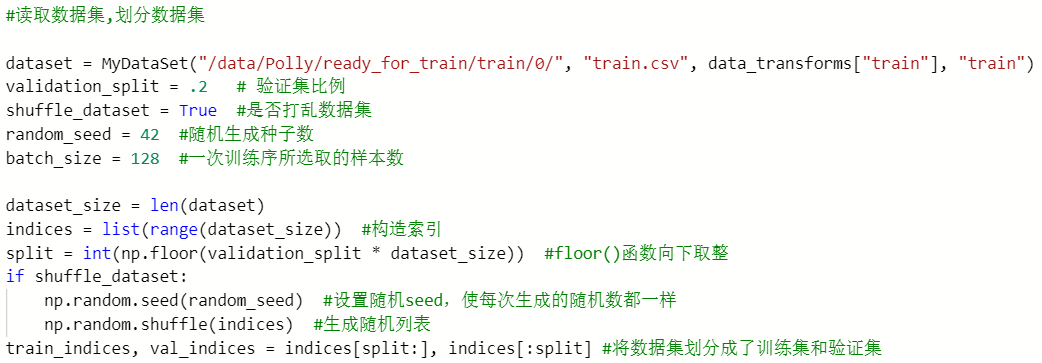


图4-2 数据集的划分

从数据集划分的代码中可以看出，划分过程中打乱了数据集，并且使用了随机生成种子，保证数据分布的随机性。

## 第二节 医学图像预处理

在肺炎医学图像判别研究过程中通常包括对图像进行预处理、对医学图像进行特征提取、将提取的特征进行优化（本质上是从提取出的众多特征中选出自己所需要的特征）、使用分类器分类以及将分类结果输出等步骤[12]。

图像预处理的方法有很多，比如：数据增强、滤波、图像锐化等，图像的预处理操作可以使特征提取过程更加方便。特征提取（Feature Extraction）一般是根据目标的分类结果，提取出我们需要的行为特征。优化特征有时也称作特征选择，从提取出的特征中选出使既定指标最优的特征。最后是使用特征分类器进行分类，特征分类器的种类有很多，不同的分类器适用于不同的神经网络，分类效果也不尽相同[24]。图像预处理的方法包括：

（1）翻转

图像翻转包括水平翻转和垂直翻转，垂直翻转可以通过先将图像旋转180度，再水平翻转得到相同的效果。我们将图像进行随机翻转的目的是为了在训练网络模型的过程中，能够保证用于训练的数据集能够不受因图像对齐而带来的问题的影响[6]。图4-1显示了肺炎X光片水平翻转后的情况：

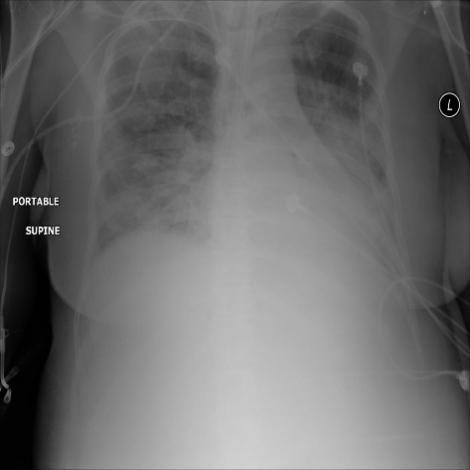
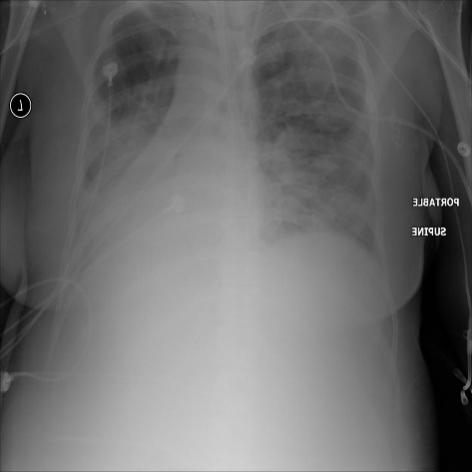
 

图4-3 水平翻转

（2）旋转

随机旋转技术能够使神经网络模型具有更强的识别能力，因为提高模型的泛化能力和算法的鲁棒性会得到不同程度的提高[25]。从理论上讲，随机旋转的角度可以是任意的，但是为了减少网络的计算量，通常我们会选择45°、90°、180°等容易计算的度数。当我们对图像进行随机旋转的时候，同时还要对其标签的值作出相应的变化，其坐标的旋转变换公式如式4-1所示：

· （4-1）

公式4-1的含义是：将的坐标沿着顺时针旋转度，然后和变换矩阵作乘法，得到一个新的坐标。如图4-2所示，将肺炎的医学图像以不同的方式旋转后得到图像：

图4-4 旋转

（3）缩放

缩放是图像几何变换方法中对空间坐标变换的一种，该操作不会改变图像的像素值大小，只是重新排列了图像的像素[17]。缩放包括向内缩放和向外缩放两种，向外缩放时，一般应用于图像处理的框架都会将得到的新图像进行剪切，使其大小等于原始图像的尺寸。向内缩放图像时，得到的新图像的尺寸小于原始图像，此时我们只能对未被新图像覆盖的边界部分作出假设。如式4-2所示为缩放变换的公式：

（4-2）

其中、、依次为、三个方向进行缩放时采用的缩放因子。缩放后可能会导致新图像的部分像素点的坐标非整数而导致无法准且定位，因此，我们需要近似处理缩放后的图像，一般使用插值法。

（4）裁剪

裁剪的过程是随机的，每次裁剪的位置几乎都不相同，这样可以保证数据的多样性[9]。本实验中的随机裁剪的尺寸为24\*24像素，将裁减后的图像作为卷积神经网络的输入[2]。数据预处理的部分代码如下：

data\_transforms = {

"train":

transforms.Compose([ #transforms是常用的数据变换方式

transforms.RandomResizedCrop(size=256, scale=(0.8, 1.0)),#随机长宽比裁剪

transforms.CenterCrop(size=224), #从中心裁剪图像，大小为224\*224

transforms.RandomHorizontalFlip(),# 随机水平翻转

transforms.ColorJitter(brightness=0.2, contrast=0.2),#随机改变图片的亮度、对比度

transforms.RandomRotation(15),#按角度旋转对象

transforms.ToTensor(),#将PIL图片或者numpy.ndarray转成Tensor类型的，并且归一化至[0-1]

transforms.Normalize([0.485, 0.456, 0.406], [0.229, 0.224, 0.225])#用均值和标准差对张量图像进行归一化

]),

"val":

transforms.Compose([

transforms.RandomResizedCrop(size=256, scale=(0.8, 1.0)),

transforms.CenterCrop(size=224),

transforms.ToTensor(),

transforms.Normalize([0.485, 0.456, 0.406], [0.229, 0.224, 0.225])

]),

"test":

transforms.Compose([

transforms.RandomResizedCrop(size=256, scale=(0.8, 1.0)),

transforms.CenterCrop(size=224),

transforms.ToTensor(),

transforms.Normalize([0.485, 0.456, 0.406], [0.229, 0.224, 0.225])

]),

}

## 第三节 LBP算法特征提取

常见的基于深度学习的特征提取的方法有LBP算法，也可以称为局部二值模式、SIFT算子，很多时候都称其为尺度不变特征变换、HOG特征提取算法又名方向梯度直方图等[11]。

LBP是一种视觉算子，在计算机视觉领域中的图像处理方向应用的最为广泛，在纹理分析和人脸识别方向的应用更加显著[23]。最早的版本起源于1994年，被定义在尺寸大小为3\*3的邻域上，用来比较图像的局部特征。对LBP算子改进后，其邻域的尺寸大小可以随机变化，基于改进后的LBP算子得到的图像特征将会更具有代表性。

LBP算法中特征向量的简单构建过程如下：

1. 首先将待检测区域划分大小相同的像素块，例如4\*4、16\*16等；
2. 分别处理每个划分好的像素块，将像素块中的每一个像素分别与相邻的8个像素（以16\*16的像素块为例）进行比较；
3. 假如当前像素的值比相邻像素值大，记为0，否则记为1，也可以执行相反的操作，比较结束后得到8位数字；
4. 在每一个像素块中计算每一个数字出现的频率，并制作频率直方图，必要的时候，可以将直方图规范化；
5. 将每一个像素块规范化后的频率直方图相互连接起来，就构成了整个图像的特征向量。

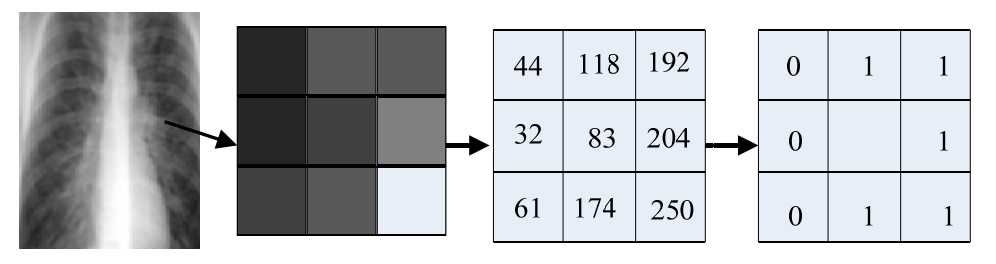


图4-5 LBP算法中特征向量的构建过程

LBP算子对灰度值的变化不敏感，并且计算速度快，此外，通过LBP算子提取到的图像特征的维度也比较低，亮度对图像表达的影响不大，这些优点是促使LBP特征被广泛使用的原因之一[23]。LBP算子的公式表达如式4-3：

（4-3）

其中s是sgn函数的缩写，表示的是符号函数。由LBP算子的公式描述可以得出该算子能够处理很小的纹理细节。

## 第四节 模型参数初始化及反向传播

从本质上讲，卷积神经网络模型的训练过程是一个迭代过程[2]，迭代过程最重要两部分是初始条件和结束条件，因此，对模型参数的初始化显得尤为重要，并且模型参数的初始化和模型最终的训练准确率息息相关[19]，参数初始化在不同程度上影响着大部分的卷积神经网络的模型算法，甚至还会影响到整个网络的收率速度。本实验的模型初始化代码如下：

#模型初始化，修改输出神经网络，

def initialize\_model(model\_name, num\_classes, feature\_extract, use\_pretrained=True):

model\_ft = models.resnet34(pretrained=False) # 使用预训练模型pretrained默认是false ##ft（feature transform）

model\_ft.load\_state\_dict(torch.load('/model/Polly/acm/resnet34-333f7ec4.pth')) #加载预训练模型（加载本地模型）

set\_parameter\_requires\_grad(model\_ft, feature\_extract)

num\_ftrs = model\_ft.fc.in\_features #提取fc层中固定的参数

model\_ft.fc = nn.Sequential(nn.Linear(num\_ftrs, num\_classes)) #修改类别，分类num\_classes=5，肺炎有5类 fc（fully connected layer全连接层）

#输入图片大小

input\_size = 224

return model\_ft, input\_size

从上面的代码中可以看出，本实验引用的预训练模型ResNet34是从本地加载的，ResNet34模型的前面的网络结构保持不变，仅仅对后面的全连接层进行更新，能够在保证运算速度的同时，加快网络的收敛速度以及达到更高的准确率

本实验引用的卷积神经网络模型是ResNet34，该模型每一个激活层后面采用的激活函数均为ReLu函数，训练模型引用了Pytorch官方训练文档。反向传播算法是一种有效的计算偏导数的方法，该算法是在梯度下降算法的基础上得出的，它主要是依靠复合的非线性函数来处理数据，因此在很大程度上能够复现函数。每一轮在反向传播时都可以获得各层各单元的误差信号，作为修正各单元值权值的根据，权值不断的修正过程就是训练模型、学习参数的过程。部分反向传播代码如下：

# forward

# track history if only in train

with torch.set\_grad\_enabled(phase == 'train'):

outputs = model(inputs)

loss = criterion(outputs, labels)

\_, preds = torch.max(outputs, 1)

# backward + optimize only if in training phase

if phase == 'train':

loss.backward()

optimizer.step()

## 第五节 肺炎病灶识别

1. 模型训练过程

训练模型使用的代码引用了Pytorch官方训练文档，训练模型时，将划分好的训练集和测试集输入到网络模型中进行训练，引用ResNet34作为预训练模型，冻结最初的几层网络，保证其在训练过程中权重不会更新，然后对模型参数初始化，修改其输出神经网络，将分类类别分为5类，从0开始，包括0、1、2、3、4共五类病灶数量类型。



图4-4 训练模型代码

1. 肺炎病灶识别过程

模型训练完毕后保存到模型文件中，在运行预测代码进行预测时，首先对用于预测的数据集进行数据增强，然后直接引用保存好的模型文件。

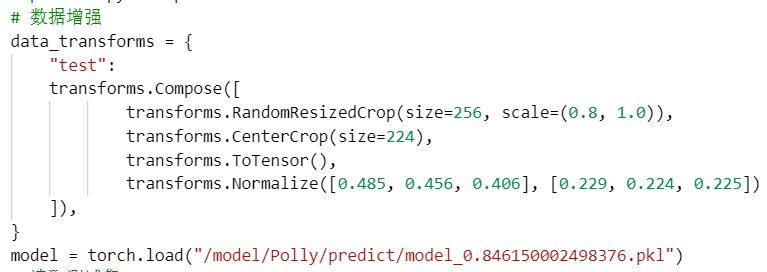


图4-5 引用训练好的模型文件

预测过程首先读取测试集，再使用一个MyDataSet类完成测试集的读取和预测结果的保存，该类继承了DateSet类程序，运行完毕后，会将运行的预测结果保存在predict.csv文件中。代码如下所示：

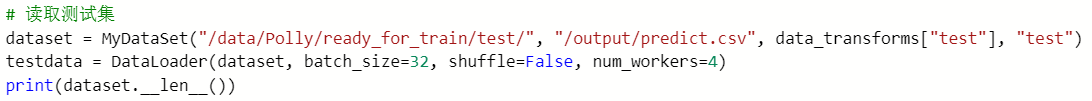


图4-6 测试集读取及保存预测结果

1. 网页交互过程

因为Bitahub服务器平台未提供公网ip，无法实现云服务器与网页之间的网络通信，因此，选择将模型代码在本地服务器运行，但运行时间长达10几个小时，预期实现将图片上传后保存到本地文件夹中，运行用于预测图像病灶数量的Python文件，从本地获取的用户上传的文件作为预测模型的输入数据，预测结束后会将预测结果保存到文件夹中，用户可直接查看。

程序代码分别在在Bitahub云服务器平台和本地服务器进行训练，云服务器平台训练20轮，由于本地服务器配置较低，仅训练4轮。最终在云服务器平台上最优的准确率达84.4151%，是一个较好的结果。最优准确率也达到达84.1%。

# 第五章 总结

经济飞速发展，国家日益富强，医疗水平也在不断地提升，但肺炎的发病率一直高居不下，给人们的生理和心理带来了极大的伤害。虽然国家医疗水平稳速提高，但也无法避免有一些偏远的地区存在专业医师不足的情况，因此，采用人工智能技术对疾病进行辅助检测能够给人们带来极大的好处。

本论文采用的模型为ResNet34，算法采用Python3.6.8语言，基于Pytorch框架编写。本实验的工作和功能概括如下：

1. 前后端设计：前端使用HTML、JavaScript等语言编写，后端使用Python语言编写，可以实现前端将肺炎X光片上传，后端接收图片并保存至本地文件夹，且在网页上实现回显的效果。
2. 搭建卷积神经网络：引用ResNet34模型进行预训练，冻结预训练模型的前面50层的网络，使其参数在模型训练过程中不发生改变，仅修改ResNet34模型的最后几层网络，保证模型具有良好的性能，运算速率和识别准确率更高。
3. 训练模型：分别在Bitahub平台和本地服务器进行模型训练，在Bitahub服务器平台时，将训练好的模型保存在模型文件中。
4. 预测肺炎X光片的病灶数量：将6000多张测试集作为预测算法的输入，最后将预测结果保存在文件中便于查看。

总的来说，本论文的主要功能均已经实现，但是还存在一些缺陷与不足，也有预期的功能未能实现，比如识别出肺炎X光片病灶后能够将病灶用边界框框出来，并实时地显示到网页，给用户更好的使用体验，但是本论文模型对肺炎病灶识别的准确率较高，希望在后期能够进一步完善算法，使其能够实现预定目标，并达到更高的准确。

参考文献

[1] 庞浩. 基于深度卷积神经网络的医学影像诊断关键技术研究[D].北京邮电大学,2019.

[2] 高震宇. 基于深度卷积神经网络的图像分类方法研究及应用[D].中国科学技术大学,2018.

[3] 郝兵. 基于深度学习的肺炎检测方法研究[D].北京工业大学,2019.

[4] 袁茂洲. X线肺部图像的特征提取与分类方法研究[D].大连理工大学,2019.

[5] 相文波. 基于卷积神经网络的肺炎类型影像判别[D].哈尔滨理工大学,2017.

[6] 金海滨. 基于改进的SSD肺炎检测识别研究[D].辽宁科技大学,2019.

[7] 张明书.基于深度学习及胸腔X光进行肺炎病变诊断[J].科学咨询(科技·管理),2019(11):60-61.

[8] 周进凡,刘宇红,张荣芬, 等.基于卷积神经网络的肺炎检测系统[J].现代电子技术,2019,42(23):35-39.

[9] 逄淑超. 深度学习在计算机视觉领域的若干关键技术研究[D].吉林大学,2017.

[10] 刘义. 基于深度学习的特征融合及其在医学图像中的应用[D].武汉科技大学,2019

[11] 王文智,王林,张儒良, 等.基于卷积神经网络的图像分类方法研究[C].//贵州省系统工程学会.贵州省系统工程学会第六届学术年会论文集.2016:44-53.

[12] 潘丽艳,梁会营.基于深度学习的儿童肺炎病原学类型判别模型[J].中国数字医学,2019,14(3):59-61.

[13] 潘亚玲,王昊,王晗琦, 等.基于深度学习的肺炎病灶分割技术用于新型冠状病毒肺炎的定量分析[J/OL].中国医学计算机成像杂志,2020,001:1-10.

[14] 邓棋,雷印杰,田锋.用于肺炎图像分类的优化卷积神经网络方法[J].计算机应用,2020,40(01):71-76.

[15] 张物华,李锵,关欣.基于多尺度卷积神经网络的X光图像中肺炎病灶检测[J].激光与光电子学进展,2020,57(08):187-194.

[16] 王庆凤. 基于多语义任务与多标签增量学习的胸部影像辅助诊断方法研究[D].中国科学技术大学,2019.

[17] 高冬冬. 基于卷积神经网络的肺炎识别与研究[D].桂林理工大学,2019.

[18] 冉静. 基于深度学习的医学影像分析与标注[D].北京邮电大学,2019.

[19] 何新宇,张晓龙.基于深度神经网络的肺炎图像识别模型[J].计算机应用,2019,39(06):1680-1684.

[20] 邹倩颖,吴宝永,王小芳.仿生模式识别在影像学肺炎判别算法的改进[J].实验室研究与探索,2020,39(02):34-38.

[21] Gaobo Liang, Lixin Zheng.A transfer learning method with deep residual network for pediatric pneumonia diagnosis[J].Computer Methods & Programs in Biomedicine, 2020.04

[22] B Pardamean, Tjeng Wawan Cenggoro, Reza Rahutomo.Transfer Learning from Chest X-Ray Pre-trained Convolutional Neural Network for Learning Mammogram Data[J].Procedia Computer Science,2018,135:400-407

[23] Ilyas Sirazitdinov,Maksym Kholiavchenko,Tamerlan Mustafaev et al.Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray dtabase[J].Computer & Electrial Engineering,2019.09,78:388-399

[24] E.J.Yates,L.C.Yates,H. Harvey.Machine learning “red dot”: open-source, cloud,deep convolutional neural networks in chest radiograph binary normality classification[J].Clinical Radiology,2018.09,73:827-831

[25] Yu Wu,Zhang Yi. Automated detection of kidney abnormalities using multi-feature fusion convolutional neural networks[J]. Knowledge-Based Systems,2020,200.

[26] Eric Ke Wang,Chien-Ming Chen,Mohammad Mehedi Hassan,Ahmad Almogren. A deep learning based medical image segmentation technique in Internet-of-Medical-Things domain[J]. Future Generation Computer Systems,2020,108.

# 致 谢

大学四年的时光走得很匆忙，如白驹过隙。我要感谢每一个出现在我生命中的人，无论是同学、老师，还是室友、朋友甚至是陌生人，他们都给了我很多帮助，在生活上，学习上对我的关怀我都感激不尽。

在毕业设计和毕业论文的实践过程中，我的指导老师杨巨成院长给我提供了很大的帮助，从选择具有挑战创新性的题目到功能的完善，无一不是耐心、认真地给予指导。由于疫情原因，学校暂缓返校，院长还是坚持每两周开一次网络会议，抽出自己宝贵的时间来听我们的进展汇报，并且提出建设性的建议，为我们指出下一阶段努力的方向。他常常教导我们，毕业设计是自己的事情，我们一定要为自己负责，遇到问题时先试着自己解决，自己寻找解决办法，这是一个成长的过程。杨院长的话总是能够为我指点迷津，将我从迷茫中拉出来。此外，杨院长还曾是我的面向对象程序设计的老师，他在课堂上授课时神情专注、目光炯炯有神像会说话一样，鼓励着你跟着他的思路一步一步找出问题的答案，讲课方式也幽默有趣，他的课堂上总是充满了同学们的笑声和积极回答问题的声音，现在想想还是意犹未尽。在日常的生活中，虽然整天都是一身正装，给人一种很威严的感觉，实际上却是一个平易近人且特别亲切的人。总之，特别感谢杨院长在毕业设计和毕业论文上给予我专业性的指导。

此外，还要感谢对我提供帮助各位老师、同学和研究生学长学姐们，他们的热心深深地打动了我，每一次向他们请教问题时，都是认真、耐心地帮我解答，让我备受感动，就算有时候问的问题很愚蠢，他们也并不会嘲笑我，而是从基本的概念上为我解疑答惑，使我身心受益。同时也向在我本科阶段教导我知识的各位老师说一声：你们辛苦了！四年来，你们风雨无阻、言传身教，倾尽自己的毕生所学授予我们知识，教我们为人处世的道理，使我受益匪浅，最后还要感谢我的家人在生活中、学习上给我提供无限的支持，他们夜以继日的努力地工作，才使得我如愿以偿地完成学业。

大学四年的生活虽然结束了，但毕业绝不是终点，而是一段新生活的开始，我们正值壮年、风华正茂，未来的生活将会更加美好，我会用更加阳光、积极向上的生活态度迎接美好明天。