

# 基于 Darknet-19 算法的肺炎 X 射线图像应用

徐蔓 李丹

四川大学锦城学院 四川 成都 611731

**【摘要】**约翰霍普金斯大学和“拯救儿童”救助组织在 2018 年根据肺炎的发展趋势做出分析表明：到 2030 年，肺炎将导致近 1100 万名 5 岁以下儿童死亡。肺炎的早期确诊有助于降低死亡率，但是在医学上通过胸部的 X 射线图像确诊肺炎是一项艰巨的任务，需要有临床经验的医生对胸片进行长时间的阅片观察，导致医疗资源匮乏的偏远地区肺炎早期确诊非常困难。除此之外，由于一些肺炎的 X 射线图像极为近似，医生容易判读错误，造成误诊。近年来先进的人工智能 (AI) 技术与放射成像的结合与应用可以帮助人们准确地检测这种疾病，同时也可以帮助解决偏远村庄缺乏专业医生的问题。本文基于深度学习算法中的 Darknet-19 算法对肺炎 X 射线图像展开了研究。首先利用改进后的 Darknet-19 算法进行特征提取，改进后的 Darknet-19 算法包含 16 个卷积层、5 个最大值池化层 (max pooling layers)，处理一张照片需要 55.8 亿次运算，较高的特征提取准确率是为了之后判断患者是否患有肺炎打下坚实的基础；随后使用二分类的方法对患有肺炎和没有患有肺炎的图像进行标记；通过模型对图片的训练，得到判断患肺炎或健康的结果，最后通过训练出来的模型就可以输入图片对是否患有肺炎进行正确的诊断。

**【关键词】**深度学习；Darknet-19 算法；肺炎；胸部 X 线图像

## 1 研究背景

### 1.1 肺炎现状

肺炎是全世界最常见的一种呼吸道疾病，细菌、病毒、支原体、衣原体等病原体细菌很容易让人感染肺炎，其中以病毒和细菌感染最为常见<sup>[1]</sup>。细菌、非细菌感染型肺炎临床表现少，所以很难根据经验做出决定并合理用药。对于应用经验治疗无效的患儿，临床上通常会对这类患者进行痰培养<sup>[2]</sup>，明确致病原因来精准用药。但是，痰培养中微生物繁殖所需时间长，肺炎患者不能及时地接受治疗。

2017 年，党的十九大报告明确提出实施健康中国战略，推动健康中国上升为国家战略。据有关部门统计，2015 年我国城镇居民非正常死亡因素的主要风险来自恶性肿瘤、心脑血管疾病、肺癌等与我们生活密切相关的疾病<sup>[3]</sup>。

在本国，每年有 430 万人被诊断为癌症，其中超过 280 万人死于癌症<sup>[4]</sup>。现阶段我国癌症死亡约占全部死因的 25%，居死因首位<sup>[5]</sup>。肺炎是肺癌的重要表现，肺癌早期的 5 年生存率为 60%，而晚期的 5 年生存率仅为 4%<sup>[6]</sup>。早期肺癌的生存率明显高于晚期肺癌。因此，“早筛查、早确诊、早治疗”是肺癌治疗的关键。

### 1.2 深度学习研究现状

深度学习是人工智能领域的一个重要支系，在视觉、NLP、语音 M 等各个领域被广泛应用。根据深度学习在

多个领域的杰出表现，将其誉为近年来科学领域的十大突破之一<sup>[7]</sup>。截至目前，深度学习的医学影像研究已经涉及到 X 射线、CT 等多种医学图像。AI 在疾病诊断中的应用越来越广泛。埃斯特瓦和其他研究人员通过神经网络模型对 120000 例皮肤癌患病情况进行训练，得到训练模型。通过验证发现，AI 在是否患皮肤癌的判断中的正确率 (69.4%) 与皮肤科医生判断的正确率 (66.0%) 相近<sup>[8]</sup>；西塞罗和其他研究人员通过深度学习对 35,038 张胸部 X 光片进行训练，以获得训练模型。模型预测胸腔积液、肺水肿、肺实变、心影增大、气胸和气胸的灵敏性和特异性分别为 91%、82%、74%、75%、81%、80%<sup>[9]</sup>。Setio 和其他研究人员从开放的数据集中提取了 800 多个肺结节的 CT 图像，并使用深度学习训练模型确定了 85% 的灵敏度<sup>[10]</sup>。哈兹利特等其他研究人员使用三层深度学习网络来减少 315 个参数的维数，检测自闭症的可能性。灵敏度 88%，特异性 95%，准确性 94%。他们还发现，大脑皮层面积增加，患自闭症的可能性也会随之增加<sup>[11]</sup>。这些令人瞩目的初步结果表明，基于深度学习的人工智能在医学影像领域有着非常光明的前景<sup>[12]</sup>。

## 2 材料和方法

### 2.1 X 射线图像数据集介绍

本项目根据由百度 AI 收集到的相关肺炎患者的肺部 X 射线图像，其中训练集数据包括患者 20013 张 X 射

线图图像，400 张正常人群的肺部 X 射线图像；验证集数据包括 6671 张病人的 X 射线图像，400 张健康人群的肺部 X 射线图像。在随后的模型探索中，我们仅使用了部分数据集。在最后实验结果和验证中，我们将训练集中的数据分为了几个不同数目的数据集对模型展开了不同情况的验证。

其中健康人的肺部 X 射线图像适宜黑白显示，胸部轮廓为主的检查胶片。上方可以显示到锁骨，以及双肩关节。下部可以清晰显示肋膈角，第 12 肋骨以及部分腰椎。首先以 12 对肋骨组成的对称性胸廓。中间为纵膈区，其内部可以显示气管、心脏形态组织密度影。双侧为低密度肺组织影，其内见细条状略高密度的肺纹理影，形似树枝分叉样，逐渐变细变淡，图 1 左侧所示。肺炎患者由于各种原因引起的肺部黏膜出现了充血水肿和炎性物质渗出的肺部炎性改变。胸片的特点显示为肺部多部位炎性浸润，增生，钙化，甚至出现空洞等肺部病理性病灶，如图 1 右侧所示：



图 1 胸部 X 射线影像

### 2.2 Darknet-19 算法介绍

Darknet-19，包括 19 个卷积层和 5 个 maxpooling 层。Darknet-19 与 VGG16 模型设计原则是一致的，主要采用 3\*3 卷积，采用 2\*2 的 maxpooling 层之后，特征图维度降低 2 倍，而同时将特征图的 channels 增加两倍。与 NIN(Network in Network) 类似，Darknet-19 最终采用 global avgpooling 做预测，并且在 3\*3 卷积之间使用 1\*1 卷积来压缩特征图 channels 来降低模型的计算量和参数。Darknet-19 每个卷积层之后使用 batch norm 层来加快收敛速度，降低模型过拟合的可能性。Darknet-19 层次结构如图 2。

C<sub>1</sub>-M<sub>1</sub>-C<sub>2</sub>-M<sub>2</sub>-C<sub>3</sub>-C<sub>4</sub>-C<sub>5</sub>-M<sub>3</sub>-C<sub>6</sub>-C<sub>7</sub>-C<sub>8</sub>-M<sub>4</sub>-C<sub>9</sub>-C<sub>10</sub>-C<sub>11</sub>-C<sub>12</sub>-C<sub>13</sub>-M<sub>5</sub>-C<sub>14</sub>-C<sub>15</sub>-C<sub>16</sub>-C<sub>17</sub>-C<sub>18</sub>-C<sub>19</sub>

图 2 Darknet-19 层次结构

### 2.3 基于 Darknet-19 算法的图像研究

首先我们会在 Darknet-19 算法的基础上搭建用于我们提取图像特征的模型，我们先来了解一下卷积层和最大池化层的作用。

卷积层的作用：卷积运算是提取输入图像的不同特

征。卷积层使用不同的卷积核的组合来实现梯度计算、尺度计算（配合 maxpooling 层）等作用，可以对显著性特征进行选择。第一卷积层只能提取边缘等主要特征，多层卷积层可以多次地从低层特征中获取更复杂的特征。但是当卷积层数过多的时候就可能造成梯度消失、梯度爆炸、过拟合等问题，甚至还会使其陷入局部最优。

最大池化层的作用：最大池化层将整个图像分割成大小相同且没有重叠的小块。在每个小块中取最大值，丢弃其他结点后，通过保持原有平面结构得到输出。maxpool 的目的在于关注每个 patch 里面的重要信息，同时减少数据量，对于计算来说是非常有效的。同时也在一定程度上避免整个模型过拟合。但是 Max pooling 也有不好的地方，当 Max pooling 对所有的 Feature Maps 进行等价操作时，其中有些重要的特征会有所遗漏。

该项目在 Darknet-19 算法的基础之上，考虑到卷积层和池化层的数量对图像识别准确率的影响，通过不同卷积层 19 层、18 层、17 层、16 层和 5 层池化层的搭配最后选择出相对最佳的卷积层和池化层数量，分别为 16 层卷积层和 5 层池化层。以训练 2 次后得到的相关参数作为批判标准。在该研究中我们选择了准确率、召回率、f-score、支持数、准确率、宏平均、加权平均作为我们的相关参数。其中准确率具体的结果如图 3 所示：

	19 层	18 层	17 层	16 层
准确率	0.846723	0.840050	0.836726	0.864977
训练时间	17: 02	16:51	12:06	11:60

图 3 不同层数准确率及对比

除此之外，还采用将 3 个卷积层用函数封装成一个块，这使得每个 3 × Conv 层都具有三次连续形式的相同设置。使用批量归一化操作对输入进行标准化，该操作有利于减少训练时间和提高模型稳定性等优点。随后使用 ADAM 优化器进行权重的更新，ADAM 优化器综合了 AdaGrad 和 RMSProp 两种优化算法好的地方。能够做到实现简单、计算高效、内存需求少、自动调整学习率并且参数的改变不受梯度的伸缩变化的影响。最后，对交叉熵损失函数和选择的学习率进行 3e-3、3e-2、3e-1 的调试，相关参数的对比，选择出相对最佳的结果。对比结果如下图 4 所示：

参数	3e-3	3e-2	3e-1	5e-3
准确率	0.857914	0.833818	0.831741	0.424595
训练时间	07:34	07: 44	07: 24	07:31

图 4 不同参数准确率及时间对比

经过不同卷积层、不同学习率的比较后，我们确定了相对最佳的卷积层数和学习率，随后我们通过对比不同训练迭代次数下得到的精确率，来选择最佳的训练次数。具体的比较过程如下图 5 所示：

参数	1	4	5	20
准确率	0.852098	0.869547	0.883673	0.863315
训练时间	07:42	07:31	07:34	08: 42

图 5 不同迭代次数准确率及时间对比

在提取特征的模型确定之后, 将我们的数据集传入模型, 其中有正常人的胸部 X 射线图像和肺炎患者的胸部 X 射像图片, 然后使用二分类的方法进行标记, 分别标记为“0”和“1”。利用模型训练出的关于正常胸部图像的特征模型和患者胸部图像的特征模型就可以对未知病情的 X 射线的胸部图像进行诊断。

### 3 实验结果及分析

本实验中通过卷积层数的变化、学习率的变化以及迭代次数的变化来训练出最佳模型。其中在学习率由  $3e-3$  变化到  $5e-3$  时, 发现训练模型的准确率从 0.86 变为了 0.42, 发生了近一半的降低, 说明并不是随着学习率的增加, 损失的减少, 模型就会训练得越来越好, 过高的学习率还有可能导致过拟合, 让模型越来越差, 准确率下降, 所以选择合适的学习率是训练好模型的关键之一。其次在训练次数的选择上, 我们从 1 次、4 次、5 次到 20 次, 从 1 次到 4 次到 5 次的过程中, 准确率都呈现上升的趋势, 但当我们增加至 20 层的时候, 准确率却发生了下降, 说明学习率和模型的好坏并不是训练次数越多模型越好, 当训练次数过多时模型出现了过拟合, 使得模型的准确率下降。所以训练次数的处理也是训练好模型的关键之一。一个好的模型搭建不能仅仅依靠一个方面的变化来决定, 而应该是模型中各部分的协调, 最终找到相对最佳的模型。

我们将通过测试不同数量的 X 射线图像, 分别为 50 张、900 张、1400 张、2000 张来进行对模型的测试, 其中获得的准确率和时间对比如图 6 所示:

	50	900	1400	2005
准确率	0.8834	0.8901	0.8899	0.8927
训练时间	07: 44	07: 65	08: 32	08: 61

图 6 不同数量图像准确率对比图

### 4 结束语

通过技术的应用, 将会提高肺炎的诊断效率和正确率, 同时不仅推动了我国医疗事业的发展, 还落实了“早筛查、早确诊、早诊疗”的国家政策, 能够提高我国的

癌症患者存活率, 保障各族人民更加健康幸福的生活。让技术不是那么发达的地区也可以通过互联网的帮助下得到高科技高效率的诊断。

该研究在一定范围内, 对相应参数进行调整获得了相对最佳的模型, 但是该论文同样存在着一些需要研究和改进的地方, 该方法训练时间比较长, 以及训练次数没有全面涉及等。可以通过相应的处理来实现, 缩短训练时间, 或采用相应的优化方案实现。除此之外, 本研究仅仅针对肺部疾病的肺炎展开了研究, 若要对肺部的多种疾病进行研究判断, 则需要在二分类的基础之上采取多分类的方法, 可以做为下一步的研究方向。

### 【参考文献】

- [1] 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 中国卫生和计划生育统计年鉴 2016 卷, 2016.
- [2] Chen W, Zheng R, Baade P D. Cancer statistics in China, 2015. CA: a cancer journal for clinicians, 2016, 66(2): 115-132.
- [3] Torre L.A ,Bray F,Siegel RL,et al.Global cancer statistics,2012. CA Cancer JClin,2015, 65(2):87-108.
- [4] 赵鑫, 强彦, 葛磊. 基于双模态深度自编码的孤立性肺结节诊断方法 [J]. 计算机科学, 2017, 44(8): 312-317.
- [5] Lecun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521(7553): 436-444.
- [6] Dermatologist-level classification of skin cancer with deep neural networks. Esteva A, Kuprel B, Novoa R A, et al. Nature, 2017.
- [7] Pulmonary nodule detection in CT images: False positive reduction using multi-view convolutional networks. SETIO A A A, CIOMPI F, LITJENS G. IEEE Transactions on Medical Imaging. 2016.
- [8] Early brain development in infants at high risk for autism spectrum disorder. Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., S-tyner, M., Wolff, J. J. Nature, 2017.
- [9] Deep Learning in Medical Image Analysis. Shen D, Wu G, Suk H I. Annual Review of Biomedical Engineering, 2017.
- [10] 刘士远, 萧毅. 基于深度学习的人工智能对医学影像学的挑战和机遇 [J]. 中华放射学杂志, 2017, (12).
- [11] Tulin Ozturk, Muhammed Talo, Eylül 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.