

# 基于 DarkCovidNet 模型自动检测肺结核的分析

张立夫 李丹

四川大学锦城学院 四川 成都 611731

**【摘要】**肺结核是目前世界上单一感染源导致人类死亡的主要原因，超过了艾滋病，每年有数百万的人罹患这种疾病，作为一种慢性传染病，尽早发现对后续治疗有着至关重要的影响。目前，对于此结核病的诊断方法中，X光胸片检查仍是临床诊断的首选方法之一，随着病例的增多，放射学医师有时很难通过肉眼观察到患者X光胸片的特征，基于机器学习的计算机辅助系统可以辅助医师做出更加准确的决策。在本项研究中，利用 DarkCovidNet 模型对肺结核病例 (abnormal/normal) 进行自动检测，分别在 5 次、10 次、15 次、100 次这四种迭代次数上训练并进行对比分析，发现迭代次数的增加有助于使模型训练的效果变得更好和趋稳，得到最好的准确率为 0.81 (迭代次数 100 次)。除此之外，还利用经典的 ResNet34 模型与 DarkCovidNet 模型分别迭代 5 次进行训练比较分析，得到 ResNet34 模型的准确率为 0.84，而 DarkCovidNet 模型仅为 0.77，更深的模型网络层次在一定程度上有助于模型训练的效果，但同时也耗费了更多的时间。

**【关键词】**肺结核；X 光胸片；机器学习；DarkCovidNet

## 1 背景介绍

### 1.1 课题背景及研究意义

肺结核作为一种传染病，是世界十大死亡原因之一，也是单一感染源导致死亡的主要原因（排名高于 HIV/ 艾滋病）<sup>[1]</sup>。1997 年是世卫组织（WHO）发布全球结核病报告的元年，从那一年起，每年都会发布一份，其目的是在全球范围内，报告全球、各区域和国家、各层面评估结核流行情况，以及对结核病疫情的处置情况进行总结<sup>[2]</sup>。2017 年，WHO 报告指出大概有 150 万感染结核病患者死亡，2018 年，报告指出大概有 180 万感染结核病患者死亡<sup>[3]</sup>。结核分枝杆菌——引发肺结核疾病的罪魁祸首，通过空气传播是它的主要传播途径，当结核病患者进行咳嗽、打喷嚏、与人近距离交流等行为活动时，细菌就此会传染给接触的人。作为一种传染性较强的慢性传染疾病，肺结核越早发现治疗效果就越好。目前，对于肺结核的诊断，有一些基于分子分析和细菌培养的高度准确的方法，但是大多数方法的成本都不允许在发展中国家大规模采用。对于结核病的新型诊断方法，目前正在研发的诊断技术依旧停留在细胞层面；而对于结核病的新型治疗药物方面，如今大多数的治疗方法是还是利用多种药物混合治疗。在 WHO 发布的 2019 结核报告中，对药物敏感结核病医治成功率为 85%，对耐药结核病的医治成功率为 56%，对广泛耐药结核病的医治成功率仅为 39%。目前，荧光定量 PCR (fluorescence quantitative PCR, FQ-PCR) 检测结核杆菌

(tubercle bacillus, TB) DNA (TB-DNA) 作为一种快速诊断肺结核的技术，在临床中已普遍应用，但是其还是存在着假阳性和不同检样中敏感性不确定等问题<sup>[4]</sup>。因此，在全球范围内，防止结核病的传染，有效诊断结核病仍是一项重大的挑战。

在早期发现肺结核疾病和诊断肺结核疾病时，一个简而易又不可或缺的方法就是常规的胸部 X 光片检查，目前它仍然是肺结核临床诊断的首选方法之一。X 线胸片检查的特点是性价比相对较高、易操作、对身体几乎不造成危害。除了能通过图像对肺部疾病定性之外，它还可以通过分析图像的特征来判断肺结核的类型和程度<sup>[5]</sup>。由于人的肉眼所见图像的范围和程度是有限的，随着肺结核病例的增多，医生有时很难辨别一些病例的特征，所以迫切需要计算机来辅助放射科医生的诊断与分析以便做出更加准确的决策。

### 1.2 计算机辅助检测肺结核研究现状

目前，用于计算机辅助检测放射学影像主要是机器学习方法，其在医学领域的自动诊断的应用也越来越多。机器学习的一个重要组成部分——深度学习，是近些年有关研究者提出的一种具有多个隐藏层的深度神经网络来完成特项学习要求与任务的方法之一。其主要通过搭建神经网络模型，利用大量的训练样本或数据来学习与训练，得到有价值的特征，从而提升网络模型的分类、预测的准确性<sup>[6]</sup>。深度学习技术已成功应用于皮肤癌分类、肺分割、胸部 X 射线肺炎检测、脑疾病分类、眼底

图像分割、乳腺癌检测、心律失常检测等诸多问题<sup>[7]</sup>。

近段时间,许多影像学图像被广泛应用于肺结核的检测中,Liu 等人<sup>[8]</sup>提出一种基于 CNN 的端到端的定位模型,该模型在不同医院提供的两个测试数据集的平均精度分别为 0.9023 和 0.9332;毋一帆<sup>[9]</sup>提出一种分类算法,它是使用肺结核病灶的检查结果所得到的胸部 X 光片的信息,通过训练 TCLD\_CXR 这一数据集,得到了 0.85 的 Acc 和 0.86 的 Auc; P.Lakhani 等人<sup>[10]</sup>,得到结果最好的分类方法是 Alex Net 和 Google Net DCNN 这两个模型的集合,其 Auc 为 0.99。

在本文中,利用了 Tulin Ozturk 等人提出的自动检测 Covid-19 的深度模型,此模型除了能诊断 Covid-19 外,它还可以诊断普通肺炎与肺结核。它具有端到端的结构,不需使用任何特征提取方法,只需要通过胸部 X 线图像来反馈诊断结果。本文利用此模型来检测肺结核,并分析在不同数据集上模型效果的差异。

## 2 材料与方法

### 2.1 X 射线图像数据集

在本文的研究中,采用从 Kaggle 获得的 X 射线图像用于肺结核的诊断。数据集 ChinaSet 是由美国马里兰州国家医学图书馆与中国深圳广东医学院深圳第三人民医院合作创建的,这些胸部 X 线图像均来自于门诊,其中包括了 336 张表现为肺结核的胸部 X 线图像以及 326 张表现为正常的胸部 X 线图像。图像文件名编码为 CHNCXR\_####\_0/1.png,其中“0”为正常肺,“1”为异常肺,如图 1 所示。

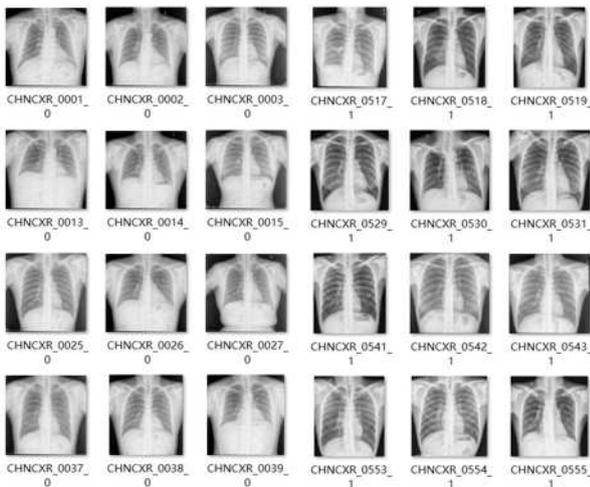


图 1 ChinaSet 中表现为“正常肺”“异常肺”部分截图

### 2.2 DarkCovidNet 模型简介

本文运用了 Tulin Ozturk 等人提出的 DarkCovidNet

模型,它是一种可以检测目前在全球蔓延的新型冠状病毒 (Covid-19) 的一种深度学习模型,值得一提的是,它同时也可以检测肺结核上展开应用。要了解该模型,首先要了解卷积神经网络。卷积神经网络 (Convolutional Neural Networks, CNN) 早在上个世纪八九十年代就被提出,目前主要运用于像图像识别、语音识别这样的领域当中。它是将深度学习和人工神经网络相结合而产生的一种新型的、先进的人工神经网络技术,并且具有层次结构化、特征抽取等特点<sup>[11]</sup>。卷积神经网络最重要的特征是增加了“卷积层”和“池化层”,除此之外,“滤波器”“填充”“步幅”这些概念也是卷积神经网络中不可或缺的。

DarkCovidNet 模型一共有 17 个卷积层,在其模型中,有一个集成的 DN 层,其中包括了一个卷积层,一个批标准化操作和一个 LeakyReLU 函数。我们熟知的激活函数是 ReLu,也就是其负类都为 0,但是这会成为一个问题,也就是说对于小于 0 的值,这个神经元的剃度永远都会是 0,很容易就“死”了,神经元不能更新参数,从而神经元不学习了。如果用 LeakyReLU 激活函数,就会保留一些负值,不会全部丢失掉。ReLU 函数与 LeakyReLU 函数的公式如下所示,对应的图像如图 2 所示。

$$ReLu(x) = \max(x,0) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (1)$$

$$Leaky Re Lu(x) = \begin{cases} x, & x \geq 0 \\ ax, & x < 0 \end{cases} \quad a \in (0,1) \quad (2)$$

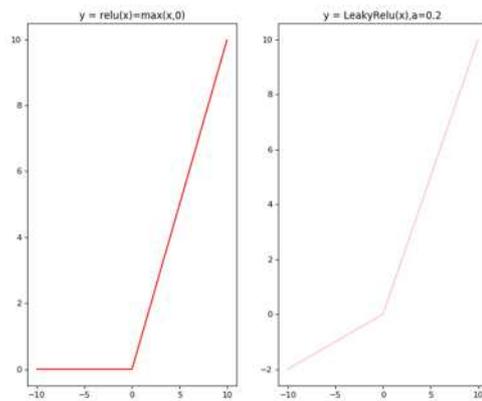


图 2 ReLu 函数与 LeakyReL 函数图像

令字母 C 表示一个卷积层, M 表示最大池化 Maxpool 层,以及 DN 表示一个集成的 DN 层, LR 表示 LeakyReLU 函数, BN 表示批标准化操作,该 DarkCovidNet 模型的层次结构如下所示:

DN1(C1,BN1,LR1)-M1-DN2(C2,BN2,LR2)-C3-C4-C5-M2-C6-C7-C8-M3-C9-C10-C11-M4-C12-C13-C14-DN3(C15,BN3,LR3)-DN4(C16,BN4,LR4)-C17

池化除了最大池化 Maxpool 外，还有平均池化 Avgpool，在图像识别领域，主要使用最大池化 Maxpool<sup>[12]</sup>，它通过获取由滤波器确定的区域内的最大值来减小输入。模型中还选择了交叉熵损失函数，它的优点是简化网络计算量<sup>[13]</sup>。

### 3 实验结果与分析

#### 3.1 数据的预处理

对于数据集 ChinaSet，一共有 662 张 X 光胸片，考虑到计算机硬件性能的问题，首先将图像本身的存储大小进行缩小处理，将原数据集的存储大小缩小了八倍。虽然存储大小变小了，但其呈现效果还是很可观的。将此数据集分割为训练集和验证集，其中训练集占 80%，验证集占 20%，训练集与验证集中分别存放了对应类别 (abnormal 与 normal) 的 X 光胸片。最终，训练集有 528 张 X 光胸片，验证集有 134 张 X 光胸片。

为了得到的更好的结果，对数据集进行了数据增强处理 `ds_tfms=get_transforms()`，此操作实质上是数据的扩充，在原图片上进行翻转、拉伸、旋转，形成一些“新”的训练数据，可供训练的有关数据越多，越不容易出现过拟合之类的问题。然后将图像的大小做了统一处理，使之都变为 256\*256 的图片尺寸，大小保持一致有助于训练的效果。最后对数据集进行批处理，将 32 张图像“弄成一捆”，由于许多用于处理数值计算的方法库都在大型数组计算这一问题上被优化过，所以这种方法使处理时间有效变短，这样一次性计算大型数据要比分开逐步计算各个小型数组速度更快一些。

#### 3.2 采用不同迭代次数的实验结果分析

分别使用不同迭代次数对 DarkCovidNet 模型下的 ChinaSet 数据集进行了训练。采用的迭代次数分别为 5 次、10 次、15 次以及 100 次，最大学习率选择的都是  $3e-3$ 。表 1 给出了训练的迭代次数以及其对应的准确率，图 3 给出了这 4 次训练的验证集的混淆矩阵图，可以看到，迭代次数越多，混淆矩阵越来越“分明”。

表 1 采用不同迭代次数训练模型的准确率

迭代次数	学习率	准确率
5	$3e-3$	0.74
10	$3e-3$	0.81
15	$3e-3$	0.78
100	$3e-3$	0.81

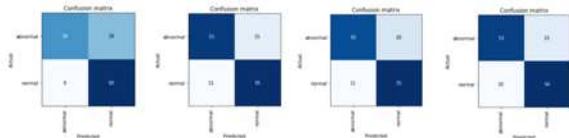


图 3 从左至右依次为迭代 5 次、10 次、15 次、100 次的验证集混淆矩阵图

由于受计算机硬件性能限制，从此表的数据中看出，迭代次数的增大，在一定程度上能够使准确率增大并趋于稳定状态。我们分别再来看下当迭代次数选择 15 次以及 100 次时，每一次迭代的具体过程及他们的损失值变化趋势，如图 4、图 5 所示。

Loss of Train, Validation, and Test					Loss of Confusion Matrix				
epoch	train_loss	valid_loss	accuracy	time	epoch	train_loss	valid_loss	accuracy	time
0	0.6697	0.9650	0.5788	0.128	0	0.7447	0.9474	0.5716	0.128
1	0.5919	0.8495	0.6220	0.127	1	0.5212	0.7247	0.6220	0.128
2	0.5610	0.8067	0.5723	0.129	2	0.5889	0.8831	0.6089	0.128
3	0.5421	0.8150	0.7305	0.124	3	0.5327	0.8324	0.5916	0.128
4	0.5120	0.7983	0.5477	0.124	4	0.4276	0.8549	0.7465	0.128
5	0.4523	0.7584	0.7204	0.123	5	0.4092	0.8627	0.7204	0.128
6	0.4523	0.7584	0.7204	0.123	6	0.4092	0.8627	0.7204	0.128
7	0.4523	0.7584	0.7204	0.123	7	0.4092	0.8627	0.7204	0.128
8	0.4523	0.7584	0.7204	0.123	8	0.4092	0.8627	0.7204	0.128
9	0.4523	0.7584	0.7204	0.123	9	0.4092	0.8627	0.7204	0.128
10	0.4523	0.7584	0.7204	0.123	10	0.4092	0.8627	0.7204	0.128
11	0.4523	0.7584	0.7204	0.123	11	0.4092	0.8627	0.7204	0.128
12	0.4523	0.7584	0.7204	0.123	12	0.4092	0.8627	0.7204	0.128
13	0.4523	0.7584	0.7204	0.123	13	0.4092	0.8627	0.7204	0.128
14	0.4523	0.7584	0.7204	0.123	14	0.4092	0.8627	0.7204	0.128
15	0.4523	0.7584	0.7204	0.123	15	0.4092	0.8627	0.7204	0.128
16	0.4523	0.7584	0.7204	0.123	16	0.4092	0.8627	0.7204	0.128

图 4 从左至右分别为迭代 15 次、100 次 (前 19 次) (后 16 次) 的具体过程图

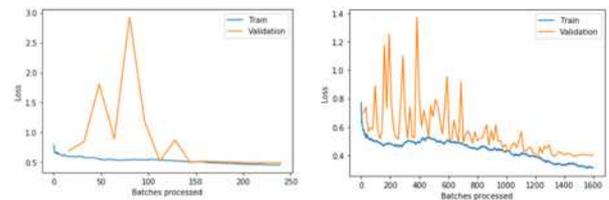


图 5 从左至右分别为迭代 15 次、100 次的损失值变化趋势图

观察图 4 中迭代 15 次的具体过程图以及图 5 中迭代 15 次的损失值变化趋势图，我们看到刚开始训练迭代时，模型的准确率比较低，在 0.5 左右，并且波动的幅度较大；但伴随着迭代次数的增加，准确率渐渐趋稳，稳定在了 0.77 左右。训练的损失值基本上从开始就依次递减，验证集的损失值刚开始波动较大，但随着训练的深入，也逐渐呈趋稳和下降的趋势。

观察图 4 中迭代 100 次的具体过程图，发现前 19 次的验证集损失值以及准确率都是不稳定的，都存在着一定程度的波动，并且效果不是很好，准确率在 0.65 左右徘徊。而随着迭代次数的增加，这里看向最后 16 次的迭代过程，无论是训练集损失值，还是验证集损失值，以及准确率，都得到了最稳定以及最好的一个效果，准确率基本上可以稳定在 0.8 左右。从图 5 中迭代 100 次的损失值变化趋势图中可以更加直观地观察到迭代次数对损失值的影响。验证集的损失值刚开始波动较大，究其原因是因为 X 光胸片的差异不是特别大，在刚开始训练时经常会“犯错”，因为肺结核病在 X 光胸片上的特征有时很难分辨，计算机也要通过迭代次数的增加来训练，得到更好的“经验”，图 6 给出了模型迭代 100 次时，在训练中最不容易分辨的 4 张图。观察第一排左边的这幅图，我们知道肺部疾病在 X 光胸片上的效果大多数呈

现“浑浊”的效果，而正常的肺部 X 光胸片大多数呈现“清澈”的效果，但实际上这幅图却是肺部正常的表现。从这个例子可以看出计算机在做分类任务时，不是看一眼就能学会的，需要大量的迭代次数以及有足够多的数据量支撑，才能使训练得到的“经验”更加老道。

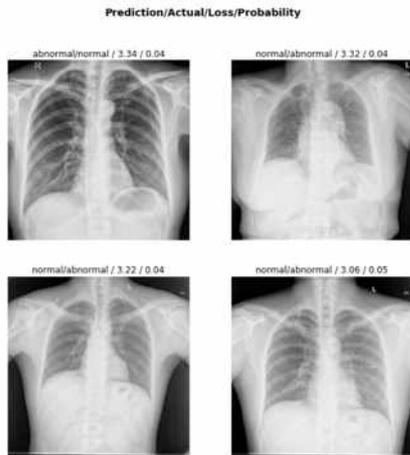


图 6 迭代 100 次时模型最不容易分辨的四张图

### 3.3 采用不同模型训练的实验结果分析

本文采用的是只有 17 层卷积层的 DarkCovidNet 模型进行训练，但在卷积神经网络中，比较有名的几个模型的层数都是很大的，比如 ResNets 或 ResNext 模型。为了探究 DarkCovidNet 模型与其他模型的训练效果，选择了 ResNet34 模型进行对比。2015 年，Facebook 公司的 AI 研究员何恺明以及他的同事提出了 ResNet(Residual Neural Network, 深度残差网络)，“退化”问题是它的关键点之一，即随着模型网络深度的增加，模型网络的学习能力逐渐增强，获得的有效的经验更多，如果再有更深的网络，不允许有比之前更浅的网络更低的模型准确率<sup>[14]</sup>。令迭代次数都为 5 次，学习率都选择 3e-3，进行两个模型的训练，两个模型的训练具体过程及其 precision 等指标对比图如图 7 所示。

epoch	train_loss	valid_loss	accuracy	time	134_005 0.7836567164:70164 [146.26] [ 9.111]	precision	recall	f1-score	support
0	0.609000	0.993407	0.492537	04:40					
1	0.542411	0.744122	0.792537	04:28	abnormal 0.94 0.83 0.76 68 normal 0.72 0.88 0.79 60	0.94	0.83	0.76	68
2	0.513669	0.550952	0.791493	04:39	abnormal 0.88 0.77 0.72 134 normal 0.49 0.77 0.47 134	0.88	0.77	0.72	134
3	0.495734	0.515570	0.760657	04:29	abnormal 0.88 0.77 0.72 134 normal 0.49 0.77 0.47 134	0.88	0.77	0.72	134
4	0.483144	0.515778	0.760057	04:33	abnormal 0.88 0.77 0.72 134 normal 0.49 0.77 0.47 134	0.88	0.77	0.72	134

epoch	train_loss	valid_loss	accuracy	time	134_135 0.6452839520895072 [134.14] [ 7.5911]	precision	recall	f1-score	support
0	1.113987	0.985096	0.701493	05:40					
1	1.004695	0.574111	0.776149	05:10	abnormal 0.83 0.79 0.54 68 normal 0.81 0.99 0.66 60	0.83	0.79	0.54	68
2	0.992232	0.432428	0.783502	05:34	abnormal 0.83 0.79 0.54 68 normal 0.81 0.99 0.66 60	0.83	0.79	0.54	68
3	0.777742	0.618043	0.828906	05:39	abnormal 0.83 0.79 0.54 68 normal 0.81 0.99 0.66 60	0.83	0.79	0.54	68
4	0.690790	0.573859	0.843264	05:40	abnormal 0.83 0.79 0.54 68 normal 0.81 0.99 0.66 60	0.83	0.79	0.54	68

图 7 两个模型的训练过程及 precision 等指标对比图 (上为 DarkCovidNet, 下为 ResNet34)

从图中可以看出，最终 ResNet34 模型的训练准确

率要高一些，达到了 0.84，在 134 张验证集图像中，有 113 张图像被正确分类，并且可以推断随着迭代次数的增加，准确率还应该会有更好的效果。而 DarkCovidNet 模型在相同条件下，最终的准确率仅为 0.77，在 134 张验证集图像中，有 99 张图像被正确分类。在 precision、recall、f1-Score 等指标的表现上，ResNet34 模型的成绩显然也是更好的。但是在训练时间上面，ResNet34 模型体现出了它的劣势，每次迭代都需要 5 分 30 秒以上的时间，而反观 DarkCovidNet 模型，一次迭代只需要 1 分 35 秒左右。得到的这样的训练结果得益于 ResNet34 更深的层次，而 DarkCovidNet 模型的层次要少很多。Tulin Ozturk 等人认为执行这种分类的模型应该有一个能够捕捉和学习微小差异的结构，而不是像 ResNets 或 ResNext 模型那样非常深，很显然，更深的层次也有助于正确分类。迭代次数的增加同时增加了 DarkCovidNet 模型的准确率，但同样需要考虑的是每次迭代的效果，这一方面 ResNet34 做得更好。从时间效率上来说，DarkCovidNet 模型更占优势。

### 4 结束语

在本研究中，我们利用 DarkCovidNet 模型分别采用不同的迭代次数对肺结核病例进行了自动检测训练，迭代次数的增加在一定程度上使模型的训练更加趋稳，但模型训练的准确率还不是非常高，如何在迭代次数增大的同时也更加提高模型的准确率是一个值得研究的问题。还将 ResNet34 模型与 DarkCovidNet 模型在 5 次迭代上做了分析比较，ResNet34 的准确率令人惊喜，但其每次迭代耗费的时间是非常多的。使用更丰富的数据集以及对模型的内核进行有关参数的调整有助于模型的准确率以及时间效率的提高，这也是后面进行有关研究的关键所在。

### 【参考文献】

- [1] Global tuberculosis report 2019. World Health Organization. [http://www.who.int/tb/publications/global\\_report/en](http://www.who.int/tb/publications/global_report/en). 2019.
- [2] 罗一婷, 翁榕星, 周芳, 余卫业, 陆普选. 2019 WHO 全球结核病报告: 全球与中国关键数据分析 [J]. 新发传染病电子杂志, 2020,5(01):47-50.
- [3] 刘洋. 肺结核病灶自动检测方法研究 [D]. 哈尔滨: 哈尔滨工业大学, 2019.
- [4] 陈曼, 胡启飞, 祝绚, 吴民沪, 代娟, 赖翼, 段佳慧, 谭积善. 肺结核快速检测技术对肺结核的早期诊断 [J]. 中华实用诊断与治疗杂志, 2010,24(03):304-305.
- [5] 赵长成. X 线胸片在肺结核诊断的应用价值分析 [J]. 影像研究与医学应用, 2020,4(12):247-248.
- [6] 张效荣. 基于卷积神经网络的深度学习算法与应用研究 [D].

- 西安：西安电子科技大学,2015.
- [7]Tulin Ozturk,Muhammed Talo,Eylul Azra Yildirim,Ulas Baran Baloglu,Ozal Yildirim,U. Rajendra Acharya. Automated detection of COVID-19 cases using deep neural networks with X-ray images[J]. Computers in Biology and Medicine,2020,121.
- [8]Liu J , Liu J , Liu Y , et al. A Locating Model for Pulmonary Tuberculosis Diagnosis in Radiographs[J]. 2019.
- [9]毋一帆 . 基于深度卷积神经网络的 X 光片肺结核检测和分类 [D]. 西安：西北大学 ,2019.
- [10]P.Lakhani,B.Sundaram, 王臣 . 胸部 X 线影像的深度学习：应用卷积神经网络进行肺结核自动分类 [J]. 国际医学放射学杂志 ,2017,40(05):601-602.
- [11]许可 . 卷积神经网络在图像识别上的应用的研究 [D]. 杭州：浙江大学 ,2012.
- [12]任进军,王宁 . 人工神经网络中损失函数的研究 [J]. 甘肃高师学报 ,2018,23(02):61-63.
- [13]刘亚 . 动漫角色智能分析算法研究及应用 [D]. 合肥：安徽大学 ,2018.