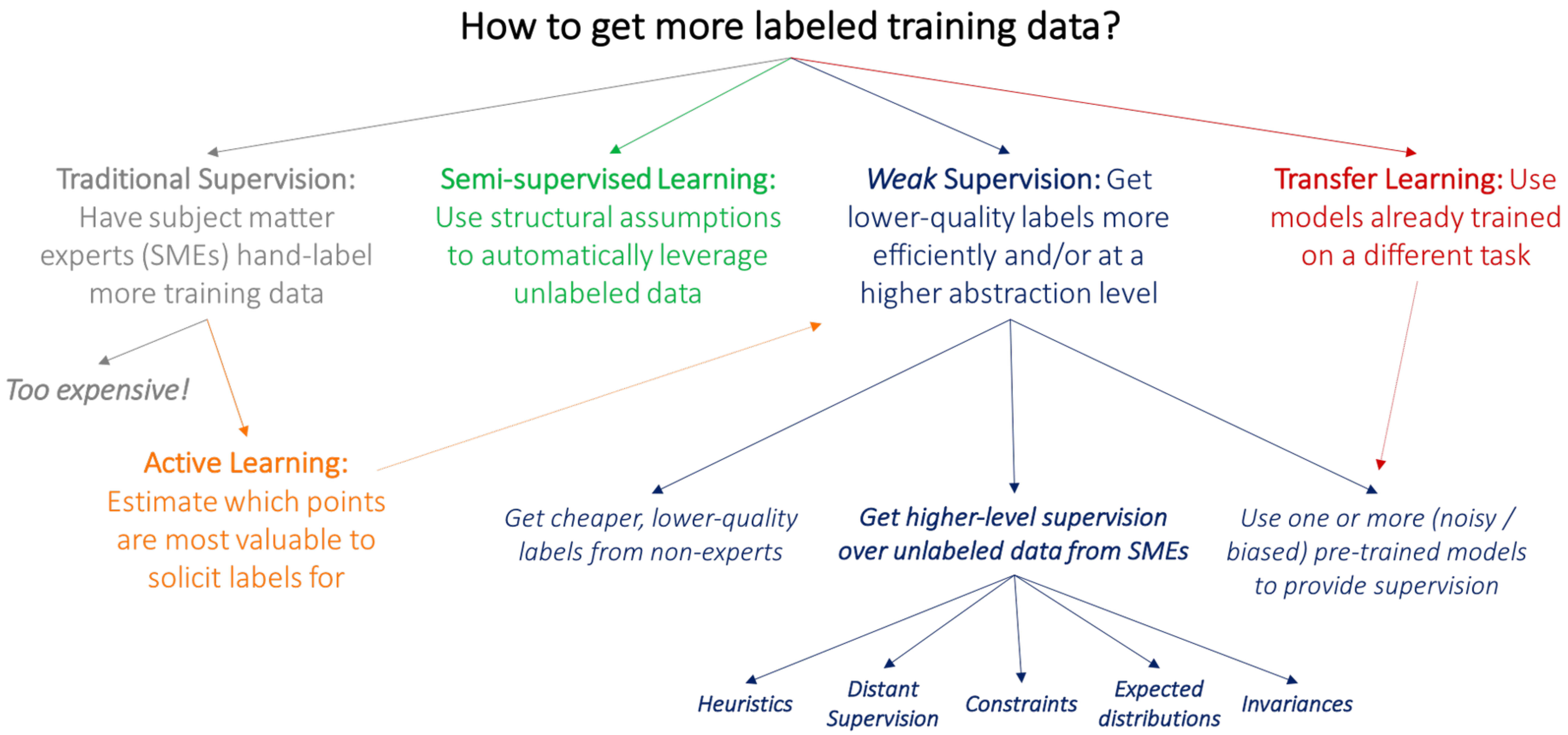
Paper List	
CVPR2020	FocalMix: Semi-Supervised Learning for 3D Medical Image Detection
CVPR2021	Instant-Teaching: An End-to-End Semi-Supervised Object Detection Framework

Introduction

What is semi-supervised



Introduction

Why we need semi-supervised

- 对应用场景一致数据的依赖: 以人体检测模型为例，在室外数据上训练的检测器，在室内场景的检测效果往往差强人意；在白天数据上训练的检测器，在夜晚场景下的效果通常也不会很好

- 对数据规模的依赖: 模型的性能严重依赖带标注的训练数据, 大数据集训练的模型往往比小数据集的精度要高

Relevant Work

半监督学习的两个主要思路

- 基于一致性约束的方案: 通过对未标注数据做弱增强(flip)，组成pair对输入给检测模型，然后对模型预测输出的pair结果进行一致性约束，从而尽可能利用到这些未标注数据 -> FocalMix: Semi-Supervised Learning for 3D Medical Image Detection

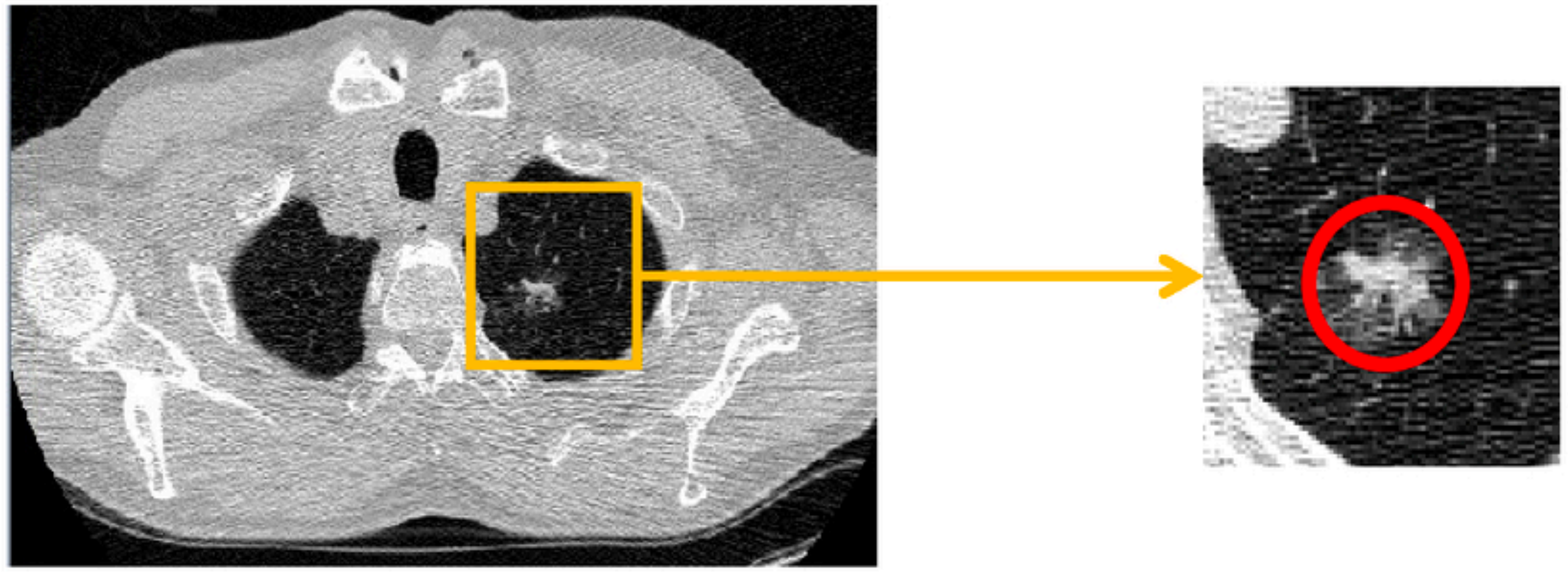
- 基于teacher-student蒸馏方案: 首先所有标注数据训练一个Teacher模型，然后在所有未标注数据上做Inference得到pseudo labels作为未标注数据的ground truth -> Instant-Teaching: An End-to-End Semi-Supervised Object Detection Framework

Information

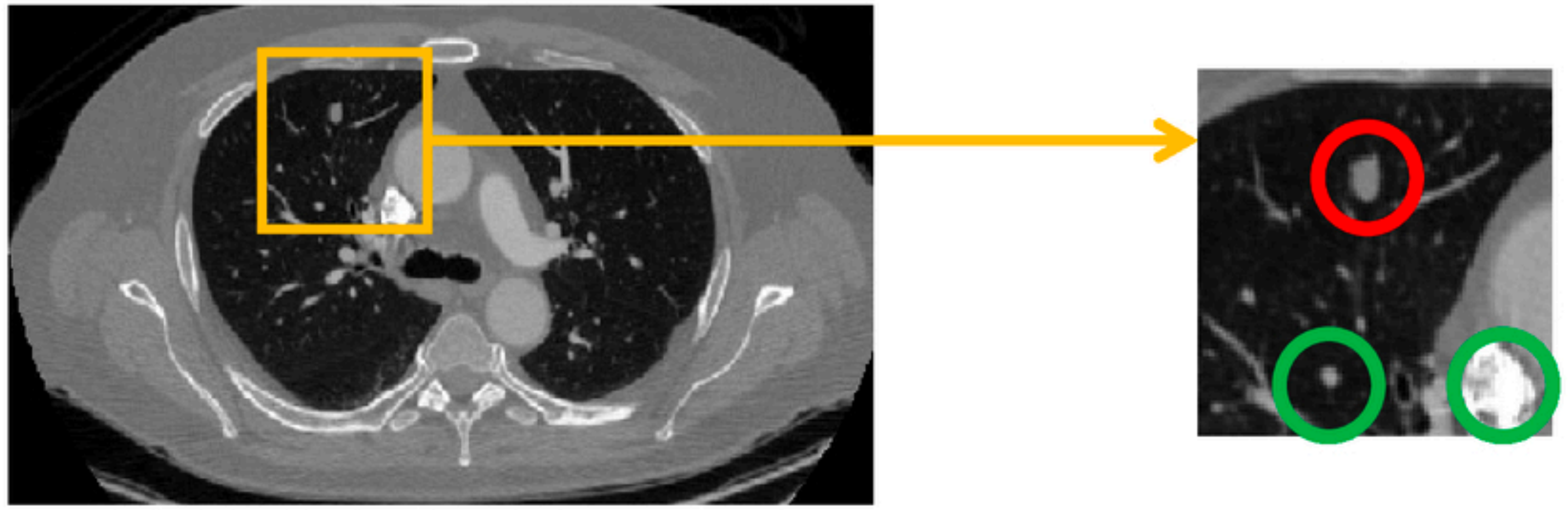
FocalMix: Semi-Supervised Learning for 3D Medical Image Detection

Dong Wang[1*] Yuan Zhang[2*] Kexin Zhang[2,3†] Liwei Wang[1,2]
[1]Center for Data Science, Peking University
[2]Key Laboratory of Machine Perception, MOE, School of EECS, Peking University
[3]Yizhun Medical AI Co., Ltd
{wangdongcis,yuan.z,zhangkexin,wanglw}@pku.edu.cn

Dataset - LUNA 16 and NLST



(a)



(b)

- Format: CT
- Dim: 3D
- Input size: 160 * 160 * 160
- Annotations: 3d bbox

Methodology

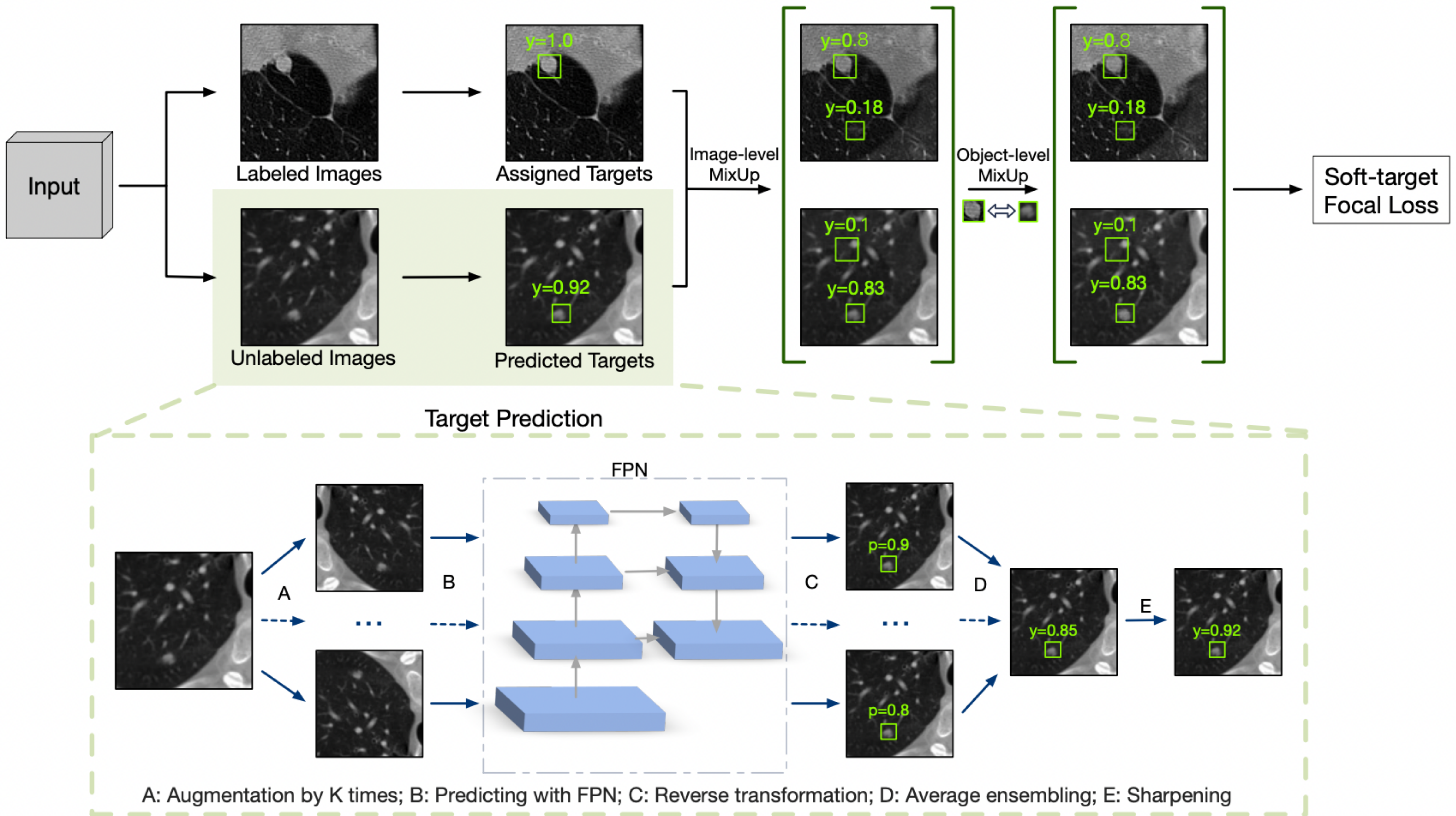
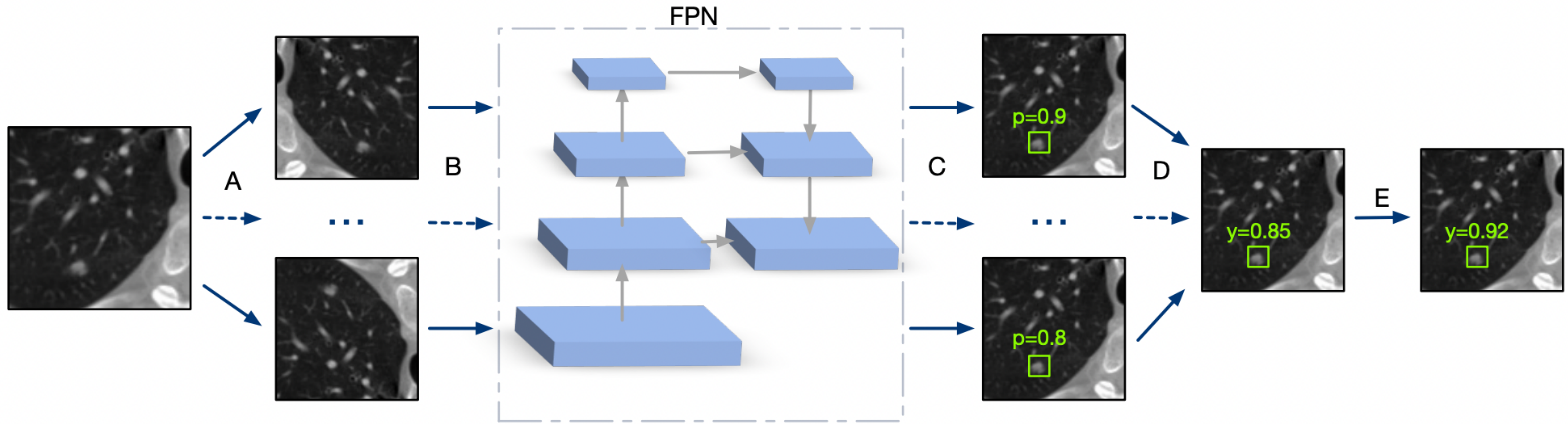


Figure 2: **Overview of our proposed method FocalMix.** For an input batch, the training targets of anchors in labeled images are assigned according to annotated boxes, while the unlabeled are predicted with the current model as shown in the lower part of the figure. After applying two levels of MixUp to the entire batch, we use the proposed soft-target focal loss to train the model. Throughout this paper, we only show a slice of each 3D CT scan with 3D anchors on it for ease of presentation.

Methodology

1. Target Prediction Augmentation

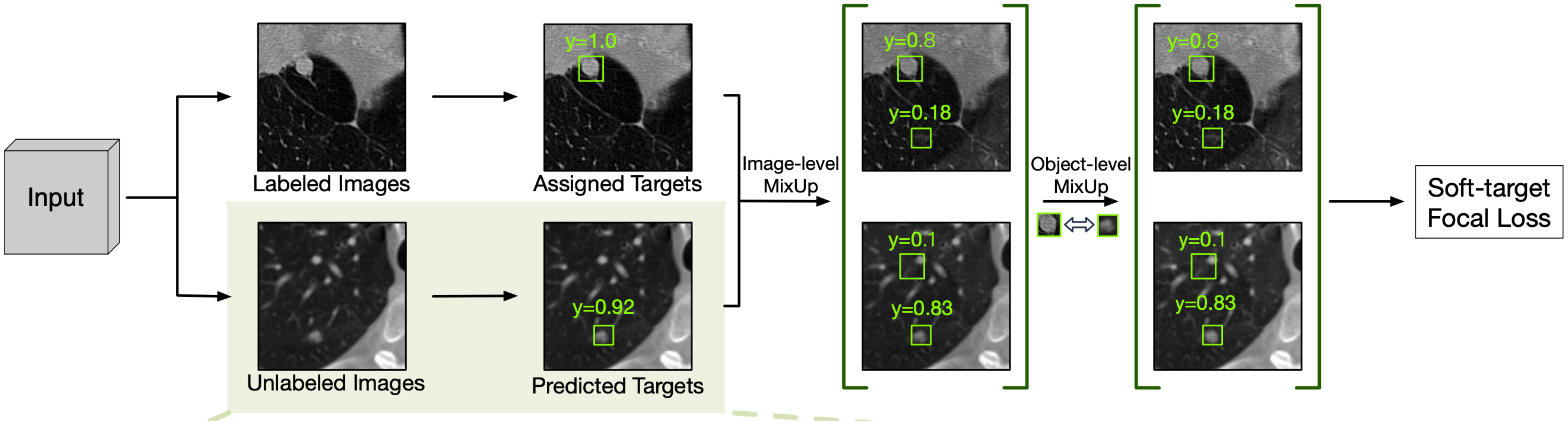


A: Augmentation by K times; B: Predicting with FPN; C: Reverse transformation; D: Average ensembling; E: Sharpening

$$\text{Sharpen}(\bar{y}, T)_i = \bar{y}_i^{\frac{1}{T}} \bigg/ \sum_{j=1}^{L} \bar{y}_j^{\frac{1}{T}},$$

Methodology

2. MixUp augmentation



- Image-level mixup

$$\lambda \sim \text{Beta}(\eta, \eta), \tag{10}$$

$$\tilde{\lambda} = \max(\lambda, 1 - \lambda), \tag{11}$$

$$\hat{x} = \tilde{\lambda} x + (1 - \tilde{\lambda}) x', \tag{12}$$

$$\hat{y}_i = \tilde{\lambda} y_i + (1 - \tilde{\lambda}) y_i', \forall i. \tag{13}$$

Results

Labeled	Unlabeled	Recall(%) @ FPs							CPM(%)	Improv.
		0.125	0.25	0.5	1	2	4	8		
25	-	46.7	54.0	60.6	68.6	74.4	79.1	82.4	66.6	**11.5 (17.3%)**
25	400	57.6	64.5	74.6	80.5	87.0	90.1	92.1	**78.1**	
50	-	57.2	65.7	71.4	77.9	82.6	85.6	87.2	75.4	**6.6 (8.8%)**
50	400	64.1	71.0	78.7	85.2	89.3	92.3	93.5	**82.0**	
100	-	64.9	73.8	79.7	85.2	89.0	92.3	94.5	82.8	**4.4 (5.3%)**
100	400	73.4	80.9	84.8	88.6	92.3	94.7	96.1	**87.2**	

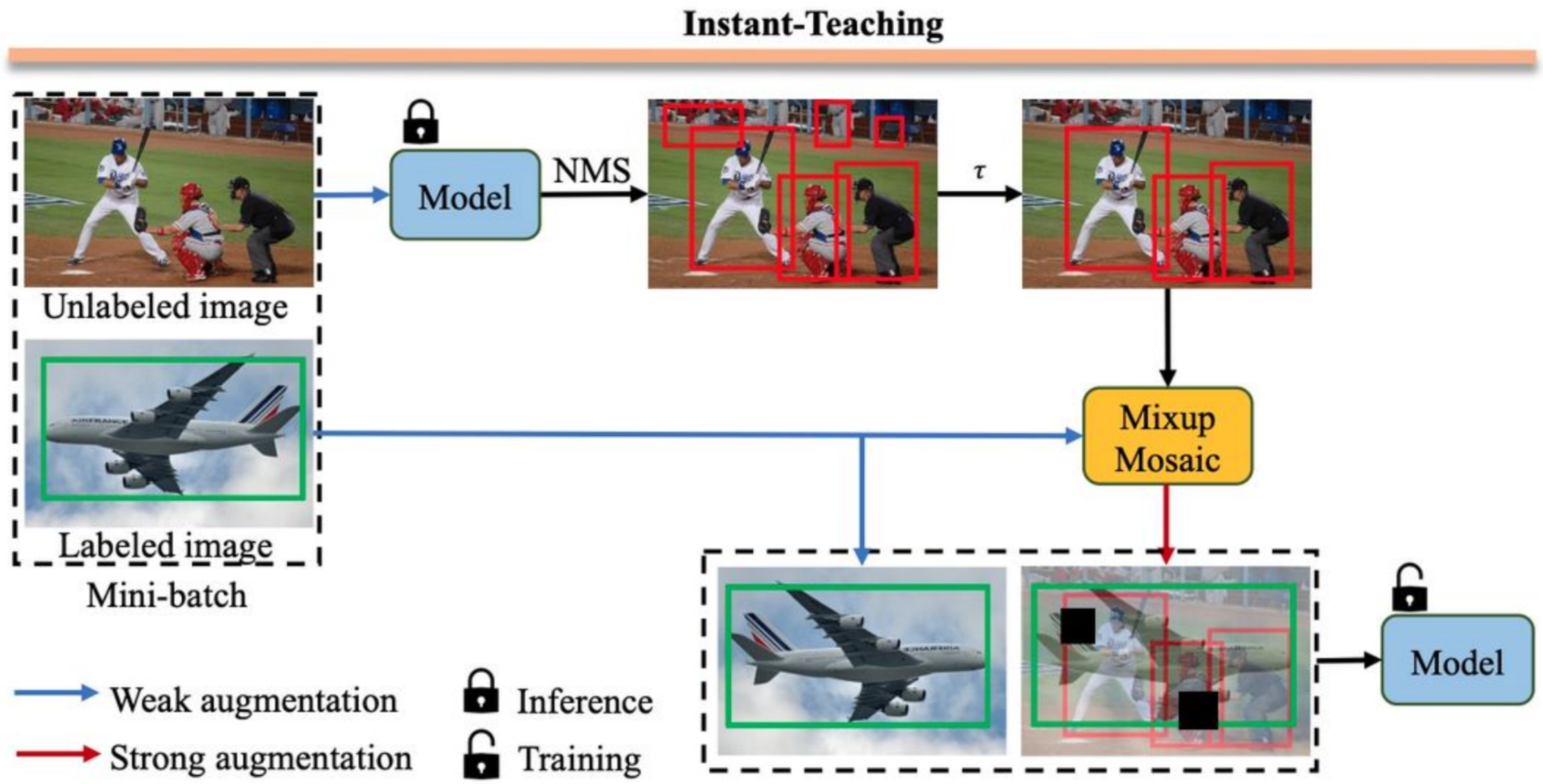
Table 1: **Main results on the LUNA16 dataset.** We evaluate FocalMix with $\{25, 50, 100\}$ labeled CT scans, respectively. *Improv.* denotes the improvements in CPM over the fully-supervised baseline (relative improvements shown in parentheses).

Information

FocalMix: Semi-Supervised Learning for 3D Medical Image Detection

Dong Wang[1*] Yuan Zhang[2*] Kexin Zhang[2,3†] Liwei Wang[1,2]
[1]Center for Data Science, Peking University
[2]Key Laboratory of Machine Perception, MOE, School of EECS, Peking University
[3]Yizhun Medical AI Co., Ltd
{wangdongcis,yuan.z,zhangkexin,wanglw}@pku.edu.cn

Methodology



1. 用已有的标签图像训练一个教师模型(teacher model)用来生成伪标签
2. 用训练好的模型推理剩余的未标注的图像，生成伪标签
3. 对未标注的数据进行增强 (Mixup Mosaic)
4. 使用半监督Loss来训练检测器

Results

Methods	Backbone	Unlabeled	$AP^{0.5:0.95}$	$AP^{0.5}$	$AP^{0.75}$
Supervised (Ours)	R50-FPN		43.60	76.70	44.50
CSD [22]	R101-R-FCN	VOC12	-	74.70	-
STAC [45]	R50-FPN		44.64 (+1.04)	77.45	-
Instant-Teaching	R50-FPN		**48.70** (+5.10)	**78.30**	**52.00** (+7.50)
Instant-Teaching*	R50-FPN		**50.00** (+6.40)	**79.20**	**54.00** (+9.50)
CSD [22]	R101-R-FCN	VOC12 & COCO	-	75.10	-
STAC [45]	R50-FPN		46.01 (+2.41)	79.08	-
Instant-Teaching	R50-FPN		**49.70** (+6.10)	79.00	**54.10** (+9.60)
Instant-Teaching*	R50-FPN		**50.80** (+7.20)	**79.90**	**55.70** (+11.20)

Table 2. Comparison of mAP for different semi-supervised methods on VOC07. We report the mAP at IoU=0.50:0.95 ($AP^{0.5:0.95}$), IoU=0.5 ($AP^{0.5}$) and IoU=0.75 ($AP^{0.75}$), which are the standard metrics for object detection [31, 7].