

基于 Faster-RCNN 的结核杆菌自动检测方法 研究与应用

卞景帅¹, 卢家品¹, 罗月童¹, 张敏²

(1. 合肥工业大学可视化与协同计算研究室, 安徽 合肥 230009;
2. 合肥思润生物科技有限公司, 安徽 合肥 230601)

摘 要: 染色处理可使结核杆菌在显微镜拍摄的医学图像中显现, 医生通过检测图像中的结核杆菌辅助诊断结核病。近年来卷积神经网络(CNN)在目标检测上取得了突破性进展, 但结核杆菌存在图像上尺度小, 构造标注数据难, 不适用迁移学习等问题, 使得基于 CNN 的目标检测方法在结核杆菌检测方面尚存在一定的困难。为此, 以 Faster-RCNN 目标检测算法为基础, 研究在医学图像上的结核杆菌检测问题。针对结核杆菌尺度小, 提出重叠子图划分策略; 针对标注数据构造难, 提出分块、迭代标注策略。实践证明, 该方法有较高的准确度以及可接受的速度, 已构建了 13 261 个结核杆菌的训练数据, 应用于合作单位的医疗检测产品, 能满足实际应用需求。

关 键 词: 小目标检测; 医学图像; 结核杆菌; CNN

中图分类号: TP 391

DOI: 10.11996/JGj.2095-302X.2019030608

文献标识码: A

文章编号: 2095-302X(2019)03-0608-08

Research and Application of Faster-RCNN Based M. Tuberculosis Detection Method

BIAN Jing-shuai¹, LU Jia-pin¹, LUO Yue-tong¹, ZHANG Min²

(1. Visualization & Cooperative Computing, Hefei University of Technology, Hefei Anhui 230009, China;

2. Hefei Sirun Biological Technology Co. Ltd, Hefei Anhui 230601, China)

Abstract: Through sputum-smear staining, mycobacterium tuberculosis can be shown on microscope image, which makes it possible to detect M. tuberculosis on the image for facilitating tuberculosis diagnosis. On the microscope image, M. tuberculosis is characterized with diverse color saturation, various shape, and undistinguishable appearance confused with background, which make it a great challenge for traditional object detection methods. As convolutional neural networks (CNN) has achieved great success in object detection recently, we study CNN-based method, for instance, Faster-RCNN for M. tuberculosis detection. Nevertheless, there are still some problems with CNN-based tuberculosis detection: a) Size of M. tuberculosis on image is too small, b) Constructing enough accurate labeled data is difficult, and c) Transfer learning does not work for tuberculosis detection. All of those make it hard to apply CNN-based method to M. tuberculosis detection directly. To overcome these problems, we adopt two strategies. We present overlapping sub-image partition strategy for the small-size problem caused by anchor structure which is component of prevalent

收稿日期: 2018-07-02; 定稿日期: 2018-07-17

第一作者: 卞景帅(1992-), 男, 安徽阜阳人, 硕士研究生。主要研究方向图像处理与模式识别。E-mail: bianjingshuai@163.com

通信作者: 罗月童(1978-), 男, 安徽合肥人, 教授, 博士, 硕士生导师。主要研究方向为数据可视化、可视分析、计算机图形学。

E-mail: ytluo@hfut.edu.cn

CNN-based object-detection method. The partition strategy overlappingly partition raw image into sub-images as per a formula presented by us. After partitioning, the proportion of M. tuberculosis on input image of model have been increased, that improving detecting accurate but reducing detecting speed. According to practice, we deem it acceptable. By cooperating with the co-author, 13 261 labeled data of M. tuberculosis have been constructed. Through a series of experiments, it has proved that our method is effective not only in improving detecting accurate and generalization of the model, but also in reducing necessary labeled data. The methods have been integrated into medical inspection products and confirmed to satisfy practical application requirements.

Keywords: small target detection; medical image; M. Tuberculosis; convolutional neural networks

据报道,我国肺结核的发病人数和死亡人数一直居甲乙类传染病之首,我国也是全球 22 个结核病负担最高的国家之一^[1],因此我国对高效的结核病检测手段有迫切需求。检测结核杆菌是诊断肺结核病的重要依据。对结核杆菌进行染色,然后观察确定其数量是进行肺结核诊断的基本方法。传统检测由医生通过显微镜观察染色结核杆菌,存在工作量大、医生容易疲劳,易漏检等问题。因此通过特殊设备拍摄结核杆菌的染色图像,基于目标检测技术自动检测、统计结核杆菌的数量,进而实现自动、智能诊断具有重要的应用价值,是医学检测的重要发展方向,也正是本文所要研究的内容。

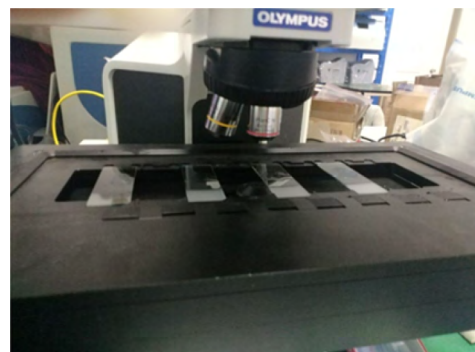
目标检测(object detection)广泛存在于人脸识别^[2]、手势识别^[3]、自动驾驶等任务中,是机器视觉领域最基础、最重要的技术之一。视觉检测包括候选区域提取和目标类别判别^[4] 2 个步骤,其中目标类别判断是各种检测方法之间的主要差别所在。在深度学习方法出现之前,研究人员构造如 HOG^[5-6] (histogram of oriented gradient)、LPB^[7] (local binary pattern)等特征,并使用规则匹配、支持向量机、人工神经网络等方法进行分类识别。但人工构造特征往往受目标图像的多变性等因素困扰,很多情况下检测效果并不理想。结核杆菌,存在尺度小、形态多变等特点,基于传统的检测方法效果不理想,因此本文探索使用卷积神经网络(convolutional neural networks, CNN)进行结核杆菌检测。

Alexnet 在 ImageNet 竞赛上,将分类的错误率从 26%降至 15%,展示出 CNN 在图像识别领域中的巨大优势^[8]。此后 CNN 逐渐应用于各种视觉识别领域中,其发展推动了目标检测技术的进步。CNN 的一个重要特点是通过训练能提取多层次特征^[9],无需人工设计特征,较为有效地避免了人工设计的特征对原始图像表达得不完全的不足,同时实践证

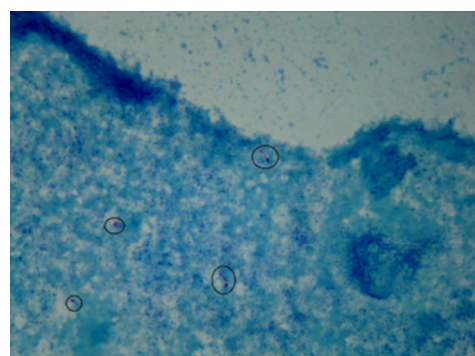
明这一特性取得更高的分类检测准确度。主流的目标检测技术有 Faster-RCNN^[10]、R-FCN(ResNet 版本)^[11]、SSD^[12]、YOLOv2^[13]。本文在分析结核杆菌检测问题和各种检测方法特点的基础上,以 Faster-RCNN 为基础开展结核杆菌的检测方法研究。

1 结核杆菌图像与检测方法特点分析

诊断结核病时,需收集就诊者的痰液并将其放入载玻片中,然后通过特定试剂对其染色,图 1(a)为使用显微设备进行拍摄制作出结核杆菌图像。本文实验所用结核杆菌图像的分辨率为 1920×1200,结核杆菌在图像中呈现偏紫的颜色,如图 1(b)中圆圈所标记的即为结核杆菌。



(a) 载玻片与拍摄设备



(b) 设备拍摄的图像

图 1 待测图像来源

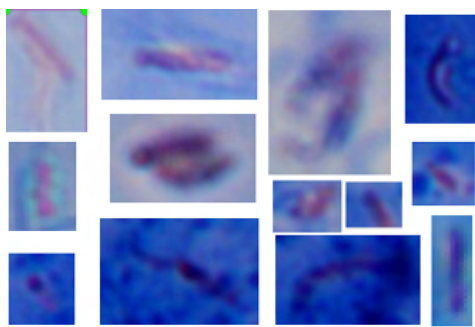
通过对大量图像的观察及与领域专家的交流, 本文总结出结核杆菌在图像中呈现如下特点:

(1) 尺度较小, 且在一定范围内变化。对 1 900 张图像的 3 657 个结核杆菌样本进行分析, 发现其在图像上的区域长度大多在 40 像素左右, 最小为 11 像素, 最大为 135 像素。其中最小 11 像素的结核杆菌相对于原始图像 1920×1200 的分辨率而言过小。

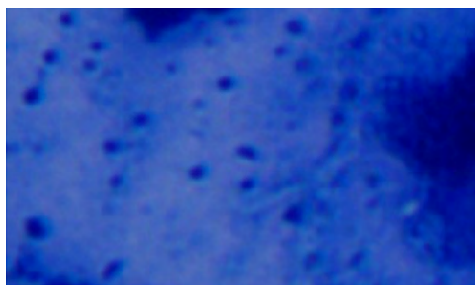
(2) 结核杆菌形状多变。图 2(a)展示了多种不同形态的结核杆菌, 并未呈现出特别的规律。

(3) 部分结核杆菌与无结核杆菌的区域类似。图 2(b)为无结核杆菌的区域, 观察易得, 图 2(a)的某些结核杆菌容易和图 2(b)部分区域混淆。

(4) 颜色是重要特征, 但形态也是判断依据。结核杆菌会呈现紫色是检测的重要依据, 但医生还会结合一些诸如形状、背景等其他特征综合判断。



(a) 结核杆菌样例



(b) 无结核杆菌区域样例

图 2 结核杆菌图像示例

因为结核杆菌图像特征复杂多变, 难以准确刻画, 使得基于颜色、形态等特征的传统检测方法效果并不理想。而基于 CNN 的方法可以从样本数据中学习特征, 能提取出与数据更相符的特征, 进而获得更高的检测准确度。

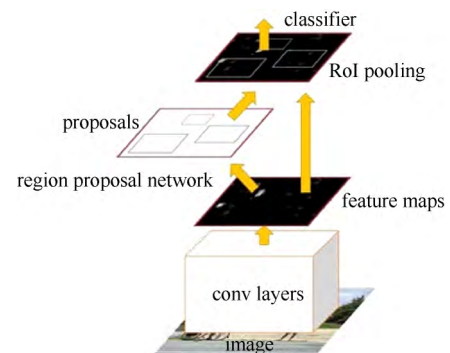
基于 ResNet^[14]的目标检测算法是目前效果最好的方法, 但其需要大量高质量训练数据, 构建大量结核杆菌训练数据非常困难。迁移学习技术是解决训练数据不足的有效手段, 但前提是需要

迁移的源数据和目的域数据比较类似^[15], 而结核杆菌图像难以满足该条件, 因此本文未使用基于 ResNet 的检测方法。VGG16^[16]网络的结构相对简单, 不需要大量的训练数据即可达到较高的准确度。文献[17]指出在没有预训练的情况下 ResNet 比 VGG16 的性能差很多, 基于以上分析, 选择 VGG16 网络更合适。SSD^[12]和 YOLO2^[13]2 种方法在小目标的检测上表现不佳。因此本文选择小目标检测上表现相对较好的 Faster-RCNN, 作为基本的目标检测方法。

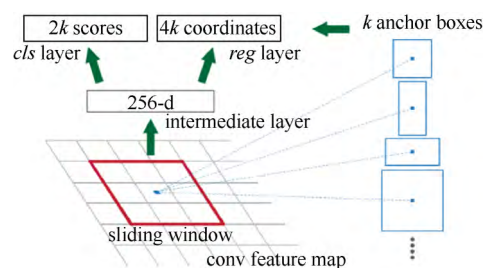
2 基于重叠子图的肺结核杆菌检测

2.1 Faster-RCNN 简介与分析

Faster-RCNN 的基本结构如图 3(a)所示, 输入图像经过底部 CNN (conv layers)提取特征得到特征图(feature map), RPN (region proposal network)利用特征图生成候选区域, 用分类器(classifier)对 RPN 提取出的候选区域进行分类, 最后通过判断候选区域中是否含有目标实现目标检测。



(a) Faster-RCNN 基本结构



(b) anchor 结构

图 3 Faster-RCNN 结构^[10]

将 Faster-RCNN 直接应用在肺结核图像检测上, 表现较差, 大部分结核杆菌均不能被识别出来。经研究发现, Faster-RCNN 的 RPN 使用名为锚(anchor)的结构生成候选区域, 该结构如图 3(b)所示。输入图像经过底部 CNN 以后, 得到长宽为 $M \times N$ 的特征图, 其蕴含的信息是其对应感受野(receptive

field)的信息^[10]。若将输入图像无重叠划分成 $M \times N$ 个小格, 则得到的特征图上每个点的感受野为该点对应的 CNN 覆盖的输入图像的区域, 感受野的中心点为划分小格的中心。Faster-RCNN 用预定义的 k 个固定大小的锚盒(anchor boxes)来决定使用某个中心点的感受野中的区域生成候选区域, 再通过分类器判断这些区域中哪些为候选区域。

图 4 中粗线代表划分网格边界, 细框代表锚盒, 粗线代表结核杆菌。当定义的锚盒尺寸小于划分格的尺寸, 如图 4 细框 a, b 所示, 则锚盒与锚盒之间会出现一定的间隙, 形成不检测区域, 造成漏检目标落入该区域。当定义的锚盒尺寸大于划分格的尺寸, 如图 4 中的细框 c 所示。则由于目标过小, 前景与背景的比例过小, 造成误判, 从而造成漏识别。

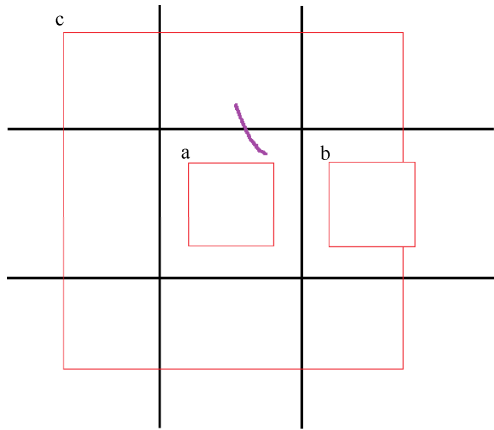


图 4 锚盒与小目标与漏识别

提高小目标的识别准确度有 2 种思路: ①缩小划分的小格尺寸。划分小格的尺寸是由 CNN 的结构所决定, 因此通过改变神经网络结构来缩小划分小格的尺寸, 必然会导致网络其他性能的改变。如何选择最适合的划分小格尺寸是一个复杂的最优化网络结构问题。②用一定的方法(如高斯金字塔)

来放大结核杆菌在输入网络图像中的占比, 通过牺牲速度来换取精度。本文选择后一种思路。

2.2 基于重叠策略的子图生成

本文考虑将大的原始的检测图片分成多个大小相等的子图片, 目的是为了便于后续的批处理。将原始图像划分成子图时, 易将某个结核杆菌被划分到 2 个子图中, 将会对结果造成影响。已知结核杆菌区域的长度有上界, 故采用重叠策略生成子图, 同时需考虑子图的分辨率问题。分辨率过大, 可能导致结核杆菌检测不到; 分辨率过小可能导致检测结果无法满足实时性要求。同时还需考虑到重叠区域, 因此需要通过一定的方法来生子图, 解决该问题。

图 5 为有重叠区域的滑动窗口提取子图过程。重叠区域的宽度略大于结核杆菌的最大尺寸, 度量尺寸的基本单位为像素。 W_r 为原始图像的宽; H_r 为原始图像的高; S_{rmax} 为结核杆菌的最大尺寸; R 为生成子图的行数; C 为生成子图的列数, M_w 为滑动窗口横向移动的步长; M_h 为滑动窗口纵向移动的步长, 计算如下

$$M_w = \frac{W_r - S_{vmax}}{C} \tag{1}$$

$$M_h = \frac{H_r - S_{vmax}}{R} \tag{2}$$

度量尺寸的基本单位为像素, 结果用进一法取整。子图的宽 W_s 和高 H_s 分别为

$$W_s = M_w + S_{vmax} \tag{3}$$

$$H_s = M_h + S_{vmax} \tag{4}$$

则第 $i+1$ 列第 $j+1$ 行个子图相对原图的坐标范围可表示为

$$\{(x_s, y_s)_{j+1, i+1} \mid x_s \in [i \times M_w, i \times M_w + W_s], y_s \in [j \times M_h, j \times M_h + H_s]\} \tag{5}$$

其中, $i \in [0, C-1], j \in [0, R-1]$ 。

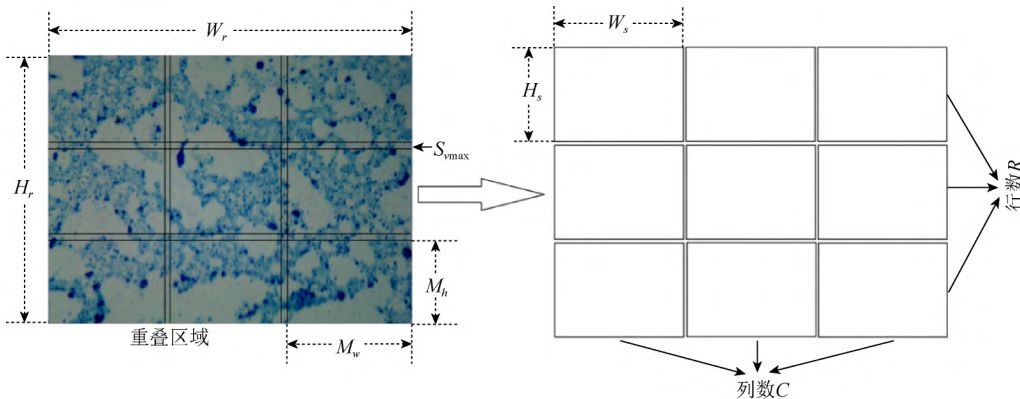


图 5 生成子图过程示意图

在实际应用中由原始图像的分辨率确定 W_r , H_r , 由检测目标的最大尺寸确定 S_{vmax} , 使用交叉验证法根据检测结果选择最合适的 R 和 C 。该方法将复杂的子图划分问题简化为 R 和 C 2 个超参数的选择问题, 极大地降低了问题的复杂性。该方法也可以推广到其他类似的问题中。

应用该方法的检测软件对子图进行检测, 得到的坐标为子图上的坐标, 需要将其还原为原图中的坐标。设对于第 $i+1$ 列第 $j+1$ 行个子图, 得到的检测结果中的某个坐标为 (x_s, y_s) , 则其在原始图像上的坐标 (x, y) 为

$$(x, y) = (x_s + i \times M_w, y_s + j \times M_h) \quad (6)$$

其中, $i \in [0, C-1], j \in [0, R-1]$ 。

由于采用了重叠策略生成子图, 得到的结果中必然会出现某些目标有多个可能的位置区域。为了消除同一个目标的多余位置区域, 使用非极大值抑制算法(non-maximum suppression, NMS)^[18]。NMS 属于同一类别目标的多个重叠检测结果, 取其中未遍历过的概率最大的那个结果, 当某一其他的检测结果与概率最大的检测结果的重叠面积大于一定的比例值, 则删除其结果, 循环迭代至遍历所有检测结果, 效果示意图如图 6 所示。

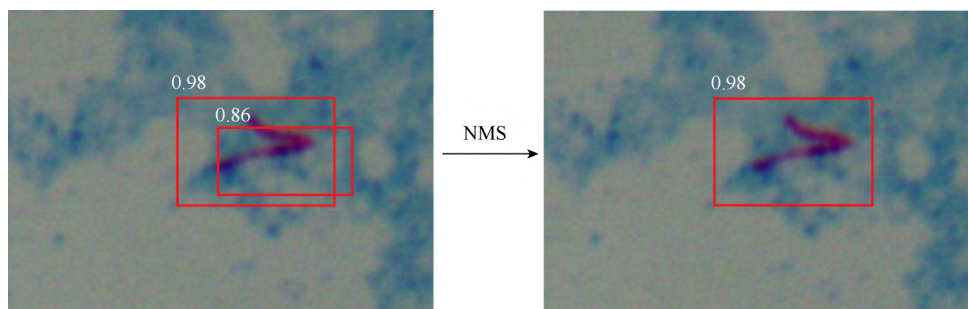


图 6 NMS 效果示意图

3 训练数据构建策略

3.1 基于分块采样的快速标注

在使用神经网络的目标检测算法之前, 需要用一定的数据训练。以往的经验表明, 神经网络检测的准确度依赖于大量、准确的训练数据。结核杆菌由人工来完成检测较为困难, 需要长期从事该工作的专家来完成标注, 且收集大量标注数据成本极高。人长期从事单调乏味的工作易疲劳, 进而使人工检测的准确度下降, 导致收集准确标注数据存在一定的困难。而常用的解决数据量不足的方法—迁移学习中的预训练法^[19]和交叉学习法^[20]不适用于本问题。

针对本问题的特殊性, 根据同一批生物样本中结核杆菌取样所得的图像及背景有较大的相似性, 本文认为该相似性导致标注工作存在很大的冗余, 实际上使用更少的标注数据就能达到近似的结果。运用局部性原理, 将大量数据等分成若干小块, 只取每小块数据的前部分进行标注作为训练集, 即可达到用更少的数据取得和大量数据接近的结果。构造数据的示意图如图 7 所示, 红色部分为需要标注的数据, 而蓝色部分则为不需要标注的数据。使用更少的数据获得了和大量数据相当的结核杆菌图

像分布律信息。合理运用此方法即可构造蕴含更全面信息的数据集, 进而获得更高的准确度和泛化性, 有效性将通过实验进行检验。

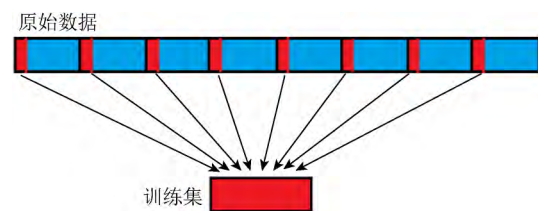


图 7 基于分块采样的快速标注

3.2 渐进性训练数据优化策略

肺结核杆菌检测图片标注难度大, 需要相关领域的专业知识, 即使是专业人员, 长时间从事枯燥的标注工作也较易犯错, 出现漏标情况, 即使要求二次校验, 错误仍旧不可避免。少量的、随机的漏标对实验结果影响不大, 但如果漏标的数据是具有某一特点的数据, 则对检测结果的影响很大。在实践中发现, 颜色较淡的和形状较小的肺结核杆菌容易漏标, 导致训练好的检测软件无法检测到这 2 类错误。针对该问题, 本文采取如下渐进性标注措施:

(1) 使用训练好的模型进行检测, 对未成功检

测的肺结核杆菌进行人工分析;

- (2) 根据分析结果再人为构造少量的此类样本;
- (3) 将新构建样本加入训练集重新训练, 更新模型;
- (4) 重复(1)~(3)过程直至检测结果满足要求。

4 实验结果

4.1 子图的划分及分块采样

实验采用 LabelImg^[21]对原始数据进行标注。标注训练数据时, 可直接在原图上完成。划分子图时, 子图的范围由式(5)确定。分割子图采用超参数 $R=3$, $C=4$ 。

在训练时, 直接舍弃超出图片范围的结核杆菌样本。由于肺结核杆菌样本有最大范围, 所以在某张子图中丢弃的目标标注会完全包含在下一张子图中。在训练过程中不会漏训标注样本, 但有可能不能识别位于图像边界的不完整的结核杆菌, 这一点在实际应用中可以接受。若有检测边界病菌的需求, 可用一些简单的方法对训练过程做一定的约束。不含结核杆菌的子图不参与训练。由于在计算子图的大小时采用了进一法, 边界区的子图可能会与原子图大小不一致, 不利于批处理。对于边界情况, 将截取框向内移位, 以保证所有子图大小一致。在测试时分别对每张子图进行检测, 所得结果根据式(6)转换为原图的结果, 再用 NMS 消除多余重复结果。

分块采取每 200 个数据分成一个数据块, 取前 50 个数据作为训练集, 后 150 个数据作为验证集。对于不足 200 个数据的数据块, 按 1:3 的比例划分训练集和测试集。

4.2 实验结果与分析

本文提出的方法在 Linux 系统下, 用 caffe 实现 CNN, 用 python 实现训练程序的编写。实验使用 I7-7700K CPU, Titan X GPU, 32 GB 内存的计算机。为验证本文各部分提出方法的有效性, 设置不同情况进行多组测试。测试结果见表 1。

表 1 实验结果

情况	分割子图	渐进标注	分块采样	训练集数目	测试集数目	准确率	运行时间(ms)
情况 1	×	×	×	4 201	500	0.588	84
情况 2	√	×	×	4 201	500	0.803	600
情况 3	√	√	×	4 201	500	0.863	600
情况 4	√	√	×	1 185	3 526	0.762	600
情况 5	√	√	√	1 185	3 526	0.853	600
情况 6	√	√	√	4 201	500	0.893	600

情况 1 和情况 2 结果对比, 可以看出将原始图片分割成多个子图可以有效解决小目标识别问题。相应的代价是运行速度的大幅下降。为满足实际应用的需要, 要求结核杆菌检测软件在 2 s 以内处理完一张图片。从实验结果可看出, 本文提出的方法, 满足该要求。情况 2 和情况 3 对比, 可以发现用渐进策略构造难识别样本来解决初始标注质量不高的问题是有效的。生物医学领域样本标注十分困难, 即使是该领域的专家, 也无法做到不漏标。而目标检测的精准度十分依赖于标注样本的质量, 因此需要渐进地进行标注。情况 4 和情况 3 表明要达到较高的检测精度需要大量的数据。情况 5 在情况 4 的基础上采用了分块采样法构造训练集, 其余数据作为测试集。结果准确度有很大的提升, 证明了分块采样法能有效降低所需训练数据量, 提升数据蕴含信息的全面性。在实际应用中, 采用分块采样法构造训练数据, 可以极大地减少标注数据的工作量, 有助于达到更高的准确度。分块采样法的核心思想是局部性原理, 对于不满足这一条件的应用场景, 其效果会相应地降低。情况 6 则是综合各种方法后达到的最终精度。

本文提出的方法的最终检测效果如图 8 所示。图 8(a)是本问题的一个检测难点, 2 个目标较小切与背景颜色相似, 图 8(b)是小目标的放大图。该检测结果表明, 本方法可以较好地检测出小目标。图 8(c)则是另一个检测难点, 即样本染色程度较轻, 不易识别, 但本文方法仍能较好地完成检测。图 8(d)左则是图 8(c)结核杆菌检测结果的放大图, 右侧则是来自其他检测图中颜色较淡样本的检测结果。本研究开始时采用传统图像处理方法, 无法解决这 2 类结核杆菌的识别问题。如果通过设定相应阈值检测出这 2 类样本, 则会极大地增加误检率。即使是人工检测, 也容易漏检, 通过使用本文方法, 能较好地检测出来。但目前本文方法漏识别的样本仍集中在特别小的目标和颜色很淡的样本中。

图 8(e)和图 8(f)则是本方法对于含有较多结核杆菌图的检测效果, 证明了本方法对于各类结核杆菌均有较好的检测效果。图 9 则是本方法误检的样本, 这 2 个检测出来的目标是将细胞的边缘效应误判成了结核杆菌。但由于这类结果极少, 且实际应用中对误判的容忍程度较高, 不影响实际使用。

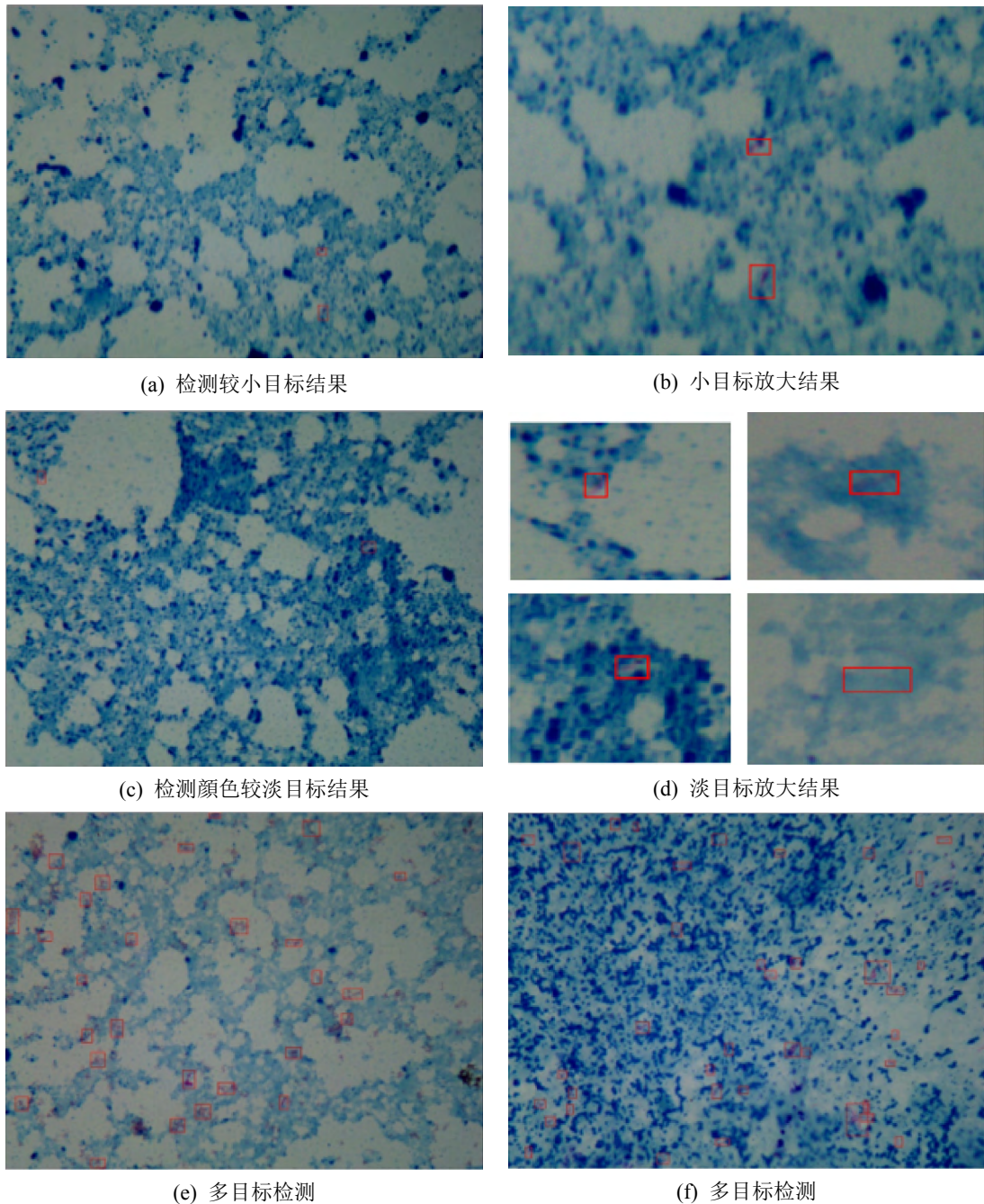


图8 最终检测效果



图9 边缘效应导致的误检

5 总结与展望

结核杆菌的检测是典型的小目标、无明显特征的目标检测问题。传统的图像检测方法无法有效解

决该问题。而基于 CNN 的目标检测方法虽能在一定程度上解决这一问题，却又要面对标注数据难的问题。本文分析了基于 CNN 的目标检测方法，提出了采用 VGG16 结构的 Faster-RCNN 技术作为基础的目标检测技术，并采用重叠子图解决小目标检测问题。由于迁移学习不适用于本问题，本文提出采用分块采样和渐进标注解决获取大量、准确标注数据难的问题。综合多种方法的优势最终解决了传统图像检测技术不易处理的结核杆菌检测问题，并集成至相关单位的医疗检测产品。本文方法有助于提高医生诊断结核病的准确度，有利于疾病的早发

现,早治疗。本文在解决问题过程中提出的相关方法易于应用到类似的场景,为解决类似问题提供一定的思路。

基于CNN的目标检测方法在小目标检测问题上乏力的根本原因是锚结构。如果能改进这一机制,或者提出更加有效的候选区域提取方法,则能从根本上解决小目标检测问题,得到更具通用性的目标检测方法,提高程序的运行速度。另外医学图像领域对于图像中病毒或细菌的判断标准有人类已知和未知标准。而基于有监督的人工标注数据的检测算法,只能学习和利用人类已知的知识来检测结核杆菌,方法很难超越人检病毒的极限准确度。但由于人长时间从事某项工作容易疲劳准确度下降,本方法仍具有较高的研究应用价值。而研究出能利用未知的知识来诊断疾病的机器学习算法,超越人的检测能力,是未来的研究方向。

参考文献

- [1] World Health Organization. Global tuberculosis report 2017 [EB/OL]. [2018-05-20]. <https://www.who.int/tb/publications/2017/en/>.
- [2] JAFRI R, ARABNIA H R. A survey of face recognition techniques [J]. *Journal of Information Processing Systems*, 2009, 5(2): 41-68.
- [3] RAUTARAY S S, AGRAWAL A. Vision based hand gesture recognition for human computer interaction: a survey [J]. *Artificial Intelligence Review*, 2015, 43(1): 1-54.
- [4] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2014: 580-587.
- [5] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). New York: IEEE Press, 2005: 886-893.
- [6] FELZENSZWALB F, GIRSHICK B, MCALLESTER D, et al. Object detection with discriminatively trained part-based models [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(9): 1627-1645.
- [7] 董恩增,魏魁祥,于晓,等.一种融入PCA的LBP特征降维车型识别算法[J].*计算机工程与科学*, 2017, 39(2): 359-363.
- [8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [9] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. *Nature*, 2015, 521(7553): 436-444.
- [10] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [11] DAI J F, LI Y, HE K M, et al. R-FCN: Object detection via region-based fully convolutional networks [C]// *Advances in Neural Information Processing Systems*. Cambridge: MIT Press 2016: 379-387.
- [12] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector [C]//*European Conference on Computer Vision – ECCV 2016*. Heidelberg: Springer, 2016: 21-37.
- [13] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2017:6517-6525.
- [14] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2016: 770-778.
- [15] PAN S J, YANG Q. A survey on transfer learning [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(10): 1345-1359.
- [16] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2014-09-04) [2018-05-03]. <https://arxiv.org/abs/1409.1556>.
- [17] SHEN Z Q, LIU Z, LI J G, et al. DSOD: Learning deeply supervised object detectors from scratch [C]//2017 IEEE International Conference on Computer Vision (ICCV). New York: IEEE Press, 2017: 1937-1945.
- [18] NEUBECK A, VAN GOOL L. Efficient non-maximum suppression [C]//*ICPR'06 Proceedings of the 18th International Conference on Pattern Recognition*. Washington, DC: IEEE Computer Society Press, 2006: 850-855.
- [19] DONAHUE J, JIA Y Q, VINYALS O, et al. DeCAF: A deep convolutional activation feature for generic visual recognition [C]//*Proceedings of the 31st International Conference on International Conference on Machine Learning*. New York: ACM Press, 2014: 647-655.
- [20] MISRA I, SHRIVASTAVA A, GUPTA A, et al. Cross-stitch networks for multi-task learning [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society Press, 2016: 3994-4003.
- [21] Tzutalin. LabelImg: A graphical image annotation tool [EB/OL]. (2017-09-03) [2018-05-03]. <https://github.com/tzutalin/labelImg>.