

doi: 10.3969/j.issn.0490-6756.2020.03.008

基于深层迁移学习的 DR 胸片肺结核病灶检测

胡恒豪¹, 王俊峰², 方智阳¹, 周海霞³

(1. 四川大学计算机学院, 成都 610065; 2. 四川大学空天科学与工程学院, 成都 610065;
3. 四川大学华西医院呼吸与危重症医学科, 成都 610041)

摘 要: 针对基于传统机器学习方法设计的 DR 胸片肺结核检测器存在着泛化能力不强, 实际检测精度低等问题, 提出了一种基于 Focal Loss 的深度学习检测方法 Tuberculosis Neural Net(TBNN). 医学图像的特殊性, 存在带标注的数据量小导致无法充分训练深层网络模型等问题. 该方法利用肺炎和肺结核同为呼吸道感染疾病且在 DR 胸片上有相似表征的特点, 基于迁移学习原理训练特征提取子网络, 减少肺结核胸片样本不足对模型训练造成的影响. 首先在大型的肺炎胸片数据集上训练特征提取网络, 以获取 DR 图像中丰富的深层图像语义信息, 然后使用样本较少的肺结核数据集微调网络参数, 并将多层卷积的输出作为 TBNN 分类子网络的输入, 得到基于 DR 胸片的肺结核病灶检测模型. 实验结果表明, 该方法生成的检测模型在分类精度和性能上均优于基于传统机器学习的肺结核检测器. 在同等训练数据量和训练周期下, 模型性能高于其他采用传统数据增强方法的深层网络肺结核检测算法, 且能标识病灶区域, 准度上有不低于放射科阅片医生的表现.

关键词: 深层迁移学习; 病灶检测; 肺结核; DR 胸片; 计算机辅助诊断

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0490-6756(2020)03-0459-10

Deep transfer learning based pulmonary tuberculosis lesions detection on DR films

HU Heng-Hao¹, WANG Jun-Feng², FANG Zhi-Yang¹, ZHOU Hai-Xia³

(1. College of Computer Science, Sichuan University, Chengdu 610065, China;
2. School of Aeronautics and Astronautics, Sichuan University, Chengdu 610065, China;
3. Department of Respiratory and Critical Medicine, West China Hospital, Sichuan University, Chengdu 610041, China)

Abstract: To find solutions to the problem that the tuberculosis detectors based on the traditional machine learning method are poor generalization ability and low detection accuracy, this paper proposed a deep learning method based on focal loss to detect tuberculosis lesions called Tuberculosis Neural Net (TBNN), which makes full use of pneumonia and tuberculosis as respiratory diseases and has similar characteristics on chest X-ray. During model training, the feature extraction sub-network is pretrained based on transfer learning to reduce the impact of insufficient tuberculosis chest X-ray samples. Firstly, the feature extraction network was trained on a large open pneumonia dataset to obtain the rich deep image semantic features. Secondly fine-tune the model parameters with the tuberculosis dataset and then outputs were used as inputs to the TBNN classification sub-network to calculate tuberculosis lesion detection results. The experimental results show that TBNN is superior to the traditional machine learning methods in classification accuracy and performance. Compared with other deep learning detectors in clas-

收稿日期: 2019-07-25
基金项目: 四川省重点研发项目(18ZDYF2039)
作者简介: 胡恒豪(1994—), 男, 四川宜宾人, 硕士研究生, 主要研究方向为医学影像处理. E-mail: 2017223040101@stu.scu.edu.cn
通讯作者: 王俊峰. E-mail: wangjf@scu.edu.cn

sification accuracy, TBNN is better, and its ability to locate lesion area has reached the level of human radiologists.

Keywords: Deep transfer learning; Lesions detection; Pulmonary tuberculosis; DR films; Computer-aided diagnosis

1 引言

结核病(Tuberculosis, TB)被列为我国法定重大传染病之一,是严重危害人民群众健康的呼吸道传染病.第五次全国结核病流行病学调查结果显示^[1],活动性肺结核乡村患病率(569/10 万)明显高于城镇(307/10 万),西部地区(695/10 万)患病率高于东部(291/10 万)、中部地区(463/10 万).由于西部地区贫困,边远和少数民族地区较多,结核病防治人力资源匮乏等因素增加了结核病防治工作难度.

计算机辅助诊断医学图像一直是国内外研究热点^[2-4].随着深度学习在自然语言处理和图像处理领域的不断发展,越来越多的研究者采用深层神经网络模型处理医学图像的病理分类和病灶检测等问题.谷歌 AI 团队设计深度学习算法分析视网膜图像并识别糖尿病视网膜病变的特征,且在验证集上该算法性能略高于 8 位眼科医生^[5].斯坦福大学研究者利用深层神经网络训练皮肤癌诊断模型,并同 21 位认证皮肤科医生对比测试,结果表明其诊断能力接近皮肤科医生^[6].谷歌和斯坦福大学的研究成果建立在大量经医生标注的数据上,而文献^[7]指出深度医学图像研究面临带标注的数据量小,不能满足大数据分析要求,不能保证模型分析结果的泛化能力等问题.

目前针对数字化 X 线摄影(Digital Radiography, DR)胸片研究的大型公开数据集有美国国立卫生研究院提供的 NIH Chest X-Ray-14^[8]和斯坦福大学吴恩达团队的 CheXpert^[9].其中 Chest X-Ray-14 包含 112 120 个来自 30 805 名不同患者的正面 DR 胸片,每张胸片由 14 种不同的胸部病理标签标注. Rajpurkar 等^[10]在 ChestX-ray14 上设计并训练了 121 层卷积神经网络用于肺炎检测. CheXpert 数据集里面有 224 316 张 DR 胸片,共涉及 65 240 名患者.胸片均是患者在斯坦福医院进行胸部 X 光检查之后的留存,并附有相关的放射学报告.研究者采用自然语言处理技术开发了一种基于规则的自动贴标机,从放射学报告中提取观察结果,为胸片做病理分类.麻省理工学院研究者成

功使用 CheXpert 的自动贴标机为 371 920 张 DR 胸片作标注,准确率能达到专业医生水平,并整理得到 MIMIC-CXR 数据集^[11].虽然以上数据集量级均达 10 万张,且有 14 种病理标签,但没有包含肺结核 DR 胸片.

2 结核胸片相关研究

在欠发达地区,基层卫生院是检测肺结核的第一关口,影像学方法被广泛用于医生临床诊断.当前的肺结核临床诊断方式中,医生通常先观察 DR 胸片来判断患者是否属于肺结核或疑似活动性肺结核,再做进一步诊断.但在实际阅片过程中,由于胸部肋骨、胸膜等重叠组织会造成伪影以及医生的经验差异、阅片水平不一等问题,临床诊断结果存在漏检、误判. Schalekamp 等^[12]研发了计算机辅助系统(Computer-Aided Diagnosis, CAD)辅助诊断肺结核,表明可减小诊断结果的误诊率,显著提高阅片医生的工作效率.兰云旭等^[13]提出基于 Spark 的并行医学图像处理方法,并以肺叶胸片分割算法为实例证明计算机辅助诊断可适用于大规模高并发的医学图像处理场景.

2.1 基于人工设计特征提取的肺结核病灶检测

基于传统机器学习的胸片肺结核病灶检测方法首先通过对 DR 胸片做预处理,提取疑似病变区域的特征,如形状特征、纹理特征、密度特征等,再使用传统机器学习方法训练的分类器来判断候选疑似区域是否为病变区域. Noor 等^[14]提出了一种统计解释技术来检测 DR 胸片中的结核病灶.他们首先对 DR 胸片进行小波变换,然后从小波系数中计算 12 个纹理测量值,再对这些纹理测量执行主成分提取以减少维数.最后,使用概率椭圆和判别函数来估计错误分类概率. Shen 等^[15]基于自适应阈值均值漂移的聚类算法先分割出候选区域,然后使用贝叶斯分类器判断候选区域是否为异常区域. Tan 等^[16]使用统计学特征在 DR 胸片上筛查肺结核取得良好效果.符尧等^[17]采用加速稳健特征算子来检测分隔后的肺实质图像特征点得到疑似病变区域,再利用支持向量机成功对疑似区域做分类判断.

针对基于人工设计特征提取的结核检测,文献[18]指出没有特定的分类器表现最佳,因为每个分类器使用自己的训练集,且纹理特征比几何特征更常用.采用基于人工的特征提取方法处理医学图像,需要算法设计者有丰厚的医学知识储备,才能提取出有足够区分性的高质量特征.而传统的机器学习分类器性能受限于训练样本的质量和分布,容易出现训练样本不够造成过拟合现象,算法鲁棒性低导致模型泛化能力不强等问题.

2.2 基于深度学习的肺结核胸片分类

由于传统的特征提取方法往往会忽略图像的深层语义,深度学习网络能随着层数加深,从图像数据中获取更高阶的特征,可以有效改善检测器的分类性能,避免人工设计特征提取算法的复杂性和局限性. Hwang 等^[19]设计了基于深度卷积网络的 CAD 系统,用于自动 TB 筛查. Lakhani 等^[20]利用 AlexNet 和 GoogLeNet 实现对肺结核胸片的分类,其中分类器在白俄罗斯结核病公共卫生计划和托马斯杰斐逊大学医院收集的非公开胸片数据集上训练 120 个周期后达到了 99% 准确度. 目前公开可用的肺结核胸片数据集^[21]数量有限,不足以充分一个训练用于检测病灶的深层网络,且数据集的正常样本是大多数,病理样本少,导致训大多数时间在训练正常样本. Ghorakavi^[22]在训练前使用 Haar 特征和 LBP 特征提取方法,通过裁剪肺部轮廓进一步关注病灶区域,以加强神经网络对突出结核病特征的关注,改善结核病诊断系统的性能和响应时间.

针对训练如 GoogLeNet 和 AlexNet 的低层卷积神经网络 120 个时期会导致数据过度拟合,且肺轮廓分割并未实际解决训练样本不充足问题,本文提出了利用深层网络实现 DR 胸片肺结核病灶检测与病变区域定位的方法及其基于 RetinaNet^[23]改进的肺结核病灶检测架构 Tuberculosis Neural Net(TBNN). 方法将迁移学习^[24]原理应用模型训练中,保留模型在大型数据集中提取胸片的深层特征,并用于小量的肺结核胸片样本检测. TBNN 模型将训练重心放在肺结核样本上,减弱大量正常样本对模型的反馈. 实验结果证明,本文提出的方法能更好的解决肺结核病灶检测问题.

3 方 法

3.1 基于 FocalLoss 的检测框架

医学数据集中的正常样本是大多数,病理样本

较少,导致训练中的大多数时间用在了正常样本上,病理样本对模型训练的影响较弱. Facebook AI Research 研究者发现造成这种情况的核心问题是检测器在训练过程中会遇到前景和后景类别极度不平衡的问题,提出 Focal Loss (FL)来替代 Cross EntropyLoss(CE)以解决类别不平衡问题,如式(1)所示.

$$FL(p_t) = (1 - p_t)^{\gamma} CE(\hat{y})_i$$
 (1)

其中, $p_t \in [0, 1]$ 代表二分类模型将类别 t 正确分类的能力,如式(2)所示.

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{otherwise,} \end{cases}$$
 (2)

$$CE(p, y) = \begin{cases} -\log(p), & \text{if } y = 1 \\ -\log(1 - p), & \text{otherwise,} \end{cases}$$
 (3)

当 $\gamma = 0$ 时, FL 就是 CE, 当 γ 增加时, 调制系数变大. FL 把训练重点放在更稀疏的样本上,可以有效降低在训练期间大量容易被正确分类的样本给目标检测器所带来抑制影响.

采用 FL 取代常见 one-stage 目标检测架构中的 CE 后,得到基于 FL 的检测框架示意图如图 1, 处理流程如下.

(1) 一个负责计算整个输入图像的卷积特征图的骨干网络. 此部分由一个自底向上前向传播的特征提取网络和一个自顶向下上采样的特征金字塔网络组成, 分别对应图中的 a) 和 b). 网络 a) 通过对输入的多次卷积, 提取出由浅至深层的图像语义, 网络 b) 通过对前馈网络 a) 的结果上采样, 再横向连接 a) 中相同大小的特征图, 并将输出送到子网络.

(2) 两个子网络分别负责对骨干网络输出的分类以及执行卷积得到预测边框的回归. 每一层的计算都是相对独立的, 低层特征和高层特征都能对最后预测结果提供信息.

(3) 使用 FL 作为分类子网络的输出.

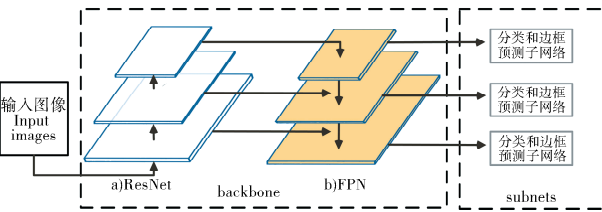


图 1 基于 Focal Loss 的检测框架
Fig. 1 The detection architecture based on focal loss

在此基础上,本文通过实验改进了特征提取网络,并提出针对 DR 胸片的肺结核病灶检测模型

TBNN.

3.2 基于深层迁移学习的病灶特征提取

传统的机器学习只有在训练集数据和测试集数据都来自同一个域和同一分布的时候才能很好运行,但是在实际应用中,这种假设可能不成立.模型的泛化能力以及迁移能力很低,这也是运用传统机器学习方法处理医学图像问题的缺陷之一.根据经验,至少需要 2 万张胸片才能充分训练一个深度神经网络.迁移学习指从一个或多个任务的域中提取知识并将知识应用于目标任务的方法.即一个域 D 由两个部分组成:特征空间 X 以及边缘概率分布 $P(X)$,其中, $X=\{x_1,\cdots,x_n\}\in X$,指某个特定的学习样本;一个任务 T 由输出空间 $Y=\{y_1,\cdots,y_n\}$ 和预测函数 $f(\cdot)=P(Y|X)$ 组成. T 不能被观察,但是可通过训练多组数据 $\{x_i,y_i\},x_i\in X,y_i\in Y$ 被学习.假设有两个域 D_S 和 D_T 以及不同的学习任务 T_S 和 T_T ,迁移学习就是利用 D_S 和 T_S 的知识来改善任务 T_T 的预测函数 $f(\cdot)$ 在域 D_T 中的表现能力.基于肺炎在 DR 影像学上表征和肺结核相似的特点,本文采用图 2 所示的深层迁移学习方法实现 DR 胸片肺结核病灶检测.

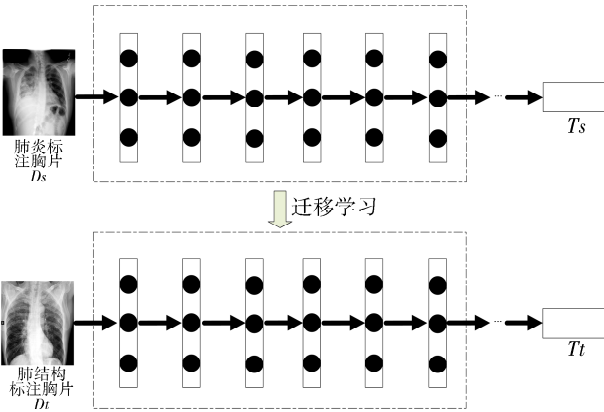


图 2 基于迁移学习的深层特征提取网络

Fig. 2 The deep feature extraction network based on transfer learning

小样本肺结核标注数据 D_T 和目标结核检测任

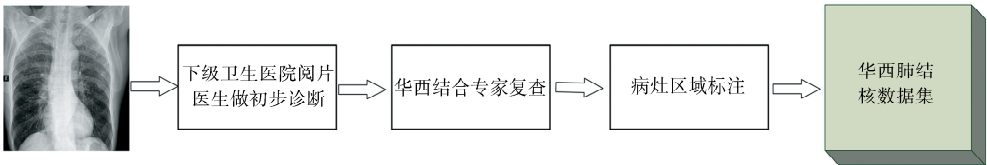


图 3 肺结核标注数据采集

Fig. 3 Labeled tuberculosis dataset collection

务 T_T ,从肺炎检测 T_S 训练肺炎标注数据 D_S 的过程中,学习知识以充分训练深层特征提取网络用于肺结核检测任务 T_T .假设 $D_S=\{(x_{s_1},y_{s_1}),\cdots,(x_{s_n},y_{s_n})\},D_T=\{(x_{t_1},y_{t_1}),\cdots,(x_{t_n},y_{t_n})\},T_S=\{Y_S,P(Y_S|X_S)\},T_T=\{Y_T,P(Y_T|X_T)\}$,其中 $x_{s_i}\in X_S,y_{s_i}\in Y_S,x_{t_i}\in X_T,y_{t_i}\in Y_T$. 深层神经网络提取特征过程如式(4),当域 D_S 中特定数据 $\{x_m,y_m\}$ 通过 $f(x_m)$ 计算后,若 $\hat{y}_m=y_m$,则认为特征提取成功.

$$\hat{y}=f\left(\sum_{i=1}^n w_i x_i-\theta\right) \tag{4}$$

每次训练中更新参数 w 和 θ 的过程看作是一次学习.学习误差函数采用 FL ,优化函数选用随机梯度下降法,如式(5)所示.

$$\begin{cases} V_{t+1}=\mu V_t+\alpha \nabla FL\left(w_t\right) \\ w_{t+1}=w_t-V_{t+1} \end{cases} \tag{5}$$

其中, w_{t+1} 表示第 $t+1$ 次学习后的网络权重; V_{t+1} 表示第 $t+1$ 次学习中网络的更新量;动量 μ 是权重更新量的系数;学习率 α 是权重系数; $\nabla FL\left(w_t\right)$ 是误差的梯度.这样特征提取网络可通过朝梯度负方向更新参数来减小误差以达到学习目的.

4 实验结果与分析

4.1 肺结核标注数据集

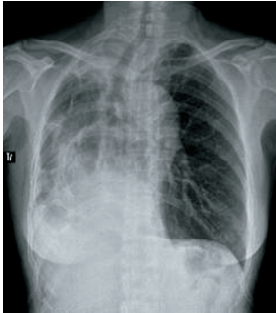
除了训练特征提取网络使用的 RSNA 肺炎数据集(26 684 张 DR 胸片训练样本,3 000 张 DR 胸片测试样本),实验所使用的华西肺结核数据集数据有以下三点特性.

1) 针对性.肺结核标注胸片数据是在重大传染病国家科技重大专项实施中由西南地区 30 余所基层乡镇卫生院在体检筛查时采集.采样地区属于肺结核多发和医疗资源相对缺乏区域.

2) 真实性.由于阅读医学图像需要具备丰富的医学理论知识,我们从收集的胸片中,按图 3 的三级递进式诊断中整理出 1 935 张疑似肺结核胸片由华西医院结核专家做标注.

3) 准确性.除了对肺结核胸片打标签外,还利用矩形框对病变区域做了像素级标注.标注前胸片

已进行了脱敏处理,只保留其图像信息,如图 4. 将标注信息和图像相匹配后,得到表 1 所示的肺结核标注数据集. 其中, X , Y , $Width$, $Height$ 分别表示标注框的左上顶点的坐标以及标注框的宽和高的像素值.



(600 px * 699 px)
图 4 DR 胸部平片 (正位)
Fig. 4 Chest X-ray (PA)

表 1 肺结核标注数据集示例

Tab. 1 The example of labeled tuberculosis dataset		标注信息			
图名	编号	属性			
		X/px	Y/px	Width/px	Height/px
图 4DR 胸部平片	1	367	27	141	220
	2	88	46	144	165

4. 2 特征提取子网络和评价指标

考虑到肺炎和肺结核同为呼吸道感染疾病,在 DR 胸片上有相似的表征,本文采用 RSNA 肺炎数据集^[25]来训练肺结核检测模型的特征提取子网络,比较不同网络对相似检测任务的检测能力. 我

们在 Keras 上选取了 9 种利用 ImageNet^[26]预训练好的网络模型,分别是 VGGNet16,VGGNet19,ResNet101, ResNet152, MobileNet128, MobileNet192, DenseNet121, DenseNet169, DenseNet201^[27-31],并以相同的学习方式,参数优化采用随机梯度下降法 (Stochastic Gradient Descent, SGD),学习率为 1×10^{-2} ,在 RSNA 肺炎数据集上训练 10 个 epoch,每个 epoch 学习 2 500 步并保留最佳的一组参数. 评价指标采用了约登指数 (Youden’s index),表示筛检方法发现真正的病人与非病人的总能力. 越大说明模型实验的效果越好,真实性越大.

$$Youden = Se + Sp - 1 \tag{6}$$

其中, $Se \in [0, 1]$ 是灵敏度 (sensitivity),代表模型将真实的病人判断为有病的能力; $Sp \in [0, 1]$ 是特异度 (specificity),代表模型将没有发病的人判断为无病的能力. 灵敏度和特异度如下式计算.

$$Sensitivity = \frac{N_{tp}}{N_{fn} + N_{tp}} \tag{7}$$

$$Specificity = \frac{N_{tn}}{N_{fp} + N_{tn}} \tag{8}$$

$$Accuracy = \frac{N_{tn} + N_{tp}}{N_{fp} + N_{tn} + N_{fn} + N_{tp}} \tag{9}$$

$$漏检率 = \frac{N_{fn}}{N_{fn} + N_{tp}} \tag{10}$$

其中, N_{tp} , N_{tn} , N_{fp} , N_{fn} 分别指代混淆矩阵中的实际发病且分类正确、实际未发病且分类正确、实际发病却分类错误、实际未发病却分类错误的样本个数,通过计算得到表 2.

表 2 多网络在不同训练周期下对 DR 胸片肺炎检测能力比较(%)

Tab. 2 Architecture’s ability to detect pneumonia on different training epochs (%)		Epoch									
Setting		1	2	3	4	5	6	7	8	9	10
VGG16	Se	802	781	802	728	802	830	857	857	734	872
	Sp	739	793	799	854	835	814	788	786	870	772
	Youden	541	574	601	582	637	644	645	643	604	644
VGG19	Se	120	237	859	784	841	813	915	845	845	837
	Sp	992	980	767	814	769	839	721	805	804	805
	Youden	112	217	626	598	610	651	636	650	649	642
ResNet101	Se	848	795	859	873	809	859	816	869	837	809
	Sp	774	831	766	767	818	763	809	753	779	805
	Youden	622	626	625	640	627	622	625	622	616	614
ResNet152	Se	816	859	848	812	837	795	837	901	820	827
	Sp	763	732	772	802	798	818	791	723	797	811
	Youden	579	591	620	614	635	613	628	634	617	638

(续表 2)

Setting		Epoch									
		1	2	3	4	5	6	7	8	9	10
MobileNet128	<i>Se</i>	763	660	767	841	837	848	855	802	806	788
	<i>Sp</i>	834	895	840	763	784	775	772	810	807	833
	<i>Youden</i>	597	555	607	604	621	623	627	612	613	621
MobileNet192	<i>Se</i>	848	491	611	583	834	827	806	859	841	770
	<i>Sp</i>	739	927	907	920	822	800	803	758	757	841
	<i>Youden</i>	587	418	518	503	656	627	609	617	598	612
DenseNet121	<i>Se</i>	3	23	735	558	756	820	855	809	792	859
	<i>Sp</i>	1000	974	766	879	708	760	723	786	801	723
	<i>Youden</i>	3	204	501	437	464	580	578	595	593	581
DenseNet169	<i>Se</i>	692	784	431	350	749	855	834	806	827	855
	<i>Sp</i>	798	742	931	954	746	740	745	809	755	753
	<i>Youden</i>	490	536	362	304	495	595	579	615	582	608
DenseNet201	<i>Se</i>	18	134	654	661	781	862	852	841	802	862
	<i>Sp</i>	1000	987	833	861	774	747	745	770	777	718
	<i>Youden</i>	18	121	487	522	555	609	597	611	579	580

表 2 中网络在每一周期得到的 $(1 - Sp, Se)$ ，即为当前参数下 ROC 曲线上的截断点. 如果将同种网络所有截断点连线，如图 5 所示，可以发现 VGG19 整体表现优于 VGG16； MobileNet192 和

MobileNet128 表现相近,前者在表 2 中的 *Youden* 指数更大 ($0.656 > 0.623$)； ResNet101 略优于 ResNet152；DenseNet 整体表现差于其他三种，且抖动幅度大，曲线不平滑.

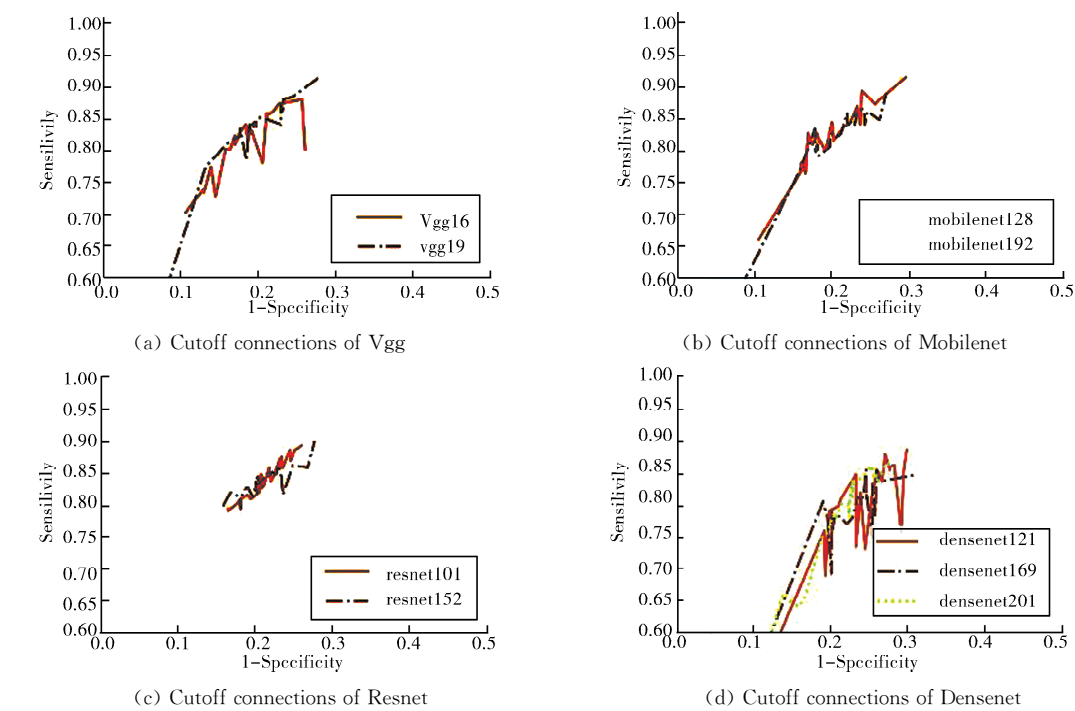


图 5 各网络在不同训练周期下的 ROC 曲线截断点连线

Fig. 5 The connections of ROC cutoffs on different training epochs with various architecture

图 5 所示 VGG19、ResNet101 以及 MobileNet192 的曲线更为平滑，性能更优，DenseNet 整体性能较差，且 MobileNet 相较于 VGG、ResNet 最大 *Youden* 值更大. 为了比较各网络在肺结核检测任务中的表现，我们分别将三种网络替换肺结核检测框架中骨架的特征提取网络部分，再通过迁移

学习的方法训练得到检测模型 TBNN-vgg、TBNN-mobile 和 TBNN-res。其中 TBNN-mobile 为图 6 所示的 TBNN,特征提取子网部分网络采用了和 MobileNet^[27]前 13 层相同的网络结构,其中 cov5,cov11 和 cov13 分别对应文献[27]中 Table1 的第 5 步,第 11 步和第 13 步卷积操作。卷积得到的 3 个特征图分别作为 RetinaNet^[23]中的 Cov3,Cov4,Cov5。在 FPN 部分得到 5 个特征图 P3-P7,并将每一个特征图送到分类和病灶区域预测子网计算结果。

此次模型迁移学习中用到的训练集和验证集图像数据全部来自于华西肺结核数据集,其中训练集样本数量 1886 张,验证集样本数量 155 张,测试集样本数量 143 张。以学习率 1×10^{-2} 训练 10 个 epoch 后得到各模型在验证集上分类结果的 ROC

曲线和 PR 曲线,如图 7 所示。其中 TBNN-mobile 的 AUC 值大于 TBNN-resnet 和 TBNN-vgg,且在 PR 曲线中的平衡点(BEP)值更高。以上实验结果表明,相较于其他主流的网络模型架构,MobileNet 网络更适合作为 TBNN 的特征提取子网。

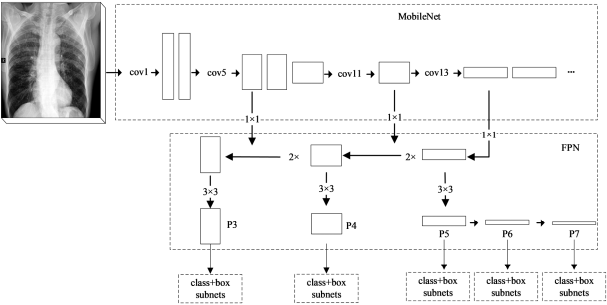


图 6 TBNN 肺结核病灶检测模型
Fig. 6 TBNN Tuberculosis lesions detection model

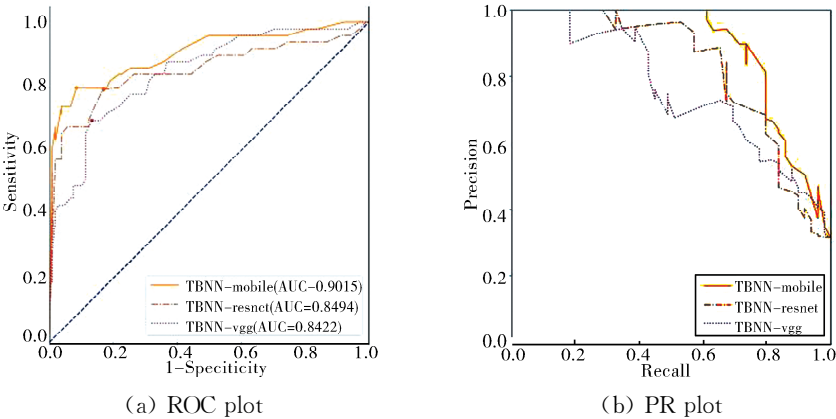


图 7 不同特征提取网络的分类结果 ROC 曲线(a)和 PR 曲线(b)

Fig. 7 ROC plot and PR plot for tuberculosis detection with different feature extraction network

4.3 迁移学习前后的检测结果与分析

避免模型在迁移学习过程中取巧地利用已经习得的肺炎特征来对肺结核做病灶检测,产生负迁移现象,我们比较了迁移学习前后的 TBNN 模型在肺结核测试集上的表现。统计结果如表 3 所示,

表 3 迁移学习前后模型对测试集的检测结果
Tab. 3 Model detection results of pre-transfer learning and post-transfer learning

时间	N_{tp}	N_m	N_{fp}	N_{fn}
迁移学习前	2	131	2	8
迁移学习后	10 ↑	125	8	0 ↓

可以看出迁移学习前仅凭借肺炎特征并不能很好检测出测试集中的肺结核胸片,漏检率达到 80%,而迁移学习后测试集中 10 张含肺结核病灶的胸片均被成功检测。

4.4 模型评估

比较了基于单个或多个手工构造的特征做特征提取的方向梯度直方图特征(HOG)算法、基于均值漂移聚类(HKG)、统计特征和 SURF 算法的检测方法以及利用神经网络做特征提取的 TBNN 检测模型对同组 DR 胸片做肺结核分类的性能,结果显示 TBNN 相比于基于人工构造特征的传统机器学习检测方法有更高的分类精度,性能上也更优。

表 5 为基于深度学习的肺结核检测算法 TBNet^[22]和 TBNN 在同等规模训练集下,训练 10 个 Epoch 后模型性能比较。其中 TBNet 使用了基于 Haar 特征和 LBP 特征的多重显著特征数据增强方法,从原始的 800 张胸片中裁剪出敏感区域和噪声加入训练,以达到充分训练模型且提高分类准确率。TBNet 的训练集、验证集和测试集直接或间接

来自同源数据集^[21],本文提出的 TBNN 在训练阶段采用和 TBN_{et} 规模相近总量更少的训练集和验证集,测试集为文献^[21]提供的所有胸片. 结果显示,相同训练周期下,TBNN 所需训练样本更小、测试集更大,模型准确性更高,且模型具有较强的泛化能力.

表 4 本文方法与基于人工构造特征的检测方法对比
Tab. 4 Compared withartificial feature extraction and detection methods

	HKG ^[15]	统计特征 ^[16]	SURF ^[17]	TBNN
Accuracy%	77. 67	79. 84	82. 27	89. 68
AUC	0. 790 2	0. 824 9	0. 851 5	0. 901 5

为了比较肺结核病灶检测模型的在实际情景中的检测能力,我们邀请了三位来自放射科的人类阅片医生和 TBNN 对同一组胸片做肺结核分类,其中三位医生都工作在图 3 所示数据采集流程中不同地区的基层卫生院,拥有丰富的阅片经验,参与测试的时间和地点相互独立,测试的 DR 胸片选自数据采集工作得到的结核标注数据集,且未参与 TBNN 训练. 为了尽量减小其他因素对阅片过程的影响,测试前没有向医生告知 TBNN 的任何细节.

表 5 本文方法和基于多重显著特征的数据增强方法对比
Tab. 5 Compared with multiple salient-feature based data augmentation method

Setting	模型	
	TBN _{et}	TBNN
训练集/张	255 7	188 6
验证集/张	150	155
测试集/张	412	800
Validation Accuracy/%	81. 33	89. 68
Testing Accuracy/%	75. 42	85. 38
是否标识病灶	×	√

表 6 本文方法和单个阅片医生检测同组 DR 胸片结果比较
Tab. 6 Compared with single radiologist

阅片医生	N_{tp}	N_m	N_{fp}	N_{fn}	Sensitivity	Specificity
医生 A	33	88	18	16	0. 67	0. 83
医生 B	33	102	4	16	0. 67	0. 96
医生 C	34	103	3	15	0. 69	0. 97
TBNN	41	98	8	8	0. 84	0. 92

统计结果由表 6 所示,基于深度迁移学习的 TBNN 肺结核检测模型在对 DR 胸片的检测上,灵敏度更高,即漏检率更低,特异性与人类医生相近. 三位人类阅片医生的 N_{tp} 数值大小几乎相同,即能正确检测结核阳性胸片的水平接近,但通过混淆矩阵分析相互检测结果差异(+表示分类阳性,-表示分类阴性),如表 7 所示,医生 A 和医生 B、医生 C 分别在 30 和 38 张胸片上检测结果不一致,整体检测结果接近的医生 B($Se=0. 67, Sp=0. 96$)和医生 C($Se=0. 69, Sp=0. 97$)也存在 16 张结论相反的胸片.

表 7 医生阅片结果混淆矩阵
Tab. 7 Confusion matrix of radiologists' reading results

阅片医生	医生 A		医生 B	
	+	-	+	-
医生 A	+	•	•	29
	-	•	•	8
医生 C	+	25	12	29
	-	26	92	8

表 8 本文方法和阅片医生组合检测同组 DR 胸片结果比较
Tab. 8 Compared with combinationsof radiologists

阅片医生	N_{tp}	N_m	N_{fp}	N_{fn}	漏检率	Accuracy
医生 A+B	40	87	19	9	0. 18	0. 82
医生 A+C	43	87	19	6	0. 12	0. 84
医生 B+C	39	100	6	10	0. 20	0. 90
医生 A+B+C	44	85	21	5	0. 10	0. 83
两人组平均	40. 7	91. 3	14. 7	8. 3	0. 17	0. 85
整体平均	41. 5	89. 8	16. 2	7. 5	0. 15	0. 85
TBNN	41	98	8	8	0. 16	0. 90

如果只要有一位医生结论是结核阳性,则将该胸片判别为阳性,以减弱单个医生检测水平经验对结果的影响,减低漏检率. 表 8 统计了多位医生所有可能组合的检测结果,其中医生 A 存在的两人组检测准确率相对较低,但能检测出相对更多的阳性患者,医生 B 和医生 C 的组合有最高的准确率,但漏检率也最高. 结合表 6 单个医生检测结果,TBNN 的 N_{tp} 大于单个医生,即 TBNN 漏检率比单个医生更低,且低于两人组平均值,略高于整体平均值,其准确率高于所有组合结果.

阅片医生 A 除了对测试胸片进行分类,还通过标注工具标注出疑似病灶区域. 图 8 所示为医生

A和TBNN对同一张DR胸片的肺结核病灶区域的标注结果比较,其中,图8(a)为华西肺结核数据集的真实结果;图8(b)为医生A通过标注工具手工标注结果;图8(c)为TBNN在对DR胸片做肺结核分类时的自动标注结果。考虑到医生会通过临床表现和生化数据等多维度来诊断,TBNN在实际中的诊断表现已经有不低于医生的水平。



(a) Ground Truth (b) Doctor A (c) TBNN
图8 医生A和TBNN标注同张DR胸片的肺结核病灶区域

Fig. 8 The boxes of labeling the same DR film by Radiologist A and TBNN

5 结 论

由于基于人工构造特征的传统机器学习的肺结核检测方法存在着泛化能力不高,无法挖掘图像深层语义等问题,本文启发性利用肺炎深度网络检测模型做迁移学习来训练针对DR胸片的特征提取子网络。为了减弱容易分类的正样本对模型训练的影响,提出了基于Focal Loss目标检测架构的肺结核病灶检测模型TBNN。实验结果证明,本文提出的基于深层迁移学习的肺结核病灶检测模型在多种分类器评价指标上均高于以往的基于人工构造特征的传统机器学习检测方法,并且在真实测试集上有不低于放射科阅片医生的诊断表现。相较于传统数据增强手段的深度学习的肺结核分类算法,我们提出的方法能充分利用样本特征和更快的收敛,并且在训练速度和模型分类精度等能力上更优。在深层神经网络训练中,数据集样本数量和质量对模型性能好坏起着决定性作用。利用小量肺结核胸片样本调参可以提高模型性能,但并不能解决网络训练问题。我们为了充分训练TBNN的特征提取子网络使用了大型肺炎数据集,造成TBNN对肺结核检测更依赖于肺炎相似的图像语义特征,深层网络的黑盒属性会造成无法解释的误检和漏检情况,如何解决这类问题将是下一步的研究方向。

参考文献:

- [1] 全国第五次结核病流行病学抽样调查办公室. 2010年全国第五次结核病流行病学抽样调查报告[M]//全国第五次结核病流行病学抽样调查资料汇编. 北京:军事医学科学出版社,2011.
- [2] 凌语,孙自强. 基于卷积神经网络的乳腺病理图像识别算法[J]. 江苏大学学报:自然科学版,2019,40: 573.
- [3] 王伟,司森森,陈芙蕖,等. 低倍率镜检图像无标记红白细胞识别方法研究[J]. 重庆邮电大学学报:自然科学版,2019,31: 579.
- [4] 李清亮,张子鹏,时玮淞,等. 基于空间像素强度的脑瘤图像检索方法[J]. 吉林大学学报:理学版,2018,56: 214.
- [5] Gulshan V, Peng L, Coram M, *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs[J]. J Am Med Assoc, 2016, 316: 2402.
- [6] Esteva A, Kuprel B, Novoa R A, *et al.* Dermatologist-level classification of skin cancer with deep neural networks [J]. Nature, 2017, 542: 115.
- [7] Greenspan H, Van Ginneken B, Summers R M. Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique [J]. IEEE T Med Imaging, 2016, 35: 1153.
- [8] Wang X, Peng Y, Lu L, *et al.* ChestX-Ray8: hospital-scale chest X-Ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos: IEEE Computer Society, 2017.
- [9] Irvin J, Rajpurkar P, Ko M, *et al.* CheXpert: a large chest radiograph dataset with uncertainty labels and expert comparison [J]. arXiv preprint, 2019, arXiv: 1901.07031.
- [10] Rajpurkar P, Irvin J, Zhu K, *et al.* CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning [J]. arXiv preprint, 2017, arXiv:1711.05225.
- [11] Johnson A E W, Pollard T J, Berkowitz S J, *et al.* MIMIC-CXR: A large publicly available database of labeled chest radiographs [J]. arXiv preprint, 2019, arXiv: 1901.07042.
- [12] Schalekamp S, Van Ginneken B, Koedam E, *et al.* Computer-aided detection improves detection of pulmonary nodules in chest radiographs beyond the

support by bone-suppressed images [J]. Radiology, 2014, 272: 252.

[13] 兰云旭, 王俊峰, 唐鹏. 基于 Spark 的并行医学图像处理研究 [J]. 四川大学学报: 自然科学版, 2017, 54: 65.

[14] Noor N M, Rijal O M, Yunus A, *et al.* A statistical interpretation of the chest radiograph for the detection of pulmonary tuberculosis [C]// Proceedings of the Biomedical Engineering & Sciences. Piscataway: IEEE, 2011.

[15] Shen R, Cheng I, Basu A. A hybrid knowledge-guided detection technique for screening of infectious pulmonary tuberculosis from chest radiographs [J]. IEEE T Biomed Eng, 2010, 57: 2646.

[16] Tan J H, Rajendra A U, Tan C, *et al.* Computer-assisted diagnosis of tuberculosis: A first order statistical approach to chest radiograph [J]. J Med Syst, 2012, 36: 2751.

[17] 符尧, 王俊峰, 高琳, 等. 基于 SURF 的肺结核 DR 图像病变区域检测[J]. 计算机工程与设计, 2015, 36: 2447.

[18] Jaeger S, Karargyris A, Candemir S, *et al.* Automatic screening for tuberculosis in chest radiographs: a survey [J]. Quant Imag Med Surg, 2013, 3: 89.

[19] Hwang S, Kim H E, Jeong J, *et al.* A novel approach for tuberculosis screening based on deep convolutional neural networks [C]// Proceedings of the Medical Imaging: Computer-aided Diagnosis. Bellingham: International Society for Optics and Photonics, 2016.

[20] Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of, pulmonary tuberculosis by using convolutional neural networks [J]. Radiology, 2017, 2017: 162326.

[21] Jaeger S, Candemir S, Antani S, *et al.* Two public chest X-ray datasets for computer-aided screening of pulmonary diseases [J]. Quant Imag Med Surg, 2014, 4: 475.

[22] Ghorakavi R S. TBNet: pulmonary tuberculosis diagnosing system using deep neural networks [J]. arXiv preprint, 2019, arXiv: 1902.08897.

[23] Lin T Y, Goyal P, Girshick R, *et al.* Focal loss for dense object detection [J]. IEEE T Pattern Anal, 2017, 99: 2999.

[24] Pan S J, Yang Q. A survey on transfer learning [J]. IEEE T Knowl Data En, 2010, 22: 1345.

[25] Radiological Society of North America. Rsn-pneumonia-detection-challengedataset [EB/OL]. (2018-02-23)[2019-03-23]. <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data>.

[26] Russakovsky O, Deng J, Su H, *et al.* Imagenet large scale visual recognition challenge [J]. Int J Comput Vis, 2014, 115: 211.

[27] Howard A G, Zhu M, Chen B, *et al.* MobileNets: efficient convolutional neural networks for mobile vision applications [J]. arXiv Preprint, 2017, arXiv: 1704.04861.

[28] Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2014.

[29] He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition [J]. arxiv Preprint, 2015, arXiv: 1512.03385v1.

[30] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [C]//International Conference on Learning Representations 2015. San Diego: Arxiv, 2015.

[31] Huang G, Liu Z, Maaten L V D, *et al.* Densely connected convolutional networks [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Computer Society, 2017.

引用本文格式:

中 文: 胡恒豪, 王俊峰, 方智阳, 等. 基于深层迁移学习的 DR 胸片肺结核病灶检测[J]. 四川大学学报: 自然科学版, 2020, 57: 459.

英 文: Hu H H, Wang J F, Fang Z Y, *et al.* Deep transfer learning based pulmonary tuberculosis lesions detection on DR films [J]. J Sichuan Univ: Nat Sci Ed, 2020, 57: 459.