

# 数学学院案例分析报告

案例名称：肺结核影像分析

报告参与人：魏丹怡 (202032164)

完成时间：2021.07.01

# 目 录

一、绪论.....	1
1、选题背景及意义.....	1
2、问题陈述.....	1
二、数据集基本情况.....	2
三、模型介绍.....	3
1、残差网络.....	3
2、FPN 网络.....	6
3、Focal Loss.....	6
四、实验算法.....	7
1、模型构建.....	7
2、模型训练.....	7
3、分类结果及评估.....	9
五、实验总结.....	10

# 一、绪论

## 1、选题背景及意义

肺结核病可谓是中国最为常见的慢性传染疾病之一。近年来，肺结核的发病率一直呈现出逐年上升的状态，及早诊断及早治疗便成为了抑制该种疾病的唯一有效方式。在当前临床诊疗的过程当中，现代医学影像诊断技术的临床价值日渐凸显，已经成为了不可替代的重要诊疗手段。

医学影像诊断技术是指以非侵入的方式对人体或人体的一部分获得内部组织图像以进行医疗或医学研究的过程。随着科学技术的飞速发展，医学影像诊断技术也有了长足的进步，已经有包括 X 射线、计算机断层扫描（CT）、核磁共振成像、超声和核医学等多种医学成像方式。然而，随着科学技术的发展和医学影像应用的推广，越来越多的医学图像需要医生解释。医学图像的解释逐渐成为一个挑战性的任务，由于缺乏经验或疲劳等因素的影响，医生可能会在解读的过程中出错，遗漏一些疾病，导致假阴性出现，也可能将非病变解读为病变或将良性病变误解读为恶性，导致假阳性出现。而无论出现这两种情况中的哪一种，都可能带来不可获知的不良影响。因此，计算机辅助检测与计算机辅助诊断的出现最终成为时代的需求。

目前，对肺结核患者进行 X 射线诊断是我国医院最常见的诊断方法，是医学影像诊断技术在肺结核诊断中的重要组成。X 射线穿过胸部，投影在胶片上，形成 X 光胸片，受检者取站立位，一般在平静吸气下屏气拍照。针对胸部 X 光片图像，通过对胸部 X 光片图像进行分类来判断该图像是否存在肺结核病是最常见的方法，也即对胸部 X 光片图像进行目标识别。

## 2、问题陈述

目前，目标识别有两大经典结构：一类是以 Faster R-CNN 为代表的两级识别方法，这种结构的第一级专注于 proposal 的提取，第二级则对提取出的 proposal 进行分类和精确坐标回归；另一类结构是以 YOLO 和 SSD 为代表的单级结构，它们摒弃了提取 proposal 的过程，只用一级就完成了识别。两类结构进行对比发现，两级结构准确度较高，但因为第二级需要单独对每个 proposal 进行分类/回归，速度较慢；一级结构虽然速度较快但准确率远远比不上两级结构。为了兼顾速度与准确度，在单级结构中实现较高的准确度呢，有人提出了 Focal Loss 解决这个问题。在本次实验中，我们使用 Retina Net（Focal Loss）来实现在给定的 X 光胸片上找到肺结核的发病位置，并标注出具体的肺结核病类别。

## 二、数据集基本情况

数据集是关于肺结核病人及正常人 X 光胸片影像数据，共包括四个文件夹，其中 annotation 为训练集的标注，test\_data\_A 为测试数据集，包含 260 个样本，train\_data 为训练数据集，包含 2999 个样本。肺结核病人的病症包含九个类别，具体展式如下：

表 1 数据集基本情况

标签	英文名	中文名
0	Consolidation	肺实变
1	Fibrosis	纤维化表现
2	Effusion	胸腔积液
3	Nodule	结节
4	Mass	肿块
5	Emphysema	肺气肿
6	Calcification	钙化
7	Atelectasis	肺不张
8	Fracture	骨折

此外，X 光胸片影像如下图所示：

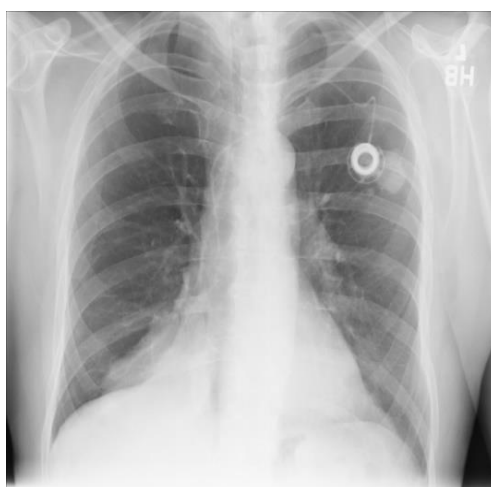


图 1 X 光胸片影像

图像信息储存在 Jason 文件中，如果是正常人则 syms 和 boxes 为空，如果患病，会给出患病类别（syms）以及所在区域（boxes）。

```
[{'file_name': '36200.png', 'syms': [], 'boxes': []},
{'file_name': '36201.png', 'syms': [], 'boxes': []},
{'file_name': '36202.png', 'syms': [], 'boxes': []},
{'file_name': '36203.png', 'syms': [], 'boxes': []},
{'file_name': '36204.png',
 'syms': ['Nodule'],
 'boxes': [[233, 666, 258, 693]]},
```

### 三、模型介绍

Retina Net 本质上是 Resnet + FPN + 两个 FCN 子网络，图 2 为 Retina Net 的结构图，我们可以看出，Retina Net 的特征提取网络选择了残差网络 ResNet，特征融合这块选择了 FPN（特征金字塔网络），以特征金字塔不同的尺寸特征图作为输入，搭建三个用于分类和框回归的子网络。分类网络输出的特征图尺寸为  $(W, H, KA)$ ，其中  $W$ 、 $H$  为特征图宽高， $KA$  为特征图通道，存放  $A$  个 anchor 各自的类别信息（ $K$  为类别数）。

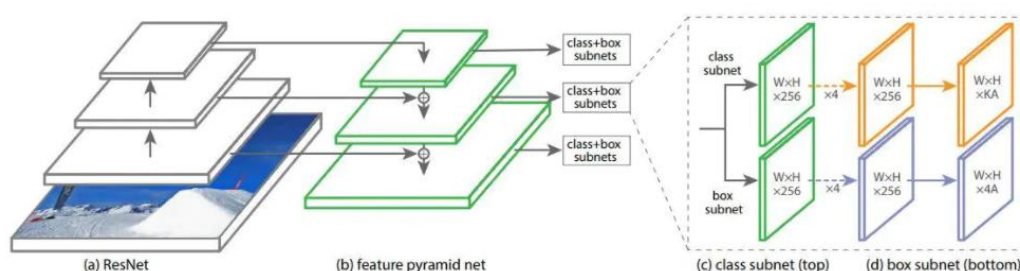


图 2 Retina Net 的结构图

#### 1、残差网络

##### (1) 残差网络的来源

残差操作这一思想起源于论文《Deep Residual Learning for Image Recognition》。这篇文章发现，如果存在某个  $K$  层的网络  $F$  是当前最优的网络，那么可以构造一个更深的网络，其最后几层仅是该网络  $F$  第  $K$  层输出的恒等映射，就可以取得与  $F$  一致的结果；也许  $K$  还不是所谓“最佳层数”，那么更深的网络就可以取得更好的结果。但是，随着网络层数的增加，网络发生了退化的现象，即随着网络层数的增多，训练集 loss 逐渐下降，然后趋于饱和，当你再增加网络深度的话，训练集 loss 反而会增大。总而言之，与浅层网络相比，更深的网络的表现不应该更差。但是如下图所示，56 层的神经网络表现明显要比 20 层的差。证明更深的网络在训练过程中的难度更大，因此作者提出了残差网络的思想。

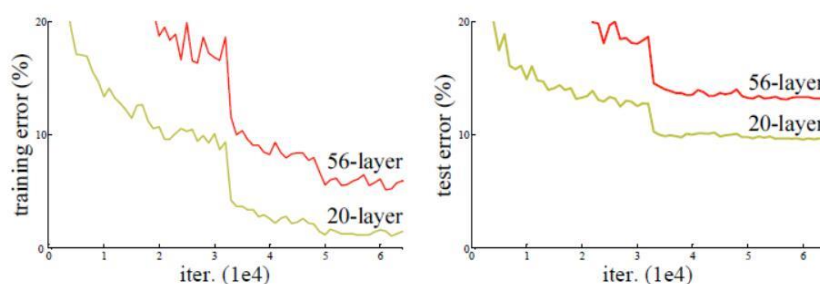


图 3 层数与 loss 关系图

## (2) 残差网络的定义

一个具体残差模块的定义如下图：

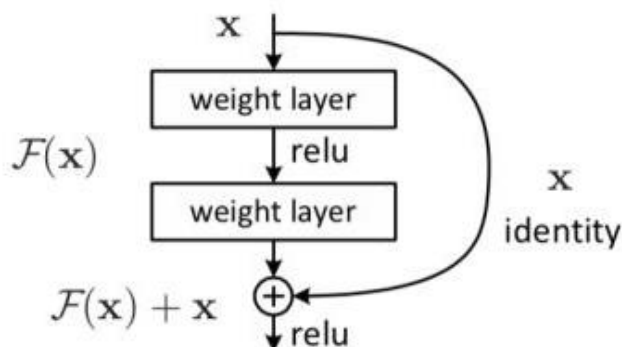


图 4 残差定义图

残差网络是由一系列残差块组成的。一个残差块可以用表示为：

$$x_{l+1} = x_l + \mathcal{F}(x_l, W_l)$$

残差块分成直接映射部分和残差部分。 $h(x_l)$ 是直接映射，反应在下图中是左边的曲线； $\mathcal{F}(x_l, W_l)$ 是残差部分，一般由两个或者三个卷积操作构成，即下图中右侧包含卷积的部分。

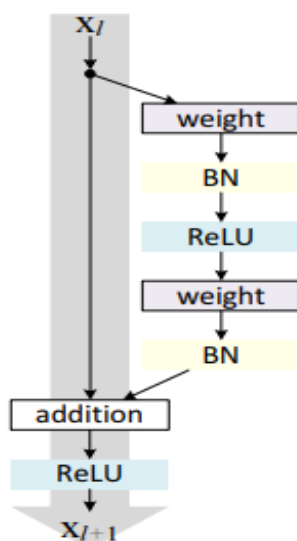


图 5 残差块

上图中的 Weight 在卷积网络中是指卷积操作，addition 是指单位加操作。这个模型能很好的解决退化问题，使得模型层数增加时，loss 在不断的减小。

## (3) 残差网络的优势

下图是 VGGNet-19、34 层深的普通卷积网络、34 层深的 ResNet 网络的对比图：

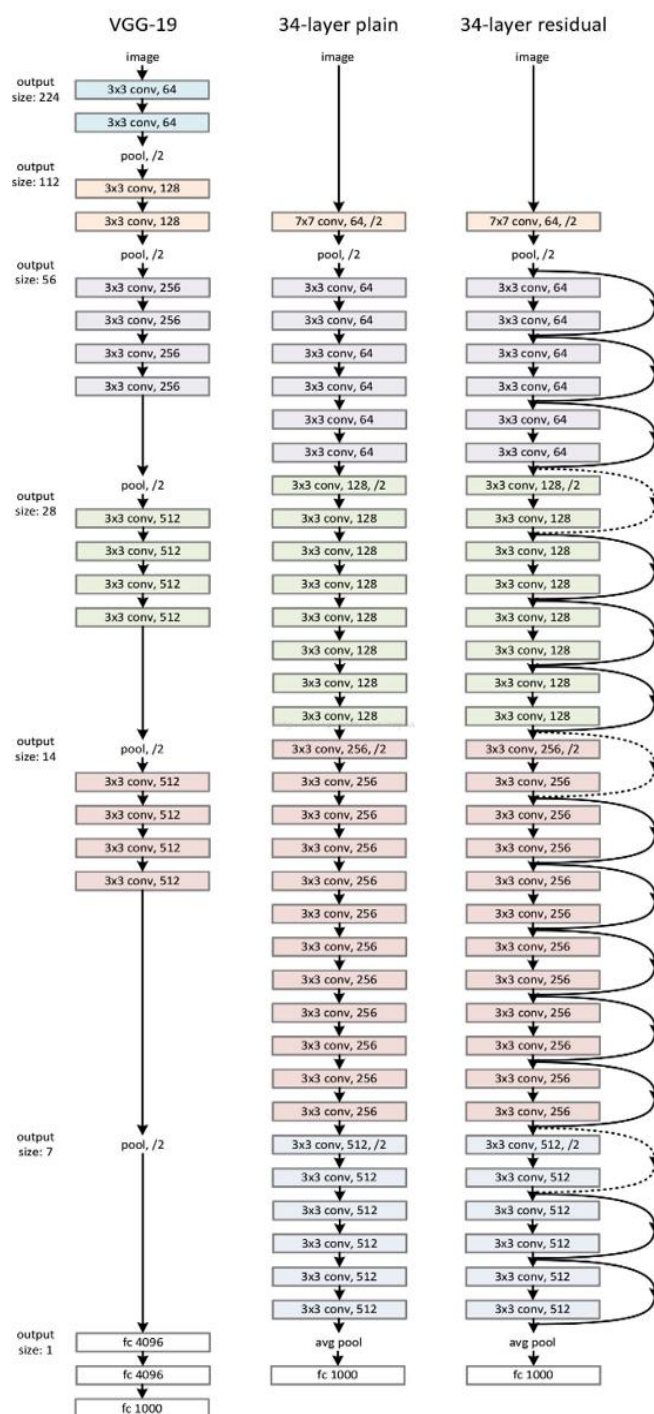


图6 神经网络对比图

可以看到普通直连的卷积神经网络和 ResNet 的最大区别在于, ResNet 有很多旁路的支线将输入直接连到后面的层, 使得后面的层可以直接学习残差, 这种结构也被称为 shortcut 或 skip connections。

传统的卷积层或全连接层在信息传递时, 或多或少会存在信息丢失、损耗等问题。ResNet 在某种程度上解决了这个问题, 通过直接将输入信息绕道传到输出, 保护信息的完整性, 整个网络则只需要学习输入、输出差别的那一部分, 简化学习目标和难度。

## 2、FPN 网络

FPN 网络把低分辨率、高语义信息的高层特征和高分辨率、低语义信息的低层特征进行自上而下的侧边连接，使得所有尺度下的特征都有丰富的语义信息。

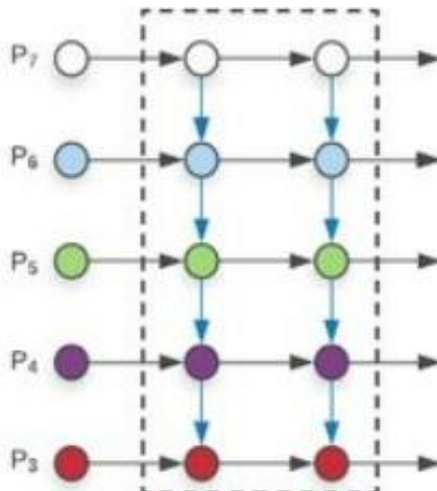


图 7 FPN 网络图

采用自顶向下的 FPN 结构，这种多尺度特征融合的目的是为了对较小物体也能够保持检测的精度。

## 3、Focal Loss

Focal loss 是在交叉熵损失函数基础上进行的修改，首先回顾二分类交叉熵损失：

$$L = -y \log(y') - (1 - y) \log(1 - y')$$

$y'$  是经过激活函数的输出，所以在 0-1 之间。可见普通的交叉熵对于正样本而言，输出概率越大损失越小。对于负样本而言，输出概率越小则损失越小。此时的损失函数在大量简单样本的迭代过程中比较缓慢且可能无法优化至最优。

Focal Loss 通过调整 loss 的计算公式使单级结构达到和 Faster R-CNN 一样的准确度。

$$FL(p_t) = a_t (1 - p_t)^\gamma \log(p_t)$$

其中， $p_t$  是不同类别的分类概率， $\gamma$  是个大于 0 的值， $a_t$  是个  $[0, 1]$  间的小数， $\gamma$  和  $a_t$  都是固定值，不参与训练。从表达式可以看出：

(1) 无论是前景类还是背景类， $p_t$  越大，权重  $(1 - p_t)$  就越小。也就是说 easy example 可以通过权重进行抑制。换言之，当某样本类别比较明确些，它对整体 loss 的贡献就比较少；而若某样本类别不易区分，则对整体 loss 的贡献就相对偏大。这样得到的 loss 最终将集中精力去诱导模型去努力分辨那些难分的目标类别，于是就有效提升了整体的目标检测准度。



(2)  $a_t$  用于调节 positive 和 negative 的比例, 前景类别使用  $a_t$  时, 对应的背景类别使用  $1 - a_t$ ;

(3)  $\gamma$  和  $a_t$  的最优值是相互影响的, 所以在评估准确度时需要把两者组合起来调节。作者在论文中给出  $\gamma=2$ 、 $a_t=0.25$  时, ResNet-101+FPN 作为 backbone 的结构有最优的性能。

## 四、实验算法

### 1、模型构建

构建的是 RetinaNet 网络模型, 输入数据是 1024\*1024 大小的 X 光胸片图像。

首先, 连接一个 ResNet 网络进行特征提取, 使用的是 Resnet 网络, 深度为 101 层。

其次, 连接一个 FPN 网络进行特征融合, 该特征金字塔网络输入为残差网络 ResNet-101 不同尺度的三张特征图; 输出为多尺度特征融合的 P3\_x, P4\_x, P5\_x, 以及由残差网络 ResNet-101 最后的一张特征图继续通过卷积层 (stride=2, 特征图尺寸继续减小) 的结果 P6\_x, P7\_x。

接下来, 在特征提取网络 ResNet-101 和特征融合网络 FPN 后, 对获得的五张特征图 [P3\_x, P4\_x, P5\_x, P6\_x, P7\_x], 通过具有相同权重的框回归和分类子网络, 获得所有框位置和类别信息。框回归子网络由四层卷积层组成, 每层卷积层的 stride=1, kernel\_size=3, padding=1, 也就是说特征图通过该网络, 长宽大小不变, 通道维变为  $4 * \text{num\_anchors}$ 。和框回归子网络类似, 分类子网络也是通过了四层卷积层, 特征图的长宽保持不变, 通道扩展为类别数  $9 * \text{anchor}$  数, 用于存放所有基于 anchor 的检测框的分类信息。

### 2、模型训练

利用 Python 的生成器, 逐个生成数据的 batch 并进行训练, batch\_size=2, 当生成器返回 steps\_per\_epoch 次数据时计一个 epoch 结束, 执行下一个 epoch, 共执行 10 次 epoch, 用时 3 天。

以下是截取的部分训练结果:

```

Num training images: 1845
Epoch0-Running loss: 1.55521: 100%|██████████| 923/923 [6:15:35<00:00, 24.42s/it]
Evaluating dataset

mAP:
Consolidation: 0.3228166221094477
Fibrosis: 5.445937681550009e-05
Effusion: 0.1918638500090007
Nodule: 0.0038177324931553507
Mass: 0.0027472527472527475
Emphysema: 0.006408145543443387
Calcification: 0.0
Atelectasis: 0.03291658762428986
Fracture: 0.001471633400666137

Epoch9-Running loss: 0.81116: 100%|██████████| 923/923 [6:22:19<00:00, 24.85s/it]
Evaluating dataset

mAP:
Consolidation: 0.4836052066794784
Fibrosis: 0.07388723465464371
Effusion: 0.292979744353457
Nodule: 0.23633689882494524
Mass: 0.14206080015727757
Emphysema: 0.2767318542166536
Calcification: 0.0959339117254969
Atelectasis: 0.20898954363976877
Fracture: 0.2785043572856528

```

通过观察训练结果，得到训练集中对于肺实变（Consolidation）的分类最为准确，最高准确率为 48.36%；分类结果较差的为纤维化表现（Fibrosis）和钙化（Calcification），最高准确率分别为 7.44%和 9.59%。

绘制学习曲线如下：

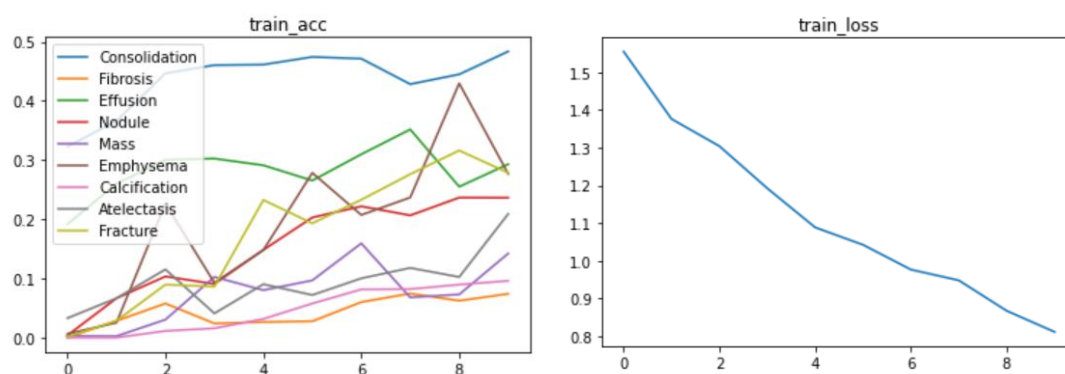


图 8 学习曲线

由上图可以看出，随着训练次数的增加，每一类病症的分类准确率总体都呈上升趋势，且损失代价在不断减小。

### 3、分类结果及评估

用训练好的模型对测试集进行预测，可视化结果如下所示：

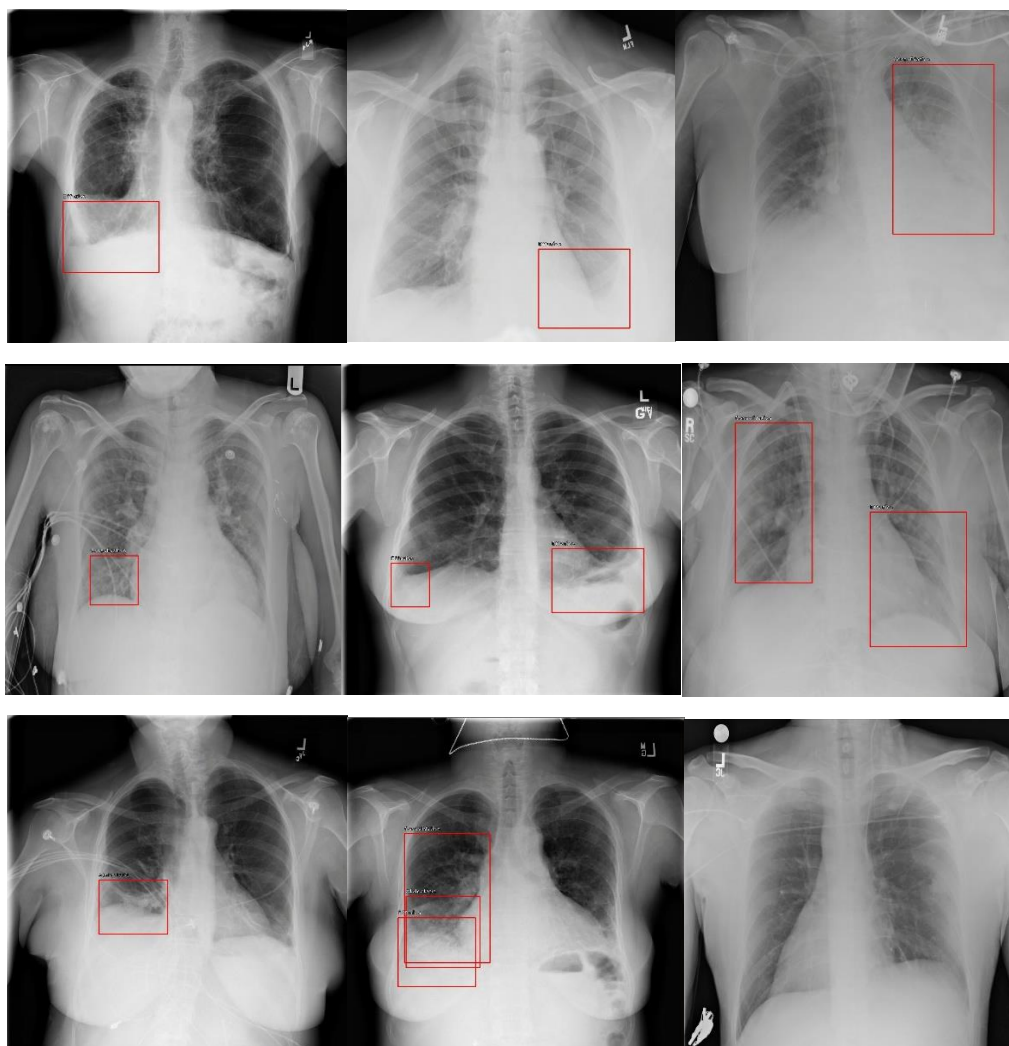


图9 肺部疾病可视化

上图为部分病人的病症分类，通过可视化，可以在 X 光胸片影像上标注出病人患病的种类，若为正常人的 X 光胸片影像，则不会出现标注。

可以看出，多数病人在出现肺实变（Consolidation）的同时也会出现胸腔积液（Effusion）。胸腔积液是以胸膜腔内病理性液体积聚为特征的一种常见临床症状。胸膜腔为脏层和壁层胸膜之间的一个潜在间隙，正常人胸膜腔内有 5~15ml 液体，在呼吸运动时起润滑作用，胸膜腔内每天有 500~1000ml 的液体形成与吸收，任何原因导致胸膜腔内液体产生增多或吸收减少，即可产生胸腔积液。肺实变是由于肺不张和肺泡积液造成的肺内气体消失，肺组织呈现实质性改变。大量的渗出物充塞于肺泡腔内，肺泡腔内气体减少或消失，肺组织质地如肝脏，称为实变。也就是说，肺实变出现的极大原因在于肺泡积液，因此，我们得出的结论具有一定的科学解释。

## 五、实验总结

本次实验利用 retinanet 网络框架完成了肺结核病症的分类。得到训练集中对于肺实变 (Consolidation) 的分类最为准确, 最高准确率为 48.36%; 分类结果较差的为纤维化表现 (Fibrosis) 和钙化 (Calcification), 最高准确率分别为 7.44% 和 9.59%。在完成训练后, 在测试集数据上进行了肺部疾病类别可视化, 用方框标注出了病症位置及病症类别。在实现分类算法的过程中, 学习了目标检测中两大类方法的不同之处, 同时加深了对残差网络、FPN 网络以及 Focal Loss 函数等深度学习知识的理解, 同时在不断的尝试与纠错中, 增加了许多编程经验。

在代码运行过程中, 我遇到了一些困难, 比如, 在运行代码时部分 csv 文件中的图片信息没有对应的 X 光胸片影像, 我删除了 CSV 文件中的多余信息。此外, 由于时间有限, 无法运行足够多的 epoch, 因此我只运行了 10 个 epoch, 以至于我的准确率不高, 如若时间充裕我可以增加 epoch 个数以提高准确率。