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

- Anchor boxes

- Focal Loss

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t)$$

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{otherwise.} \end{cases}$$

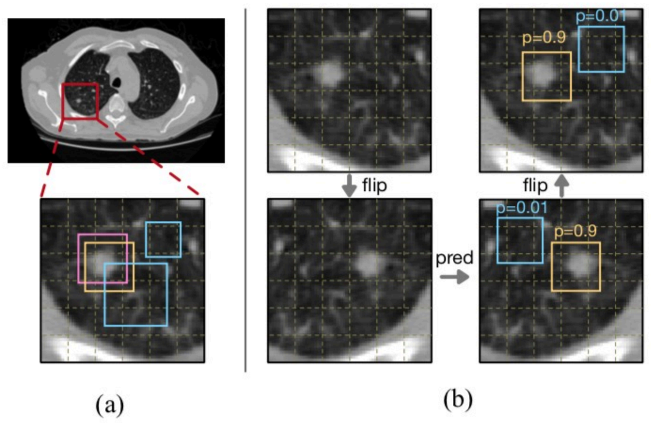


Figure 1: (a) is an example of assigning targets to anchors. The dashed grids represent output feature maps where anchor boxes are defined, and each bin in the grids corresponds to a point in the feature map. The pink box is a ground-truth bounding box. The orange box is a positive anchor and the blue boxes are negative anchors. (b) is an example of our augmentation method used for target prediction. We use flip augmentation for the image patch and predict the probability for each anchor with the model. After that, an inverse transformation is applied to the patch and anchors. We only show two example anchors for illustration purposes and use consistent coloring for each anchor. Note that anchors in 3D images are also three-dimensional, of which we only show 2D slices for better visualization.

- Semi-supervised Learning

  Mix-Match: target prediction & MixUp Augmentation

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

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{otherwise.} \end{cases}$$

$$\alpha(y) = \alpha_0 + y(\alpha_1 - \alpha_0).$$

$$\beta(y,p) = |y-p|^\gamma.$$

Soft-target focal loss for SSL:

$$SFL(p) = [\alpha_0 + y(\alpha_1 - \alpha_0)] \cdot |y-p|^\gamma \cdot CE(y,p), \quad CE(y,p) = -y\log p - (1-y)\log(1-p)$$

focal loss is a special case of our proposed soft-target focal loss when y ∈ {0, 1}.

$$
\begin{aligned}
\lambda &\sim \mathrm{Beta}(\eta, \eta), \\
\tilde{\lambda} &= \max(\lambda, 1 - \lambda), \\
\hat{x} &= \tilde{\lambda} x + (1 - \tilde{\lambda}) x', \\
\hat{y}_i &= \tilde{\lambda} y_i + (1 - \tilde{\lambda}) y_i', \forall i.
\end{aligned}
$$

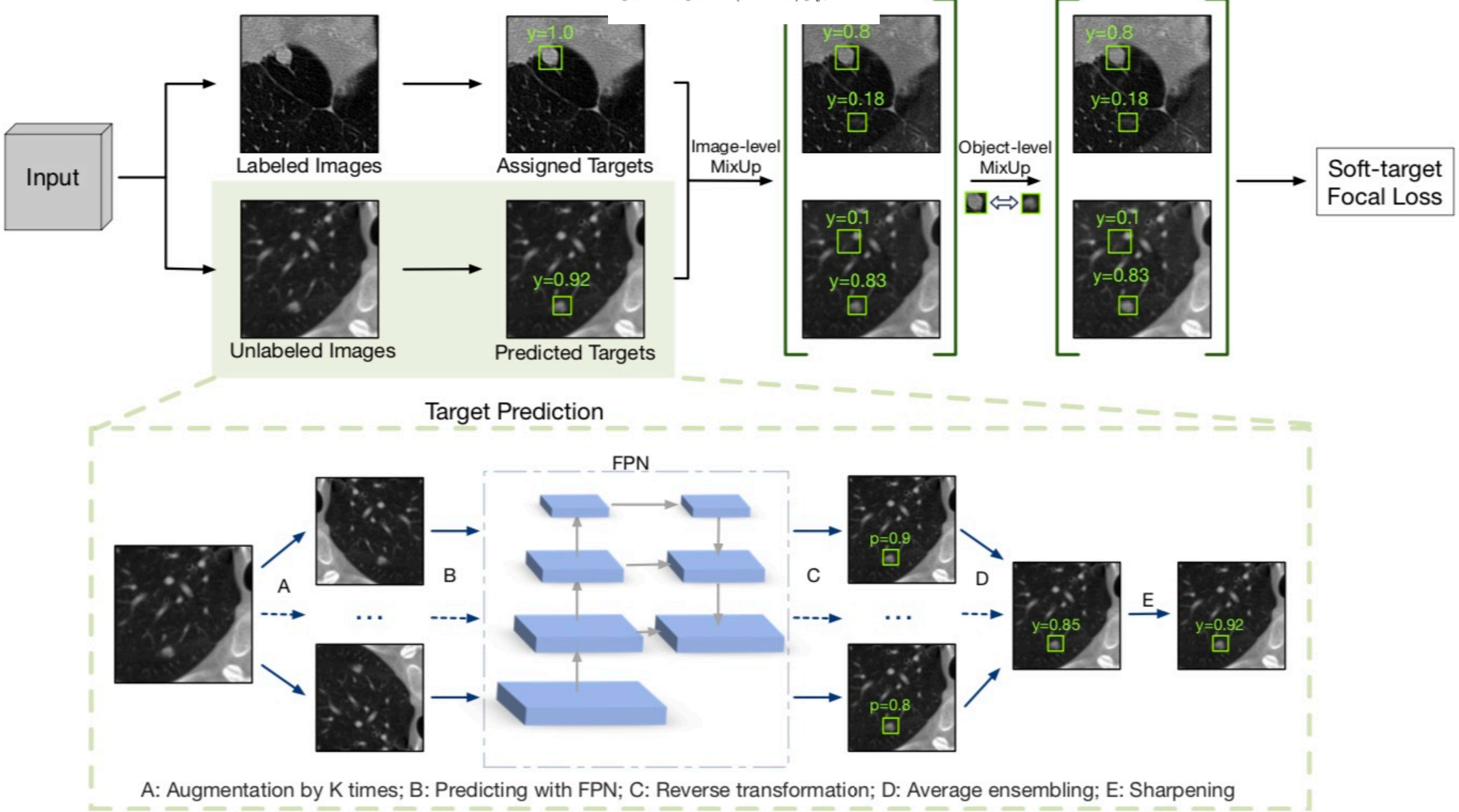


Figure 2: **Overview of our proposed method FocalMix.** For an input batch, the training targets of anchors in labeled images are assigned according to annotated boxes, while the unlabeled are predicted with the current model as shown in the lower part of the figure. After applying two levels of MixUp 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.

Experiment

- Dataset: LUNA16, NLST
- Evaluation: FROC, CPM

- Detection Model：A 3D variant of FPN
- Semi-supervised Learning
- Full-Supervised Baseline Performance

Method	Data Split	CPM(%)
DeepLung [41]	10-fold	84.2
DeepSeed [19]	10-fold	86.2
S4ND [14]	10-fold	89.7
3D FPN [23]	10-fold	91.9
Our base model	10-fold	91.2
Our base model	533/355	89.2

Table 2: **Performance of the base model used in our experiments.** Our re-implemented 3D FPN is comparable with state-of-the-art single-stage nodule detection models.

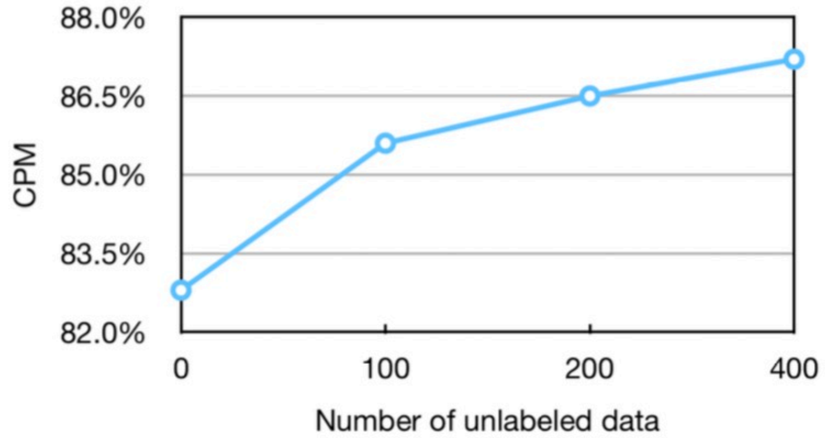


Figure 3: **Performance with different amounts of unlabeled data on LUNA16.** We use 100 labeled images.

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).

Contributions

- Proposed FocalMix, a novel semi-supervised learning framework for 3D medical image detection.
- First to investigate the problem of semi-supervised learning for medical image detection.
- Demonstrated that the proposed semi-supervised approach can significantly improve the performance of fully-supervised learning approaches.

SOS: Selective Objective Switch for Rapid Immunofluorescence Whole Slide Image Classification

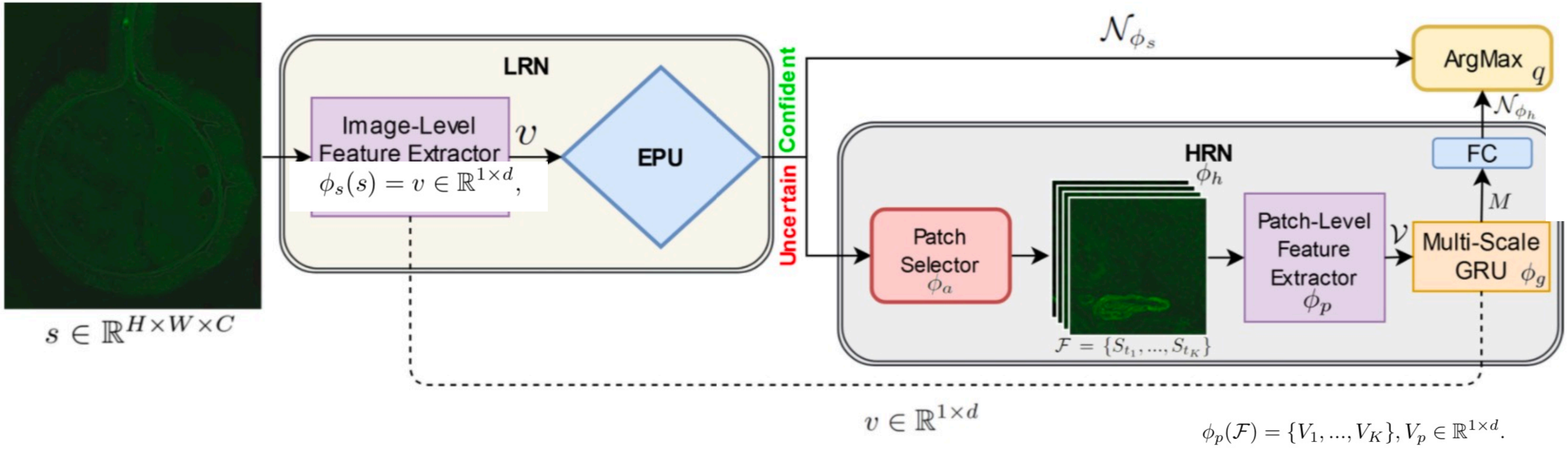
Sam Maksoud[1,2] Kun Zhao[1] Peter Hobson[2] Anthony Jennings[2] Brian C. Lovell[1]
[1]The University of Queensland, St Lucia QLD 4072, Australia
[2]Sullivan Nicolaides Pathology, Bowen Hills, QLD 4006, Australia

Avoid excessive high resolution patch-level for WSIs that can be classified at image level

estimates a set of class probabilities $\mathcal{N}_{\phi_s} = \{N_{s1}, ..., N_{sn}\}$, where n is number of WSI classes. To compute $\mathcal{N}_{\phi_s}$, we apply a linear transformation to v followed by the softmax function σ:

$$\mathcal{N}_{\phi_s} = \sigma\left(vA_s^T + b_s\right), \tag{2}$$

where $A_s \in \mathbb{R}^{n \times d}$ and $b_s \in \mathbb{R}^n$ are parameters learned by



$$\mathcal{N}_{\phi_h} = \sigma\left(MA_m^T + b_m\right),$$

$$q = \arg\max\left(\mathcal{N}_{\phi_h}\right),$$

Figure 2: Framework of the SOS protocol. Dashed lines indicate the residual connection between the LRN and HRN.

Classification Loss:

$$L_{ce_1} = \frac{1}{B}\sum_{o=1}^{B}\left(-\sum_{i=1}^{n} y_{o,i}\log(N_{s\,o,i})\right),$$

Paradoxical Loss:

$$L_2 = \frac{1}{B}\sum_{o=1}^{B}\max\left(N_{sx,o} - N_{hx,o}, 0\right),$$

Executive Loss:

$$L_{he} = \sum_{o=1}^{B} y_{s,o}\max\left(\left((c+\epsilon) - \max\left(\mathcal{N}_{\phi_s}\right)\right), 0\right),$$

$$L_{hu} = \sum_{o=1}^{B} y_{h,o}\neg(y_{s,o})\max\left(\left(\max\left(\mathcal{N}_{\phi_s}\right)\right) - c, 0\right);$$

$$L_3 = \frac{1}{B}\left(\lambda_1 L_{he} + \lambda_2 L_{hu}\right),$$

Results

Method	TA% ↑	RS ↓	IT(s) ↓	SB ↑	LP
Image-Level	81.95	**1.00**	**8.37**	**14.59**	-
Patch-Level	69.27	1.50	94.10	1.3	-
Multi-Scale	85.37	2.17	122.10	1.00	-
RDMS	88.78	3.83	57.30	2.13	0.55
SOS (ours)	**90.73**	2.17	15.78	7.74	0.94

Table 3: Comparison of Total Accuracy (TA), Relative Size (RS), Inference Time (IT) and Speed Boost (SB) metrics. The ratio of low resolution predictions (LP) is also provided for the dynamic multi-scale classification methods.

- Image-Level
- Patch-Level
- Conventional Multiscale
- Reinforced Dynamic Multiscale

- Processing Speed:

Inference Time & Speed Boost

- Model Size:

Relative Size

- Classification Accuracy:

Total Accuracy

Method	F1 ↑	PR ↑	RE ↑	SP ↑
Image-Level	0.8800	0.8115	**0.9612**	0.7745
Patch-Level	0.7967	0.6853	0.9515	0.5588
Multi-Scale	0.9083	0.8609	**0.9612**	0.8431
RDMS	0.9300	0.9588	0.9029	**0.9608**
SOS (ours)	**0.9406**	**0.9596**	0.9223	**0.9608**

(a) **Negative.** Evaluation of Negative classification performance.

Method	F1 ↑	PR ↑	RE ↑	SP ↑
Image-Level	0.8989	0.9090	0.8889	**0.9750**
Patch-Level	0.8471	0.9000	0.8000	**0.9750**
Multi-Scale	0.8706	**0.9250**	0.8222	0.9813
RDMS	0.9149	0.8776	**0.9556**	0.9625
SOS (ours)	**0.9348**	0.9149	**0.9556**	**0.9750**

(b) **AMA.** Evaluation of AMA classification performance.

Method	F1 ↑	PR ↑	RE ↑	SP ↑
Image-Level	0.6667	0.7368	0.6087	0.9371
Patch-Level	0.2353	0.3636	0.1739	0.912
Multi-Scale	0.7778	0.7955	0.7609	**0.9434**
RDMS	0.8367	0.7885	**0.8913**	0.9308
SOS (ours)	**0.8542**	**0.8200**	**0.8913**	**0.9434**

(c) **SMA-V.** Evaluation of SMA-V classification performance.

Method	F1 ↑	PR ↑	RE ↑	SP ↑
Image-Level	0.1667	**1.000**	0.0909	**1.000**
Patch-Level	0.0000	0.0000	0.000	**1.000**
Multi-Scale	0.4706	0.6667	0.3636	0.9897
RDMS	0.5556	0.7143	0.4545	0.9897
SOS (ours)	**0.7000**	0.7778	**0.6364**	0.9897

(d) **SMA-T.** Evaluation of SMA-T classification performance.

Table 4: Evaluation of F1 scores, Precision (PR), Recall (RE) and Specificity (SP) for each of the four WSI classes.

Novel Lover-Kidney-Stomach Dataset

Set	Neg	AMA	SMA-V	SMA-T	Total
Train	239	106	107	27	479
Test	103	45	46	11	205

(a) The distribution of classes in the train and test set.

Size	Resolution	Objective	Format
300GB	$40000 \times 40000 \times 1$	$\times 20$	TIFF

(b) Meta-Information pertaining to the LKS dataset.

Table 1: Structure of the Liver-Kidney-Stomach Dataset.

A team of trained medical scientists manually labelled the slides into one of four classes: Negative (Neg); Anti-Mitochondrial Antibodies (AMA); Vessel-Type Anti-Smooth Muscle Antibodies (SMA-V) and Tubule-Type Anti-Smooth Muscle Antibodies (SMA-T).

Contributions

- The first to propose a Dynamic Multi-Scale WSI classification network which regulates the use of high-resolution image streams via the uncertainty of predictions at low resolution;
- Introduced a novel learning constraint, the paradoxical loss, to discourage asynchronous optimization of the LRN and HRN during training;
- Will release our novel dataset1 of 684 LKS WSIs to the community. This will be the first publicly available dataset for multi-tissue IIF WSI analysis.