Det-Y: A Multi-center Dataset and Benchmark for Efficient Detection of Mycobacterium Tuberculosis in Sputum Smears

Zijian Xie[1], Yongjie Liang[1][0009−0001−1059−088X], Yawen Huang[2], Xian Wu[2], Bizhong Wei[3,4,5][0009−0000−1523−1125], and Yuexiang Li[1,4,5][0000−0001−8076−2619]*

[1] Life Sciences Institute, Guangxi Medical University, Nanning, 530021, China
[2] Tencent Jarvis Lab, Shenzhen, 518054, China
[3] School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, 541004, China
[4] Guilin University of Electronic Technology Nanning Research Institute, Nanning, 530201, China
[5] Guangxi Academy of Artificial Intelligence, Nanning, 530201, China

Abstract. Tuberculosis (TB) remains a significant global health threat. The detection of Mycobacterium tuberculosis in sputum smears is an effective approach for TB diagnosis; however, traditional manual examination is laborious, time-consuming, and susceptible to inter-observer variation. Furthermore, the absence of annotated datasets has hindered the application of efficient deep learning-based models in this domain. To address this gap, we introduce Det-Y—a multi-center dataset and benchmark designed to facilitate the development of efficient deep learning frameworks for detecting Mycobacterium tuberculosis in sputum smears. Det-Y comprises 5,444 annotated bacilli from 1,970 sputum smear images collected across two clinical centers. In addition, we propose YOLO-Y, a lightweight model optimized for accurately detecting slender tubercle bacilli. Its core innovation is the Adaptive RoiStrip Block, which employs strip-shaped convolutional kernels to enhance spatial perception along both horizontal and vertical orientations, integrated with PANet for multi-scale feature analysis. Experiments on the Det-Y dataset demonstrate that YOLO-Y achieves superior detection accuracy while maintaining a lightweight architecture with low computational cost. The complete Det-Y dataset and corresponding code are publicly available at https://github.com/FanFan902/Det-Y.

Keywords: Mycobacterium Tuberculosis · Adaptive Spatial Attention · Tuberculosis Detection · YOLO.

* Corresponding author: Yuexiang Li (Email: yuexiang.li@sr.gxmu.edu.cn)
Zijian Xie and Yongjie Liang contribute equally to this work.

1 Introduction

Tuberculosis (TB), caused primarily by Mycobacterium tuberculosis [4], continues to impose a significant burden on global public health, with pulmonary TB representing the most common manifestation [13]. The extended treatment regimen not only diminishes patients' productivity but also places considerable economic and social strain on families and communities. Despite ongoing global control efforts, effective TB management remains hindered by substantial diagnostic limitations. Conventional diagnostic techniques—including chest CT, interferon-gamma release assays (IGRAs), molecular testing, and culture—are frequently prohibitively expensive, time-intensive, or limited in specificity [18]. These shortcomings underscore the urgent need for the development of affordable and efficient screening tools suitable for widespread implementation.

Microscopic smear examination remains a cornerstone of TB diagnosis due to its low cost and relatively high sensitivity. However, the procedure is labor-intensive and highly operator-dependent. Technicians must maintain prolonged focus to identify sparse, slender bacilli, a process requiring 10–20 minutes per smear [2]. This intensive manual effort often leads to fatigue, reduced diagnostic efficiency, and an elevated risk of false-negative results, particularly in high-burden settings where caseloads are high. The automation of this process using deep learning introduces several technical challenges: (1) the absence of publicly available annotated datasets for Mycobacterium tuberculosis detection; (2) the small size, slender morphology, morphological variability, and frequent clustering of tubercle bacilli, which complicate small-object detection and introduce substantial variation in object size; and (3) the necessity for any deployable solution to be computationally lightweight to ensure feasibility in resource-constrained primary healthcare environments.

To address these challenges, this study introduces Det-Y, a large-scale Mycobacterium tuberculosis dataset collected from two clinical centers. In addition, we propose YOLO-Y, a lightweight deep learning model optimized for detecting Mycobacterium tuberculosis. Specifically, we selected the widely used YOLOv11 [9] as the backbone architecture due to its strong overall performance, architectural flexibility, and low computational requirements, as demonstrated in preliminary evaluations on the Center-1 subset of our Det-Y dataset [2]. To improve feature extraction for the slender morphology of tubercle bacilli, we introduce a novel Adaptive RoiStrip Block and integrate it into the YOLO architecture. This module employs vertical and horizontal strip convolution kernels to construct a cross-shaped spatial attention mechanism, enabling adaptive adjustment of the receptive field and enhancing sensitivity to elongated targets without altering spatial dimensions. Experiments conducted on the Center-1 subset of Det-Y demonstrate that YOLO-Y achieves superior performance compared to other YOLO-based networks. Moreover, the model exhibits strong generalization

The dataset is named after the characteristic **Y**-shaped appearance of Mycobacterium tuberculosis in sputum smears, a morphology that makes its **Det**ection particularly challenging.

capability on the independent Center-2 subset of Det-Y [17], which comprises auramine-O-stained smear images. The main contributions of this work are:

- We publicly release the reconstructed and annotated multi-center Det-Y dataset to support and advance research in computer-aided detection of Mycobacterium tuberculosis.
- We improve upon existing YOLO-based architectures by proposing the Adaptive RoiStrip Block, a novel module purpose-built to overcome the inherent difficulties in detecting slender, size-variable objects like Mycobacterium tuberculosis. The module enhances the model's spatial awareness along critical orientations through the strategic use of adaptable strip convolutions within a directional attention mechanism. This targeted approach yields substantial improvements in both the precision and reliability of detecting these challenging small and elongated targets.
- Extensive experiments on the Det-Y dataset demonstrate the superior performance of our YOLO-Y model, which outperforms comparable models such as YOLOv12 [1], YOLOv10 [5], and YOLOv8 [8]. In addition, ablation studies further confirm the contribution of each proposed component.

2 Method

In this section, we introduce the multi-center Det-Y dataset and benchmark framework YOLO-Y in details.

2.1 Det-Y Dataset

We enlisted experienced pathologists to annotate a total of 630 and 4,814 Mycobacterium tuberculosis bacilli from the Center-1 [2] and Center-2 [17] datasets of our Det-Y collection, respectively. It is worthwhile to mention that the original public datasets had no bounding boxes for individual bacilli. Several challenges arose during the annotation process. In the Center-1 dataset [2], false-positive signals introduced by staining required repeated cross-validation to ensure the correct identification of true Mycobacterium tuberculosis bacilli. In the Center-2 dataset [17], high background noise due to inadequate washing during smear preparation complicated target distinction. To address this, we carefully delineated lightly stained bacilli and those blending into the background, resulting in a precisely annotated dataset to facilitate robust model training. In addition, when creating the Det-Y dataset, data splitting was performed by patient ID, and no cropping was conducted before the split. Overlapping patches were then generated independently within each split, which guarantees zero information leakage. Exemplars can be found via https://github.com/FanFan902/Det-Y.

2.2 YOLO-Y

Overview. The proposed benchmark framework, *i.e.,* YOLO-Y, is constructed upon the YOLOv11 backbone [9], which was chosen for its efficient end-to-end detection framework, lightweight design, and robust performance. As shown

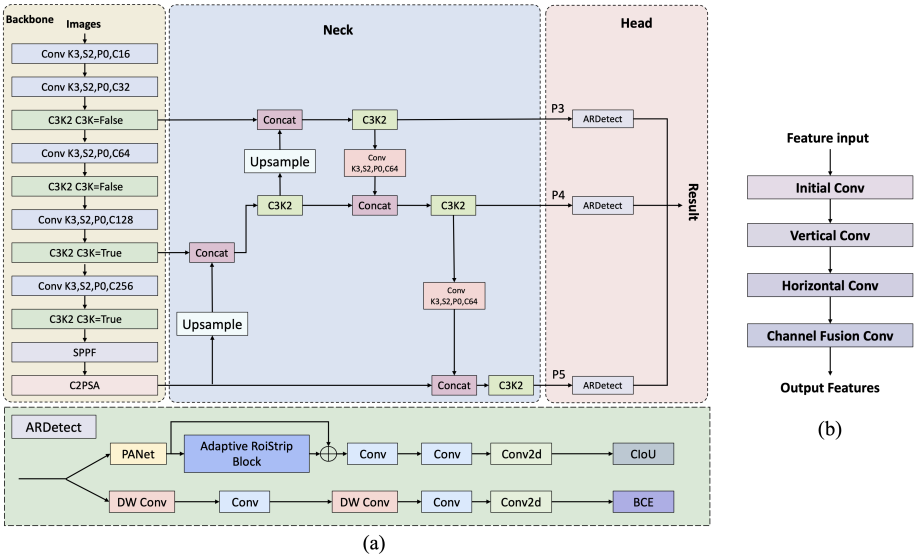

Fig. 1. Network architecture of proposed YOLO-Y (a) and Adaptive RoiStrip Block (b). Built upon the YOLOv11 [9] framework, this model integrates the Adaptive RoiStrip Block module within the detection head, forming a cross-shaped spatial attention mechanism specifically designed to capture the slender and minute morphological features of mycobacterium tuberculosis. Similar to YOLOv11 [9], our YOLO-Y consists of C3K2, SPPF, and C2PSA, which refer to Cross-Stage Partial Bottleneck, Spatial Pyramid Pooling Fast, and Cross-Stage Partial Bottleneck combined with Position-Sensitive Attention, respectively.

in Fig. 1 (a), our YOLO-Y processes input images by the backbone and neck networks, and the resulting multi-scale feature maps from the PANet [12] are directed to the ARDetect head. In the ARDetect head, we integrate a novel module, termed Adaptive RoiStrip Block, to specifically address the challenges inherent in detecting slender Mycobacterium tuberculosis organisms. Particularly, the embedded Adaptive RoiStrip Block augments spatial feature extraction along both the vertical and horizontal axes within the regression branch. This design optimizes performance for small, morphologically variable targets while maintaining the efficiency essential for clinical deployment. For network supervision, our YOLO-Y adopts the same loss functions as vanilla YOLOv11 [9], *i.e.,* complete IoU (CIoU) loss and binary cross-entropy (BCE) loss.

Adaptive RoiStrip Block. The main difference between our YOLO-Y and vanilla YOLO is the novel Adaptive RoiStrip Block, which is a lightweight, cross-shaped spatial attention module designed to capture the rod-shaped morphology of bacilli. In this section, we introduce this block in details. Its pipeline is illustrated in Fig. 1 (b). Specifically, the proposed Adaptive RoiStrip Block consists of a hierarchical cross-convolution structure ($1 \times k + k \times 1$). Assuming

the feature maps yielded by backbone and neck networks are P3, P4 and P5, respectively, *i.e.,* the feature maps fed to the three ARDetect modules as shown in Fig. 1 (a). Different kernel sizes are applied at each feature level: $1 \times 7 + 7 \times 1$ at P3 to preserve small-target details, $1 \times 5 + 5 \times 1$ at P4 to balance local and global information, and $1 \times 3 + 3 \times 1$ at P5 to prevent localization blur. This design reduces the number of parameters by over 50% compared to standard 2D convolution. While prior work employs strip kernels for directional feature enhancement [16], our module is distinct in its hierarchical, level-specific kernel adaptation to accommodate variability in clinical scenes. The mechanism of our Adaptive RoiStrip Block are detailed below. Feature maps are firstly fed to the basic feature extraction component (Initial Conv) of the block:

$$X_1 = \text{Conv2d}(P, \text{kernel\_size} = 3, \text{padding} = 1). \tag{1}$$

The resulting features are then passed to the vertical strip convolution (Vertical Conv) for vertical feature extraction:

$$X_2 = \text{Conv2d}(X_1, \text{kernel\_size} = (1, k), \text{padding} = (0, k//2)), \tag{2}$$

where k denotes the vertical receptive field size (chosen per level: P3=7, P4=5, P5=3). Such a vertical receptive field enhances the model's ability to extract features of mycobacterium tuberculosis in the vertical direction. The output is then fed to the horizontal strip convolution (Horizontal Conv):

$$X_3 = \text{Conv2d}(X_2, \text{kernel\_size} = (k, 1), \text{padding} = (k//2, 0)). \tag{3}$$

where k denotes the horizontal receptive field size (chosen per level: P3=7, P4=5, P5=3). This horizontal receptive field jointly with the vertical convolution can effectively extract features from targets of varying sizes in the image. Finally, channel fusion is performed to fuse the horizontal and vertical information via Channel Fusion Conv:

$$X_4 = \text{Conv2d}(X_3, \text{kernel\_size} = 1, \text{stride} = 1, \text{padding} = 0). \tag{4}$$

This yields the final cross-shaped spatial attention feature map $X_4 \in \mathbb{R}^{C \times H \times W}$. As shown in Fig. 1 (a), the original features are then combined with the obtained features via:

$$X_{\text{enhanced}} = X + s \cdot (X_4 - X), \tag{5}$$

where s is a learnable intensity parameter. The enhanced feature map X_{enhanced} is sent to yield the final prediction of Mycobacterium tuberculosis detection.

3 Experiments

Det-Y Dataset. To address the issue of varying image dimensions resulting from the use of different microscopes across healthcare settings, we employed the clinical data from two centers [2,17] with different image sizes to form our

Det-Y dataset. Particularly, the primary subset (Center-1 data) [2] consists of $1,000$ high-resolution (2448×2048 pixels) images from culture-confirmed TB patients. We selected 92 well-stained positive images for processing. To adapt these images to the model's input size (640×640 pixels) and prevent bacilli fragmentation, each image was divided into 25 overlapping sub-images using horizontal and vertical strides of 452 and 352 pixels, respectively. After filtering out background-only patches, 393 positive sub-images were annotated and randomly split into training, validation, and test sets (275, 78, and 40 images) in a $7:2:1$ ratio. To evaluate the model's generalization, we utilized the secondary subset (Center-2 data) [17] comprising 510 auramine O-stained images (1296×966 pixels). Applying the same segmentation and filtering procedures produced $1,577$ annotated images, which were subsequently divided into $1,103$ for training, 315 for validation, and 159 for testing.

Evaluation Metrics. We employed multiple metrics to evaluate model performance, *e.g.,* GPU memory size (GPU men.), Precision (P), Recall (R), mAP@0.5, and mAP@0.5:0.95. The definitions of evalution metrics can be found in [9].

Implementation Details. Our experiments were conducted on a workstation equipped with an NVIDIA RTX A6000 GPU, utilizing the PyTorch deep learning framework. To enhance model robustness and adapt to variations in clinical settings—such as image heterogeneity, differences in smear staining quality, variations in microscope focal length, and varying lighting intensities—we applied standard data augmentation techniques during model training. These augmentations included adjustments to HSV hue, saturation, and value; translation; scaling; horizontal flipping; and Mosaic augmentation.Training was performed with an initial learning rate of 0.6, a batch size of 16, and continued for 2000 epochs. The model with the best validation IoU was selected as the final model. The learning rate (0.6) was selected via grid search over [0.1, 0.9]. Training ran for up to 2000 epochs with cosine decay and early stopping (patience=100). Loss curves converged by 500 epochs with no overfitting gap between training and validation loss.

3.1 Comparative Results

We conducted a comprehensive evaluation of YOLO-Y against eight other YOLO variants, including YOLOv12 [1], YOLOv11 [9], YOLOv10n [5], YOLOv9s [3], YOLOv8 [8], YOLOv6 [10], YOLOv5 [7], and YOLOv3 [14]. The results are summarized in Table 1. It is worthwhile to mention that this study is primarily built upon the previous state-of-the-art YOLOv11 [9]. The performance of the more recent YOLOv12 [1] is evaluated mainly for reference. On **Center-1 data**, YOLO-Y achieved the highest recall. Its mAP@0.5 matched that of YOLOv9s [3] while using only half the computational resources, demonstrating superior efficiency. Its mAP@0.5:0.95 was also competitive, trailing YOLOv9s [3] by only 0.001. Although its precision was slightly lower, the higher recall is more critical

Table 1. Comparison results of YOLO-Y on Center-1 [2], Center-1-2 [15] and Center-2-2 data [17], where **bold** text indicates the best result and underlined text denotes the second-best result.

Dataset	Model	GPU mem.	Parameters(M)	GFLOPs	Inference latency(ms)	Precision	Recall	mAP@50	mAP@50:95	F1 Score
Center-1	YOLOV26n [6]	2.88	2.38	**5.20**	2.60	0.852	0.842	**0.919**	**0.577**	0.847
	YOLOV12 [1]	1.95	2.56	6.30	2.50	0.868	0.831	0.908	0.575	0.849
	YOLOV11n [9]	1.43	2.58	6.30	1.80	0.852	0.839	0.909	0.573	0.845
	YOLOV10n [5]	1.68	2.27	6.50	1.70	0.856	0.831	0.896	0.562	0.843
	YOLOV9t [3]	3.10	**1.97**	7.60	4.10	0.753	0.895	0.908	0.551	0.818
	YOLOV9s [3]	2.99	7.17	26.70	3.70	**0.905**	0.846	0.918	0.574	0.875
	YOLOV8 [8]	1.25	3.01	8.10	1.60	0.899	0.823	0.913	0.572	0.859
	YOLOV6 [10]	**0.98**	4.23	11.70	**1.40**	0.797	0.823	0.889	0.550	0.810
	YOLOV5 [7]	1.33	2.50	7.10	1.50	0.865	0.869	0.908	0.562	0.867
	YOLOV3 [14]	8.34	103.67	282.20	10.03	0.876	0.808	0.897	0.555	0.841
	RT-DETRV4-S [11]	6.03	10.18	24.82	30.85	0.819	**0.992**	0.818	0.516	**0.897**
	YOLO-Y(Our)	1.44	3.00	7.80	3.00	0.827	0.877	0.918	0.573	0.851
Center-1-2	YOLOV26n [6]	2.88	2.38	**5.20**	3.60	**0.702**	0.477	0.522	**0.362**	0.568
	YOLOV12 [1]	1.95	2.56	6.30	3.50	0.657	0.544	0.587	0.320	0.595
	YOLOV11n [9]	1.43	2.58	6.30	4.20	0.681	0.496	0.557	0.309	0.574
	YOLOV10n [5]	1.68	2.27	6.50	2.70	0.602	0.516	0.496	0.286	0.556
	YOLOV9t [3]	3.10	**1.97**	7.60	4.30	0.697	0.447	0.474	0.276	0.545
	YOLOV9s [3]	2.99	7.17	26.70	7.80	0.666	0.536	0.529	0.304	0.594
	YOLOV8 [8]	1.25	3.01	8.10	**2.60**	0.640	0.476	0.507	0.279	0.546
	YOLOV6 [10]	**0.98**	4.23	11.70	**2.60**	0.607	0.551	0.544	0.303	0.578
	YOLOV5 [7]	1.33	2.50	7.10	2.90	0.588	0.576	0.56	0.316	0.582
	YOLOV3 [14]	8.34	103.67	282.20	23.20	0.639	0.535	0.559	0.311	0.582
	RT-DETRV4-S [11]	6.03	10.18	24.82	68.69	0.425	**0.732**	0.425	0.243	0.538
	YOLO-Y(Our)	1.44	3.00	7.80	4.90	0.633	0.568	**0.607**	0.353	**0.599**
Center-2-2	YOLOV26n [6]	2.87	2.38	**5.20**	2.40	0.974	0.909	0.964	0.821	0.940
	YOLOV12 [1]	3.72	2.56	6.30	3.20	**0.987**	0.941	**0.981**	**0.895**	0.963
	YOLOV11n [9]	2.69	2.58	6.30	2.50	0.971	0.905	0.964	0.836	0.937
	YOLOV10n [5]	3.25	2.27	6.50	3.90	0.963	0.925	0.967	0.828	0.944
	YOLOV9t [3]	3.10	**1.97**	7.60	2.80	0.952	0.952	0.977	0.849	0.952
	YOLOV9s [3]	5.58	7.17	26.70	6.70	0.967	0.948	0.963	0.825	0.957
	YOLOV8 [8]	2.49	3.01	8.10	**1.80**	0.971	0.941	0.972	0.832	0.956
	YOLOV6 [10]	**2.04**	4.23	11.70	**1.80**	0.957	0.945	0.974	0.837	0.951
	YOLOV5 [7]	2.51	2.50	7.10	2.00	0.957	0.951	0.978	0.813	0.954
	YOLOV3 [14]	15.30	103.67	282.20	10.50	0.972	0.939	0.974	0.835	0.955
	RT-DETRV4-S [11]	6.03	10.18	24.82	31.18	0.960	**0.989**	0.960	0.849	**0.974**
	YOLO-Y(Our)	2.68	3.00	7.80	2.00	0.962	0.949	0.977	0.878	0.955

for sensitive screening. On **Center-2 data**, YOLO-Y demonstrated strong robustness. Compared to its predecessor, YOLOv11 [9], it improved precision by 0.016, with only minor reductions in recall (0.01) and mAP@0.5 (0.004), while achieving a higher mAP@0.5:0.95 (+0.004). This favorable trade-off reduces false positives, as confirmed by visual results.

To evaluate generalization under identical staining, we tested models trained on Center-1 on the external ZNSM-iDB dataset (both Ziehl-Neelsen) [15]. As shown in Table X, YOLO-Y (mAP@0.5=0.607, F1=0.599) outperforms YOLOv12 [1] (0.587, 0.595) and YOLOv11n [9] (0.557, 0.574). Note that cross-center testing between Center-1 (Ziehl-Neelsen) and Center-2 (Auramine-O) is clinically unrealistic due to different staining protocols, and is therefore not reported. Additionally, we trained RT-DETRv4-S [11] (10.2M) and YOLOv26n [6] (2.4M) on Center-1 and evaluated on ZNSM-iDB [15]. They achieve mAP@0.5 of 0.425 and 0.522 respectively, both below YOLO-Y (0.607), suggesting that task-specific architectural design matters more than raw model scale for TB detection. We repeated all main experiments five times (different random seeds) and conducted paired t-tests (normality verified via Shapiro-Wilk, all $p>0.05$). YOLO-Y achieves mAP@0.5:0.95 of 0.573 ± 0.002 vs. YOLOv11n's [9] 0.571 ± 0.003 ($p=0.03$) on Center-1, and 0.728 ± 0.003 vs. 0.724 ± 0.004 ($p=0.01$) on Center-2. While mar-

Table 2. Ablation studies on Center-1 data [2] and Center-2 data [17], where **bold** text indicates the best result and underlined text denotes the second-best result.

Dataset	Adaptive RoiStrip Block	GPU mem.	Parameters(M)	GFLOPs	Inference latency(ms)	Precision	Recall	mAP@50	mAP@50:95	F1 Score
Center-1	×	**1.43**	2.58	**6.3**	**4.2**	**0.681**	0.496	0.557	0.309	0.574
	✓	1.44	**3.00**	7.8	4.9	0.633	**0.568**	**0.607**	**0.353**	**0.599**
Center-2	×	2.69	2.58	**6.3**	2.5	**0.971**	0.905	0.964	0.836	0.937
	✓	**2.68**	**3.00**	7.8	**2.0**	0.962	**0.949**	**0.977**	**0.878**	**0.955**

gins are modest, statistically significant gains translate to fewer missed bacilli in screening.

Visualization. Due to the limitation of paper length, please refer to our GitHub repository for the visualization results. YOLO-Y's predictions align precisely with the ground truth, whereas other models exhibit notable false positives and missed detections. This visual evidence highlights the limitations of models such as YOLOv9s [3] and YOLOv11 [9] in capturing fine morphological features, despite their strong metric performance. The results confirm that the Adaptive RoiStrip Block effectively enhances directional feature extraction, enabling YOLO-Y to better leverage target morphology and distinguish bacilli from complex backgrounds, thereby directly addressing the challenge of detecting slender Mycobacterium tuberculosis.

3.2 Ablation Studies

We conducted ablation studies on Center-1 and Center-2 data of our Det-Y, respectively, to evaluate the contribution of the Adaptive RoiStrip Block, which provides cross-shaped spatial attention (see Table 2, where "✓" and "×" indicate the inclusion and exclusion of the module, respectively). The results demonstrate that the module consistently enhances performance. On Center-1 data, it improved both Recall and mAP@0.5. On Center-2 data, it achieved a better balance between Precision and Recall, significantly increasing Precision and mAP@0.5:0.95 with only a minimal decrease in Recall and mAP@0.5, indicating stronger performance under more stringent criteria. These findings validate our design, *i.e.,* the module enhances feature extraction along both horizontal and vertical axes, thereby improving the detection of slender targets. Ablation studies further confirm that the Adaptive RoiStrip Block plays a key role in achieving superior detection performance.

4 Conclusion

In this work, we presented two key contributions to the field of automated TB detection. First, we introduced Det-Y, a large-scale, multi-center dataset of annotated Mycobacterium tuberculosis bacilli, designed to fill a critical gap and enable the development of data-driven models. Second, we proposed YOLO-Y,

a lightweight detection architecture optimized for this task. Its core innovation—the Adaptive RoiStrip Block—leverages strip convolutions to enhance axial feature perception within a PANet [12], leading to robust multi-scale analysis. Evaluated on Det-Y, YOLO-Y outperforms existing YOLO-based frameworks in detection accuracy while preserving a computationally efficient, lightweight design. This work establishes a strong benchmark and a practical tool for future research in microscopic TB detection. For clinical deployment, patch-level detections can be aggregated to patient-level decisions via a simple rule: a patient is classified as smear-positive if ≥ 1 bacillus is detected across all patches.

Acknowledgments. This study was funded by the Guangxi Natural Science Foundation (Grant No. 2025GXNSFAA069680), the "Yongjiang Plan" Leading Talent Team Program of Nanning City (Grant No. 2024020), the Basic Ability Enhancement Program for Young and Middle-aged Teachers of Guangxi (Grant No. 2025KY0157), the Innovation Project of Guangxi Graduate Education (Grant No. YCSW2026388), the Innovation Project of GUET Graduate Education (Grant No. C25YJH01JZ01, GDZZ03250014), the Innovation Project of Guangxi Academy of Artificial Intelligence Graduate Education (Grant No. SA2600000501), and the Innovation Project of GXMU Biomedical Engineering Graduate Education (Grant No. SG202608, SG202609).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Akre, P.D., Ghavghave, Y.G., Pacharaney, U.: A YOLOv12-based approach for automatic detection of cephalometric landmarks on 2D lateral skull X-ray images. Scientific Reports (2026), https://doi.org/10.1038/s41598-026-43250-z
2. Chen, W.C., Chang, C.C., Lin, Y.E.: Pulmonary tuberculosis diagnosis using an intelligent microscopy scanner and image recognition model for improved acid-fast bacilli detection in smears. Microorganisms **12**(8), 1734 (2024), https://doi.org/10.3390/microorganisms12081734
3. Chng, S.Y., Li, J.W., Trinh, H.H., Lim, J.L.X., Kan, M.R.X., Vadapalli, D.S., Toh, P.Z.J., Ang, T.L.: Polyp detection using YOLOv9 on real and synthetic colonoscopy images. Journal of Gastroenterology and Hepatology (2026), https://doi.org/10.1111/jgh.70391
4. Ghosh, M., Lahiri, M., Dalal, A., Parida, K.K., Kalia, N.P.: Advancements in tuberculosis diagnostics: An update. Microbial Pathogenesis p. 107843 (2025), https://doi.org/10.1016/j.micpath.2025.107843
5. Guo, Y., Huang, X., Chen, H., Chen, W., Wu, Y., Jiang, M., Wu, J.: Bi-directional YOLOv10 with average convolution for brain tumor detection in MRI. Brain Research Bulletin **237**, 111780 (2026), https://doi.org/10.1016/j.brainresbull.2026.111780
6. Huaman-Pacco, I.P., Sacoto-Cabrera, E.J., Silva-Alvarado, V.L., Ahmad, A., Sendra, S., Lloret, J., Moreno-Cardenas, E.: Proposal for computationally efficient fog computing system for coffee berry borer detection via optimized YOLOv26. Sensors **26**(7), 2212 (2026), https://doi.org/10.3390/s26072212
7. Jocher, G.: Ultralytics YOLOv5 (2020), https://github.com/ultralytics/yolov5

8. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics YOLOv8 (2023), https://github.com/ultralytics/ultralytics
9. Jocher, G., Qiu, J.: Ultralytics YOLO11 (2024), https://github.com/ultralytics/ultralytics
10. Li, C., Li, L., Geng, Y., Jiang, H., Cheng, M., Zhang, B., Ke, Z., Xu, X., Chu, X.: YOLOv6 v3.0: A full-scale reloading (2023), https://doi.org/10.48550/arXiv.2301.05586
11. Liao, Z., Zhao, Y., Shan, X., Yan, Y., Liu, C., Lu, L., Ji, X., Chen, J.: RT-DETRv4: Painlessly furthering real-time object detection with vision foundation models. arXiv preprint arXiv:2510.25257 (2025), https://doi.org/10.48550/arXiv.2510.25257
12. Liu, S., Qi, L., Qin, H., Shi, J., Jia, J.: Path aggregation network for instance segmentation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8759–8768 (2018), https://doi.org/10.1109/CVPR.2018.00913
13. Long, B., Liang, S.Y., Koyfman, A., Gottlieb, M.: Tuberculosis: a focused review for the emergency medicine clinician. The American Journal of Emergency Medicine **38**(5), 1014–1022 (2020), https://doi.org/10.1016/j.ajem.2019.12.040
14. Redmon, J., Farhadi, A.: YOLOv3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018), https://doi.org/10.48550/arXiv.1804.02767
15. Shah, M.I., Mishra, S., Yadav, V.K., Chauhan, A., Sarkar, M., Sharma, S.K., Rout, C.: Ziehl–neelsen sputum smear microscopy image database: a resource to facilitate automated bacilli detection for tuberculosis diagnosis. Journal of Medical Imaging (2017), https://doi.org/10.1117/1.JMI.4.2.027503
16. Yuan, X., Zheng, Z., Li, Y., Liu, X., Liu, L., Li, X., Hou, Q., Cheng, M.: Strip R-CNN: Large strip convolution for remote sensing object detection. arXiv preprint arXiv:2501.03775 (2025), https://doi.org/10.48550/arXiv.2501.03775
17. Zachariou, M.: Towards fully automated analysis of sputum smear microscopy images. Ph.D. thesis, The University of St Andrews (2024), https://doi.org/10.17630/sta/858
18. Zheng, R., Xu, F., Huang, X., Wang, J., Feng, Y., Huang, J., Qin, L.: Evaluation of aptamer fluorescence microscopy in the diagnosis of pulmonary tuberculosis. Microbiology Spectrum **10**(4), e02602–21 (2022), https://doi.org/10.1128/spectrum.02602-21