

学号：20123101

# 《机器学习基础》 课程论文

基于 CNN 和 ResNet 系列模型的 X 光影像肺炎分类

姓 名	李昀哲
学 号	20123101
论文评分	

2023 年 3 月 1 日

# 目 录

基于 CNN 和 ResNet 系列模型的 X 光影像肺炎分类.....	2
1、引言.....	2
2、分类网络构建.....	3
2.1 CNN 卷积神经网络.....	3
2.2 ResNet 预训练模型.....	5
2.3 数据增强.....	7
2.4 优化器的选择.....	8
3、数据实验.....	9
3.1 数据集描述.....	9
3.2 数据处理.....	10
3.3 对比实验.....	11
3.4 超参数设置体会.....	16
4、结论与展望.....	17

# 基于 CNN 和 ResNet 系列模型的 X 光影像肺炎分类

[摘要]针对 X 光影像由医学专家人工识别肺部炎症容易造成误诊的问题,本文对数据集中正常、细菌性肺炎、病毒性肺炎三类别的影像图片进行考察。充分对比传统机器学习算法、CNN 卷积神经网络和深度残差网络 ResNet 系列模型。实验表明,ResNet 凭借其泛化性能强、分类能力强的特点,较好地应用于“肺炎/正常”,“细菌性肺炎/病毒性肺炎/正常”的分类问题;CNN 卷积神经网络对于“肺炎/正常”二分类问题也有较高准确度。

[关键字]X 光肺部影像分类,卷积神经网络,深度残差网络

## Classification of pneumonia in X-ray images based on CNN and ResNet series models

**Abstract:** In view of the problem that lung inflammation can be easily misdiagnosed by medical experts in X-ray images, this paper investigates the images of normal pneumonia, bacterial pneumonia and viral pneumonia in the data set. The traditional machine learning algorithm, CNN convolutional neural network and deep residual network ResNet series models are fully compared. The results showed that ResNet could be applied to the classification of "pneumonia/normal" and "bacterial pneumonia/viral pneumonia/normal" by virtue of its strong generalization and classification ability. The CNN convolutional neural network also has high accuracy for the "pneumonia/normal" binary classification problem.

**Key words:** X-ray pulmonary classification; Convolutional Neural Network; Deep Residual Neural Network

## 1、引言

肺炎是一种肺部的炎症性疾病,主要影响被称为肺泡的小气囊。典型症状包括咳嗽或干咳、胸痛、发烧和呼吸困难。病情的严重程度是可变的。肺炎通常由病毒或细菌感染引起,少数由其他微生物、某些药物或自身免疫性疾病等疾病引起。风险因素包括囊性纤维化、慢性阻塞性肺病(COPD)、哮喘、糖尿病、心力衰竭、吸烟史、中风后咳嗽能力差以及免疫系统弱。诊断通常基于症状和体检。胸部 x 光片、血液检查和痰培养有助于确诊。该疾病可根据感染地进行分类,如社区或医院获得性肺炎或卫生保健相关肺炎。

肺炎主要分为细菌感染和病毒感染,确诊往往根据 x 光胸片影像,通过炎症部位和阴影密度进行判别。目前,大量的胸片影像检查工作由医学专家靠经验进行判断,这个过程需要耗费大量精力且对于医学专家的经验知识有较高要求,误

诊情况时有发生。因此，随着计算机技术的发展，基于图像处理和机器学习技术，肺炎的确诊逐步由计算机完成。

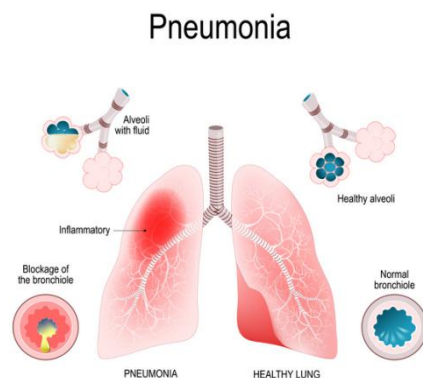


图 1. 肺炎示意图

国内外已经做了大量研究：基于图像处理技术的方法通过提取肺炎纹理 LBP 特征作为识别特征，用 BP 神经网络识别<sup>[1]</sup>，但识别率太低，BP 神经网络结构的选择到目前为止仍无统一的设计方案；方向梯度直方图结合朴素贝叶斯方法对肺炎类型识别<sup>[2]</sup>，通过计算和统计图像局部区域的梯度方向直方图来构建特征，但对于阴影密度和炎症部位的识别同样效率不高；SVM 算法对肺炎纹理识别<sup>[3]</sup>，AdaBoost 算法对肺炎部位识别<sup>[4]</sup>等，上述机器学习算法都在不同程度上对前人工作做了改进，但整体识别率仍较低，无法投入实际应用。

基于现有的、经过医学专家判别后的肺炎类型数据集，自然地想到可以使用深度学习技术。深度学习的概念由 Hinton 等人于 2006 年提出，其本质是通过大量有标签的训练数据，学习输入到输出的映射关系，而不需要精确的数学表达式，和本任务当前数据和需求非常契合。其中，卷积神经网络是深度学习技术的一种模型，通过提取数据视觉特征自主学习。课程中深度学习部分主要介绍了 CNN 卷积神经网络，因此本文以 CNN 为切入点，对模型进行优化和改进，再利用加入残差块的深度 CNN 模型 ResNet 进行训练和实验。

## 2、分类网络构建

### 2.1 CNN 卷积神经网络

卷积层的出现主要解决仅使用 MLP 造成的需要学习的参数过多的情况，由于图像存在变换不变性（Translation invariance）和本地性（Locality），以上两个特性保证了无论物体在图像中的什么位置，都能获得相似的输出；像素的特征和它周围的邻近像素更为相关，它的输出仅由其周边一个  $k \times k$  的窗口计算得到。CNN 卷积神经网络<sup>[5]</sup>的三个基本思想：局部感受野（local receptive fields）、权

值共享（shared weights）、池化（pooling）也是基于上述两个特性。局部感受野和权值共享减少了参数数量，池化则简化卷积层的输出，减少了像素信息，仅保留重要信息以降低计算量。简化后的公式如(1)所示， $a^1$  表示隐藏层输出， $a^0$  表示隐藏层输入。

$$a^1 = \sigma(b + w * a^0) \quad (1)$$

如图 2 所示，对于一个原先需要对所有输入加权才能得到的输出，如图 2(a)，卷积后仅需要  $k \times k$ ，图 2(b)中为  $2 \times 2$ ；对于图 2(b)右上角蓝色点的权重参数，由于变换不变性，其参数和右下角橙色点相同，因此不需要再训练一套新的权重，实现了权值共享。

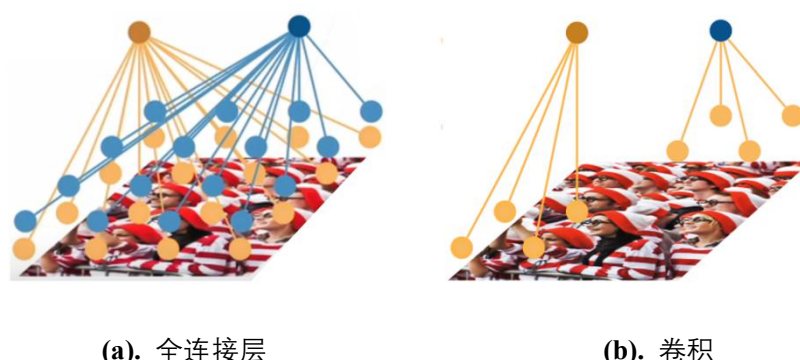


图 2. 全连接层和卷积

上述  $k \times k$  的窗口常被称为卷积核（kernel），用以学习数据以识别一种模式，如图 3 展示了卷积核（kernel）识别红色、绿色、纹理、阴影等模式。窗口大小的决定变量  $k$  也是一个可以调整的超参数，需要注意的是，一个窗口只能学到一种模式，因此，图像分类时，我们往往需要更多窗口以识别更多模式或特征。



图 3. 卷积核（kernel）模式识别

卷积神经网络本质上还是一个神经网络，同样需要激活层等神经网络必需的组件，通过卷积层的堆叠抽取图片中的空间信息。卷积层可以认为是一个特殊的全连接层，本质还是线性变换，只是限制了部分的权重必须是 0，因此无论堆叠多少层，没有激活层，仍旧是一个线性模型。同时，卷积层对位置很敏感，需要池化层得到一些对于位置没有那么敏感的输出。综上，CNN 的模型基本架构可以由图 4 概括。



图 4. CNN 卷积神经网络基本组成

## 2.2 ResNet 预训练模型

ResNet 残差神经网络在传统的卷积神经网络中加入残差学习<sup>[6]</sup>（Residual Learning）的思想而得名，以处理深度神经网络（Deep Neural Network）模型层数加深时，难以训练和学习退化的问题。

其思想可以通过函数来解释，图 5 中  $F_i$  表示函数复杂度（即模型深度）， $f^*$  表示最优值，分别利用非嵌套函数和嵌套函数观察：随着函数复杂度提升，函数区域内到最优值的距离变化以评估精度。不难发现，非嵌套模型随着函数复杂度提升至  $F_6$ ，到最优值的距离还不如  $F_3$ 。ResNet 的出现正是利用图 5(b) 的思想，解决复杂网络随着层数加深精度未必提高的情况。

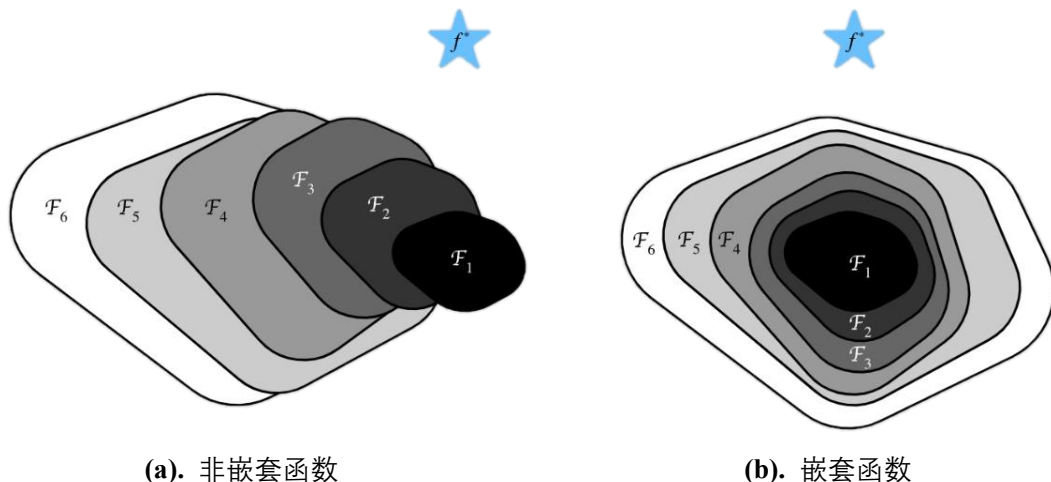


图 5. 函数复杂程度对最优值的影响

CNN 的思想是通过堆叠“层”解决问题，ResNet 希望解决堆叠层产生的复杂度不会对最优值产生影响，做法采用加入残差块产生一种“快速通道（shortcut connection）”得到

$$f(x) = x + g(x) \quad (2)$$

的结构。其中，假设  $g(x)$  是新加入的“层”，即使新加入的“层”没有学到任何信息，也能得到原来  $x$  的值。ResNet 块的细节如图 7 所示，残差块的添加位置并不固定，激活层、卷积层、批标准化层可以进行不同组合。

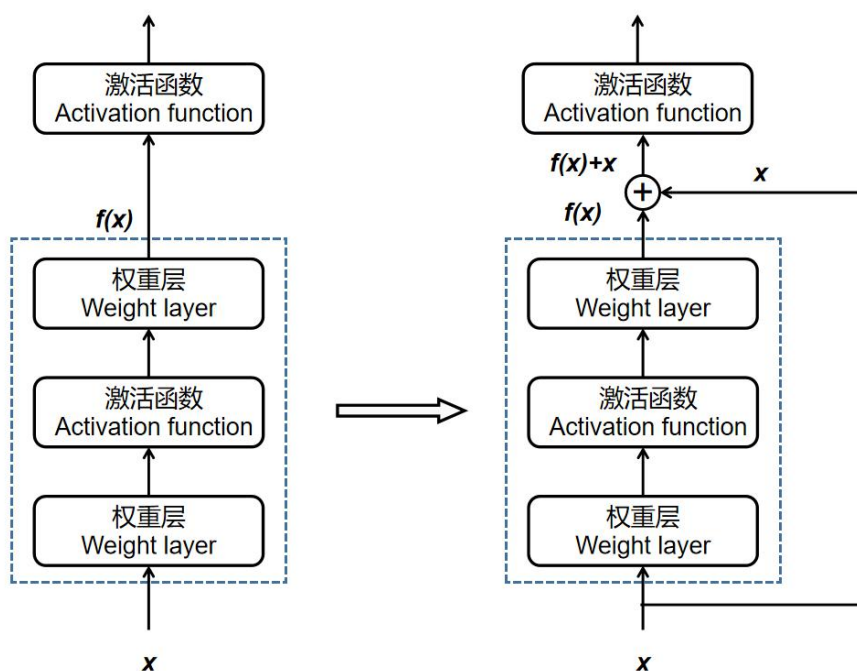


图 6. 函数角度理解残差块

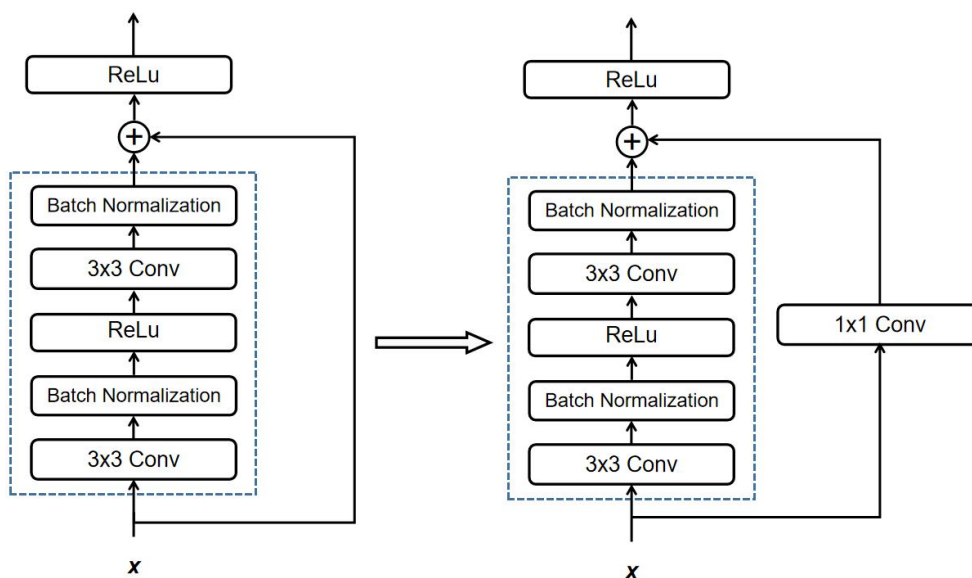


图 7. 残差块细节

ResNet 的前两层同 GoogLeNet<sup>[7]</sup>中的一样：在输出通道数为 64、步幅为 2 的  $7 \times 7$  卷积层后，接步幅为 2 的  $3 \times 3$  的最大汇聚层。不同之处在于 ResNet 每个卷积层后增加了批量规范化层（Batch Normalization）。

ResNet-18 架构如图 8 所示，每个模块有 4 个卷积层（不包括恒等映射的  $1 \times 1$  卷积层）。加上第一个  $7 \times 7$  卷积层和最后一个全连接层，共有 18 层。因此，这种模型通常被称为 ResNet-18。通过配置不同的通道数和模块里的残差块数可以得到不同的 ResNet 模型，例如更深的含 152 层的 ResNet-152。虽然 ResNet 的主体架构跟 GoogLeNet 类似，但 ResNet 架构更简单，修改也更方便。这些因素都

导致了 ResNet 迅速被广泛使用。

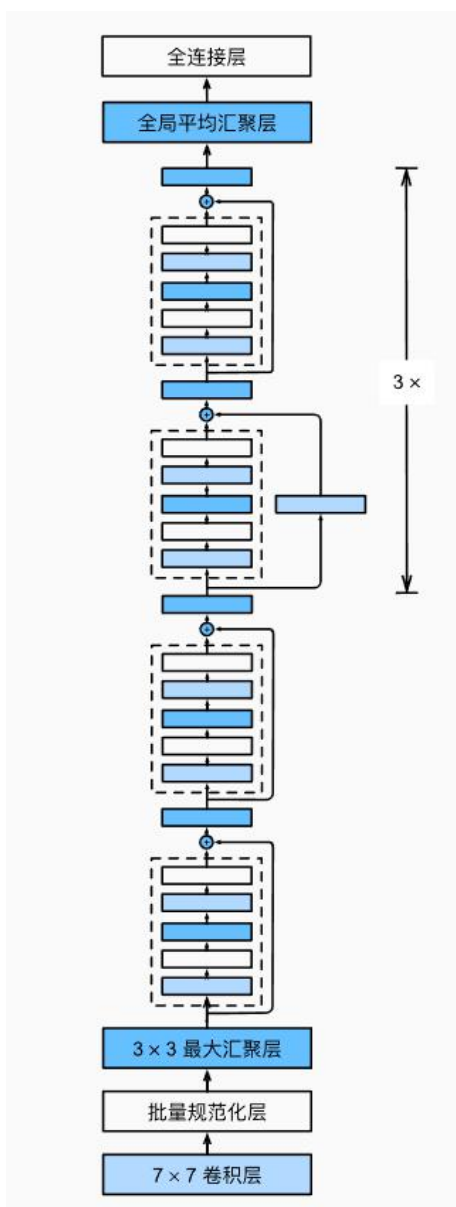


图 8. ResNet-18 架构

## 2.3 数据增强

为了避免过拟合问题，我们需要人为地扩展数据集，使当前数据集更大。这个想法是通过一些转换（transformation）来改变训练数据，以产生和原始数据区别。通过改变数组表示的方式改变训练数据，同时保持标签相同的方法称为数据增强。常见的一些增强是灰度、水平翻转、垂直翻转、随机裁剪、颜色抖动、平移、旋转等等。通过对当前训练数据应用这些转换，可以较为便捷地将训练数据的数量增加一倍或三倍，降低了模型对特定特征的依赖，从而提升模型的泛化能力，增强模型的鲁棒性。



## 2.4 优化器的选择

本实验优化器的选择上，CNN 卷积神经网络采用 RMSprop；ResNet 使用 Adam Optimizer<sup>[8]</sup>，以较低的学习率（learning rate）进行调优（fine-tuning）。一般的深度学习优化器（Optimizer）有：Momentum 梯度下降算法，使用指数加权平均数来计算梯度，从而更新参数也就是将可学习参数的梯度作为序列数据（sequential data），通过迭代找到目标函数的最小值；RMSprop 则是加入了衰减系数来控制历史信息的获取多少，以防止某一参数方向上，更新的梯度过大或过小；

$$\begin{aligned} E[g^2]_t &= \gamma \cdot E[g^2]_{t-1} + \eta \cdot g_t^2 \\ \theta_{t+1} &= \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \varepsilon}} \cdot g_t \end{aligned} \quad (3)$$

Adam Optimizer 即拥有 Momentum 梯度下降算法一样保持了过去梯度的指数衰减平均值，

$$\begin{aligned} m_t &= \beta_1 \cdot m_{t-1} + (1 - \beta_1) g_t \\ v_t &= \beta_2 \cdot v_{t-1} + (1 - \beta_2) g_t^2 \end{aligned} \quad (4)$$

又和 RMSprop 一样存储过去梯度的平方的指数衰减平均值，

$$\begin{aligned} \hat{m}_t &= \frac{m_t}{1 - \beta_1^t} \\ \hat{v}_t &= \frac{v_t}{1 - \beta_2^t} \end{aligned} \quad (5)$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t + \varepsilon}} \cdot \hat{m}_t \quad (6)$$

几种优化算法<sup>[9], [10]</sup>的特点如表所示：

优化算法	特点
Momentum	<p>参数下降初期，若前后 2 次下降方向一致，能够很好的加速；</p> <p>参数下降中后期，在局部最小值附近来回震荡时；</p> <p>梯度方向改变时，降低参数更新速度，从而减少震荡；</p> <p>梯度方向相同时，加速参数更新，从而加速收敛；</p>
Adagrad	<p>前期 regularizer 较大，能够放大梯度</p> <p>后期 regularizer 较小，能够缩小梯度</p> <p>中后期，分母梯度平方的累加越来越大，使得训练提前结束；</p>
Adadelta	<p>解决了 Adagrad 的第三条问题</p> <p>训练初中期，加速效果不错，很快。</p>

	训练后期，反复在局部最小值附近抖动
Adam	每次迭代学习率都有固定范围，使得参数比较平稳；
	善于处理非平稳目标的优点
	为不同的参数计算不同的自适应学习率；
	适用于大多非凸优化问题——适用于大数据集和高维空间；
	依赖于全局学习率；
RMSprop	效果介于 Adagrad 和 Adadelta 之间；
	适合处理非平稳目标——对于 RNN 效果很好；

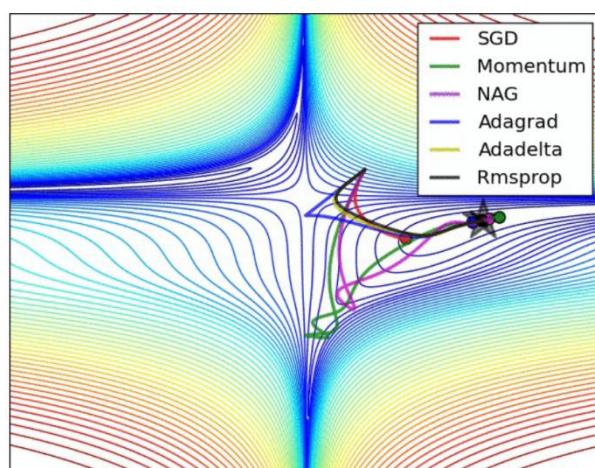


图 9. 不同优化器性能测试

### 3、数据实验

#### 3.1 数据集描述

默认数据集划分训练集、验证集和测试集，存放于 3 个文件夹中(train, test, val)，并包含每个图像类别的子文件夹(Pneumonia/Normal)。共有 5,840 张 x 射线图像(JPEG)和 2 个类别(肺炎/正常)。绝大多数的图片均是单通道灰度图，极个别彩色图像。在 Pneumonia 文件夹中，有两个类型的肺炎图片，分别为细菌性肺炎和病毒性肺炎，区分标准为若名字中带有 bacteria 即为细菌性肺炎 x 光图片，带有 virus 即为病毒性肺炎 x 光图片。

表 1. 数据集划分

	训练集	测试集	总计
<b>Bacteria</b>	2356	416	2772
<b>Normal</b>	1341	234	1575
<b>Virus</b>	1269	224	1493

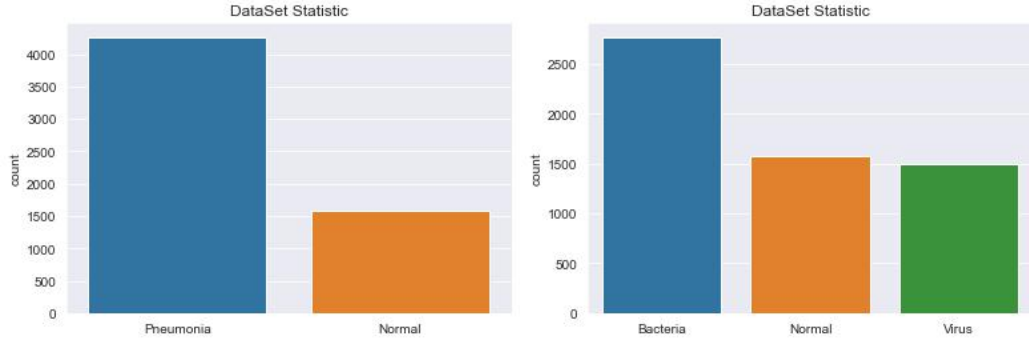


图 10. 数据集分布

### 3.2 数据处理

图像数据的处理主要包括随机旋转、水平翻转、亮度饱和度等属性该改变和标准化，部分数据展示结果如下图所示，样本中不存在缺失值情况故无需做插值处理。

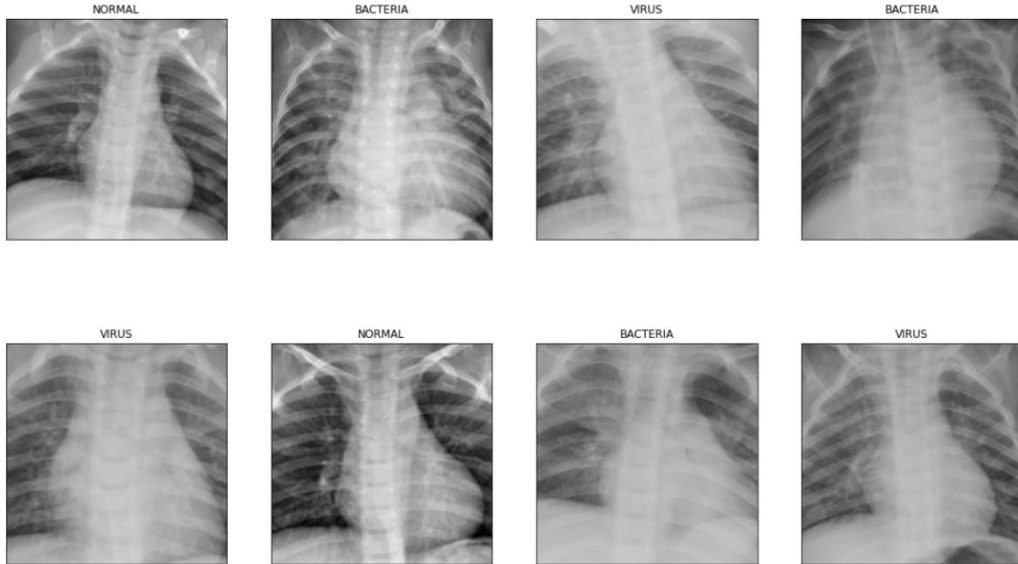


图 11. 部分数据展示

其中，由于原始数据的大小不统一，在训练前必须进行归一化处理，避免某个特征值过大导致的网络训练性能问题。标准化对每个 **channel** 逐一处理，参数采用 ImageNet 的均值和标准差，加快模型的收敛速度。线性变换的映射函数如 (7) 式所示

$$x = \frac{x - mean}{std} \quad (7)$$

*mean* 是样本均值，*std* 是样本标准差，标准差能反映一个数据集的离散程度，标准偏差越小，这些值偏离平均值就越少，反之亦然。标准偏差的大小可通过标准偏差与平均值的倍率关系来衡量。平均数相同的两个数据集，标准差未必相同。

### 3.3 对比实验

根据本文所需解决的下游任务，共进行两组实验，分别为“肺炎/正常”二分类实验和“细菌性肺炎/病毒性肺炎/正常”三分类实验。二分类实验中，包括决策树算法二分类实验、CNN 卷积神经网络二分类实验和 ResNet-34 二分类实验；三分类实验中，包括 ResNet-34 三分类实验和 ResNet-152 三分类实验。数据集的划分为 8.5:1.5，此外，本文采用精确率(Precision)、召回率(Recall)和 F1 值(F1-score)来评估模型预测精度。其中，F1 值作为精确率和召回率的调和平均数，在性能评估中起主导作用。

#### 3.3.1 “肺炎/正常”二分类实验

二分类任务首先使用决策树算法作为基线模型，验证算法的同时评估其他模型的性能，再利用卷积神经网络和 ResNet 系列模型进行实验。

##### 3.3.1.1 决策树二分类实验

课程实验过程中，发现决策树是机器学习中较为常用的算法之一，因此，利用该算法对本任务进行实验，对于二分类任务的实验结果如表 2 所示，可以发现，决策树算法能实现基本的分类，但准确度并不高，猜测是因为决策树对于特征的提取并不高效，因此在后续小节采用神经网络进行实验。

表 2. “肺炎/正常”二分类决策树实验结果

	样本量	Precision	Recall	F1-score
<b>Pneumonia (Class 0)</b>	390	0.72	0.93	0.81
<b>Normal (Class 1)</b>	234	0.77	0.40	0.52
<b>accuracy</b>	624	/.	/	0.73

##### 3.3.1.2 CNN 卷积神经网络二分类实验

CNN 卷积神经网络的输入数据经过归一化处理，以减少照明差异的影响，且 CNN 在[0-1]数据比[0-255]收敛速度更快。同时对输入尺寸和图像属性进行变换，具体变换如下：

- 1: 随机旋转训练图像 30 度；
- 2: 随机放大 20%的训练图像；
- 3: 随机将图像水平移动 10%的宽度；
- 4: 随机垂直移动图像高度的 10%；
- 5: 随机水平翻转图像；

利用 *Sequential()* 初始化模型，并添加卷积层、池化层等，模型架构如下表所示，总参数为 1,246,401，可训练参数为 1,245,313，不可训练参数为 1,088。

表 3. “肺炎/正常”二分类卷积神经网络模型结构

层类型	输出形状	参数个数
conv2d_56 (Conv2D)	(None, 150, 150, 32)	320
batch_normalization_56	(None, 150, 150, 32)	128
max_pooling2d_44	(None, 75, 75, 32)	0
conv2d_57 (Conv2D)	(None, 75, 75, 64)	18496
dropout_45 (Dropout)	(None, 75, 75, 64)	0
batch_normalization_57	(None, 75, 75, 64)	256
max_pooling2d_45	(None, 38, 38, 64)	0
conv2d_58	(None, 38, 38, 64)	36928
batch_normalization_58	(None, 38, 38, 64)	256
max_pooling2d_46	(None, 19, 19, 64)	0
conv2d_59 (Conv2D)	(None, 19, 19, 128)	73856
dropout_46 (Dropout)	(None, 19, 19, 128)	0
batch_normalization_59	(None, 19, 19, 128)	512
max_pooling2d_47	(None, 10, 10, 128)	0
conv2d_60	(None, 10, 10, 256)	295168
dropout_47	(None, 10, 10, 256)	0
batch_normalization_60	(None, 10, 10, 256)	1024
max_pooling2d_48	(None, 5, 5, 256)	0
flatten_12 (Flatten)	(None, 6400)	0
dense_23 (Dense)	(None, 128)	819328
dropout_48 (Dropout)	(None, 128)	0
dense_24 (Dense)	(None, 1)	129

在给定数据上对上述卷积神经网络进行训练并在测试集上验证模型性能，表 4 展示了卷积神经网络对“肺炎/正常”两种类别分类的实验结果，对于“肺炎”类和“正常”类的 F1-score 分别为 0.93 和 0.89，准确率为 0.92，实现了分类的基本要求。

表 4. 卷积神经网络“肺炎/正常”二分类实验结果

	样本量	Precision	Recall	F1-score
<b>Pneumonia (Class 0)</b>	390	0.93	0.94	0.93
<b>Normal (Class 1)</b>	234	0.89	0.89	0.89
<b>accuracy</b>	624	/.	/	0.92

### 3.3.1.3 ResNet-34 二分类实验

经过卷积神经网络的实验，猜测准确度为 92%的主要原因在于随着网络层数的加深，逐渐出现退化的现象，因此考虑加入残差块的 ResNet 系列模型，本节的二分类实验，选择 ResNet-34。实现结果如表 5 所示，相较于 CNN 卷积神经网络，加入残差块后随着深度增加并未出现明显的网络退化，准确度上升，验证了残差块的作用。

表 5. ResNet-34“肺炎/正常”二分类实验结果

	样本量	Precision	Recall	F1-score
<b>Pneumonia (Class 0)</b>	390	0.96	0.95	0.96
<b>Normal (Class 1)</b>	234	0.92	0.93	0.93
<b>accuracy</b>	624	/.	/	0.94

### 3.3.1.4 ResNet-152 二分类实验

通过 ResNet-34 的结果，自然地认为加深网络层数有助于提升精度，故使用 ResNet-152 大幅提升模型复杂度，实验结果如表 6 所示。结果并未如期提升，反而有所下降。虽然 ResNet 系列模型加入了残差块可以缓解由于模型深度增加导致的梯度消失，但深度过深的情况下，仍会产生退化和过拟合。因此对于本数据集的二分类任务而言，ResNet-152 性能并不佳。

表 6. ResNet-152“肺炎/正常”二分类实验结果

	样本量	Precision	Recall	F1-score
<b>Pneumonia (Class 0)</b>	390	0.91	0.90	0.91
<b>Normal (Class 1)</b>	234	0.86	0.87	0.85
<b>accuracy</b>	624	/.	/	0.88

## 3.3.2 “细菌性肺炎/病毒性肺炎/正常”三分类实验

本节将主要对不同层数的 ResNet 系列模型进行三分类考察，通过实验研究层数和模型精度间的关联，同时仍先用决策树模型对三分类任务进行实验，以作为基线模型对比同一种模型对于不同下游任务的性能。

### 3.3.2.1 决策树三分类实验

决策树三分类任务的实验结果如表 7 所示，F1-score 为 0.56，其中 Virus 类的 Precision 和 Recall 均较低，猜测决策树对于细菌和病毒性肺炎图片的特征辨别并不敏感，因此该模型并不适合此三分类任务。

表 7. 决策树“细菌性肺炎/病毒性肺炎/正常”三分类实验结果

	样本量	Precision	Recall	F1-score
<b>Bacteria (Class 0)</b>	242	0.55	0.74	0.63
<b>Normal (Class 1)</b>	234	0.78	0.47	0.59
<b>Virus(Class 2)</b>	148	0.39	0.41	0.40
<b>accuracy</b>	624	/.	/	0.56

### 3.3.2.2 ResNet-34 三分类实验

ResNet-34 三分类任务的实验结果如表 8 所示，相较于决策树模型，F1-score 有了显著的提升，Bacteria, Normal, Virus 三个类别的 F1-score 全面提升。但和决策树模型类似，Virus 类的 Precision 和 Recall 仍较低，可以看出两类肺炎类型的特征不太容易学习，因此考虑加深网络层数。

表 8. ResNet-34“细菌性肺炎/病毒性肺炎/正常”三分类实验结果

	样本量	Precision	Recall	F1-score
<b>Bacteria (Class 0)</b>	242	0.77	0.91	0.84
<b>Normal (Class 1)</b>	234	0.93	0.64	0.76
<b>Virus(Class 2)</b>	148	0.58	0.69	0.63
<b>accuracy</b>	624	/.	/	0.75

### 3.3.2.3 ResNet-152 三分类实验

ResNet-152 三分类任务的实验结果如表 9 所示，

表 9. ResNet-152“细菌性肺炎/病毒性肺炎/正常”三分类实验结果

	样本量	Precision	Recall	F1-score
<b>Bacteria (Class 0)</b>	242	0.81	0.75	0.78
<b>Normal (Class 1)</b>	234	0.89	0.91	0.91
<b>Virus(Class 2)</b>	148	0.59	0.62	0.60
<b>accuracy</b>	624	/.	/	0.78

在二分类任务中发生的过拟合问题在三分类任务中并未出现，如图 12 所示。实验结果中，F1-score 有了一定提升，观察结果发现，总体 F1-score 的提升主要在于 Normal 类的 F1-score 的提升，另外两类肺炎类型甚至有所下降，由此推断网络仍存在退化问题，但仅针对一部分特征，因此总体准确度的评价不可一概而论，需根据实际数据进行调整。

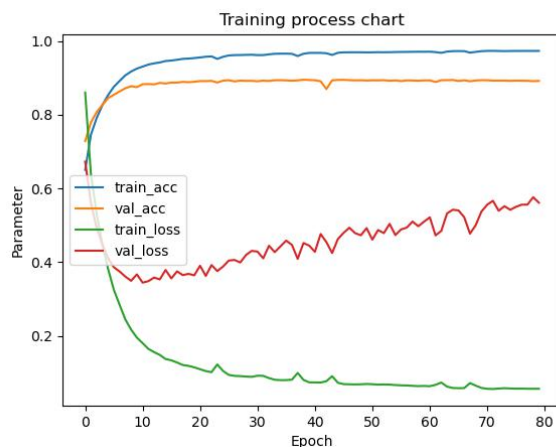


图 12. ResNet-152 训练过程

### 3.3.3 实验评估

不同的下游任务，模型的选择起到关键作用。对于“肺炎/正常”二分类任务，ResNet-34 和 CNN 卷积神经网络效果较好；对于“细菌性肺炎/病毒性肺炎/正常”三分类任务，ResNet-152 的性能稍高于 ResNet-34。

表 10. 二分类模型评估

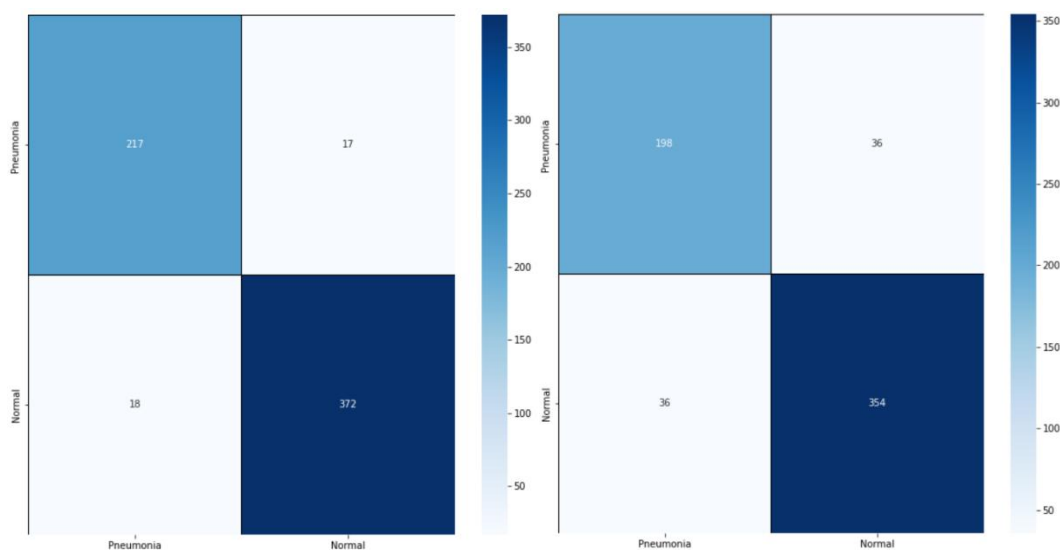
模型	准确度
决策树	0.73
CNN 卷积神经网络	0.92
ResNet-34	<b>0.94</b>
ResNet-152	0.88

表 11. 三分类模型评估

模型	准确度
决策树	0.56
ResNet-34	0.75
ResNet-152	<b>0.78</b>
/	/

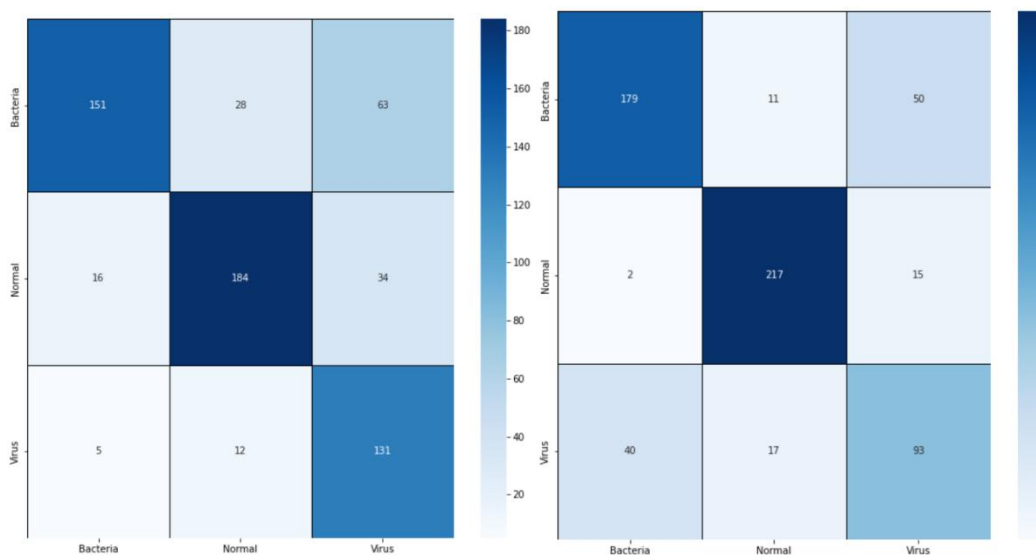
ResNet 系列模型对于分类任务效果都较好，两类分类任务的混淆矩阵如图 13 所示，CNN 卷积神经网络对二分类任务的错误率在可接受的范围内。相比之下，传统的机器学习算法在对两类任务中对特征的敏感度上稍逊一筹。实际使用中 ResNet 系列模型是较好的选择。





(a). 二分类 ResNet-34

(b). 二分类 ResNet-152



(c). 三分类 ResNet-34

(d). 三分类 ResNet-152

图 13. ResNet 系列模型分类任务混淆矩阵

### 3.4 超参数设置体会

实验中对于神经网络训练而言，超参数的调整会影响模型训练的质量从而对最终模型的泛化性能产生影响。我们日常的科研和学习实验中，往往可以先从质量较高的工具包中入手，其中设置的参数可能并非对于当前的下游任务是最佳的，但是一个不错的开始点。超参数的调整有时并不具备科学依据，会逐渐产生一种直觉（insight），获知对模型好坏影响较大的超参数。例如优化器的选择上，当 SGD 参数调的比较好的情况下，性能是优于 Adam 的，但 Adam 并没有 SGD 中参数那么敏感，因此对于调参比较友好。

同时，调参过程中需要勤记录，把 logs 输出为一个 txt 或 excel 便于观察调参前后的变化，本文中对于调参的记录有所欠缺，通过本文的多个实验认识到了这一点。

## 4、结论与展望

本文基于 X 光影像数据集，利用多种方法进行肺炎判别和分类，充分对比了传统机器学习算法、CNN 卷积神经网络和 ResNet 系列模型对于二分类、三分类任务的模型性能，并对结果进行分析，实验结果表明 ResNet 系列模型有效地解决了 CNN 卷积神经网络随着网络深度增加产生的梯度消失、学习退化的问题，其对于分类任务泛化性能较好，有较高的准确度。医学专家借助该技术，可以有效减轻工作负担和减少出现误诊的可能。同时，实验发现模型的性能可以通过数据规模的扩大得到提升，因此，未来若投入实际使用，可以补充更大规模的数据集提供给模型进行训练，相信可以得到更高的精度。

ResNet 模型实验中同样发生了当网络层数较深时梯度消失的问题，因此，残差块的引入仅能缓解该问题并不能完美解决。当然，对于本文中的任务，并不需要更复杂的模型，但对于其他下游任务，如何解决残差块无法处理的复杂模型，仍然是个问题。

随着深度学习技术和迁移学习的发展，对于下游任务可以尝试更多模型。如 NLP 领域性能较好的 BERT<sup>[15]</sup>、CLIP<sup>[16]</sup>等，也越来越多地应用到 CV 领域。但每种下游任务都有其特定的需求和局限，充分分析下游任务和模型特点，才能更好地将计算机应用于实践，相信未来计算机技术会是人类的得力助手！

## 参考文献

- [1]张亮,王齐凯.基于 bp 神经网络的肺炎分类[J].华南理工大学学报(自然科学版),2015,42(1):72-76.
- [2]李庆华,王芳.基于朴素贝叶斯的肺炎分类方法研究[J].科技大学学报(自然科学版),2014,12(1);32-36.
- [3]Ji S,Xu W,Yang M,et al.3D convolutional neural networks for human action recognition[A].Proceedings of the 28th International Conference on Machine Learning[C].2010:195-502.
- [4]赵宏,杨绍强.基于 svm 的肺炎分类识别研究[J].系统仿真学报,2014,22(3):638-641
- [5]Kim Y . Convolutional Neural Networks for Sentence Classification[J]. Eprint

Arxiv, 2014.

[6]K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” arXiv:1512.03385, 2015.

[7]Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 1–9.

[8]D. P. Kingma and J. L. Ba, Adam: A Method for stochastic Optimization. San Diego: The International Conference on Learning Representations (ICLR), 2015

[9]Sixin Zhang, Anna Choromanska, and Yann LeCun. Deep learning with Elastic Averaging SGD. Neural Information Processing Systems Conference (NIPS 2015), pages 1–24, 2015.

[10]Ruder S . An overview of gradient descent optimization algorithms[J]. 2016.

[11] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical image Database. In CVPR09, 2009.

[12] A. Khamparia, G. Saini, D. Gupta, A. Khanna, S. Tiwari, V.H.C. de Albuquerque, Seasonal crops disease prediction and classification using deep convolutional encoder network, Circ. Syst. Sign. Process. 39 (2) (2020) 818–836.

[13] S.P. Mohanty, D.P. Hughes, M. Salathé, Using deep learning for image-based plant disease detection, Front. Plant Sci. 7 (2016) 1419.

[14] P. Gong, C. Zhang, M. Chen, Deep learning for toxicity and disease prediction, Front. Genet. 11 (2020) 175.

[15] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.

[16] Radford A , Kim J W , Hallacy C , et al. Learning Transferable Visual Models From Natural Language Supervision[J]. 2021.