



澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU



Revisiting Computer-Aided Tuberculosis Diagnosis

Yu Jiening
Yu Jiening@umac.mo

um 澳大

介绍

- 结核病（TB）是一种普遍的传染病，TB患者的死亡率非常高。
- TB检测的金标准需要对痰样本和细菌培养物进行显微镜检查，以确定结核分枝杆菌，培养结核分枝杆菌需要一个BSL-3实验室，需要几个月时间，发展中国家缺乏必要的基础设施来建立BSL-3实验室。
- 胸部x射线（CXR）可以迅速发现肺结核，然而人眼很难在CXR图像中识别TB区域。



- 深度学习可以24小时自动定位TB的感染位置，但其依赖于大量数据，CXR图像收集困难

文章贡献

- 建立了一个大型CTD数据集TBX11K，比从前的TB数据集更大、注释更好、更真实，使深度神经网络训练能够同时进行CXR分类和TB感染区检测
- 提出了一个简单有效的CTD框架SymFormer，由SymAttention和SPE组成，利用CXR图像的双边对称性改善CTD基准模型
- 在TBX11K数据集上建立了一个CTD基准，通过引入评估标准，评估从现有目标探测器改造的几个baseline，并运行在线挑战，为将来的研究奠定一个良好的开端

TBX11K数据集

深度神经网络依赖于大量的训练数据，而现有的公共数据集规模并不大，为了解决这个问题，我们建立了一个全面且大规模的数据集

建立数据集的步骤：

- 1、数据的收集和注释
- 2、放射科医生的标注

TABLE 1

Summary of publicly available TB datasets. The size of our dataset is about $17\times$ larger than that of the previous largest dataset. Besides, our dataset annotates TB infection areas with bounding boxes, instead of only image-level labels.

Datasets	Pub. Year	#Classes	Annotations	#Samples
MC [25]	2014	2	Image-level	138
Shenzhen [25]	2014	2	Image-level	662
DA [6]	2014	2	Image-level	156
DB [6]	2014	2	Image-level	150
TBX11K (Ours)	-	4	Bounding box	11,200

TBX11K数据集：数据的收集与注释

- 建立分类：健康、患病非TB、活动性TB、潜伏性TB
- 数据收集：与中国顶级医院合作，收集到11200张CXR图像，其中包括5000例健康病例，5000例患病非TB病例，1200例TB病例，其中924例为活动性TB，212例为潜伏性TB，54例同时包含活动性和潜伏性TB，每张CXR图像对应一个唯一的个体。

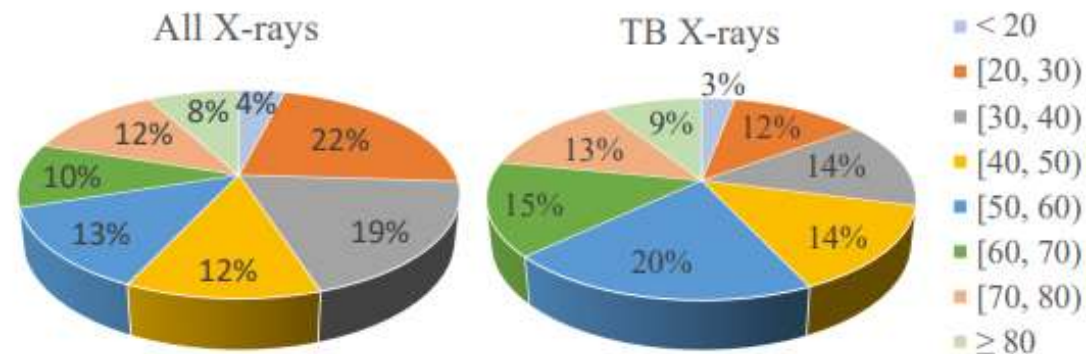


Fig. 1. Age distributions for the entire TBX11K dataset and specifically for TBX11K TB X-rays.

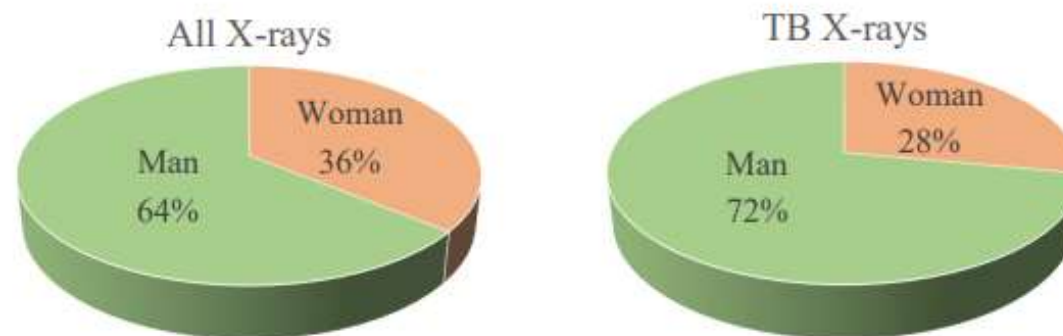


Fig. 2. Gender distributions for the entire TBX11K dataset and specifically for TBX11K TB X-rays.

TBX11K数据集：数据的收集与注释

4

- 专业的数据注释：数据的注释由经验丰富的放射科医生进行。数据集中的每张TB CXR图像首先由具有5-10年TB诊断经验的放射科医生进行标注，然后由另一位有超过十年经验的放射科医生审查方框注释。放射科医生将为每个框确定TB类型，为保证一致性，标记的TB类型将与黄金标准图像进行对照检查，若不匹配，则此CXR图像将进行重新注释。如果一个CXR图像被错误的标记了两次，注释者会讨论如何重新对其进行注释。

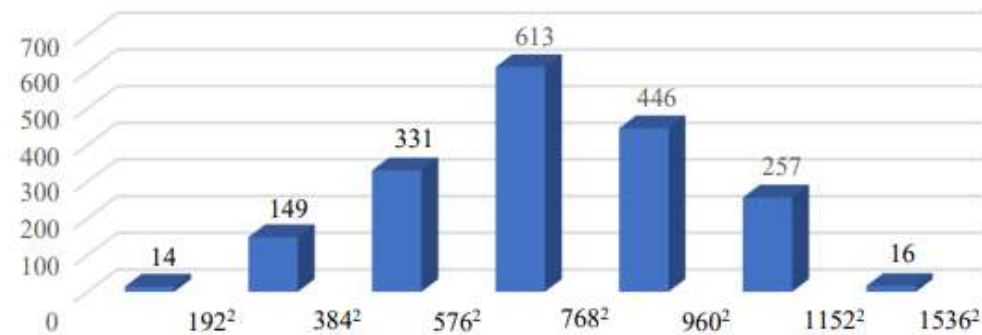


Fig. 3. **Distribution of the sizes of TB bounding boxes in the TBX11K dataset.** Each bin represents a specific range of bounding box areas. The left and right values of each bin correspond to its area range, and the height of the bin represents the number of TB bounding boxes within that range. It should be noted that the CXR images in TBX11K have a resolution of about 3000×3000 .

TBX11K数据集：数据的收集与注释

- 数据集分割：将数据集分为三个子集：训练集、验证集、测试集。
- 考虑四种不同的病例：仅具有活动性的TB CXR图像；仅具有潜伏性的TB CXR图像；同时具有活动性和潜伏性的TB CXR图像；不确定TB类型的CXR图像。对于每个TB病例，我们将训练集、验证集、测试集中的TB CXR图像保持3: 2: 1的比例。不确定TB类型的CXR图像被分配到测试集。

TABLE 2

Split for the TBX11K dataset. “Active & Latent TB” refers to CXR images with both active and latent TB; “Active TB” refers to CXR images with only active TB; “Latent TB” refers to CXR images with only latent TB; “Uncertain TB” refers to TB CXR images where the type of TB infection cannot be recognized using current medical conditions.

	Classes	Train	Val	Test	Total
Non-TB	Healthy	3,000	800	1,200	5,000
	Sick & Non-TB	3,000	800	1,200	5,000
TB	Active TB	473	157	294	924
	Latent TB	104	36	72	212
	Active & Latent TB	23	7	24	54
	Uncertain TB	0	0	10	10
Total		6,600	1,800	2,800	11,200

TBX11K数据集：放射科医生的研究

- 从新的TBX11K测试集中选取400张CXR图像，包括140张健康图像，140张患病非TB图像，120张TB CXR图像，其中63张为活动性TB，41张为潜伏性TB，15张为活动性和潜伏性TB，1张为不确定TB；接下来邀请一位没有去参与过前面标注工作的具有10年以上工作经验的放射科医生对这些图像进行标注。
- 研究发现与黄金标准产生的真实值相比，放射科医生的准确率达到了68.7%，如果忽略活动性TB和潜伏性TB的区别，达到84.8%。

TBX11K数据集： 潜在的研究主题

- 同时进行CXR图像分类和TB感染区域检测
- 数据分布不平衡： 潜伏性TB患者没有生病迹象和传染性， 因此不太可能治疗， 导致潜伏性TB病例非常少。这种数据不平衡对未来CTD方法提出了挑战。
- 利用TBX11K数据集进行增量学习

SymFormer框架

- **SymFormer**包括特征提取、对称异常搜索和结核病诊断头三个部分，采用FPN生成特征金字塔。

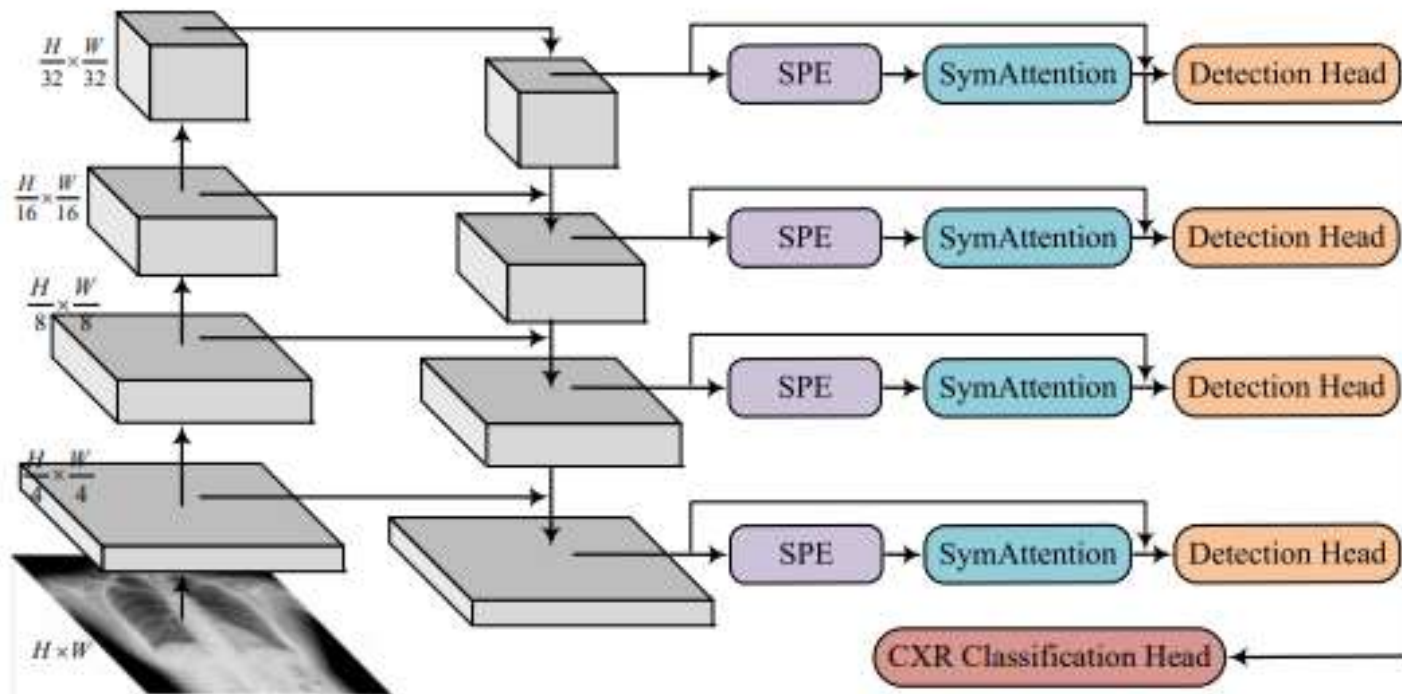


Fig. 4. Illustration of the proposed SymFormer framework. FPN [61] is applied to generate the feature pyramid.

SymFormer框架：特征提取

- 以ResNets作为Backbone网络示例进行特征提取，当给定CXR图像作为输入时，Backbone网络分四个阶段输出特征，分别按1/4，1/8，1/16和1/32的比例缩小。在Backbone网络上应用FPN，生成一个特征金字塔，即不同尺度的特征映射。对网络最后一层的feature map做上采样(图像由小分辨率到大分辨率的映射)，上采样的结果再与对应尺寸的feature map进行融合，把高层的特征传下来，补充低层的语义，得到高分辨率、强语义的特征。
- 将特征金字塔记为 $F=\{F_1, F_2, F_3, F_4\}$ $\mathbf{F}_i \in \mathbb{R}^{C \times \frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}}}$ ($i \in \{1, 2, 3, 4\}$)
- C为特征维度，H和W分别为输入CXR图像的高度和宽度

SymFormer框架：对称异常搜索

- SAS模块由SPE、SymAttention、前馈网络组成
- SAS：利用双边对称性来增强CXR图像的特征表示，增强的特征金字塔表示为 $\hat{\mathbf{F}} = \{\hat{\mathbf{F}}_1, \hat{\mathbf{F}}_2, \hat{\mathbf{F}}_3, \hat{\mathbf{F}}_4\}$ w.r.t. $\hat{\mathbf{F}}_i \in \mathbb{R}^{C \times \frac{H}{2^i+1} \times \frac{W}{2^i+1}} (i \in \{1, 2, 3, 4\})$
- CXR图像中的肺部可能不是严格对称的，为了解决这个问题，SAS首先结合SPE进行特征重新校准，然后由SymAttention使用重新校准的特征图来搜索特征图中每个空间位置的对称邻近区域（给定位置的双边对称位置的邻近区域）

SymFormer框架：对称异常搜索

对称位置编码 (SPE)

- 位置编码有两种类型：绝对位置编码和相对位置编码，SPE基于绝对位置编码。采用公式：
$$\mathbf{P}[pos, 2j] = \sin(pos/10000^{\frac{2j}{c}}),$$
$$\mathbf{P}[pos, 2j + 1] = \cos(pos/10000^{\frac{2j}{c}}),$$
- 其中pos表示空间位置，j索引特征维度 c是模型的嵌入维度
- CXR不严格遵循双边对称性，可能有轻微的旋转和平移。SPE通过特征重新校准来解决这个问题：首先将 \mathbf{P}_i 分成两边 \mathbf{P}_i^{left} 和 \mathbf{P}_i^{right} ，然后使用空间变换网络 (STN) 和水平翻转将 \mathbf{P}_i^{right} 转移到左侧，最后将转换后的左侧位置编码和 \mathbf{P}_i^{right} 沿x维度连接起来形成输出 \mathbf{P}_i^{sym}

$$\mathbf{T}_i = \text{STN}(\mathbf{F}_i; \Theta)$$

$$\mathbf{P}_i^{trans} = \text{Flip}_x(\mathcal{T}_\Theta(\mathbf{P}_i^{right}; \mathbf{T}_i)),$$

$$\mathbf{P}_i^{sym} = \text{Concat}_x(\mathbf{P}_i^{trans}, \mathbf{P}_i^{right}),$$

Θ 表示STN的权值， \mathbf{T}_i 为仿射变换矩阵， \mathcal{T}_Θ 为仿射变换， Flip_x 表示水平翻转， Concat_x 表示沿x维度的拼接

- 使用SPE P_i^{sym} 重新校准输入特征: $\mathbf{F}_i^{recalib} = \mathbf{F}_i + \mathbf{P}_i^{sym}$.
- STN的微设计: 将输入特征 \mathbf{F}_i 输入到一个小型的STN网络中, 来预测仿射变换矩阵 \mathbf{T}_i , 然后将然后将其用于单侧位置编码 P_{iright} 的仿射变换。这个小网络包括两个交叠的max-pooling和Conv-ReLU层, 然后在空间维度上进行平坦化操作, 再使用MLP预测放射矩阵。初始化MLP以保证初始化放射矩阵的仿射变换等价于一个相同的映射

SymFormer框架: SymAttention

- Self-attention是通过计算每个参考位置的attention权重来进行全局关系建模的，它融合了所有位置的特征。对于CXR图像是冗余的。
- 提出了SymAttention，它利用双边对称特性来帮助self-attention 识别CXR图像中的相关位置。SymAttention在左右肺的对称模式中搜索特征，允许每个参考位置只关注参考位置两侧对称位置周围的位置。给定特征映射 $\mathbf{F}_i^{recalib}$

$$\Delta \mathbf{p}_i^x = \mathbf{W}_x^{pos} \mathbf{F}_i^{recalib}, \quad \Delta \mathbf{p}_i^y = \mathbf{W}_y^{pos} \mathbf{F}_i^{recalib}, \quad (4)$$

in which $\mathbf{W}_x^{pos}, \mathbf{W}_y^{pos} \in \mathbb{R}^{(M \times K) \times C}$ are trainable parameter matrices. The attention \mathbf{A}_i and value \mathbf{F}_i^v are simply calculated using

$$\begin{aligned} \mathbf{A}_i &= \text{Softmax}(\text{Reshape}(\mathbf{W}^{att} \mathbf{F}_i^{recalib})), \\ \mathbf{F}_i^v &= \mathbf{W}^{value} \mathbf{F}_i^{recalib} \end{aligned} \quad (5)$$

where $\mathbf{W}^{att} \in \mathbb{R}^{(M \times K) \times C}$, $\mathbf{W}^{value} \in \mathbb{R}^{C \times C}$ are trainable parameter matrices and the softmax function is performed along the dimension of K . Then, we reshape \mathbf{F}_i^v like

$$\mathbf{F}_i^v \in \mathbb{R}^{C \times \frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}}} \rightarrow \mathbf{F}_i^v \in \mathbb{R}^{M \times \frac{C}{M} \times \frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}}}. \quad (6)$$

Next, SymAttention can be formulated as

$$\begin{aligned} \mathbf{F}_i^{att} &= \text{Concat}_{m=1}^M \left(\sum_{k=1}^K (\mathbf{A}_i[m, k] \cdot \mathbf{F}_i^v[m, :, \mathbf{p}_i^y + \Delta \mathbf{p}_i^y[m, k], \right. \\ &\quad \left. \underbrace{\frac{W}{2^{i+1}} - (\mathbf{p}_i^x + \Delta \mathbf{p}_i^x[m, k]) + 1}}_{\text{symmetric range}} \right]), \end{aligned} \quad (7)$$

in which $\text{Concat}_{m=1}^M$ means to concatenate all the results generated by setting m from 1 to M . The term with the wavy underline projects the sampled locations onto the bilaterally symmetric locations by taking the vertical centerline as the line of symmetry, which is the core of the proposed SymAttention. Finally, to ease optimization, a residual connection is connected, followed by an MLP:

$$\hat{\mathbf{F}}_i^{att} = \mathbf{W}^{proj} \mathbf{F}_i^{att} + \mathbf{F}_i, \quad \hat{\mathbf{F}}_i = \text{MLP}(\hat{\mathbf{F}}_i^{att}) + \hat{\mathbf{F}}_i^{att}, \quad (8)$$

where we have $\mathbf{W}^{proj} \in \mathbb{R}^{C \times C}$ and $\hat{\mathbf{F}}_i$ is the enhanced output as in §4.1.

- 根据经验设置 $M=8$, $K=4$, $N = \frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}}$,

SymFormer框架：TB诊断头

- 将两种类型的结核病诊断头连接到特征金字塔 \hat{F} ，该金字塔由SAS模块增强，分别用于进行结核病感染区检测和CXR图像分类。特征金字塔 \hat{F} 中的每个特征图被送入检测头，每个检测到的边界框预计覆盖一个结核病感染区域。然而存在假阳性的风险，为了解决这个问题，我们将增强特征金字塔的顶层映射 \hat{F}_4 送到反馈头中，以确定CXR是否含有结核病
- 检测头基于RetinaNet，RetinaNet是一个著名的单阶段目标检测器，由两个分支组成，用于边界框分类和位置回归。系统中每个边界框被设计为覆盖一个结核病感染区域。
- 在临床结核病筛查期间，大多数CXR病例都没有结核病感染，这使得引入假阳性的概率增大。因此增加了一个CXR图像分类头，同时进行CXR图像分类和结核病感染区域检测，若CXR图像被分为非结核病，则丢弃。

SymFormer框架：两阶段训练图

- SymFormer框架由两个磁头组成，分别用于CXR图像分类和结核病感染区域检测。由于肺结核病病例的数量大大超过结核病病例的数量，直接进行感染区域检测会导致纯背景检测过多。
- CXR图像仅描绘胸部结构和器官，如果先训练Backbone网络和分类头，用于特征提取的backbone网络会变得过拟合。
- 提出一种方法，训练Backbone网络和感染区域检测头最初只使用TBCXR图像，然后利用所有CXR图像来训练分类头。

实验设置：SymFormer的实施细节

- 使用PyTorch和开源mmdetection框架实现了SymFormer。
- 第一阶段的训练使用TBX11K中的TB CXR图像
- 第二阶段的训练使用TBX11K中的CXR图像，MC和深圳数据集的随机一半以及DA和DB数据集的训练集。使用MC和深圳的另一半以及TBX11K、DA、DB的测试集来评估CXR图像的分类性能。
- 将FPN特征通道数设为256.记为C，与RetinaNet一致。其他设置参照RetinaNet

实验设置： Baseline Models

- 结合图像分类头可以减少临床TB筛查中的假阳性，然而现有目标检测器往往忽略没有边界框对象的图像。在实践中，由于有大量非TB CXR图像，直接使用检测器检测CTD会导致大量假阳性，为了解决这个问题，引入一个分类头，可以同时进行CXR图像分类和TB感染区检测，其中CXR图像分类结果用于过滤掉检测的假阳性。
- 为实现这一目标，本文重新制定了几种著名的目标检测器，包括SSD、RetinaNet、Faster R-CNN、FCOS和Deformable DETR，用于同时进行CXR图像分类和TB感染区检测。具体操作：在这些检测器的Backbone的最后一层之后，将SymFormer中使用的相同的图像分类头添加到这些检测器中。

实验设置：评价指标

ROC曲线以 $FPR = \frac{FP}{FP+TN}$ 为横坐标, $TPR = \frac{TP}{TP+FN}$ 为纵坐标

- CXR图像分类

Accuracy: 在所有三个类别中正确分类的图像的百分比 $\frac{TP+TN}{TP+TN+FP+FN}$

曲线下面积 (AUC) : AUC计算ROC曲线下的面积, ROC曲线绘制了TB类别的真假阳性率。

Sensitivity: 被准确识别为TB的TB病例的百分比。

Specificity: 被正确识别为非TB的非TB病例的百分比。

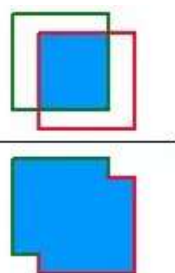
平均精度 (AP) : AP计算每个类别的精度, 并取平均值。提供了一个整体的精度测量。 $\frac{TP}{TP+FP}$

平均召回率 (AR) : AR计算每个类别的召回率并取平均值。提供了召回率的整体度量。 (eg: TP被预测为TP占总TP的比例) $\frac{TP}{TP+FN}$

混淆矩阵: 报告真阳性 (TP)、真阴性 (TN)、假阳性 (FP)、假阴性 (FN) 的数量, 以及其相对于测试CXR图像总数的比率。

实验设置：评价指标

- TB感染区检测
- 利用COCO benchmark提出的 AP^{bb} 进行评估，默认 AP^{bb} 是通过IoU阈值取平均值计算的，范围从0.5-0.95，步长为0.05.

$$IOU = \frac{\text{area of overlap}}{\text{area of union}} = \frac{\text{area of intersection}}{\text{area of union}}$$


- 引入两种评估模式：利用TBX11K测试集中的所有CXR图像，仅考虑TBX11K中的TB CXR图像

实验结果：CXR图像分类

TABLE 3

CXR image classification results (%) in terms of accuracy, AUC, sensitivity, specificity, AP, and AR on the TBX11K test data. The "Backbone" column indicates the specific backbone network used.

Methods	Backbones	Accuracy	AUC (TB)	Sensitivity	Specificity	Ave. Prec. (AP)	Ave. Rec. (AR)
SSD [72]	VGGNet-16	84.7	93.0	78.1	89.4	82.1	83.8
RetinaNet [69]	ResNet-50 w/ FPN	87.4	91.8	81.6	89.8	84.8	86.8
Faster R-CNN [74]	ResNet-50 w/ FPN	89.7	93.6	91.2	89.9	87.7	90.5
FCOS [73]	ResNet-50 w/ FPN	88.9	92.4	87.3	89.9	86.6	89.2
Deformable DETR [68]	ResNet-50 w/ FPN	91.3	97.6	89.2	95.3	89.8	91.0
SymFormer w/ Deformable DETR	ResNet-50 w/ FPN	94.3	98.5	87.3	97.3	93.2	93.2
SymFormer w/ RetinaNet	ResNet-50 w/ FPN	94.5	98.9	91.0	96.8	93.3	94.0
SymFormer w/ RetinaNet	P2T-Small w/ FPN	94.6	99.1	92.1	96.7	93.4	94.2

TABLE 4

CXR image classification results (%) in terms of the F_1 score and confusion matrix on the TBX11K test data, as well as the number of FLOPs, the number of parameters, and FPS of each model. "#Total" denotes the total number of test CXR images. We test FPS on a single TITAN XP GPU. For the ground truths, the ratio of positives (TP + FN) is 19.6%, and the ratio of negatives (TN + FP) is 80.4%.

Methods	Backbones	#FLOPs	#Params	FPS	F_1 score \uparrow	TP/#Total \uparrow	TN/#Total \uparrow	FP/#Total \downarrow	FN/#Total \downarrow
SSD [72]	VGGNet-16	90.58	38.69	32.9	70.5	15.3	71.9	8.5	4.3
RetinaNet [69]	ResNet-50 w/ FPN	55.41	48.97	35.3	73.1	16.0	72.2	8.2	3.6
Faster R-CNN [74]	ResNet-50 w/ FPN	66.27	53.98	30.3	78.5	17.9	72.3	8.1	1.7
FCOS [73]	ResNet-50 w/ FPN	53.33	44.69	39.9	76.3	17.1	72.3	8.1	2.5
Deformable DETR [68]	ResNet-50 w/ FPN	54.07	52.67	23.0	85.6	17.5	76.6	3.8	2.1
SymFormer w/ Deformable DETR	ResNet-50 w/ FPN	54.08	52.69	22.5	87.9	17.1	78.2	2.2	2.5
SymFormer w/ RetinaNet	ResNet-50 w/ FPN	59.14	50.03	24.3	89.0	17.8	77.8	2.6	1.8
SymFormer w/ RetinaNet	P2T-Small w/ FPN	55.46	45.10	17.9	89.6	18.1	77.7	2.7	1.5

FLOPs: 浮点运算次数

Params: 参数量: 模型训练中需要训练的参数总数, 用来衡量模型的大小

- 将SymFormer应用到RetinaNet和Deformable DETR中分别可以显著提高RetinaNet和Deformable DETR的性能
- 采用的默认模型是RetinaNet的SymFormer, 性能略低于Deformable DETR, 但在对象检测方面优于后者
- 所有方法的研究性能都超过了达到84.8%的放射科医生
- 在表3中, 我们观察到SymFormer和基准模型在敏感度方面的差异要比特异性方面的差异小得多。通过表4, 我们发现基准模型更倾向于做出积极预测 (TP+FP) 而不是负面预测 (TN+FN) 倾向于将测试CXR图像分为阳性。
- 在表4中, 对比原始和经过SymFormer增强的Deformable DETR, 发现SymFormer具有与Deformable DETR相似的FLOPs、Params和FPS, 因为SymFormer只向Deformable DETR引入了可忽略不计的计算

TABLE 5
CXR image classification results (%) in terms of accuracy, AUC, sensitivity, specificity, AP, and AR on the DA+DB test data [6].

Methods	Backbones	Accuracy	AUC (TB)	Sensitivity	Specificity	Ave. Prec. (AP)	Ave. Rec. (AR)
SSD [72]	VGGNet-16	51.0	53.8	100.0	1.9	75.3	51.0
RetinaNet [69]	ResNet-50 w/ FPN	50.0	50.0	100.0	0.0	25.0	50.0
Faster R-CNN [74]	ResNet-50 w/ FPN	50.0	51.9	100.0	0.0	25.0	50.0
FCOS [73]	ResNet-50 w/ FPN	50.0	52.1	100.0	0.0	25.0	50.0
Deformable DETR [68]	ResNet-50 w/ FPN	68.6	69.7	84.3	52.9	70.7	68.6
SymFormer w/ Deformable DETR	ResNet-50 w/ FPN	82.4	78.4	86.3	78.4	82.6	82.4
SymFormer w/ RetinaNet	ResNet-50 w/ FPN	78.4	74.7	90.2	66.7	80.1	78.4
SymFormer w/ RetinaNet	P2T-Small w/ FPN	84.3	89.4	84.3	84.3	84.3	84.3

TABLE 6
CXR image classification results (%) in terms of accuracy, AUC, sensitivity, specificity, AP, and AR on the MC+Shenzhen test data [25].

Methods	Backbones	Accuracy	AUC (TB)	Sensitivity	Specificity	Ave. Prec. (AP)	Ave. Rec. (AR)
SSD [72]	VGGNet-16	50.8	50.4	100.0	3.4	74.9	51.7
RetinaNet [69]	ResNet-50 w/ FPN	49.3	49.7	100.0	0.5	74.6	50.3
Faster R-CNN [74]	ResNet-50 w/ FPN	49.0	49.5	100.0	0.0	24.5	50.0
FCOS [73]	ResNet-50 w/ FPN	48.8	49.0	99.5	0.0	24.4	49.7
Deformable DETR [68]	ResNet-50 w/ FPN	81.3	83.5	92.9	70.1	83.0	81.5
SymFormer w/ Deformable DETR	ResNet-50 w/ FPN	82.0	84.7	89.3	75.0	82.7	82.1
SymFormer w/ RetinaNet	ResNet-50 w/ FPN	82.8	86.3	91.8	74.0	83.8	82.9
SymFormer w/ RetinaNet	P2T-Small w/ FPN	85.8	87.4	93.4	78.4	86.6	85.9

- 基准模型达到100的敏感度和0的特异性，进一步证实了假设：基准模型倾向于将测试CXR图像分类为阳性
- 在没有微调的情况下使用上述训练模型评估公共数据集上的CXR图像分类，表5为DA+DB测试数据的结果，表6为MC+深证测试数据的结果。表中SSD、RetinaNet、Faster R-CNN和FCOS的灵敏度约为100，特异性约为0，这表明基准模型很难学习CTD的鲁棒特征表示，经常将测试CXR数据错误地分类为TB病例。

实验结果：TB感染区域检测

TABLE 7

TB infection area detection results (%) on our TBX11K test set. The "Test Data" column specifies whether the evaluation was performed using all CXR images in the test set or only TB CXR images in the test set. The "Backbone" column indicates the specific backbone network used.

Methods	Test Data	Backbones	Category-agnostic TB		Active TB		Latent TB	
			AP_{50}^{bb}	AP^{bb}	AP_{50}^{bb}	AP^{bb}	AP_{50}^{bb}	AP^{bb}
SSD [72]	ALL	VGGNet-16	52.3	22.6	50.5	22.8	8.1	3.2
RetinaNet [69]		ResNet-50 w/ FPN	52.1	22.2	45.4	19.6	6.2	2.4
Faster R-CNN [74]		ResNet-50 w/ FPN	57.3	22.7	53.3	21.9	9.6	2.9
FCOS [73]		ResNet-50 w/ FPN	46.6	18.9	40.3	16.8	6.2	2.1
Deformable DETR [68]		ResNet-50 w/ FPN	51.7	22.0	48.9	21.2	7.1	1.9
SymFormer w/ Deformable DETR		ResNet-50 w/ FPN	57.0	23.3	52.1	22.7	7.1	2.0
SymFormer w/ RetinaNet		ResNet-50 w/ FPN	68.0	29.5	62.0	27.3	13.3	4.4
SymFormer w/ RetinaNet		P2T-Small w/ FPN	70.4	30.0	63.6	26.9	11.4	4.3
SSD [72]	Only TB	VGGNet-16	68.3	28.7	63.7	28.0	10.7	4.0
RetinaNet [69]		ResNet-50 w/ FPN	69.4	28.3	61.5	25.3	10.2	4.1
Faster R-CNN [74]		ResNet-50 w/ FPN	63.4	24.6	58.7	23.7	9.6	2.8
FCOS [73]		ResNet-50 w/ FPN	56.3	22.5	47.9	19.8	7.4	2.4
Deformable DETR [68]		ResNet-50 w/ FPN	57.4	24.2	54.5	23.5	7.6	2.3
SymFormer w/ Deformable DETR		ResNet-50 w/ FPN	60.8	24.5	55.2	23.8	9.2	2.6
SymFormer w/ RetinaNet		ResNet-50 w/ FPN	73.4	31.5	67.1	29.2	14.7	4.8
SymFormer w/ RetinaNet		P2T-Small w/ FPN	75.7	32.1	68.9	28.9	13.0	4.7

观察到 AP_{50}^{bb} 的性能通常优于 AP^{bb} ，表明虽然检测模型能够识别目标区域，但其定位精度不是很准确， AP_{50}^{bb} 比 AP^{bb} 更重要，因为IoU为0.5的预测框与目标TB区域足以帮助放射科医生识别TB感染区域。

- 为确保准确评估，丢弃了CXR图像中所有被CXR图像分类头分类为非结核的预测框
- 检测结果如图，具有Deformable DETR的SymFormer和具有RetinaNet的SymFormer相较于各自的基本方法都有了显著的改进，而具有RetinaNet的SymFormer比具有Deformable DETR的SymFormer的性能要好得多，这表明SymFormer更适合与RetinaNet框架集成。因此，选择RetinaNet与SymFormer作为CTD的默认模型。
- 所有方法都难以准确预测潜伏性TB区域，许多不知名类别TB的评价优于活动性TB，这表明许多潜伏性TB被正确定位，但被归类为活动性TB，这可能是因为TB CXR图像过少

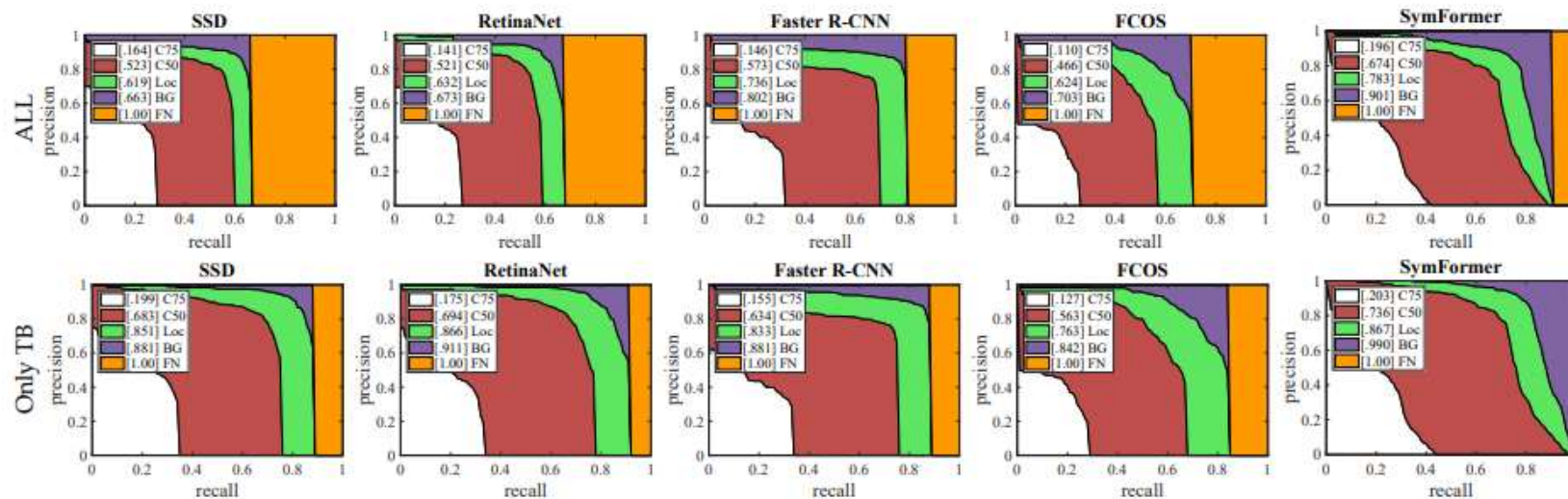


Fig. 5. Error analyses of category-agnostic TB area detection using baseline models and SymFormer w/ RetinaNet. The first row is evaluated using all CXR images, while the second row only uses TB CXR images. C50/C75: PR curves under IoU thresholds of 0.5/0.75. Loc: the PR curve under the IoU threshold of 0.1. BG: removing background false positives. FN: removing other errors caused by undetected targets (false negatives). SymFormer largely outperforms other methods in all metrics, e.g., obtaining a remarkable 99% BG score when only using TB CXR images.

- 图中我们给出了检测误差分析的精确召回率（PR），重点是未知类别的TB检测，很明显当IoU阈值从0.75到0.5转变时，所有方法都有改进。这表明由于有限的对象定位能力，所有方法在较高的IoU阈值下都将受到挑战。
- 将使用所有CXR图像获得的结果与仅使用TB CXR获得的结果进行对比，发现当使用所有CXR图像时，FN（假阴）的面积更大。这表明基于图像分类的过滤过程忽略了许多正确检测到的TB区域，SymFormer的FN区域明显小于其他，突出了它的优越能力。

实验结果：消融研究

TABLE 8

Ablation study for TB infection area detection on our TBX11K validation set. We only use TB CXR images to evaluate category-agnostic TB area detection. The "Symmetry of SPE" column indicates whether SPE transfers the right side of positional encoding to the left side, or vice versa. APE: absolute positional encoding; RPE: relative positional encoding.

Attention	Positional Encoding	Symmetry of SPE	AP ₅₀ ^{bb}	AP ^{bb}
No	No	-	72.7	31.0
Vanilla	APE	-	73.4	30.6
Vanilla	RPE	-	72.7	29.7
Vanilla	SPE w/o STN	left → right	74.0	30.5
Vanilla	SPE w/o STN	right → left	74.3	30.8
Vanilla	SPE	left → right	75.1	30.4
Vanilla	SPE	right → left	75.7	29.6
SymAttention	APE	-	74.9	30.0
SymAttention	RPE	-	73.6	29.1
SymAttention	SPE w/o STN	left → right	75.3	31.4
SymAttention	SPE w/o STN	right → left	75.5	30.7
SymAttention	SPE	left → right	76.3	30.9
SymAttention	SPE	right → left	76.6	31.7

- 使用TBX11K数据集的训练集来训练模型，并在验证集上对它进行评估。基准模型是RetinaNet，与表8中第一个模型对应，不含任何attention和位置编码
- Vanilla Attention指的是在Deformable DETR中使用的Deformable attention。
- SPE默认将位置编码的右侧传输到左侧，在这里也评估了左侧传输到右侧时的性能。
- 表8表示相对位置编码的性能不如绝对位置编码，所以使用绝对位置编码来构建SPE。
- STN的加入进一步提高了SPE的性能

TABLE 9

Evaluation results (%) of the 4-fold cross-validation. The used model is our SymFormer w/ RetinaNet using ResNet-50 w/ FPN as the backbone. We split the TBX11K `trainval` set into 4 folds, each of which has a similar class distribution. “#Total” denotes the total number of test CXR images in each fold. We only use TB CXR images to evaluate category-agnostic TB area detection.

Fold	Accuracy	AUC	Sensitivity	Specificity	AP	AR	F_1 score	TP/#Total	TN/#Total	FP/#Total	FN/#Total	AP_{50}^{bb}	AP^{bb}
1	94.7	98.9	92.2	97.5	91.1	94.6	87.6	11.0	85.8	2.2	0.9	74.7	31.7
2	95.2	98.9	92.6	97.6	91.7	95.0	88.1	11.1	86.0	2.1	0.9	75.2	30.4
3	94.6	99.1	92.9	96.9	90.5	94.6	86.4	11.1	85.3	2.7	0.8	74.1	29.5
4	95.1	99.3	92.9	97.3	91.2	94.7	87.4	11.1	85.7	2.4	0.8	75.4	33.3

- 为了证明SymFormer的鲁棒性，我们进行了四折交叉验证，将TBX11K训练数据划分为四层，确保每层保持相似的类分布。每次实验中，我们在其中三个折叠上训练SymFormer，并评估其在剩余折叠上的性能。可以观察到这些数据非常一致，证实了SymFormer的鲁棒性

实验结果：可视化

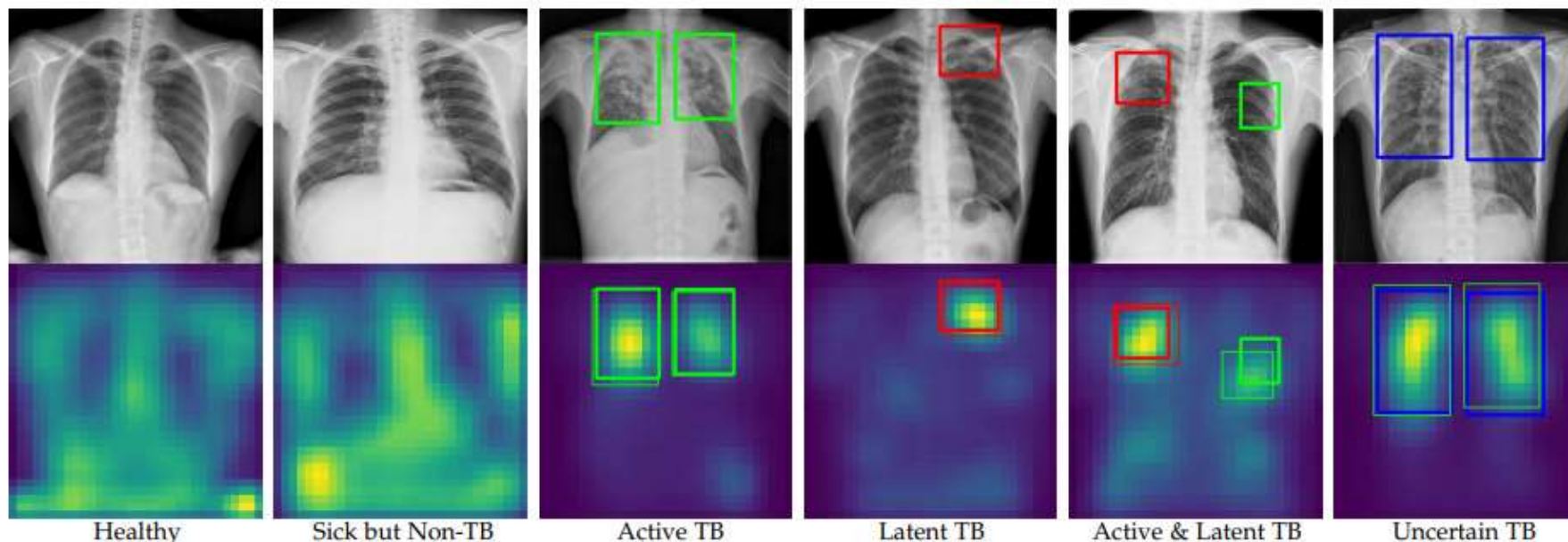
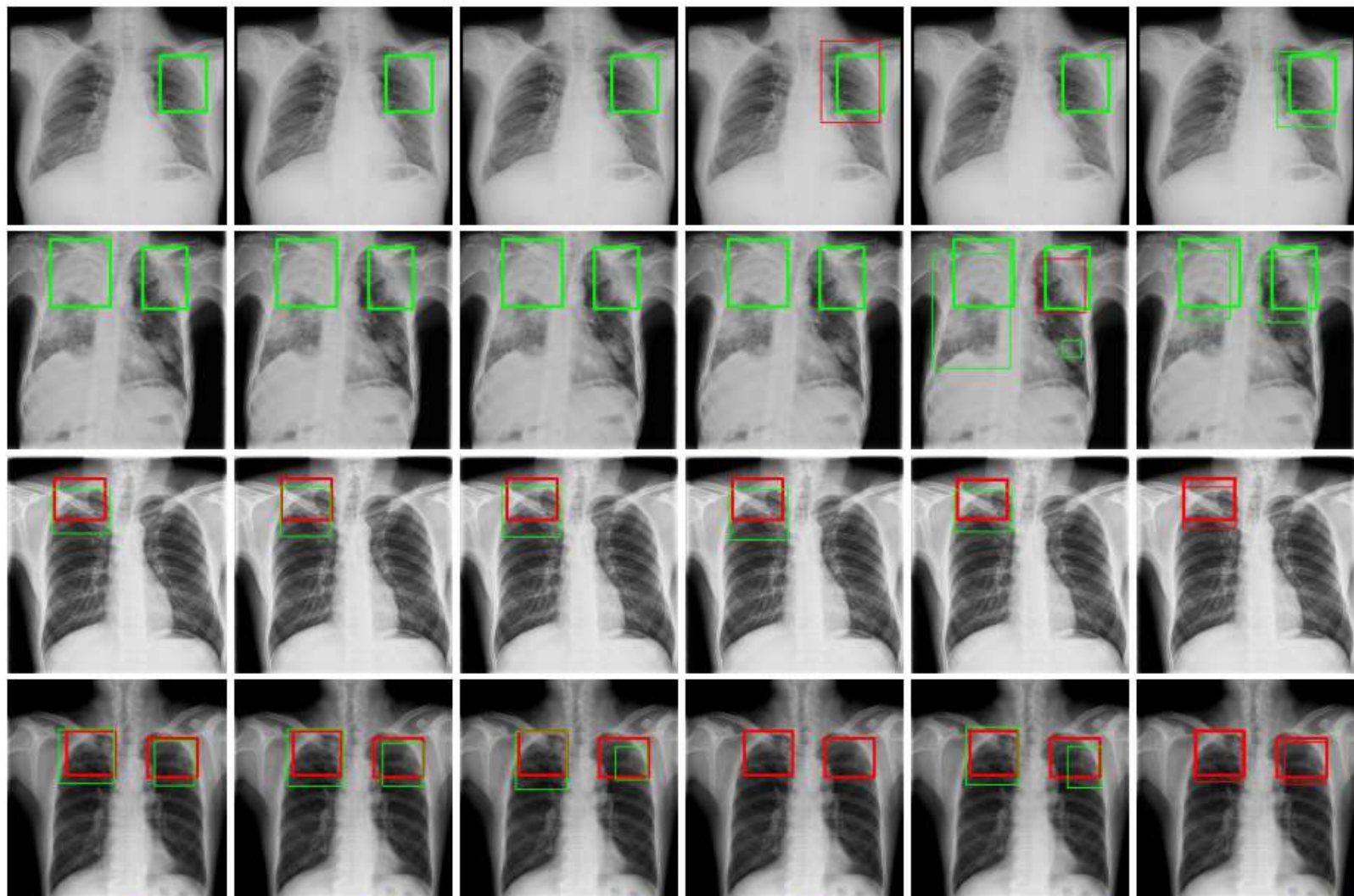


Fig. 6. **Visualization of the learned deep features from CXR images using SymFormer w/ RetinaNet.** We randomly select CXR images from the TBX11K test set, and for each class mentioned in Table 2, we provide one example. In each example, the infection areas of active TB, latent TB, and uncertain TB are indicated by boxes colored in green, red, and blue, respectively. The ground-truth boxes are displayed with thick lines, while the detected boxes are shown with thin lines.

- 图6展示了学习到的特征的可视化以及相应的检测结果，观察到健康病例的可视化表现出不规则的特征模式，表明没有明显的异常。相比之下患病但非TB的可视化显示出一些明显的两点，可能代表病变存在。对于TB病例，可视化地图中的亮点与标注的TB感染区域很好地对齐，表明所提出的SymFormer在学习TB检测的深度特征方面是有效的。



SSD [72]

RetinaNet [69]

Faster R-CNN [74]

FCOS [73]

Deformable DETR [68]

SymFormer w/ RetinaNet

Fig. 7. **Qualitative comparison between the proposed SymFormer and baseline methods.** In each example, the infection areas of active TB and latent TB are indicated by green boxes and red boxes, respectively. The ground-truth boxes are displayed with thick lines, while the detected boxes are shown with thin lines. For all examples, SymFormer can detect all TB infection areas with true categories.

- 对提出的SymFormer和基准模型进行了定性比较，SymFormer始终提供卓越的定性检测结果。

结论

- 缺乏标注数据阻碍了深度学习在计算机辅助结核病诊断方向的进展，所以本文引入了TBX11K，一个带有边界框注释的大型TB数据集，这个数据集不仅方便了训练CTD的深度神经网络，也是专门为TB检测设计的第一个数据集。
- 提出了一个简单有效的框架SymFormer，用于同时进行CXR图像分类和TB感染区域检测。利用CXR图像固有的双边对称性，SymFormer结合SymAttention来提取独特的特征表示。由于CXR图像可能不会表现出严格的对称性，本文引入了对称位置编码SPE，通过特征重新校准来增强SymAttention的性能。
- 为了提供CTD研究的基准，引入了评估指标，评估了基于现有目标检测器的Bsaeline Modles，并发起了在线挑战。

Thank You!

Avenida da Universidade, Taipa, Macau, China
Tel : (853) 8822 8833 Fax : (853) 8822 8822
Email : info@um.edu.mo Website : www.um.edu.mo