

UDC _____ 编号 _____

雲南大學
碩士研究生學位論文

題 目 基于卷积神经网络的肺炎CT图像分
割算法研究

Title Research on CT image segmentation
algorithm of pneumonia based on
convolution neural network

学院（所、中心） 数学与统计学院

专业名称 应用统计专硕

研究方向 机器学习

研究生姓名 XXX 学号 12020213031

导师姓名 XX 职称 副教授

2023 年 5 月

云南大学 硕士研究生 学位申请 简况表	论文 预 审	论文预审结果：			
		专家姓名	职称	所在单位（校内：学院/校外：所在单位）	
	论文 送 审	专家姓名	职称	所在单位（校内：学院/校外：所 在单位）	结果
	论文 答 辩	答辩结果：			
		答辩专家	职称	所在单位（校内：学院/校外：所在单位）	

论文独创性声明及使用授权

本论文是作者在导师指导下取得的科研成果。除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，不存在剽窃或抄袭行为。与作者一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

现就论文的使用对云南大学授权如下：学校有权保留本论文（含电子版），也可以采用影印、缩印或其他复制手段保存论文；学校有权公布论文的全部或部分内容，可以将论文用于查阅或借阅服务；学校有权向有关机构送交学位论文用于学术规范审查、社会监督或评奖；学校有权将学位论文的全部或部分内容录入有关数据库用于检索服务。

（内部或保密的论文在解密后应遵循此规定）

研究生签名：_____ 导师签名：_____ 日 期：_____

目录

中文摘要.....	I
Abstract.....	II
第一章 绪论.....	4
1.1. 选题背景、目的及意义.....	4
1.1.1. 选题背景.....	4
1.1.2. 选题目的及意义.....	5
1.2. 文献综述.....	7
1.2.1. 语义分割国内外文献综述.....	7
1.2.2. 图像识别国内外文献综述.....	8
1.3. 研究的主要内容.....	9
1.4. 论文组织结构.....	9
第二章 相关理论基础.....	11
2.1. 深度学习.....	11
2.2. 卷积神经网络.....	12
2.2.1. CNN 的结构.....	12
2.2.2. 全卷积神经网络.....	14
2.2.3. ResNet 网络.....	16
2.2.4. GoogLeNet 网络.....	19
第三章 基于 Inception-Resnet-v2 的肺炎 CT 图像识别研究.....	23
3.1. 基于 Inception-Resnet-v2 网络.....	23
3.1.1. 卷积神经网络模型架构.....	23
3.1.2. 卷积神经网络模型设计路线.....	25
3.2. 数据来源及预处理.....	26
3.3. 实验环境搭建.....	28
3.4. 实验过程.....	29
3.5. 实验结果.....	29
3.6. 结果对比与分析.....	31
3.7. 小结.....	32

第四章 基于 U-net 改进网络的肺炎 CT 图像语义分割研究.....	34
4.1. 改进 U-Net 网络框架设计.....	34
4.1.1. 损失函数的改进.....	34
4.1.2. 迁移学习和 EfficientNet 模块.....	34
4.1.3. 网络结构改进.....	36
4.1.4. Attention Gate 模块.....	39
4.1.5. 损失函数优化.....	40
4.2. 实验数据与环境.....	41
4.2.1. 实验数据来源.....	41
4.2.2. 实验运行环境.....	41
4.2.3. 评价标准.....	41
4.3. 实验结果与分析.....	42
4.4. 小结.....	44
第五章 结论与展望.....	46
5.1. 本文工作总结.....	46
5.2. 未来研究方向.....	46
参考文献.....	48

摘要

鉴于当今我国人口规模庞大的基本国情，医疗系统一直承受着巨大压力。特别是新冠病毒的袭来，医院人满为患，医护人员疲于奔命，不仅自身健康得不到保证，工作效率还会随着疲劳增加而降低。虽然抗原试剂的出现可以帮助人们在家自我检测，但抗原试剂有检测结果滞后性这个弊端。对于那些本身就患有基础性疾病的老年患者来说，越早发现，生命也就多一分保障。因此结合计算机辅助诊断势在必行也迫在眉睫。本文基于卷积神经网络技术搭建一套肺部 CT 图像识别模型。以下几个方面为本文主要进行的工作：

首先是在肺部 CT 图像分类任务中，过往的模型由于参数众多，计算量大而导致等待时间过长，不适合应用在实际中。本文采用的 Inception-ResNet 模型结合 GoogLeNet 和 ResNet 两大网络的优点，通过多个不同尺寸的卷积核进行信息提取的同时，借鉴残差模块的思想，防止模型出现过拟合。

然后在图像分割任务方面，本文主要选用 U-Net 网络作为基础网络框架，因为在医学图像的分割效果上，U-Net 取得了不错的效果。本文吸收了迁移学习的思想，选择 effentnet 族 CNN 模型作为编码器模块。利用预训练得到的信息解决新问题，同时需要的数据量更少。由于传统 U-Net 网络提取的浅层特征图有很多的冗余信息，会浪费大量的计算时间，引入注意力机制后，网络模型会在大量信息传输给解码器之前，找到当前任务中最重要的信息，通过抑制非相关信息的特征去提高网络的性能。本文改进后的 U-Net 模型与 CNN、FCN 等模型进行比较，可以发现模型精度提高的同时切割出的病灶部位特征也更明显。正好能弥补 Inception-ResNet 模型对于某些病灶部位模糊而导致误判普通肺炎和新冠肺炎的缺点。

关键词：肺炎；卷积神经网络；图像语义分割；肺炎 CT 图像识别

Abstract

Given the basic conditions of our country today with a large population, the healthcare system has been under tremendous pressure. Especially with the onset of the new coronavirus, hospitals are overcrowded and health care workers are overwhelmed, not only their own health is not guaranteed, but their work efficiency also decreases as fatigue increases. Although the advent of antigen reagents can help people to self-test at home, antigen reagents have the disadvantage of lagging test results. For elderly patients with underlying diseases, the earlier they are detected, the more life is guaranteed. Therefore, it is imperative and urgent to incorporate computer-aided diagnosis. In this paper, we build a lung CT image recognition model based on convolutional neural network technology. The following aspects are the main work carried out in this paper:

The first is that in the lung CT image classification task, the past models are not suitable for application in practice because of the long waiting time due to numerous parameters and large computation. The Inception-ResNet model used in this paper combines the advantages of two major networks, GoogLeNet and ResNet, to extract information through multiple convolutional kernels of different sizes while drawing on the idea of residual modules to prevent the model from overfitting.

Then, for the image segmentation task, this paper mainly selects the U-Net network as the basic network framework, because U-Net has achieved good results in the segmentation effect of medical images. In this paper, we absorb the idea of migration learning and choose the effnet family CNN model as the encoder module. The information obtained from pre-training is used to solve the new problem while requiring less data. Since the shallow feature map extracted by the traditional U-Net network has a lot of redundant information, which will waste a lot of computation time, after introducing the attention mechanism, the network model will find the most important information in the current task before a large amount of information is transmitted to the decoder, and go to improve the performance of the network by suppressing the features of non-relevant information. In this paper, the improved

U-Net model is compared with CNN, FCN and other models, and it can be found that the model accuracy is improved while the features of the cut-out lesion sites are more obvious. The Inception-ResNet model is able to compensate for the disadvantage of misclassification of common pneumonia and neocoronary pneumonia due to the ambiguity of some lesion sites.

Key Words: Pneumonia ; Convolution neural network ; Image semantic segmentation; Pneumonia CT image recognition

第1章 绪论

本章主要讲述新冠肺炎 CT 图像检测的意义，以及深度学习与新冠肺炎检测相结合的研究现状和成果，提出了本文需要做的任务，阐述了本文的主要研究内容和组织结构。

1.1. 研究背景及意义

1.1.1. 研究背景

新型冠状病毒肺炎，简称“新冠肺炎”或 Corona Virus Disease 2019（COVID-19），被世界卫生组织谭塞德于 2020 年 2 月 11 日在瑞士日内瓦正式命名，并于 2022 年 12 月 26 日由国家卫健委将新型冠状病毒肺炎更名为新型冠状病毒感染。该病毒感染是一种急性传染病，其病原体在过去从未被人类发现的单链 RNA 病毒，主要通过飞沫进入呼吸道、气溶胶体以及其他密切接触传播。传染源主要是新冠病毒感染患者，潜伏期即具备传染性，发病三天内感染性较强。即使无症状患者也具备同样传染性。人群普遍易感染，根据患病程度不同有不同的身体表现。轻型症状主要以呼吸道感染为主，包括咽干、咳嗽、发热、咽喉疼痛等，重症患者表现为气促表现、多器官衰竭、肺部影像学显示 24~48 小时内病灶进展明显大于 50%等状况。

2019 年 12 月由于国外输入带有新冠病毒的动物，武汉开始出现了新冠病例，但当时并未引起重视。随着春节临近，全国各地人民开始大规模返乡过年，人口流动的增大使得病毒以极快的速度传播，在短时间内迅速扩散至全国。2020 年 3 月进入病毒高峰期，截止 3 月底，全国累计确诊超过 8 万例；国家各级部门迅速做出响应，钟南山院士不顾自身安危，带队进入武汉，深入最初的病毒传播地——武汉华南海鲜批发市场进行调查，迅速了解了病毒的相关特征，国家卫健委做出精准、及时、科学的防疫措施。之后全国进入“动态清零”的防疫政策，健康码、行程卡应运而生。各地隔离点、核酸检测点如雨后春笋般冒出，一场艰难的、旷日持久的抗疫战斗拉开了序幕。这场战斗基本持续了两年了，中国的各个行业，尤其是旅游行业、教育行业和餐饮行业，是备受影响最重的。各个行业不景气了，大环境也就没落了。

在这个科技飞速发展的社会，计算机正在逐渐进入各个领域，它改变着人们的生活。现在“大数据”一词已经屡见不鲜了。我们的生活、工作、一切的行为都在这个社会上留下“足迹”，而许许多多“足迹”集合在一起，就成为了一种可利用的宝贵财富。比如随着大数据与医疗行业融合，许多医院经营多年，有大量的医疗数据，这些数据过去没有被有效的利用起来，而是随着数量越来越多，反而成为了医院的负担。事实上，大部分医院都有一个或多个医疗数据库，只不过这些数据库的功能都比较精简，常用于复制、查找、删除、添加等基本操作。

虽然数据库的数量正在增加，但是能从浩如烟海的数据海洋中获取有用的数据并不是一件容易的事。因此，如何高效率的挖掘医学图像数据以及精准、科学的分析相关数据成为了如今的热门话题。虽然随着我国疫情政策的放开，抗原试剂面向市场，人人都可以在居家过程中进行检测。但对于那些身患其他基础性疾病或者老龄患者，越快确诊意味进一步保证自身的生命安全，因此通过人工智能对肺部 CT 图像进行诊断成为快速确诊新冠病毒的一种不可或缺的途径。

1.1.2. 研究意义

目前通过电子计算机断层扫描技术（Computed Tomography, CT）对肺部扫描成像，可以得到 2D 成像的肺部 CT 图像。根据肺部 CT 图像中病变部位的外形大小、颜色等特点可以准确的判断病变的程度，也是判断新冠病毒病重程度的一个重要判定标准。成像基本原理因为 X 线束穿透力强，用 X 线扫描人体检查部位，然后从接收器上接收 X 线，最后转换为数字信号给计算机。过去传统的人工标记分辨 CT 图像已经不适合当今数据繁杂、多样的数字时代了，通过深度学习中各类算法对数据进行分类、筛选、分析等任务，以满足各个领域内与时俱进的需求是时代的洪流，是大势所趋。计算机相较于人类来说，拥有永不疲倦的优势，同时能发现一些人们不曾留意的细节。目前医学领域中出现计算机辅助系统已经司空见惯的事了。但医学领域中的图像数据与其他领域的图像数据又有以下的区别：

（1）医学图像具有个体差异性。由于人体的组织器官结构复杂，血管、神经、肌肉组织等结构错综复杂的组合一起，不同个人的同一器官并非一模一样，同一个体的器官在不同年龄阶段也存在一定的不同。因此某一类的医学图像本身

包含了太多的样本类型，为计算机的识别增大了难度。

（2）图像画面参差不齐。在图像成像过程中可能会受到许多非人为控制的因素影响，导致图像成像的质量鱼龙混杂，部分模糊的图像为计算机的识别带来了挑战。

（3）器官、组织在身体内的分布并不均匀。而且由于医学图像是 2D 图像，而人体是三维结构，导致选择不同角度的成像会对计算机判定结果形成影响。

（4）数据量过于庞大。比如肺部 CT 图像，拍摄的一个时间段内肺部各个位置的图像，这样就导致一个病人的某一个时期的肺部 CT 图像数量都特别多，但能用于计算机算法学习的图像就只有中间时段的某一部分图像。

鉴于卷积神经网络在执行计算机视觉任务中的优越表现，故选用卷积神经网络作为肺炎 CT 图像识别相关研究的基础框架。将深度学习的技术运用到新冠病毒的诊断方面上，能有效提高医生诊断的效率，降低人力资源的损耗同时尽可能的把人力用在其他更需要被使用的地方。虽然目前来看新冠病毒仿佛已经离我们而去，但纵观世界各国的情况仍旧能发现它的身影。它就像躲在灌木丛中的野兽，对我们和平美好的生活虎视眈眈。因此我们应对提前做好迎接它新一轮的攻势的准备。不同于传统人工检测，它具有高准确率、高效率、低失误率的特点。卷积神经网络将输入的图片转换为数字矩阵，然后通过卷积核进行卷积运算，将数字矩阵转换成不同通道的向量，降维的同时保留图像本身的特点，通过对大量不同类型图片的“学习”，计算机最终能实现图像分类等视觉任务。只要算法本身没有问题，那么通过大量的、有效的样本，模型识别图像的准确率可以得到保证，且基于如今迁移学习的思想——即一类图像识别的模型参数可以作为另一类相似模型的预训练模型参数，计算机能在短时间内对大量的、不同需求的图像任务进行处理。相比较之下人工识别因需要经过相关知识学习的传授以及学习过程而会花掉大量的时间。另一方面，一台计算机因故障不能使用了不代表它所进行的计算机视觉任务暂时无法完成，可以通过另一台计算机重新开始任务，但医生一旦因其他原因无法进行工作后，并不能保证还有其他医生可以立即代替他进行工作。综合上述深度学习的两大优势，更能表现出本文研究的意义。本文就以深度学习为基础，建立一套识别新冠肺炎 CT 图像的模型。

1.2. 文献综述

1.2.1. 语义分割国内外文献综述

这些年，深度卷积神经网络在计算机视觉任务中的表现越发优异，但由于算法学习需要用到大量的训练集以及需要考虑网络参数的数量大小（因为参数的数量大小决定消耗的资源），卷积神经网络的前进之路并非一帆风顺的。自从 Krizhevsky 等人^[1]在一个数据集包含一百万张训练图像的样本上对一个多层、超大参数量的大型网络进行监督学习开始，神经网络的模型走向了更大、更深的层次。

卷积神经网络主要运用在计算机图像分类的领域中，图像输出的结果一个表明类别的向量，向量上每一个值表明图像属于该类别的概率。但是在生物医学图像处理的任务中，我们不仅需要知道图像的类别，还希望能得到病灶区域的定位，也就是把结果中向量重新转换为一张像素图像并对图像中每一个像素进行分类。而且，由于某些疾病的发病率较低，导致该病医学图像的样本量相对较少，并不支持算法模型学习所需要的样本量。因此，Ciresan 等人^[2]滑动窗口的形式训练网络，他们将一张图像细分成一个个像素，以某一像素周围的局部区域的像素类别来预测该像素的类别，这个网络可以本地化，而且也解决了训练数据量不足的问题。他们凭此网络获得了当年 ISBI 的 EM 细分挑战。但这种方法存在两个弊端：一方面，由于每个补丁都必须独立运行，且大补丁的重叠会导致网络出现大量计算冗余，这就势必导致网络的计算速度缓慢；另一方面，感受野大小与定位精度之间的取舍。较大的补丁不能精确的表达像素的定位，因为它通过了更多的池化层操作。而较小的补丁虽然能精确表达其在像素图像中的位置信息，但它只允许网络拥有较小的感受野。为了解决上述问题，Ronneberger^[3]等人在 2015 年提出了著名的 U-Net 网络，该网络是医学图像处理领域的“神兵利器”。

当把人工智能具体应用到本次所面临的新冠病毒上后，医院诊断的工作效率和准确度都有了明显的提升。Lessmann^[4]设计了（CO-RADS）人工智能系统，该系统能对胸部 CT 图像进行正确分类从而确诊新冠患者，并且对病症的轻重程度进行智能评分。Zhang^[5]等人也做了类似的探究，他们使用一个包含三千多名患者肺部图像的大型 CT 图像数据库，成功设计了一套新冠病毒智能诊断系统。该系统可以对疑似病例进行初步筛选，且准备率非常高，能为医生减轻一部分压

力，具有十分重大的现实意义。

新冠病毒智能诊断系统主要分为两大类：新冠确诊以及病灶部位的分割。通常采用肺部 CT 图像作为判定的材料。Bai^[6]等人采用 EfficientNet^[7]网络设计出了一一种使用胸部 CT 图像来诊断新冠病毒以及其他肺炎的人工智能系统。他们得出结论：放射科医生在使用人工智能诊断系统后工作效率有了明显的提高。Wang^[7]等人建立一个基于弱监督学习框架的新冠病毒预测系统，通过对输入后的肺部图像进行分割，最终预测感染新冠的概率并且采用弱监督病灶区域定位法对病灶部位进行定位。此后 Wang^[8]等 12 人又提出一种通过有效学习各类具有分布差异的数据集来改进新冠诊断的新型联合学习框架。

1.2.2. 图像识别国内外文献综述

近年来，深度学习的话题在社会上沸沸扬扬。随着深度学习在计算机视觉任务中的卓越表现，医学图像分类领域内也渐渐出现了它的身影，越来越多先驱者进行了伟大的尝试。Hua^[9]团队基于深层卷积神经网络的理论，在孤立肺结节的 CT 图像分类任务中使用深层信念网络模型。David^[10]等人对视网膜病方面进行了相关探索，他们首先使用一系列图像增强技术对原始图像进行增强处理，然后把处理完后的图像放入预计设计且已经训练好的网络模型中，网络模型会根据图像的特征判断患者视网膜的病变情况。Ramirez^[11]等人使用大脑 CT 图像在预计训练好的卷积神经网络中分析阿兹海默症患者的各项身体情况以及患病程度。Derya Avcı^[12]等人在卷积神经网络中运用自适应小波熵能量理论，设计出一个专门用于尿液细胞分类的网络模型。沈^[13]对深度学习在医学方面的应用情况进行了详细的概括分析，整篇综述囊括了深度学习里的各类经典模型，跟随模型问世的时间线，从卷积神经网络（CNN）到 LeNet-5^[14]网络和更深层次的 AlexNet^[15]网络，最后是 ResNet 和 GoogLeNet。其中，CNN 可以作为卷积神经网络鼻祖，虽然计算速度、图像特征提取等方面存在一定的问题，但为卷积神经网络在医学图像中的应用打下了基础。GoogLeNet 网络提供了一个新的思路：将输入图片通过不同尺度的卷积核进行卷积，能降低模型的参数数量，提升计算速度。ResNet 网络的现世为神经网络家族又添加了一种全新的网络模块—残差模块，可以有效的解决梯度消失等情况。深度学习发展到如今，层出不穷的模型如雨后春笋般冒

出，一个又一个突破准确率的上限，使得将这项技术与社会中各个领域相结合提供了更多的可能性。

1.3. 研究的主要内容

本文将从图像分类和图像分割两大计算机视觉任务的角度来处理肺炎 CT 医学图像。从图像分类的角度出发，本文采用 3000 张正常人肺部 CT 图像、3000 张新冠患者肺部 CT 图像、3000 张普通肺炎患者肺部 CT 图像作为实验的对象。其中每个类型图像各取 2100 张作为训练集，600 张作为测试集，剩余 300 张作为最终模型的验证集。本文选取当前主流的几个网络，如 ResNet、GoogLeNet、以及改进的 Inception-ResNet 网络，采用角度旋转、比例缩放、翻转变换等方式对数据集中的图像进行增强处理，然后将处理后的图像输入 ResNet、GoogLeNet 以及改进后的 Inception-ResNet 网络中，最后通过一些的评判指标，对各个模型的结果进行判定以验证实验结构。

从图像分割的角度出发，虽然 Inception-ResNet 网络可以对输入的肺部 CT 图像进行分类，且准确率较高。但对于一些画面模糊的图像、由于年龄等因素导致患者的病灶区域不突出的图像以及图像边缘像素缺失的图像，上述模型就有点心有余而力不足了。因此图像语义分割就显得格外重要了。

本文将在 U-Net 网络的基础上，结合注意力机制、迁移学习以及多维度特征提取来对实验对象进行图像语义分割，“注意力”机制的加入可以分配给图像中那些不是很重要的空间位置较低的权重，以便提升计算速度。接下来，在改进后的 U-Net 网络的基础上，结合迁移学习的思想，把普通卷积核改为 effentnet 族 CNN 模型作为编码器块。最后就是引用多维度特征提取的思想，把不同尺度的上采样和下采用特征图采用相乘，之后把按照不同的权重把不同尺度相乘后的特征图按照不同权重融合起来，以便更精确的提取特征。并将最终结果与原本的 CNN、FCN 网络进行对比，证明改进的网络的可行性。

1.4. 论文组织结构

本文结构如下所示：

第一章：绪论，主要介绍研究背景以及意义，国内外的研究现状，论文框架以及创新点。

第二章：关键理论和技术，注意介绍深度学习、图像分类中 GoogLeNet、ResNet、VGG 以及图像语义分割中 U-Net 网络的基础理论。

第三章：基于改进 Inception-ResNet 网络的肺炎 CT 图像分类研究。首先，把图像数据进行基本的预处理，比如归一化，无量纲等。然后带入各个网络模型，最后将得到的混淆矩阵结果进行对比，验证本文改进的网络的性能。

第四章：基于改进的 U-Net 网络的肺炎 CT 图像语义分割研究。首先对图像进行增强，之后将增强后的数据带入改进的网络里。最后将得到的结果与 CNN、FCN 网络的结果进行对比，验证改进的网络的性能。

第五章：全文总结与展望。

第2章相关理论基础

2.1. 深度学习

深度学习是一种使用基于模仿人脑结构建立的神经网络，通过学习和分析来处理相关问题并解释结果的一种机器学习技术。何谓神经网络？简而言之即类似人类大脑处理信息——通过感受器接收信息、通过神经传导信息给大脑进行处理并最终输出结果的一种结构网络。该网络最基本的成分称为神经元模型，其结构如下图 2-1 所示。神经网络结构如图 2-2 所示，由输入层输入数据，输出层输出数据，在输入层和输出层之间的叫隐藏层，隐藏层也可以理解为神经网络中的“黑盒子”，是除了输入和输出以外其余功能的总称。当隐藏层的层数大于两层时，称这种网络为深度神经网络。

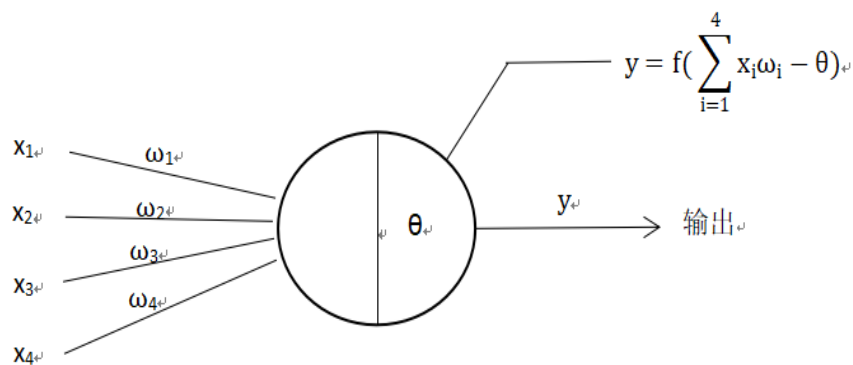


图 2-1 神经元模型

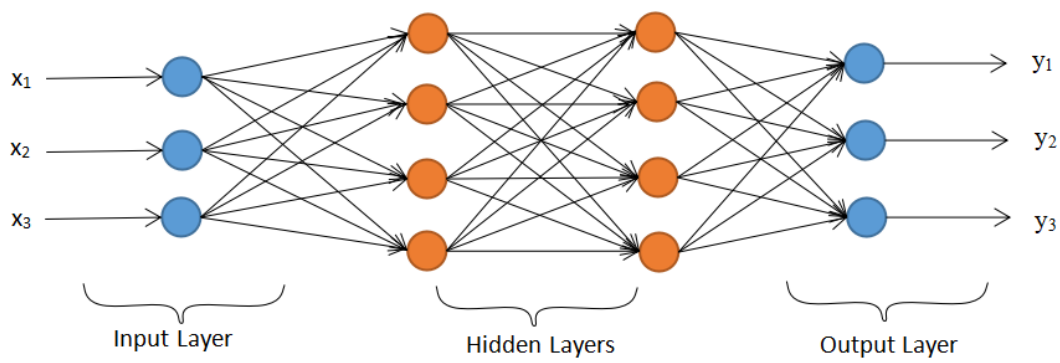


图 2-2 神经网络模型

2.2. 卷积神经网络

2.2.1. CNN 的结构

深度神经网络(以下简称 DNN)包括卷积神经网络(以下简称 CNN)、递归神经网络(RNN)和强化学习(RL)。

卷积神经网络(CNN)由每一个不同学习特征的卷积核进行图像特征提取,经由非线性激活函数对提取出来的数据分布进行重新映射,经由池化层操作对数据进行降维处理。一张图片输入进网络后,计算机首先会将它转换为相应的特征矩阵,接下来会根据通道数的不同(比如常见的 RGB 图像有 Red、Green、Blue 三个通道),对应相同数量的卷积核 $\{w_i\}_k$ (也被称为滤波器)进行卷积运算从而得到 feature map。

卷积神经网络的组成结构为:输入层、卷积层、池化层、激活层、全连接层、输出层。每一层都有每一层相应的作用,例如卷积层的主要作用为提取特征,池化层的主要作用为降维等。层与层之间大致上按照上述的顺序组成,除了卷积层和池化层可以呈现互相交替连接,即一层卷积层后面连接一层池化层,一层池化层后面连接一层卷积层。下面对各个层的作用进行详细的解释:

①输入层:输入层的工作相对简单,它可以将大部分形式的数据(数字图像、音频领域内的二维波数据等)输入到卷积神经网络中。为了解决数值问题、优化计算速度、防止神经元出现饱和现象,计算机会在这一层对数据进行归一化处理。

②卷积层:是卷积神经网络命名由来的核心,卷积是一种线性运算,由卷积核对特征图像不同位置赋予的不同权重,进行加权运算构成。卷积运算的过程如图 2-3 所示,原始图像为 5×5 的原始图像,中间的数字矩阵代表 3×3 的卷积核,设置步长为 1(即中间的数字矩阵在原始输入图像上每次卷积仅移动一格),通过卷积获得右边 3×3 的特征图。

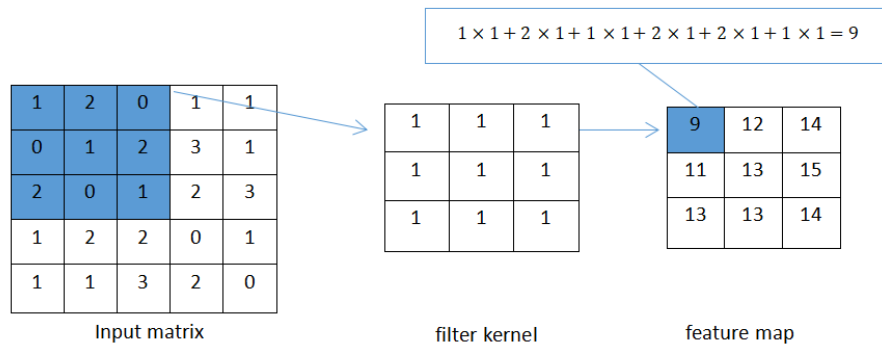


图 2-3 卷积运算过程

③激活层：激励层的作用可以理解为把卷积层的结果做非线性映射，神经网络的一个优点是它们能够以非线性的方式学习，能解决一些线性模型解决不了的问题。如果不用数学术语解释，它们的意思是能够发现不太明显的图像中的特征——树上的苹果，阳光下的，阴影下的，或厨房柜台的碗里的。

④池化层：池化层的操作是将一个窗口内的像素按照平均值加权或选择最大值来作为输出，一个窗口内仅有一个输出数据。因此，当经过池化层后，图像的尺寸会变小，计算量也会变小，相比于在池化前使用卷积，池化后同样的卷积大小具有更大的感受野。目前最常用的池化方式标准池化、重叠池化和空间金字塔池化。标准池化包括广为人知的最大池化和平均池化两种。重叠池化则是特征图中的某些区域会参与多次特征提取工作。空间金字塔对不同输出尺度采用不同的滑窗大小和步长以确保输出尺度相同，然后将不同尺度提取的特征融合起来。最大池化和平均池化的具体过程如图 2-4 所示。

⑤全连接层：是一个线性特征映射的过程，它接收前一层输出的多维特征向量，并通过对特征向量上每一个维度的数值赋予一个权重，将整个多维特征向量转变一个 $n \times 1$ 特征向量（其中 n 代表总体标签数量，例如二分类模型情况下 $n=2$ ），由于转变后的 $n \times 1$ 特征向量中每一个数值形成都是前一个多维特征向量中所有数值共同作用的结果。因此称为“全连接”。

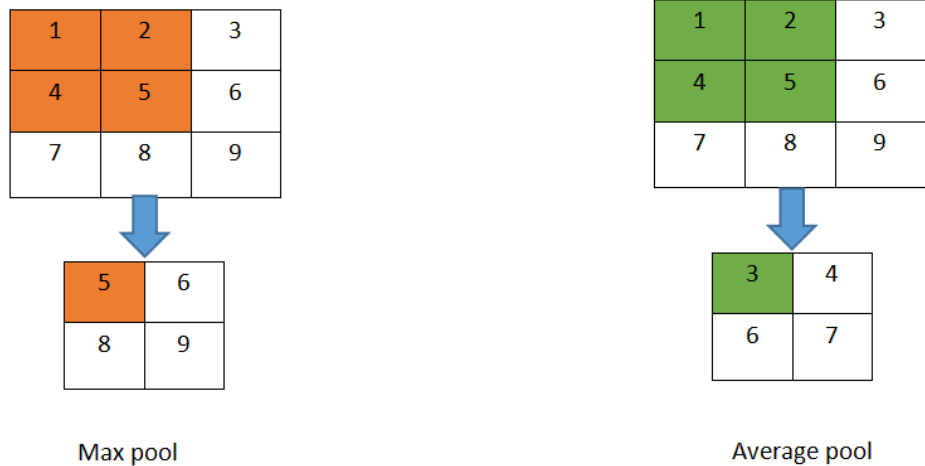


图 2-4 池化过程

2.2.2. 全卷积神经网络

VGG 网络的出现为全卷积神经网络的诞生提供了灵感。牛津大学的 Visual Geometry Group 团队于 2014 年最早提出 VGG^[15]模型的概念。VGG 模型常常应用于图像识别等计算机视觉任务中，主流上一般使用 VGG-16 和 VGG-19 两个模型（当然它们两个模型没有本质上的区别，只是网络深度不同），同时，它们的卷积核也被 ResNet、Unet 等后起之秀使用。在传统 VGG 网络模型中，图像信息在经过卷积和池化等操作后，最终会被全连接层映射成一个长度固定的向量。这个向量上的每一个值就是图像被分类成每一种类别的概率。但由于全连接层只接受预先设定好的特征图尺度，这就导致 VGG 网络应用性受到一定的限制。而 Jonathan Long^[16]等人提出的全卷积神经网络（Fully Convolutional Networks, FCN）丢掉了所有的全连接层，极大地减少了内存的占用情况，减少卷积计算冗余，在原有的基础上增加一个上采样过程，因此被称为全卷积神经网络。VGG 网络在下采样时，通过第一次池化操作会获得原图 1/2 大小的特征图，之后进入池化层会使得特征图的大小再次减半，重复上述操作会得到 1/8、1/16、1/32 大小的特征图。将 1/32 大小的特征图通过反卷积、反池化以及“双线性差值法”等方式（统称为上采样过程），又会变成 1/16 大小的特征图，重复 5 次操作，可获得一个同原图同样大小的特征图，这种分割结果被称为 FCN32s。但由于卷积、池化亦或者反卷积、反池化的过程都会丢失一些信息，也就是意味着同一张特征图经过一

次卷积再经过一次反卷积后，哪怕大小跟之前一样，但其中包含的信息会减少。因此为了获得更好的分割结果，可以先对 1/32 大小的特征图进行一次上采样过程获得 1/16 大小特征图，将之与下采样过程 1/16 大小的特征图进行融合，图像特征融合的基础方式包括相加或拼接。将融合后的图像再通过上采样操作获得与原始特征图同等大小的分割结果，这样操作获得的分割结果效果更佳，因此被称为 FCN16s。同理可得精确度更高的 FCN8s。

FCN 网络能对图像进行像素级别的分类，因此常被用于图像语义分割领域。在 CNN 网络中，由于全连接层的权值参数的固定的。因此为了获得一个固定长度的特征向量，一般我们输入的图片大小也是固定的。而在 FCN 网络中，由于我们通过数次反卷积操作，可以把特征图恢复到原有图像的大小，因此理论上 FCN 网络可以接受任意大小尺寸的输入图片。虽然 FCN 网络的精度较高，但在某些小型图像的高精度需求下仍然显得有些不足，通过反卷积操作（也称为转置卷积）获得的分割结果往往比较模糊，对细节的把握也略小粗糙。反卷积的计算公式如下所示：

$$w_p = (w_q - 1) * s - 2p + k$$

其中， w_p 代表输出特征图上的值， w_q 代表输入特征图上的值， s 代表步长， p 代表填充像素， k 代表转置卷积矩阵上的值。

基于 FCN 模型的理论基础，U-Net^[17] 模型诞生了。该模型在语义分割领域取得了非常显著的成果。如图 2-5 所示，U-Net 模型主要由编码器—解码器结构组成，因为形状像 U，固称为 U-Net 网络。编码器部分通过 2 个 3×3 的卷积和 ReLU 函数，再通过 2×2 的下采样，池化操作让特征通道的数量会翻倍。该过程也被称为下采样。解码器部分，特征图每经过一次 2×2 的 Up-Conv，特征通道就会减半，之后再与上采样过程获得的相同尺度的特征图进行融合，融合方式为拼接，再通过滤波器大小为 3×3 的卷积操作和 ReLU 激活函数。最终使用 1×1 卷积核将特征向量进行映射。惩罚函数的权重计算如下所示：

$$E = \sum_{x \in \Omega} f(x) \log(p_{k(x)}(x))$$

其中， k 为像素点的真实标签， f 为权重分布图，每一个像素点对应的权重都不同。

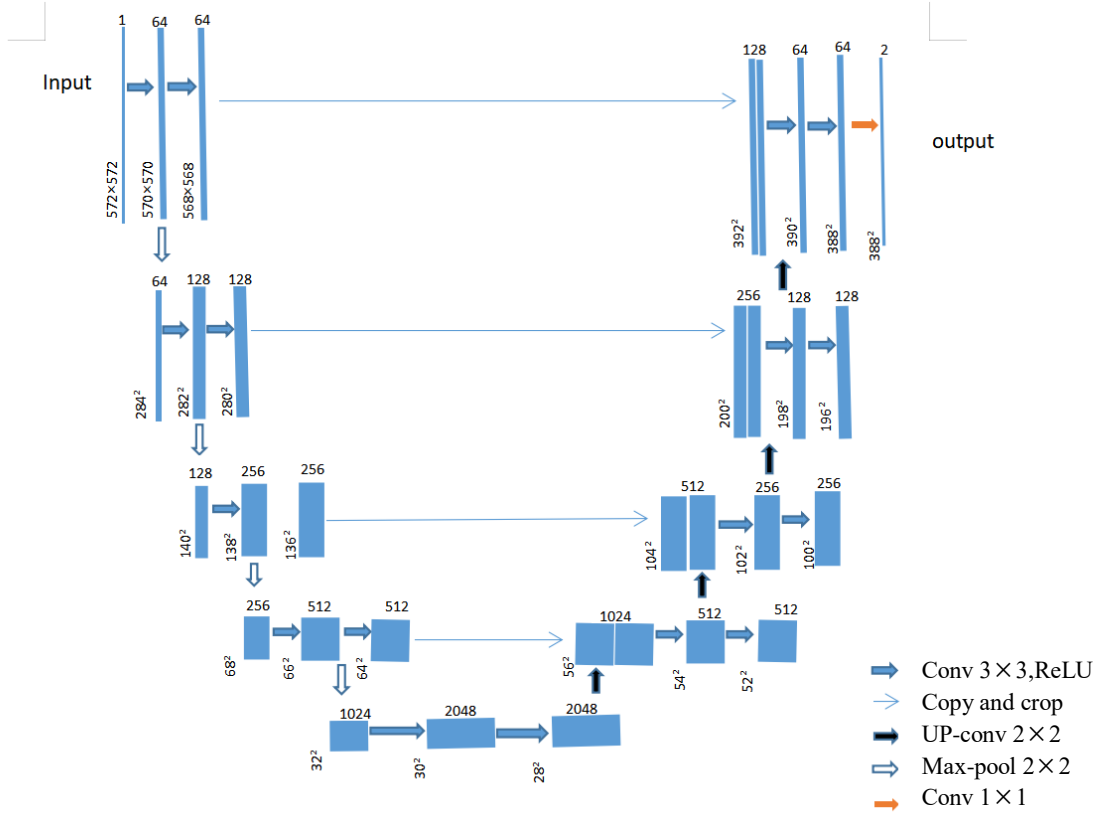


图 2-5 U-Net 网络模型结构示意图

2.2.3. ResNet 网络

在 ResNet 问世以前，就例如上文介绍的 VGG 模型，其结构中包含卷积层和池化层的交替组合，过去人们认为随着网络层数的增加，模型的性能会逐渐提升。但经过实验发现并非如此，随着层数的增加，会出现梯度消失或者梯度弥散两个问题。所以 VGG 网络只有 19 层结构组成，但 ResNet 网络层数则有 152 层，究其原因是因为 ResNet 网络采用的残差模块很好的解决了上述问题。残差网络的核心思想是：每一个附加层都应该很容易的包含原始函数作为其元素之一。以下为残差网络的主要计算公式：

$$\begin{cases} F(x_l) = \omega * x_l + b \\ y_l = R(F) + h(x_l) \\ x_{l+1} = R(y_l) \end{cases}$$

其中， x_l 为上一层输入的变量， ω 为该像素位置对应的权重， b 为偏置参数（用来加速神经网络的拟合）， y_l 为两个函数之和， R 为激活函数（通常采用 ReLU

激活函数）， $F(x_l)$ 为卷积运算过程， $h(x_l)$ 为对输入 x_l 的简单线性变换， x_{l+1} 为残差模块输出。可以看出残差网络的计算是将从输入直接引入一个短连接到非线性层的输出上。ReLU 是线性整流函数（Linear rectification function），又称修正线性单元。其优势在于：更加有效率的梯度下降以及反向传播同时缓解了梯度爆炸和梯度消失问题，ReLU 表达式如下：

$$R(x) = \max(0, x)$$

当 $x > 0$ 时， $R(x) = x$ ，导数是 1；当 $x \leq 0$ 时， $R(x) = 0$ ，导数是 0。残差模块结构示意图如下：

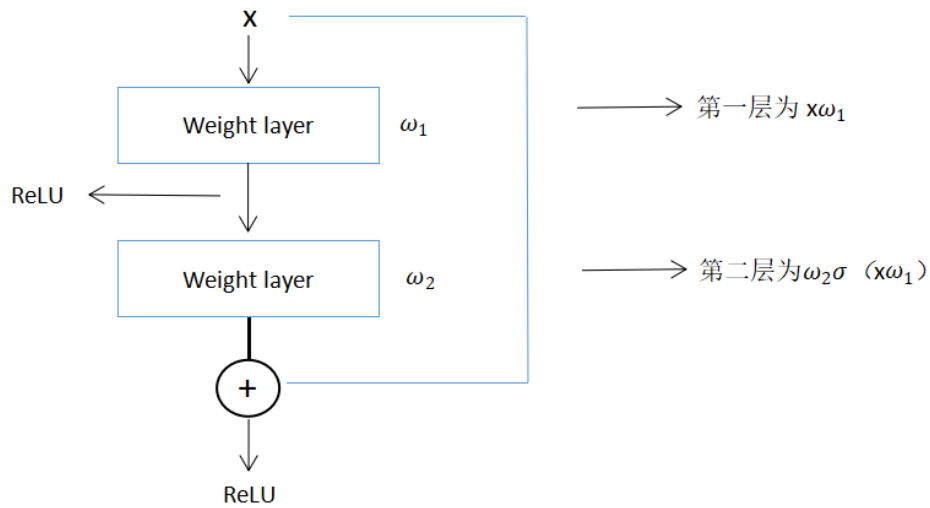


图 2-6 残差网络结构示意图

在神经网络的学习中除了梯度弥散还会出现梯度消失的情况，何谓梯度消失？图 2-7 为一个简单神经网络示意图。该简单网络每层神经网络中的神经元数目为 1， b_1, b_2, b_3, b_4 代表 4 个神经节点中的阈值， $\omega_1, \omega_2, \omega_3, \omega_4$ 代表四个权重， C 代表一个损失函数。

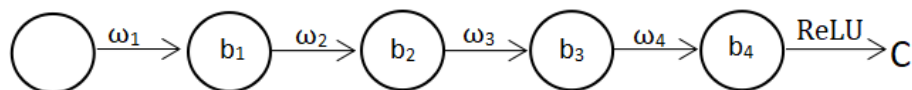


图 2-7 简单神经网络示意图

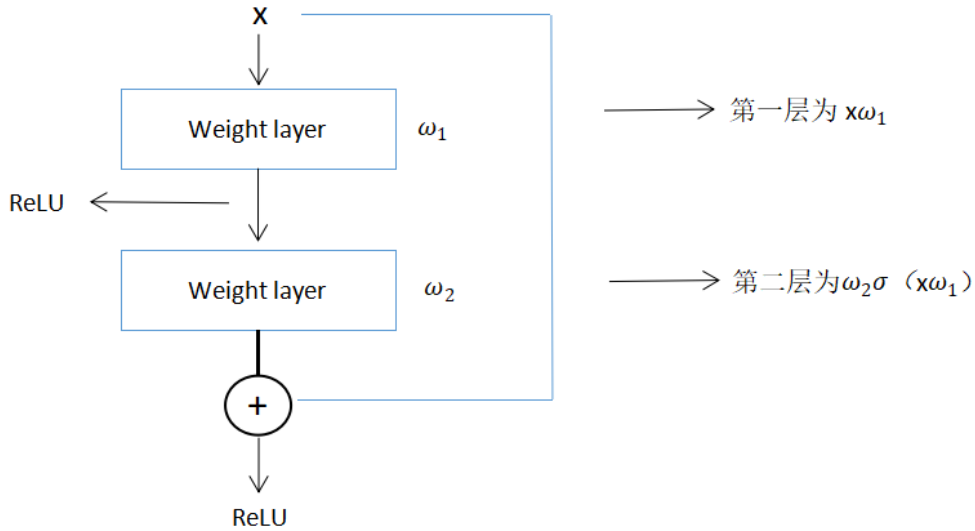
接下来令：

$$\begin{cases} z_i = \omega_i \alpha_{i-1} + b \\ \alpha_i = \sigma z_i \\ \sigma(\omega_1 \alpha_0 + b_1) = \alpha_1 \\ \sigma(\omega_2 \alpha_1 + b_2) = \alpha_2 \\ \sigma(\omega_3 \alpha_2 + b_3) = \alpha_3 \\ \sigma(\omega_4 \alpha_3 + b_4) = \alpha_4 \end{cases}$$

现在假设我们需要更新 b_1 的参数，即我们需要求出损失函数对其的偏导数。根据链式求导法则，得到如下公式：

$$\frac{\partial C}{\partial b_1} = \frac{\partial C}{\partial \alpha_4} \cdot \frac{\partial \alpha_4}{\partial \alpha_3} \cdot \frac{\partial \alpha_3}{\partial \alpha_2} \cdot \frac{\partial \alpha_2}{\partial \alpha_1} \cdot \frac{\partial C}{\partial \alpha_1} \cdot \sigma'(z_4) \cdot \omega_4 \cdot \sigma'(z_3) \cdot \omega_3 \cdot \sigma'(z_2) \cdot \omega_2 \cdot \sigma'(z_1)$$

由于 $\omega_1, \omega_2, \omega_3, \omega_4$ 这些参数设置的都比较小，所以一旦层数太多的时候，某一层的梯度就会无限逼近 0，这个现象就被称为梯度消失。那么 ResNet 为何能解决梯度消失这种情况呢？在形如下图的残差模块中：



$$\begin{cases} F = \omega_2 \sigma(\omega_1 x) \\ y = F(x_i, \{\omega_i\}) + x \end{cases}$$

其中 $x_L = x_i + \sum_{l=i}^{L-1} F(x_i, \{\omega_l\})$ ，对损失函数关于求偏导：

$$\frac{\partial Loss}{\partial x_i} = \frac{\partial Loss}{\partial x_L} \cdot \frac{\partial x_L}{\partial x_i} = \frac{\partial Loss}{\partial x_L} \cdot \left(1 + \frac{\sum_{l=i}^{L-1} F(x_i, \{\omega_l\})}{\partial x_i}\right)$$

上式具有优秀的反向传播效应。而且由于括号中存在“1”，因此确保梯度不会逼近 0，故有效的防止了梯度消失的情况。同时每一层模块都是残差函数的和，而过去的模型全是 $\prod_{i=1} \omega_i$ 的形式。

2.2.4. GoogLeNet 网络

对于深度神经网络来说，提升其性能的手段是增加其深度以及宽度（深度指神经网络的层数，宽度指神经网络每层的通道数），这意味着网络的参数数量会非常大，计算量也会非常大。然而，过大的参数往往会造成过拟合现象，为了解决这种困难，GoogLeNet^[19]网络应运而生了。GoogLeNet 网络解决上述问题的办法是通过将全连接层甚至一些卷积层都进行稀疏化处理，跳过许多不必要的计算，包括剪枝等操作对模型大小进行压缩，同时保证模型准确度不会下降太多。GoogLeNet 网络采用了 Inception-v1 模块，该模块的核心思想是通过将多个卷积层并联到一起的方式形成一个更深的矩阵，同时，Inception 模块可以对尺寸较大的矩阵进行降维后再与其他卷积层处理的矩阵进行拼接，这种降维操作一般是通过 1×1 的卷积核进行卷积。总结来说，GoogLeNet 网络具有以下几个优点：

- ① GoogLeNet 网络能在参数不多的情况保证计算速度。
- ② GoogLeNet 网络通过多种卷积核进行卷积，从而能得到多种尺度的信息，尽量保证信息的获取，通常是 $1 \times 1, 3 \times 3, 5 \times 5, 7 \times 7$ 卷积核。
- ③ Inception 模块将卷积层和池化层融合在一起，具有稀疏性，计算速度快等特点。

以下为 Inception-v1 到 Inception-v3 模块的组成结构。

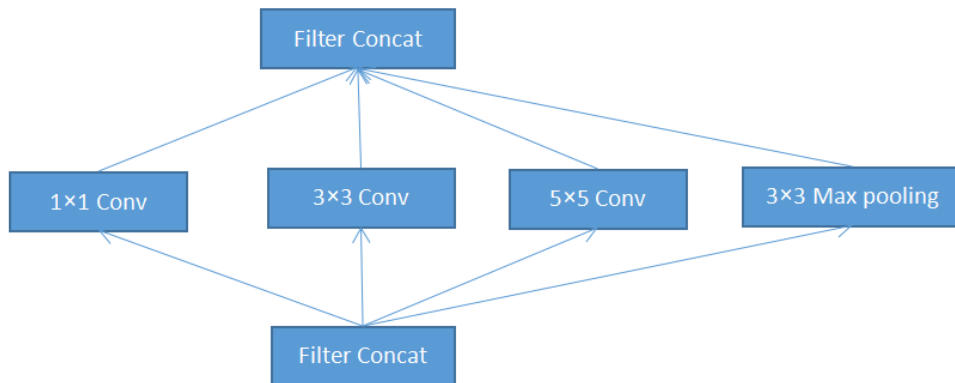


图 2-8 Inception 原始模块示意图

Inception 原始模板结构如上图所示，在收到上一层的信息后，通过三个由不同大小的卷积核组成的卷积层以及一个池化层对信息进行特征提取并连接在一起。但计算量任然比较大，因此有了 Inception-v1 模块。

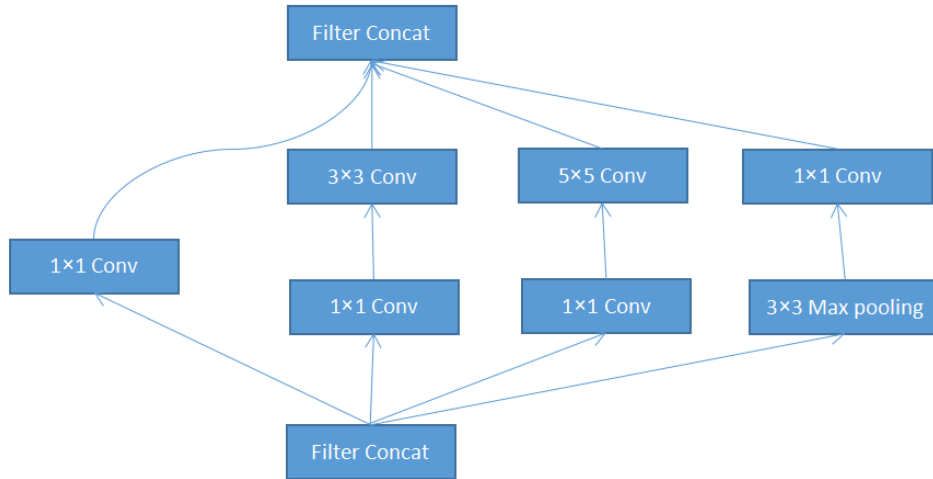


图 2-9 Inception-v1 模块示意图

相比较与原始模块，Inception-v1 模块在不改变感受野大小的前提下，使用了更多 1×1 卷积核，其作用是能获得更好的图像信息特征。传统的卷积层的输入数据只和一种尺寸的卷积核进行运算，而 Inception-v1 结构是 Network in Network(NIN)，即先进行一次普通的卷积运算(比如 5×5)，再依次经由激活函数(比如 ReLU)、 1×1 卷积运算、激活函数。 1×1 卷积实际上是对每个像素点，在不同的 channels 上进行线性组合(信息整合)，且保留了图片的原有平面结构，调控 depth，从而完成升维或降维的功能。

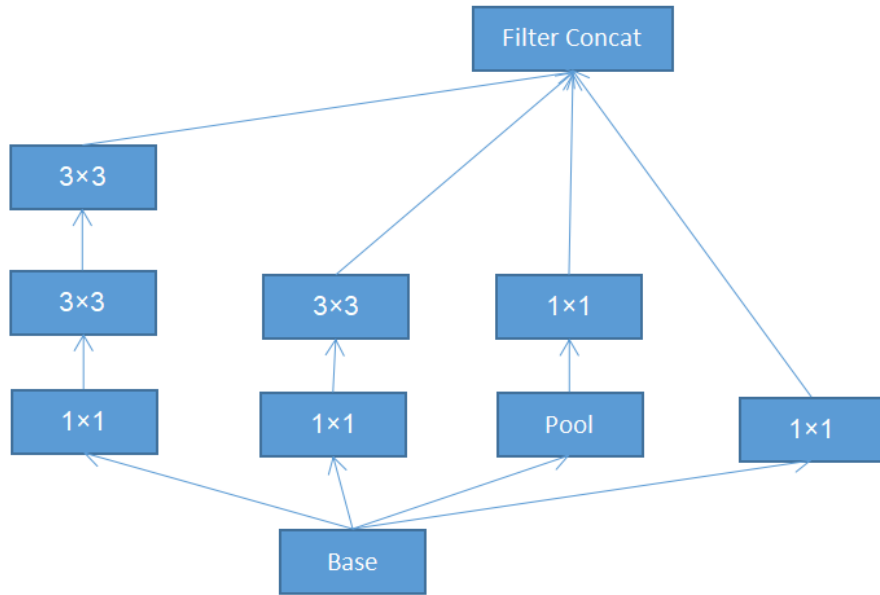


图 2-10 Inception-v2 模块结构图

Inception-v2 模型则是在 Inception-v1 模型的基础上参考 VGG，将 5×5 的卷积核替换成了两个小的 3×3 卷积核，这样可以减少模型的参数，降低计算量，增加函数的适应性，提高运行的速度。

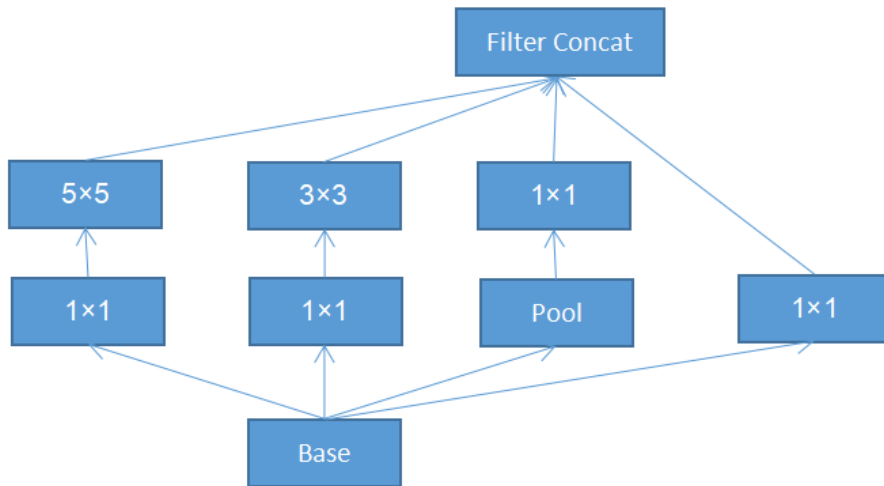


图 2-11 Inception-v3 模块结构图

Inception-v3: 针对 Inception-v2 的升级，增加了以下内容：(1). RMSProp 优化器。(2). 分解为 7×7 卷积。(3). 辅助分类 BatchNorm。(4). 标签平滑 (Label

Smoothing，添加到损失公式中的正则化组件类型，防止网络过于准确，防止过度拟合）。

第3章基于 Inception-Resnet-v2 的肺炎 CT 图像识别研究

3.1. 基于 Inception-Resnet-v2 网络

在新冠病毒爆发的初期，确诊是否患有新冠病毒需要照肺部 CT，然后由医生进行人工辨认，但医生的数量远远比不上病患的数量，这就导致每个医生的工作量都特别大，过大的工作量会导致医生产生疲惫感从而出现失误，所以采用深度学习技术为医生提供帮助的事件就变得尤为重要。

3.1.1. 卷积神经网络模型架构

本章节采用 Inception-Resnet-v2^[20]模型，该模型结合了 Inception 模块和 Resnet 模块，组成结构为：输入层、Stem层、Inception-resnet-A、Inception-resnet-B、Inception-resnet-C、Reduction-A、Reduction-B、Dropout 层以及最后用 Softmax 函数把结果转化为概率输出。Inception-resnet-A、Inception-resnet-B、Inception-resnet-C 三个模块没有本质区别，差异在于卷积核数量以及卷积核的尺寸。同时可以明确看出，上述三个模块都是采用多尺度融合的同时使用残差网络。其结构如图 3-1、3-2 所示；该模型具有以下优点：

- 引入残差模块，可以有效避免梯度消失的情况出现，同时可以提高计算的速度，能使模型更快达到收敛。
- 通过多种尺度的卷积核，能更有效的收集信息，且降低参数的数量。
- 卷积层和池化层的融合让模型具有稀疏性。

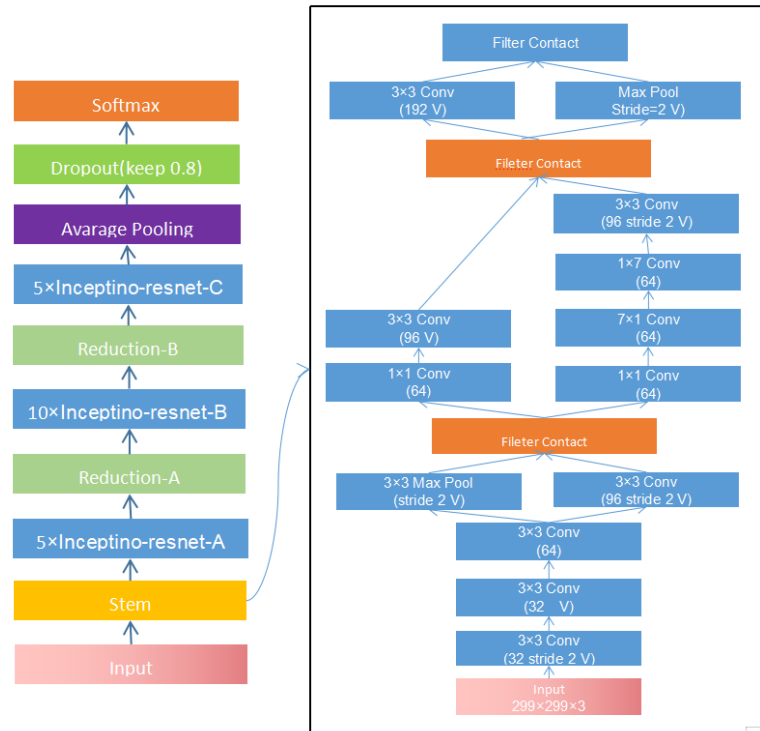
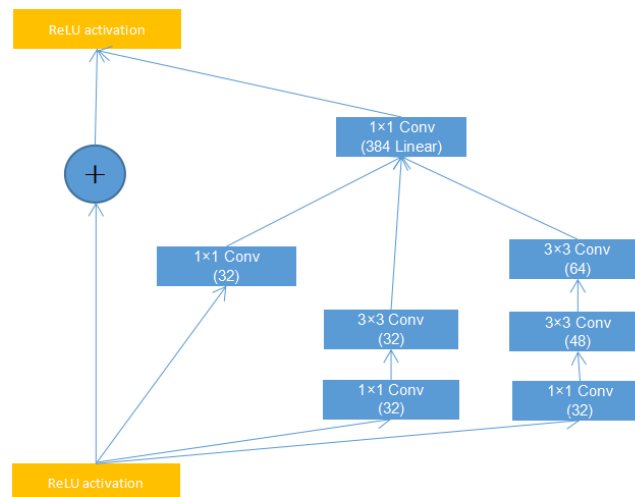


图 3-1 Inception-Resnet-v2 模块结构图

以下是 Inception-ResNet-A、Inception-ResNet-B、Inception-ResNet-C 三个模块的结构。



Inception-ResNet-A 结构图

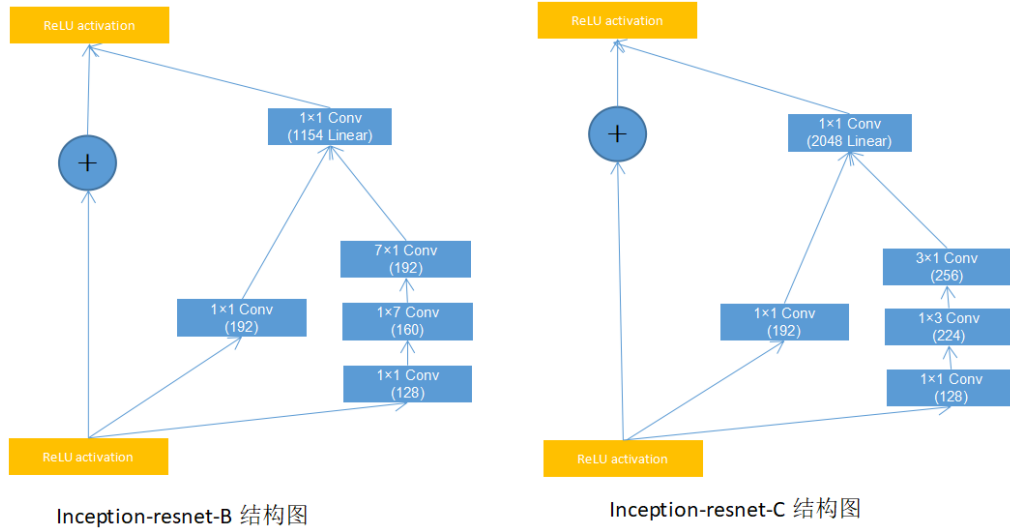


图 3-2 Inception-ResNet 各模块结构示意图

3.1.2. 卷积神经网络模型设计路线

本文的 Inception-ResNet 模块正如下图所示。

对于加入残差模块的 Inception 网络，本文使用较原始 Inception 模块更简易的模块，每个 Inception 模块之后是卷积扩展层（ 1×1 无激活函数的卷积核），作用是缩放过滤器矩阵的维度。这是用于补偿由于 Inception 模块引起的维度减少。整个网络结构的最后三层为全局平均池化、全连接层、和 Softmax 函数。整个网络模型运行流程主要有以下步骤：第一步：全局平均池化（Global pooling）会计算出每个特征图上所有像素值的均值，然后将所有计算出的均值组成一个特征向量 \vec{x} ；全连接层接受上一步的输出结果 \vec{x} ，并根据具体情况输出一个 n 维特征向量 \vec{y} 。（ n 表示类别的数目）。Softmax 函数可以将 \vec{y} 上的代表每一个分类的值转换属于这一类的概率。Softmax 函数公式如下：

$$y_k = \exp(a_k) / \sum_{i=1}^n \exp(a_i)$$

a_k 表示 \vec{y} 上第 k 个位置的数值，也可以理解为图像被识别为第 k 类的数值， y_k 表示被识别为第 k 类的概率。

在分类的图像处理问题中一般使用的是交叉熵作为损失函数^[21]，规范两个概

率分布 a 、 b ，让 b 来形容 a 的交叉熵为：

$$H(a,b) = -\sum a(x) \log b(x)$$

交叉熵满足概率分布函数：

$$\forall x a(X-x) \in [0,1] \text{ 且 } \sum a(X-x) = 1$$

任何事情发生的概率都在 0 到 1 之间，且始终概率和为 1。

3.2. 数据来源及预处理

表 3-1 表 肺部 CT 图像数据集

分类	正常肺部 CT 图像	普通肺炎 CT 图像	新冠肺炎 CT 图像
总计	3500	3500	3500
训练集	2500	2500	2500
验证集	500	500	500
测试集	500	500	500

本文数据来源于 GitHub 社区，图像画面模糊、图像大小不一、图像是否具有代表性等问题可能会影响模型学习的效率，为了规避这些问题。我们在收集图像数据时专门收集的那些图像清晰的“完美数据”。本数据集图片清晰，没有模糊区域，完整度较高。数据集总共有 194,922 张肺部 CT 图片，如表 3-1 所示，本次实验从正常肺部 CT 图像、普通肺炎 CT 图像及新冠肺炎 CT 图像集中各自随机抽取 2500 张组成总容量 7500 张的数据集作为训练集，1500 张作为验证集（各 500 张），1500 张作为测试集（各 500 张）。所有图像均是 png 格式，下图为三种类型肺部 CT 的原始图像。

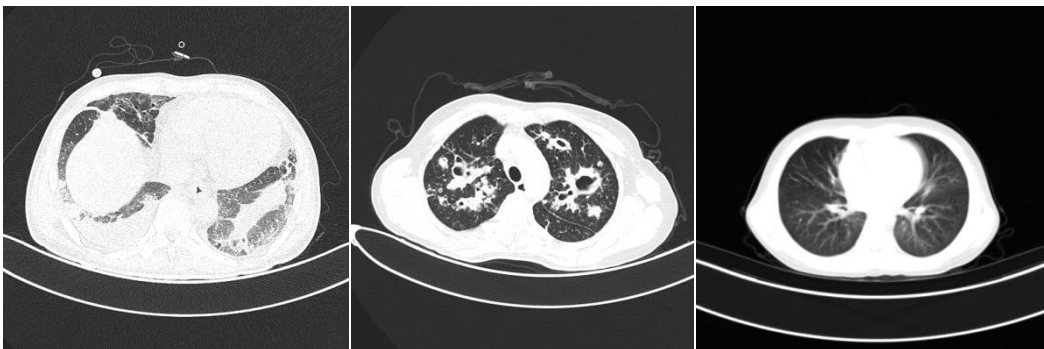


图 3-3 普通肺炎 CP



图 3-4 新冠肺炎 NCP

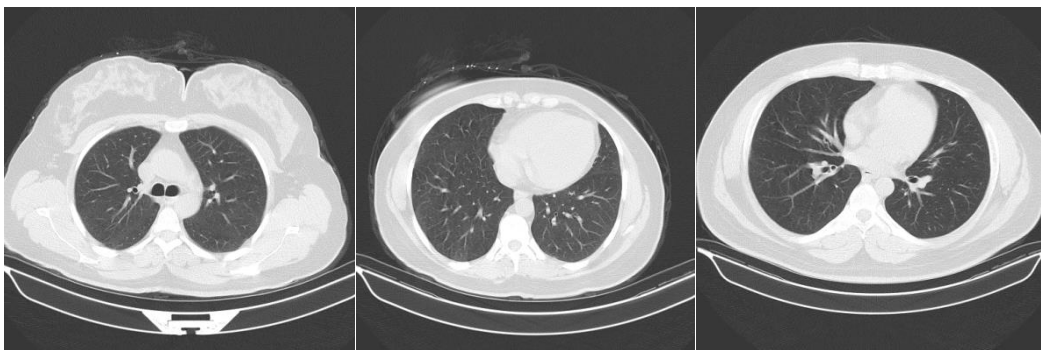


图 3-5 正常肺部 Normal

为了提高模型的准确性，尽量减少过拟合的发生，数据的预处理过程主要通过通过对图像进行数据增强^[22]的操作以达到更好提取特征的目的。本文采取的图像增强办法具体有：平移变化、缩放变换、翻转变化、角度旋转。角度旋转是根据旋转的角度来变换之前检测到的坐标，以及保存变换后的图像。图 3-6 为图像增强角度旋转增强示例。

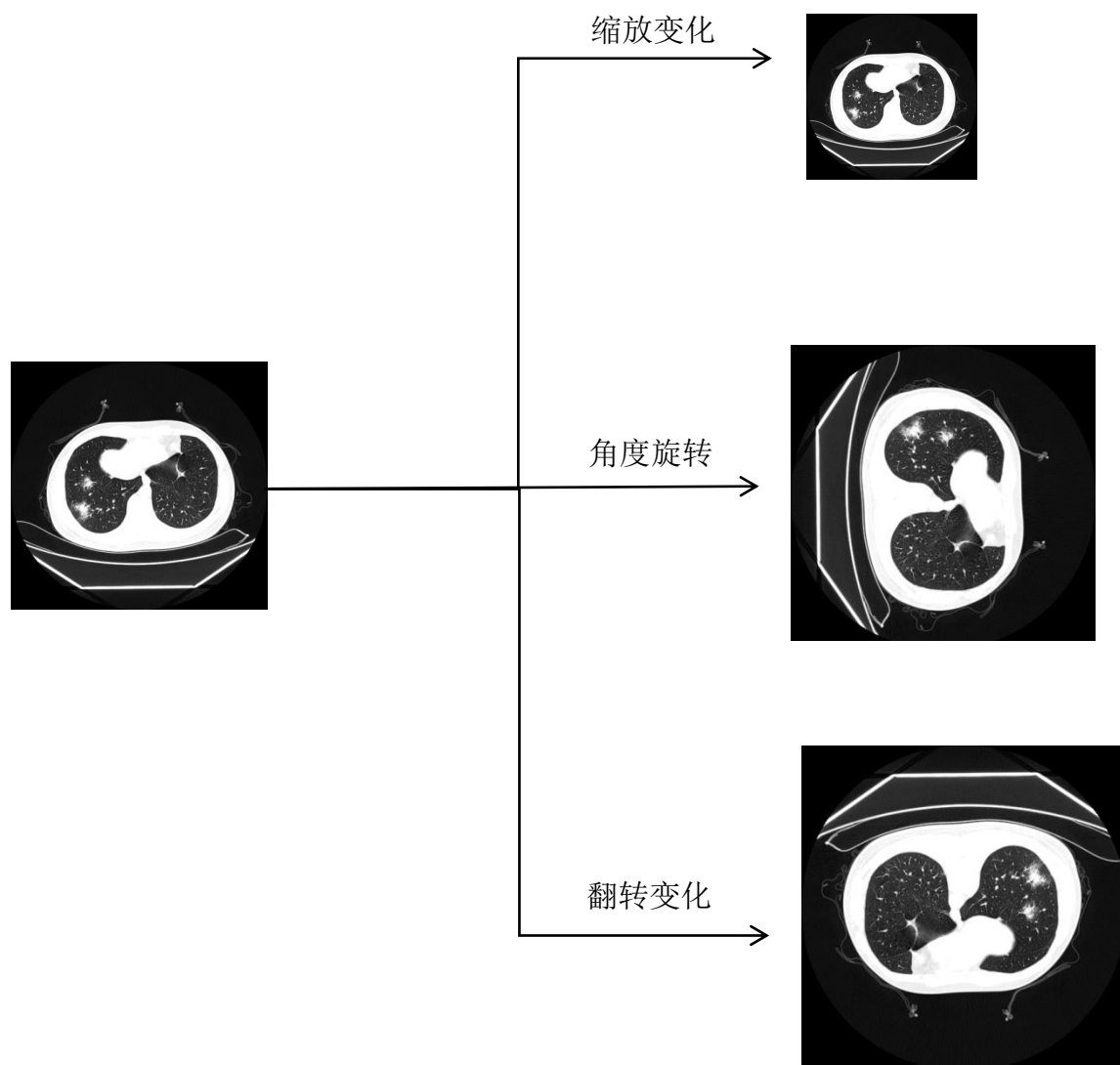


图 3-6 图像增强示例

3.3. 实验环境搭建

本章实验环境为：处理器为：Intel(R) Core(TM) i7-10870H；机带 RAM 为 16G；显卡为：NVIDIA GeForce RTX 3060；操作系统为 Windows11 家庭版；Python 版本为 3.7；

3.4. 实验过程

实验过程主要步骤如下：首先将把收集到图像数据进行预处理（旋转、放缩等）并丢入网络模型中。然后将经过训练集的模型重新投入验证集中，此步骤的目的是调整模型中的参数。最后将修正参数后的模型放入测试集中得到实验结果，并记录相关结果。本文的初始学习率为 0.001。

表 3-2 各模型收敛准确率

模型	epoch	train_loss	valid_loss	accuracy	accuracy_valid
GoogLeNet	12	0.2572	0.3077	0.9250	0.9113
ResNet50	14	0.0544	0.2461	0.9812	0.9260
Inception-Resnet	11	0.1260	0.1976	0.9610	0.9373

从表 3-2 看出，Inception-Resnet 网络达到收敛只需要 11 轮，测试集上的准确率达到 93.73%；GoogLeNet 网络达到收敛需要 12 轮，测试集上的准确率为 91.13%；ResNet50 网络达到收敛需要 14 轮，测试集上的准确率为 92.6%。Inception-Resnet 网络相较于其他两个模型它的收敛速度的确更快，同时在测试集上它的准确率也高于 GoogLeNet 以及 ResNet50。

3.5. 实验结果

通过上述的实验流程，我们得到了 ResNet50、GoogLeNet 以及 Inception-ResNet-V2 三个模型的实验结果，CP 代表普通肺炎患者，NCP 代表新冠病毒患者，Normal 代表正常人。TP: True Positive, 分类器预测结果为正样本，实际也为正样本，即正样本被正确识别的数量。FP: False Positive, 分类器预测结果为正样本，实际为负样本，即误报的负样本数量。TN: True Negative, 分类器预测结果为负样本，实际为负样本，即负样本被正确识别的数量。FN: False Negative, 分类器预测结果为负样本，实际为正样本，即漏报的正样本数量。Precision、Recall、Specificity 三个指标的计算公式如下：

$$\begin{cases} precision = TP / (TP + FP) \\ recall = TP / (TP + FN) \\ specificity = TN / (TN + FP) \end{cases}$$

以下是三种模型各自的实验结果：

表 3-3 ResNet50 三分类结果

	Precision	Recall	Specificity
CP	0.934	0.965	0.967
NCP	0.970	0.968	0.985
Normal	0.964	0.936	0.982

表 3-4 GoogLeNet 三分类结果

	Precision	Recall	Specificity
CP	0.944	0.969	0.985
NCP	0.970	0.971	0.985
Normal	0.970	0.943	0.984

表 3-5 Inception-resnet 三分类结

	Precision	Recall	Specificity
CP	0.956	0.979	0.990
NCP	0.978	0.980	0.990
Normal	0.980	0.955	0.989

为了验证模型的精确度等性能是否如上述理论般优秀。本文选取三大客观指标作为评判标准：准确率、灵敏度和特异性。对照模型为：ResNet50 和 GoogLeNet。本文虽然是三分类模型，但 TN、TP、FN、FP 各个指标的计算方式和二分类模型一样。三大指标的计算公式如下：

$$\begin{cases} ACC = (TP + TN) / (TP + TN + FP + FN) \\ SEN = TP / (TP + FN) \\ SPE = TN / (TN + FP) \end{cases}$$

根据公式得到每个网络模型的三大指标具体数值，Inception-ResNet-v2 网络模型的准确度为 93.7%、灵敏度为 96.2%、特异性为 98.2%。三大指标均高于 ResNet50 和 GoogLeNet 网络模型的指标值。当然，本文实验所选用的模型在本文数据集上的表现都很不错，这是由于本文所选的数据集画面清晰，没有模糊不

清等现象。

表 3-6 不同模型分类结果比较

	ACC	SEN	SPE
ResNet50	91.1	94.8	97.3
GoogLeNet	92.6	93.7	96.8
Inception-Resnet	93.7	96.2	98.2

3.6. 结果对比与分析

图 3-7、图 3-8、图 3-9 分别是 ResNet50、GoogLeNet、Inception-ResNet-v2 网络模型的混淆矩阵。根据结果可以明显得出结论：Inception-RseNet-v2 的性能的确是优于 ResNet50 和 GoogLeNet 网络模型。其对于正常人肺部 CT 的识别率能达到 98%，尽可能的保证了正常人被误诊这类情况的发生。因此，可以将之具体应用到新冠检测自动识别的领域，为我国乃至世界的防疫建设添砖加瓦。通过上述的结果可以得出以下结论：Inception-ResNet-V2 模型的确在保证准确度的同时，减少了网络中的参数数量，以保证节省相应资源。

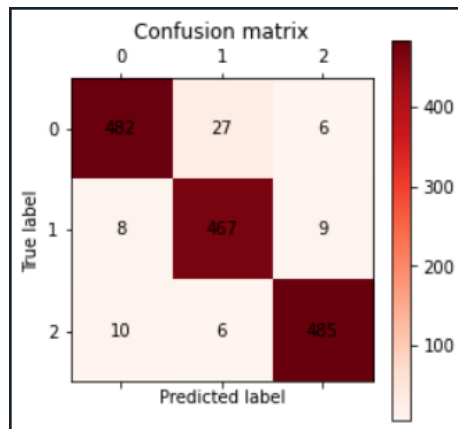


图 3-7 ResNet50 混淆矩阵

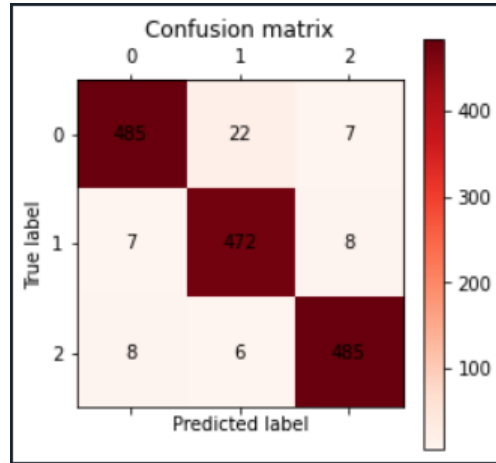


图 3-8 GoogLeNet 混淆矩阵

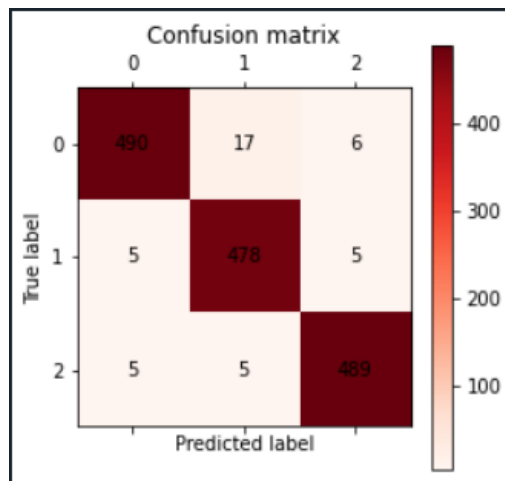


图 3-9 Inception-ResNet 混淆矩阵

3.7. 小结

本文首先介绍了 Inception-ResNet 网络的具体结构，并且分析了每个模块的优点，然后设计了相关损失函数和激活函数。将从 Github 社区上下载的数据集进行训练，不断进行优化模型参数。最终对收集到的实验数据进行分析，通过其较高的准确度证明该模型具体应用到肺炎 CT 图像分类任务上的可行性。但在本次研究中发现了一些不足：普通肺炎患者的识别度明显低于正常人和新冠病毒患者的识别度。究其原因是因为普通肺炎的病灶特征相对新冠病毒来说没有那么明显；而且本章的探索只能对图片进行分类，无法定位肺部病灶部位的精确位置，

就实用性而言会大打折扣。

因此，本文后续引出图像语义分割的模块，通过对图像进行语义分割后，有效去除多余图像部分，提高模型的精准度。为临床研究提供更加精准的医疗建议和图像辅助提示功能。

第4章基于 U-net 改进网络的肺炎 CT 图像语义分割研究

通过上一章节的图像分类研究，我们完成了对肺部 CT 图像进行分类的任务，但也指出了存在无法准确定位肺部具体病灶区域等问题。因此本章将通过图像语义分割解决以上问题。

4.1. 改进 U-Net 网络框架设计

4.1.1. 损失函数的改进

对于处理二值图像分割的问题，二值交叉熵（BCE）函数被广泛地用作损失函数。然而，当我们关注的目标的尺寸远远小于背景图像的尺寸时，使用二值交叉函数会导致偏差效应^[23]。因此，本文采用一种较为新颖的损失函数——BCE-SD 损失函数^[24]，其定义如下所示：

$$L_{BCE-SD} = \lambda_1 L_{BCE} + \lambda_2 L_{SD}$$

其中， λ_1 、 λ_2 分别是二值交叉熵 L_{BCE} 和形状距离 L_{SD} 的平衡系数。设 $y \in [0,1]$ 为预测图， $\hat{y} \in [0,1]$ 作为对应的标签图， P 为图像的像素值。二元交叉熵损失函数表示为：

$$L_{BCE} = -\frac{1}{P} \sum_{i=1}^P y \log y_i + (1-y) \log(1-y_i)$$

形状距离损失函数如下定义：

$$L_{SD} = \frac{1}{P} \sum_{i=1}^P (1-y_i)y_i + y_i(1-y_i)$$

4.1.2. 迁移学习和 EfficientNet 模块

卷积神经网络（ConvNets）一般情况下是在一个相对固定的环境中进行，环境中的资源总量保持不变。所以无法通过不断的增加算力等资源来获得更好的效果。要想增加网络的性能，通常是从网络深度(depth)、网络宽度(width)和输入图像分辨率(resolution)大小三个方面入手。但是具体对于每个方面投入多少的资源，这个问题通常无法通过人为调整来解决。。基于上述背景，本论文引用

了一种新的模型缩放方法——EfficientNet^[25]，如图 4.1 所示，它使用一个复合系数来从三个方面同时进行调整，而并非拘泥于传统的方式——随机系数改变网络的维度。基于神经结构搜索技术可以获得最优的一组参数（复合系数）。神经网络搜索是一种超参数优化算法，它可以用来调整神经网络的结构和参数，以获得更好的性能。它的原理是将超参数空间划分为一个网格，然后依次对每个超参数组合进行评估，以找到最佳的参数组合。

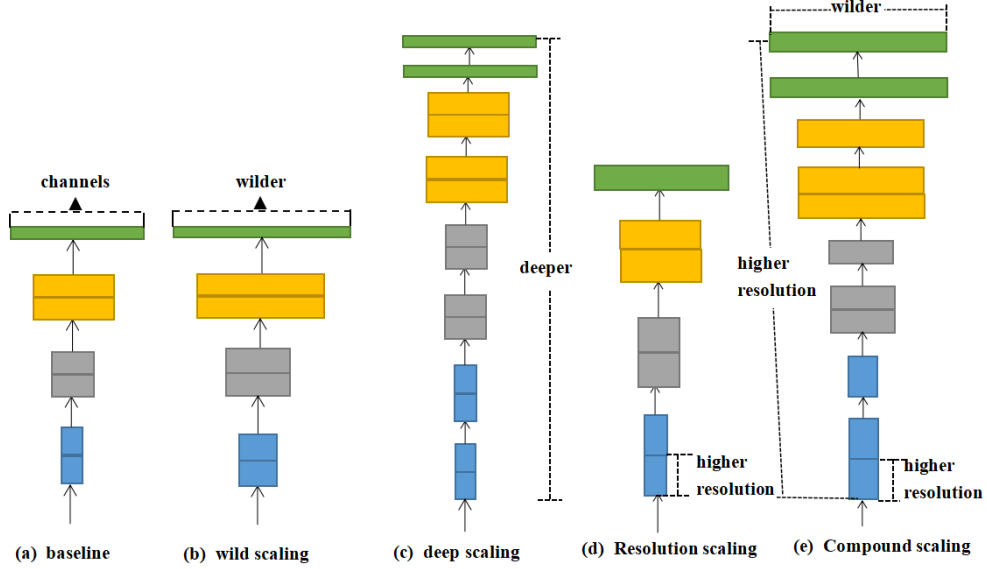


图 4-1 模型缩放复合方法

我们将一个卷积网络定义为如下式子：

$$N = \bigodot_{i=1 \dots s} F_i^{L_i}(X_{\langle H_i, W_i, C_i \rangle})$$

其中 N 为整个卷积网络， F_i 表示对第 i 层的卷积运算， $\langle H_i, W_i, C_i \rangle$ 表示第 i 层的张量。 $F_i^{L_i}$ 表示第 i 个 stage（通常会将多个结构相同的卷积层称为一个 stage），它由卷积层 F_i 重复 L_i 次构成。EfficientNet 的设想是一个卷积网络所有的卷积层必须通过相同的比例常数进行统一扩展，所以，优化目标就是在资源有限的情况下，要最大化 Accuracy，优化目标的公式表达如下：

$$\begin{aligned} & \max_{d, w, r} \text{Accuracy}(N(d, w, r)) \\ \text{s.t.} \quad & N(d, w, r) = \bigodot_{i=1 \dots s} \hat{F}_i^{d \cdot \hat{L}_i}(X_{\langle r \cdot \hat{H}_i, r \cdot \hat{W}_i, w \cdot \hat{C}_i \rangle}) \\ & \text{Memory}(N) \leq \text{target_memory} \end{aligned}$$

$$\text{FLOPS}(N) \leq \text{target_flops}$$

其中 d 、 w 和 r 分别表示网络深度、宽度和分辨率。该式表示为在给定计算内存和效率的约束下，通过优化参数 d 、 w 和 r 来实现最优的模型准确率。

EfficientNet 的规范化复合调参方法使用了一个复合系数 Φ 对三个参数进行复合调整：

$$\begin{cases} \text{depth: } d = \alpha^\phi \\ \text{width: } w = \beta^\phi \\ \text{resolution: } r = \gamma^\phi \\ s.t. \alpha \geq 1, \beta \geq 1, \gamma \geq 1 \\ \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \end{cases}$$

其中 α ， β ， γ 作为常数可以由小型的网格搜索来确定。以上就是 EfficientNet 的复合扩展方式。

4.1.3. 网络结构改进

为了提高 U-Net 的性能，本文对 U-Net 架构做了一些改进，提出适用于肺部分割的改进 U-Net 网络^[26]，其结构如图 4-2 所示。

对于编码器路径，我们利用迁移学习的思想将预训练的 effentnet 族 CNN 模型作为编码器块，特别是 EfficientNet-B4^[27]。这是因为训练一个包含大量参数的复杂模型，必须要有一个庞大的数据集。然而，收集大量的标记数据是一件较为麻烦的事。另一方面，深度学习模型一般情况是为了集中处理特定的任务。因此为了克服孤立学习的形式，迁移学习的概念被提出。它利用预训练模型得到信息，并使用它来解决新的问题，而且仅仅需要更少的数据。在这个想法的启发下，本文参考 Efficientnet 网络第一个模块设计了编码器模块，让我们在训练中使用预训练模型。更具体地说，编码器路径中的第一层包括卷积和参数标准化处理，参数批量归一化，五个被改进的移动反向瓶颈卷积（MBConv6）层。

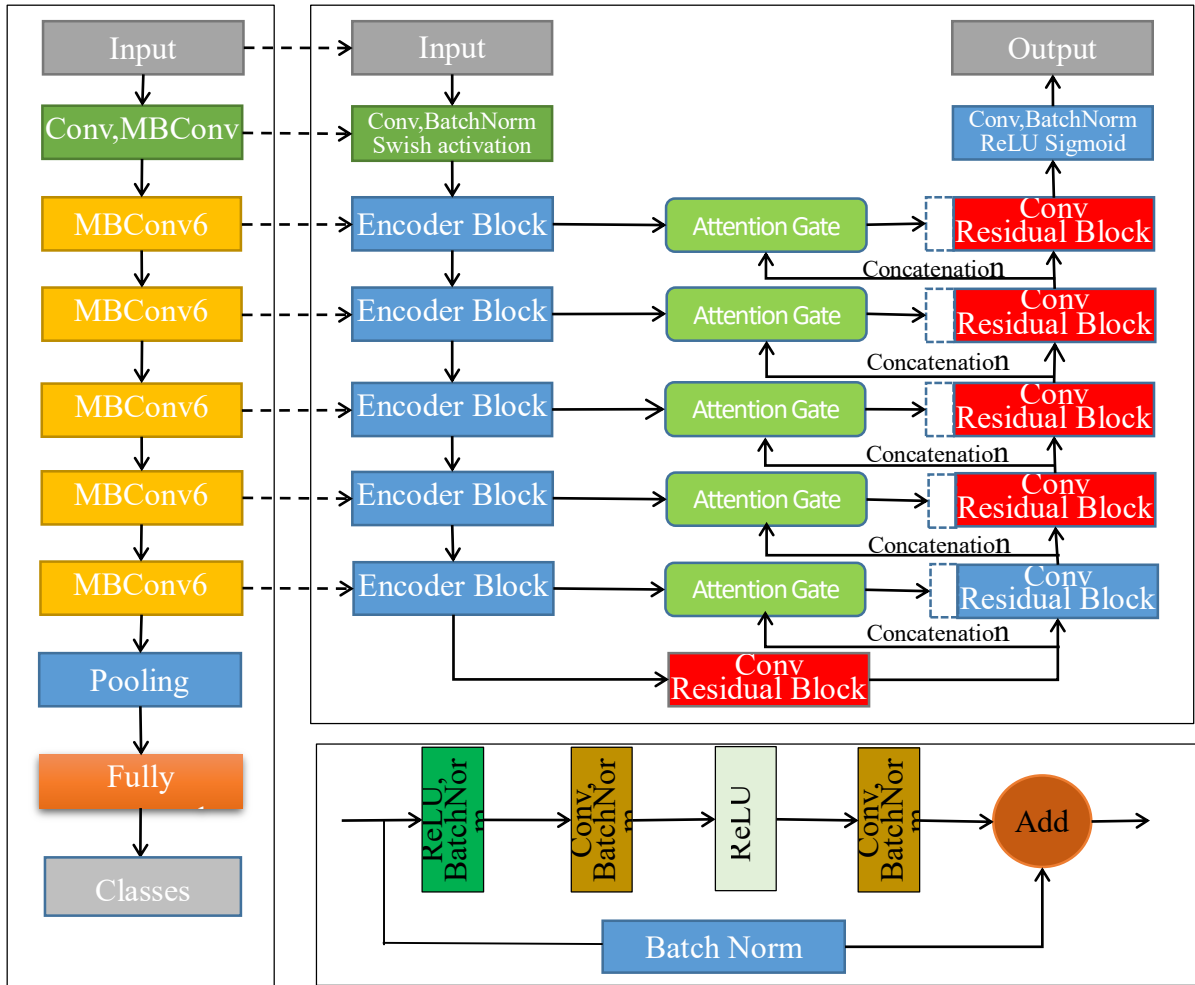


图 4-2 左图为移动效率网结构，右图为改进的 U-Net 网络，采用 MBConv6 作为编码器。

每个编码器由两个连续的标准化 3×3 卷积层组成。然后，一个 2×2 内核的最大池化层采用步长为 2 的方式对输入图像进行下采样，得到更大的图像信息和空间信息。上采样路径也同样使用编码器中的模块，除了在通过标准 3×3 的卷积层之前使来自下采样路径的类似分辨率的特征图与上采样过程中类似分辨率的特征图连接，还必须对特征图进行裁剪和对齐处理。跳跃连接用于获取在中间阶段里可能丢失的信息。上采样过程中每通过一个解码器，对应的特征图分辨率会提高 2 倍。解码器将编码器学习到的低分辨率特征投射到高分辨率的像素空间上，以获得密集的像素级分类。然而，由于下采样过程导致特征图显著减少，进而使得特征图的空间分辨率损失，边缘信息模糊化。为了克服这些问题，Ronneberger^[14]等人在 U-Net 中引入了跳跃式连接，允许解码器恢复各阶段学习到的编码过程中因池化操作损失掉的相关信息。

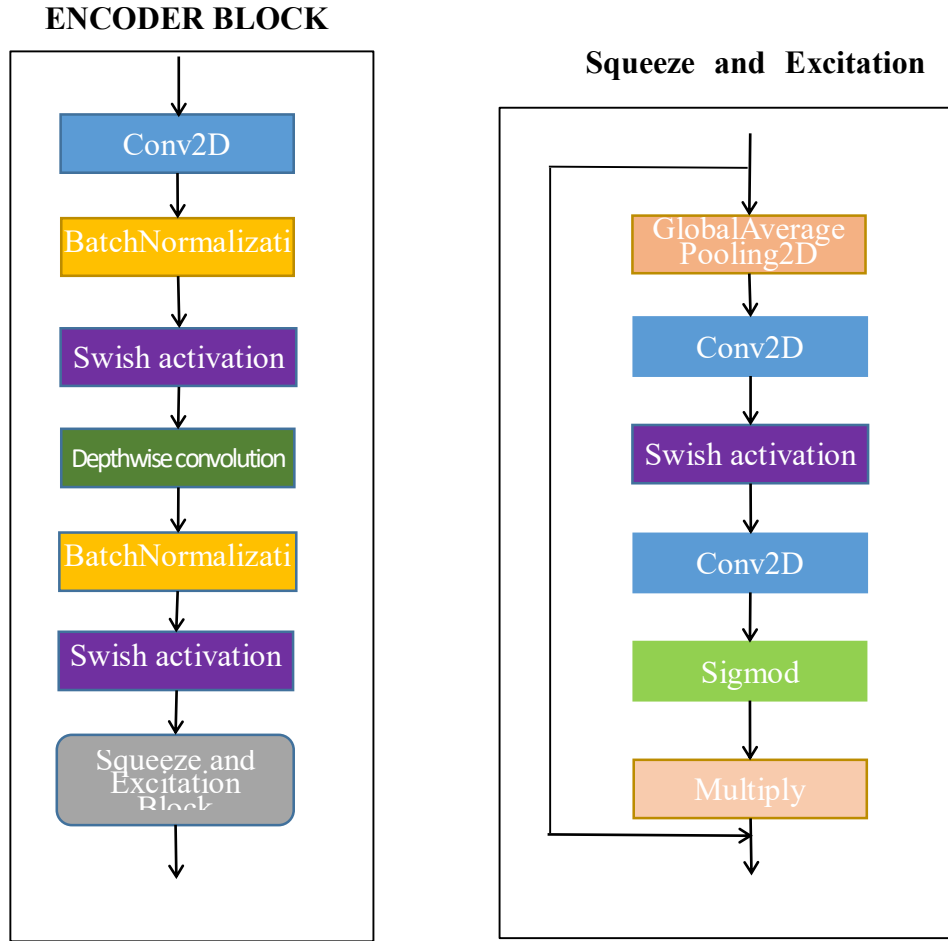


图 4-3 解码器模块结构

本文采用 MobileNetV2^[27] 版本的卷积模块为编码器。该模块结构如图 4-3 所示,在编码器块中,本文采用 Hu 等推导出的挤压-激发阶段提高网络的表示能力,使它能够执行动态通道的特征重新校准。同时,本文通过删除最后一个全局变量来修改 Effiientnet 的头部部分平均池化层和所有密集层,并添加卷积过滤器。然后是批处理规范化和 ReLU 激活,这是采用 ResNet 的残差模块设计。

对于该模型的解码器部分,在解码过程中的每一个步骤,都会对上一层输出执行一次上采样过程。在本研究中,解码器模块采用 ResNet 网络中的残差模块,旨在避免过拟合和梯度消失。

对于编码器与解码器的中心部分,该中心模块具有与解码器相同的结构,包括卷积层和残差模块。这部分负责调整编码器制作的特征信息以适应 U-Net 编码器和解码器之间的对称结构建筑,从而促进了浅层与深层的特性融合。在标准 U-Net 中,映射在压缩路径的对应特征通过跳跃连接被裁剪并复制到解码路径。

然后这些特征图被连接到上采样函数的输出。

4.1.4. Attention Gate 模块

注意力控制机制最早被引入 UNet 模型中，在传统的 UNet 中，为了避免在解码过程时丢失大量的空间精确细节信息，因此使用了 skip 的手法，直接将编码过程中提取的特征图直接连接到解码过程相对应的层。但是，提取的浅层特征图有很多的冗余信息。注意力机制的核心目的是在将大量信息传输给解码器之前，找到当前任务中最重要的信息，更具体来说，注意力机制通过抑制非相关信息的特征去提高网络的性能。注意力门（Attention Gate, AG）^[28]结构如图 4.4 所示。

令 x^l 为第 l 层输出的特征图， g 为控制门信号的矢量。 $\gamma \in [0,1]$ 为门系数。有以下关系式：

$$x_{out}^l = x^l \gamma$$

注意力门的特征包括一组线性变化在内的参数 W_x, W_g, ψ 和截距项 b_g, b_ψ 。注意力参数满足以下关系式^[29]：

$$\gamma_i = f_2(\psi^T (f_1(W_x^T x^l + W_g^T g_i + b_g)) + b_\psi)$$

其中 f_2 是 Sigmoid 函数， f_1 是 ReLU 函数。它们分别如下定义：

$$f_2(x) = \frac{1}{1 + e^{-x}}, f_1(x) = \max(0, x)$$

对于一张输入为 $1 \times 3 \times 512 \times 512$ （1 是 batchsize，3 是通道）的 2D 图，执行到 x_5 的时候（经过五次下采样）已经是最小的特征图了（ $1 \times 1024 \times 32 \times 32$ ）。对其进行上采样得到 d_5 （ $1 \times 512 \times 64 \times 64$ ）。图 4-4 中的 x^l 为 x_4 ， g 代表 d_5 。

表 4.1 注意力机制执行步骤

注意力机制执行步骤
第一步：对 g 做 1×1 卷积得到 $1 \times 256 \times 64 \times 64$ 特征图
第二步：对 x^l 做 1×1 卷积得到 $1 \times 256 \times 64 \times 64$ 特征图
第三步：为了突出特征，将第一步和第二步的结果相加
第四步：对第三步结果进行 ReLU 函数激活
第五步：对第四步结果做 Conv (256, 1) 卷积, 将 256 通道降到 1 通道。得到 $1 \times 1 \times 64 \times 64$ 的特征图
第六步：对第 5 步结果进行 sigmoid, 使得值落在 $(0, 1)$ 区间, 值越大, 越是重点。这个该值也就是注意力权重。
第七步：对图片使用 resampler 使得图片大小一样
第八步：最后和 x^l 相乘, 把注意力权重赋到 low-level feature 中。

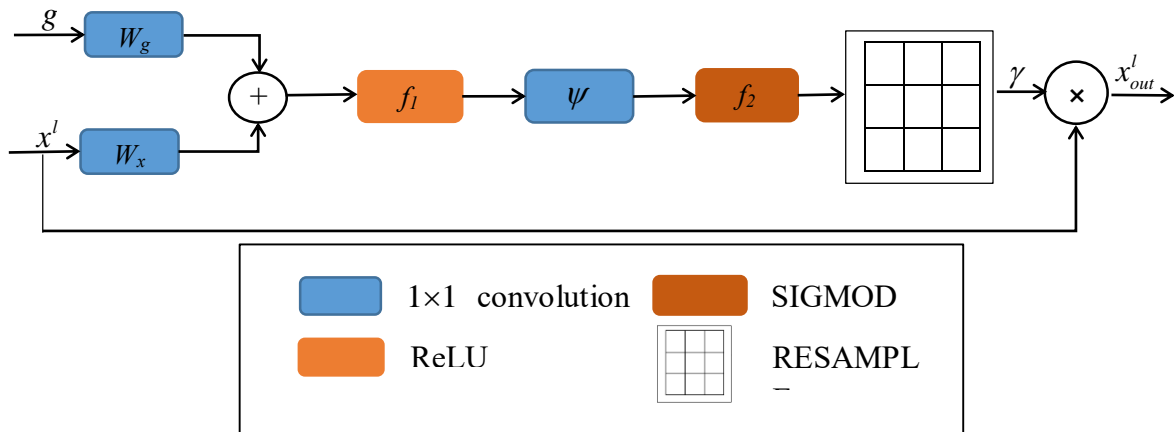


图 4-4 注意力门结构

4.1.5. 损失函数优化

与其他深度神经网络一样，FNS 通常是基于随机梯度下降 (SGD) 损失函数^[30]优化变量与衰减的学习率相结合，直到收敛。然而，传统的基于 SGD 的学习方案遇到训练数据小，不平衡，或复杂的情况时会出现过拟合。因此我们采用随机加权平均 (SWA) 方法而不是传统的 SGD 优化。SWA 的核心思想是对沿着 SGD 轨迹的多个点去平均值，具有周期性和恒定的学习率。通过这种方式，它是的以更低的损失获得全局解决方案，因此会获得优于常规的 SGD 训练方案^[31]。

采用 SWA 循环学习和余弦退火学习率下降方式来训练模型，学习率^[32]计算公式如下：

$$a(t) = \frac{a_0}{2} (\cos(\frac{\pi \bmod(t-1, [T/M])}{[T/M]}) + 1)$$

a_0 表示最初的学习率， t 表示训练到第几轮， T 表示总训练轮数， M 表示总共循环的次数。训练集和测试集图像大小都调整为 $256 \times 256 \times 3$ 像素。利用 BCE-SD 损失函数对模型进行优化。损失函数中的参数设为 $\lambda_1=1; \lambda_2=1$ ，在训练开始的时候的内核初始化来自 ImageNet 的预训练权值编码器。

4.2. 实验数据与环境

4.2.1. 实验数据来源

本章数据下载自 GitHub 社区，共计 7500 张肺部 CT 医学图像，正常肺部 CT、普通肺炎患者肺部 CT、新冠病毒患者肺部 CT 各 2500 张。每个类的数据集中 1600 张用于训练模型，400 张用于验证集进行参数调整，500 张用于测试集得出最终结果。

表 4.2 肺部 CT 图像数据集

分类	正常 CT 图像	普通肺炎 CT 图像	新冠肺炎 CT 图像
总计	2500	2500	2500
训练集	1600	1600	1600
验证集	400	400	400
测试集	500	500	500

4.2.2. 实验运行环境

本章实验环境为：处理器为：Intel(R) Core(TM) i7-10870H；机带 RAM 为 16G；显卡为：NVIDIA GeForce RTX 3060；操作系统为 Windows11 家庭版；Python 版本 3.7，TensorFlow 学习框架 2.2.0，Keras 版本 2.3.1。

4.2.3. 评价标准

为了具体衡量改进版 U-Net 网络模型与 CNN、FCN 网络模型在图像分割任务中的具体性能差异。本文选取了三个评判指标^[33]，分别是 Dice 相似系数（Dice

Similarity Coefficient, DSC)、Jaccard 相似系数 (Jaccard Similarity Coefficient)、CCR 正确分类率 (Correct Classification Rate)。Dice 相似系数是一种集合相似度量指标, 其计算公式如下:

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \Leftrightarrow Dice = \frac{2TP}{FP + 2TP + FN}$$

A 表示通过计算机算法自动分割的结果, B 表示图像正确的分割结果。Dice 系数的数值范围在 0 到 1 之间。数值为 1 表示分割结果最好, 即算法分割结果和正确的分割结果完全一样。数值为 0 表示结果最差, 即算法分割结果和正确的分割结果完全不一样。

Jaccard 相似系数主要用于计算符号度量或布尔值度量的个体间的相似度。给定两个集合 A, B, Jaccard 系数定义为 A 与 B 交集的大小与 A 与 B 并集的大小的比值, 当集合 A, B 都为空时, Jaccard 系数定义为 1。

$$Jaccard = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

正确分辨率的公式如下所示:

$$CCR = \frac{n}{N}$$

其中, 分割正确的像素数目用 n 表示, 图像像素总数用 N 表示。本文使用上述三项客观参数对模型进行评估。

4.3. 实验结果与分析

肺部 CT 图像分割多用于病情初期的诊断阶段, 主要是辅助医生确定病灶的部位, 以便后续观察病灶形状、颜色等特征。输入原始肺部 CT 图像后, 改进版 U-Net 网络模型会将图像进行像素级别的分类, 可以理解为把整张图像划分出很多个小片段, 判断小片段里的像素内容是属于哪一类标签。然后根据实验需求输出相应的分割结果。训练神经网络的时候通常会使用梯度下降优化权重。在每一次迭代中使用反向传播计算损失函数对每一个权重的导数, 并从当前权重减去导数和学习率的乘积。学习率决定了更新权重 (参数) 值的快慢。我们在本次实验中发现, 随着迭代次数一轮轮的增加, 模型的学习率变化幅度逐渐减小, 直至小于我们预先设定的阈值, 模型达到收敛状态^[34]。

从表 4.3 可以看出, 改进版 U-Net 网络的 Dice 相似系数为 0.951、Jaccard 相似系数为 0.935, CCR 正确分类率为 0.965。CNN 模型的三项指标依次是 0.937、0.866、0.945。FCN 模型的三项指标依次是 0.943、0.898、0.949。改进版 U-Net 网络的各项性能都是由于 CNN 和 FCN 两个模型的。

结合图 4-5 所展示的分割结果, 我们可以得出以下结论: 改进版 U-Net 网络模型能较好的完成肺部 CT 图像的语义分割任务, 能精准进行肺部病灶区域的目标定位, 分割后的结果图像清晰, 边缘未见模糊现象, 分类结果准确度较高。因此, 该模型具有较好的分割性能。

表 4.3 不同分割算法结果比较

算法	Dice	Jaccard	Precision
CNN	0.937	0.866	0.945
FCN	0.943	0.898	0.949
改进版 UNet	0.951	0.935	0.965

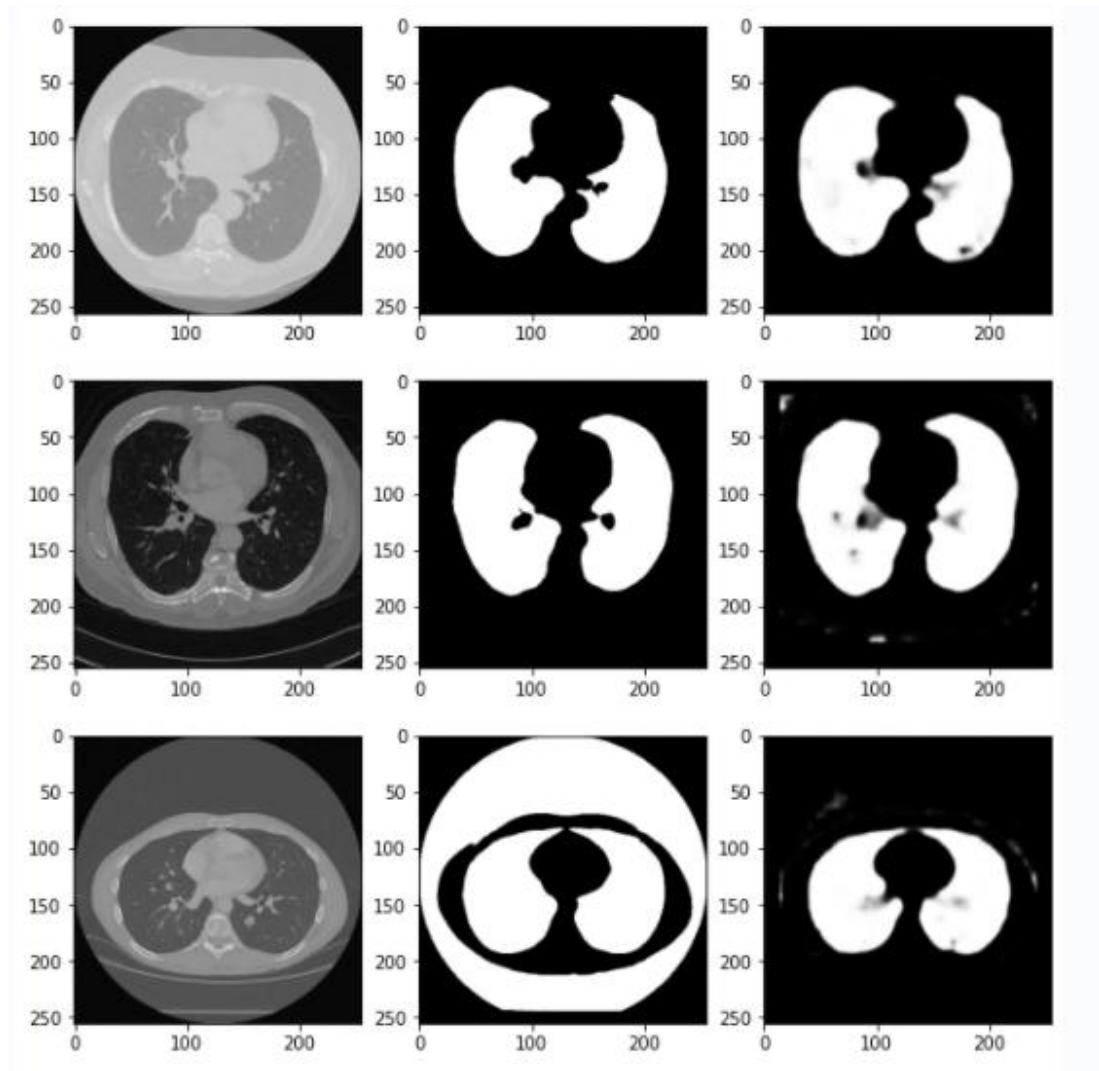


图 4-5 分割结果展示

4. 4. 小结

本章应用一种基于改进 U-Net 的肺部分割方法设计思路框架并进行实验流程及实验结果的描述,对该 U-Net 分割网络模型的分割和优化结果从 Dice 系数, Jaccard 系数和 CCR 系数三项指标来进行定量分析。结果表明,本文模型在 U-Net 替换编码器、增加注意力门机制后,肺部 CT 图像分割效果有明显的提升。通过对照实验可以发现,与 CNN 和 FCN 算法相比,改进 U-Net 网络的正确分类率远远大于前两者。从实验分割结果也可以直观地看出分割效果很好。改进 U-Net 网络模型能够准确的找出肺部病灶区域的准确位置,很好地解决第三章图像分类任务中存在的问题,增加了将卷积神经网络技术应用在新冠病毒人工智能检测上的实

用性。

第5章结论与展望

5.1. 本文工作总结

随着实际情况的不断变化，我国防疫相关政策也做出了改变。由全城静默、动态清零到全面放开。在全民经历了一波病毒的洗礼后，当今社会看似歌舞升平，新冠病毒已经离我们而去。但实际上，新冠病毒一直在变异，我们只是获得了新冠病毒亚变异体中某些分支病毒的抗体。人们仍旧面临着再次被感染的风险。鉴于抗原试剂检验的滞后性，对于带有基础病以及老年人来说，及早发现及早治疗是保证个人生命安全的最优解。而我国老年化人口众多，医疗系统应对起来较为吃力，为了提高医生的工作效率，释放医疗系统的部分压力，通过人工智能对新冠病毒进行诊断是非常有必要的。本文采用 Inception-Resnet 网络架构和改进 U-Net 网络架构建立模型，通过实验数据验证了模型的性能。本文的总结如下：

（1）研究了卷积神经网络中各种模型，分析了它们的核心特点、实现方式、本身优缺点。通过归纳总结，形成了在基于深度学习视角下对于图像识别这一部分较为完整的论述。

（2）本文在图像分类部分选取 Inception-Resnet 网络，该网络模型结合了 GoogLeNet 以及 ResNet 两个模型的优点，因此在数据集上各项结果都优于后两者。但本质上来说，该网络并没有从底层逻辑上创新。

（3）本文在图像分割方面选取的改进 U-Net 模型，核心思想是想从编码器、损失函数等方面对模型进一定的改进，迁移学习的思想能使得我们能在固定计算硬件等条件下更好的提高模型的准确率。注意力机制的效果也是类似，由于浅层特征图中提取了多余且冗长的信息，因此通过标注重要信息的方式提高模型的性能。

5.2. 未来研究方向

本文虽然最终完成了新冠病毒肺炎识别系统的搭建，但存在以下问题需要完善：

（1）本文实验结果较好，其可能原因是本文选取的图像都较为清晰，即数据本身较好。在实际应用过程由于现实各种因素，可能得到的图像不会这么清晰，

其结果较实验结果可能会略有差别，未来可通过使用更大、更全的数据作为样本对模型进行一定调参。

（2）由于新冠病毒不停的变异，未来新的毒株会造成什么样的影响无法预测，新的毒株对于肺部的影响也可能和之前的毒株造成的 CT 图像不同，该系统存在无法识别新的变异毒株引起的病症的可能性。因此未来可通过深度学习继续发展，创造出性能更好的模型解决此问题。

（3）本文最终完成的新冠病毒 CT 图像识别模型还未曾应用到手机 APP 程序上，这样在操作上不够便利。未来应该考虑尽量减轻模型的量级，以便能适用在手机 APP 上。

参考文献

- [1] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [2] Ciresan D, Giusti A, Gambardella L, et al. Deep neural networks segment neuronal membranes in electron microscopy images[J]. Advances in neural information processing systems, 2012, 25.
- [3] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241.
- [4] Lessmann N, Sánchez C I, Beenen L, et al. Automated assessment of CO-RADS and chest CT severity scores in patients with suspected COVID-19 using artificial intelligence[J]. Radiology, 2020.
- [5] Zheng Y Y, Ma Y T, Zhang J Y, et al. COVID-19 and the cardiovascular system[J]. Nature reviews cardiology, 2020, 17(5): 259-260.
- [6] Bai H X, Wang R, Xiong Z, et al. Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT[J]. Radiology, 2020, 296(3): E156-E165.
- [7] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C]//International conference on machine learning. PMLR, 2019: 6105-6114.
- [8] Wang S H, Govindaraj V V, Górriz J M, et al. Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network [J]. Information Fusion, 2021, 67: 208-229.
- [9] Hua K L, Hsu C H, Hidayati S C, et al. Computer-aided classification of lung nodules on computed tomography images via deep learning technique[J]. OncoTargets and therapy, 2015: 2015-2022.
- [10] David C, Nöhammer B, Solak H H, et al. Differential x-ray phase contrast imaging using a shearing interferometer[J]. Applied physics letters, 2002, 81(17): 3287-3289.
- [11] Martinez-Murcia F J, Górriz J M, Ramírez J, et al. Convolutional neural networks for neuroimaging in Parkinson's disease: Is preprocessing needed?[J]. International journal of

neural systems, 2018, 28(10): 1850035.

[12]Avci D, Leblebicioglu M K, Poyraz M, et al. A new method based on adaptive discrete wavelet entropy energy and neural network classifier (ADWEENN) for recognition of urine cells from microscopic images independent of rotation and scaling[J]. Journal of medical systems, 2014, 38: 1-9.

[13]Wei G, Li G, Zhao J, et al. Development of a LeNet-5 gas identification CNN structure for electronic noses[J]. Sensors, 2019, 19(1): 217.

[14]Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[15]Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[16]Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.

[17]Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241.

[18]He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[19]Falk T, Mai D, Bensch R, et al. U-Net: deep learning for cell counting, detection, and morphometry[J]. Nature methods, 2019, 16(1): 67-70.

[20]Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning[C]//Proceedings of the AAAI conference on artificial intelligence. 2017, 31(1).

[21]Kim B C , Choi J S , Suk H I . Multi-Scale Gradual Integration CNN for False Positive Reduction in Pulmonary Nodule Detection[J].2018,24(7):10581

[22]康牧. 图像处理中几个关键算法的研究[D].西安电子科技大学,2009.

[23]Hasana MK, Dahala L, Samarakoon PN, Tushara F, Martía R. DSNet: Automatic dermoscopic skin lesion segmentation. IEEE Int Conf Comput Adv Bio Med Sci 2020:103738.

InPress.

- [24]Tran TT, Pham VT, Shyu KK. Image segmentation using fuzzy energy-based active contour with shape prior. *J Vis Commun Image Represent* 2014;25:1732–45.
- [25]Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks [C]//International conference on machine learning. PMLR, 2019: 6105-6114.
- [26]Xu X, Wen Y, Zhao L, et al. CAREs-UNet: Content-aware residual UNet for lesion segmentation of COVID-19 from chest CT images[J]. *Medical Physics*, 2021, 48(11): 7127-7140.
- [27]Hu J, Shen L, Sun G. Squeeze-and-Excitation networks. *Proc Conf Comput Vis Pattern Recognit (CVPR)* 2018:7132–41.
- [28]Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention UNet: learning where to look for the pancreas. *Proc 1st Conf Med Imaging with Deep Learning* 2018. p. Available
- [29]Pham V T, Tran T T, Wang P C, et al. EAR-UNet: A deep learning-based approach for segmentation of tympanic membranes from otoscopic images[J]. *Artificial Intelligence in Medicine*, 2021, 115: 102065.
- [30]王驰. 面向随机森林与随机梯度下降的分解学习算法研究[D].哈尔滨工业大学,2021.DOI: 10.27061/d.cnki.ghgdu.2021.001948.
- [31]Izmailov P, Podoprikin D, Garipov T, Vetrov D, Wilson A. Averaging weights leads to wider optima and better generalization. *34th Conference on Uncertainty in Artificial Intelligence* 2018:876–85.
- [32]Tang P, Liang Q, Yan X, Xiang S, Sun W, Zhang D, et al. Efficient skin lesion segmentation using separable-Unet with stochastic weight averaging. *Comput Methods Programs Biomed* 2019;178:289–301.
- [33]吴云峰. 基于深度学习的肺炎医学CT图像分类算法研究[D].福建中医药大学, 2021. DOI: 10.27021/d.cnki.gfjzc.2021.000011.
- [34]代宇涵. 基于深度学习的肺部CT辅助诊断算法研究[D].重庆邮电大学, 2021. DOI:10.270675/d.cnki.gcydx.2021.000334.

致谢

这三年时光匆匆流逝，如白驹过隙般，回首昨日仿佛还在 2020 年 9 月的那个夏天。仔细回想这段记忆，它的确是我想要拥有的，也无比庆幸当年自己坚持考研的那个决定。不知不觉已经到了写致谢的时候，直到落笔这一刻心中思绪万千而又不知所措，想写些东西出来但又不知道该从何说起。就闲谈一些自己这三年时光的经历、感受。

首先很感激我的导师 XX 老师。每个老师可以教给学生的东西很多，XX 老师并没有逼着我们去读博，去走学术研究这条路。而是告诉我，自己想要做什么，那就去做，朝着自己的目标前进。正是 XX 老师给我的这种支持，让我能顺利找到一份满意的工作。研二的时候参加数学建模竞赛，跟着师兄、学姐一起学到了很多，而这个宝贵的机会也恰恰是 XX 老师给我的。此外感谢 XXX 院长允许我参加他的学术讨论班，让我的见识有了更进一步提高。

感谢我的父母能支持我读研，他们披星戴月，日夜奔波，让我能在如此安逸、舒适的环境下度过我这充实的三年求学时光。感谢我的室友梁瀚文、邹欣卫、潘渊，隔壁寝室的唐弋超和戎思宇以及像半个室友一样的黄治国，你们在日常生活和学习中给予了我太多太多的帮助。感谢同组的冀伟、李孟航、孙佳丽、于群，和你们相处的时光总是让人放松、惬意。

最后，我非常感谢我的女朋友龙家英，其实要说感谢或许也有点不当，请原谅我的辞藻不够描述此刻的心情。很庆幸和她相识、相恋。我们携手走过春华，抓过夏蝉，采过秋实，赏过冬雪。我希望未来我的一生波澜壮阔也好，风平浪静也罢，只要携手一生的人是她就足矣。

研究生这三年对我来说非常宝贵，但已经走到了尾声。新的征程即将开启，愿所有我爱的人和爱我的人，大家都得偿所愿，幸福安康。