IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 21 June 2024, accepted 15 August 2024, date of publication 20 August 2024, date of current version 30 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3446613

RESEARCH ARTICLE

Enhancing Gastrointestinal Stromal Tumor (GIST) Diagnosis: An Improved YOLOv8 Deep Learning Approach for Precise Mitotic Detection

HAOXIN LIANG[1,2], ZHICHUN LI[3], WEIJIE LIN[2], YUHENG XIE[2], SHUO ZHANG[4,5,6], ZHOU LI[1], HONGYU LUO[7], TIAN LI[3], AND SHUAI HAN[1]

[1]Department of General Surgery, Zhujiang Hospital, Southern Medical University, Guangzhou 510091, China
[2]The Second Clinical College, Southern Medical University, Guangzhou, Guangdong 510095, China
[3]Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong, SAR, China
[4]School of Instrument Science and Engineering, Southeast University, Nanjing, Jiangsu 210096, China
[5]State Key Laboratory of Digital Medical Engineering, Southeast University, Nanjing, Jiangsu 210096, China
[6]School of Biological Science and Medical Engineering, Southeast University, Nanjing, Jiangsu 210096, China
[7]Department of General Surgery, The Sixth People's Hospital of Huizhou City, Huizhou 516100, China

Corresponding authors: Hongyu Luo (hongyu_luo@163.com), Tian Li (litian.li@polyu.edu.hk), and Shuai Han (gzhanbo0624@smu.edu.cn)

This work was supported in part by Guangdong Science and Technology Programme under Grant 2023A0505010038 and in part by Southern Medical University 2024 College Student Innovation Training Plan Project under Grant 202412121238.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Medical Ethics Committee of Zhujiang Hospital, Southern Medical University, under Approval No. 2024SL0042.

ABSTRACT Gastrointestinal stromal tumor (GIST) is the most common malignant tumor originating from interstitial cells in the gastrointestinal tract. Different grades require various surgical interventions and adjuvant treatments, which are closely linked to the patient's prognosis. The current clinical risk stratification method relies heavily on the identification and counting of mitotic figures, which serve as important criteria. However, manual evaluation of pathological slides in clinical practice is often limited by the shortage of experienced clinicians and the subjectivity in interpreting results. Therefore, in this paper, we propose an enhanced YOLOv8 network framework for the automatic detection of mitotic cells in GIST. We substituted the backbone network with VanillaNet, known for its simplified model complexity in feature extraction. This change facilitated the identification of specific targets and improved network performance. Additionally, we introduced the Advanced Feature Pyramid Network (AFPN) to further enhance the model's accuracy. Experimental results show that the proposed model achieved an accuracy of 0.816, a recall rate of 0.858, and an F1-score of 0.837 on the test dataset. It demonstrates superior efficacy in identifying mitotic cells, outperforming the original YOLOv8 model in overall performance. This augmented model has the potential to significantly reduce reading time while ensuring consistent diagnostic results, thereby greatly improving diagnostic efficiency. Future large-scale validation is necessary for the clinical adoption of this model.

INDEX TERMS Artificial intelligence, gastrointestinal stromal tumors, mitotic detection, mitotic figures, YOLOv8.

I. INTRODUCTION

Gastrointestinal stromal tumor (GIST) is a rare tumor originating from the digestive system and accounts for the majority of mesenchymal tumors in the digestive tract, with an estimated global incidence of 1.1-1.5/100,000 [1], of which approximately 60% occur in the stomach, while 20-30% are found in the jejunoileum [2]. Various grades of GIST require different surgical treatments and adjuvant therapeutic approaches. Typically, low-risk GISTs are managed through complete surgical resection, while intermediate-risk GISTs undergo surgical intervention with the consideration of targeted therapeutic agents. High-risk GISTs usually require both surgical resection and targeted therapies, primarily utilizing agents such as imatinib, which have demonstrated efficacy in enhancing patients' overall survival (OS)

The associate editor coordinating the review of this manuscript and approving it for publication was Massimo Cafaro.

and prognosis [3]. Consequently, the precise pathological diagnosis of GIST assumes paramount significance.

In pathological diagnosis, the 2008 version of the enhanced National Institutes of Health (NIH) risk grading scheme (Table 1) is commonly used to evaluate the risk of recurrence in primary GIST cases [4]. This scheme primarily considers factors such as primary tumor site, volume and number of mitotic figures, which reflect different stages of mitotic activity and are closely related to tumor malignancy and prognosis. These parameters not only assess the patient's risk of recurrence and metastasis but also guide physicians in selecting candidates who may benefit from postoperative adjuvant therapy, ultimately improving prognosis and facilitating personalized treatment [5]. Currently, the identification and counting of mitotic figures are predominantly carried out by pathologists in clinical settings. This process is labor-intensive and requires significant human resources and time investment [6]. Furthermore, manual interpretation introduces subjectivity, which can result in findings of questionable reliability and reproducibility. There remains controversy regarding whether treatment plans guided by the risk classification derived from pathological diagnosis offer the utmost benefit to patients. Therefore, there is a compelling need to explore new methods for more efficient and objective pathological diagnosis of GIST images.

In recent years, deep learning has found extensive applications in the field of medical oncology, particularly for tasks such as segmenting and classification of diverse organs and tumors, as well as the classification of alterations in tumor size or texture. Furthermore, it has been employed for predicting patterns indicative of high-risk and low-risk scenarios for potential cancer images development [7], [8]. Target detection emerges as a pivotal task within the expansive realm of computer vision, serving to identify and locate all instances of an object from specified object classes when present in an image [9]. This is particularly relevant in the context of mitotic detection in medical, especially in mitotic figures, where precise identification is crucial [10]. Pantanowitz et al. [11] used 40 times magnification to train and validate deep learning algorithms by labeling mitotic figures in Whole Slide Images (WSIs) using hematoxylin and eosin (H&E) stained slides from 320 cases of invasive breast cancer. The results showed that with algorithm support, 21 readers (87.5%) identified an increased number of mitotic figures, 13 reviewers (54.2%) decreased the incidence of mislabeled mitosis, and it is worth noting that due to algorithm support, a 27.8% reduction in overall time was observed.

Currently, in the field of object detection, several frameworks have achieved unparalleled performance on various benchmark datasets [12], including Faster Region-based Convolutional Neural Network (Faster R-CNN), You Only Look Once (YOLO), and Real-Time Detection Transformer (RT-DETR). Faster R-CNN uses a two-step detection strategy, first generating candidate boxes and then classifying and regressing them, which has high detection accuracy but is slow and requires several independent steps [13]. The YOLO series, which includes YOLOv5 and YOLOv8, uses one-step detection, which treats object detection as a regression problem, mapping directly from the pixel level to bounding boxes and category probabilities [14]. This approach achieves fast detection speed and real-time performance but may compromise detection accuracy. RT-DETR introduces a transformer mechanism and achieves real-time object detection through an end-to-end attention mechanism, with good accuracy and speed [15]. However, it may face challenges when dealing with small targets and complex scenes.

Therefore, this study comprehensively compared the accuracy, recall, and F1 score of Faster R-CNN, YOLOv5, YOLOv8, and RT-DETR, and selected YOLOv8 as the basic framework, with the introduction of VanillaNet and Advanced Feature Pyramid Network (AFPN). The former simplifies the model complexity in feature extraction, while the latter constructs a feature pyramid for multi-scale feature extraction, further improving the accuracy of the model. In addition, we meticulously identified and annotated GIST mitotic figures in digital H&E stained pathological slides, thus creating a dataset of mitotic figures for model training. The purpose of this study is to enhance the detection performance of mitotic figures in GIST pathological slides by developing an improved object detection model to evaluate the risk of GIST. In addition, it will provide valuable insights for patients, help develop personalized treatment plans, and introduce new perspectives for clinical diagnosis and treatment.

In this study, several innovations were undertaken: (1) Manual annotation and review by pathologists were used to compile a dataset of mitotic figures of GIST. (2) A new deep learning framework was proposed for the rapid and accurate detection of GIST mitotic figures. This framework includes enhancements to the backbone network replacement and a multi-scale feature extraction and fusion strategy for

TABLE 1. NIH risk grading scheme(2008 version).

	Tumor Size (cm)	Mitotic Count/50HPF	Primary Tumor Site
Very Low Risk	≤2	≤5	Any sites
Low Risk	2.1~5.0	≤5	Any sites
Intermediate Risk	2.1~5.0	6~10	Stomach
	<2	6~10	Any sites
	5.1~10.0	≤5	Stomach
High Risk	Any	Any	Rupture
	>10	Any	Any sites
	Any	>10	Any sites
	>5	>5	Any sites
	2.1~5.0	>5	Not Stomach
	5.1~10.0	≤5	Not Stomach

YOLOv8. (3) Using the proposed model, which was trained and fine-tuned on the created dataset, it demonstrated the highest overall performance compared to other commonly used deep learning object detection models, thus facilitating risk classification.

II. RELATED WORK

Artificial intelligence methods in the medical field are used to shorten the time process of clinical decision-making and improve the accuracy of diagnosis. Frameworks such as Faster Region-based Convolutional Neural Network (Faster R-CNN), You Only Look Once (YOLO), and Real-Time Detection Transformer (RT-DETR) have achieved unparalleled performance on various benchmark datasets. This success is attributed to the adoption of an end-to-end training approach within these frameworks [12]. This approach enables models to directly learn features and target detection tasks from raw data, thereby enhancing efficiency through process simplification. These approaches introduce innovative techniques aimed at bolstering model performance. Specifically, Faster R-CNN integrates the Region Proposal Network (RPN) for automatic extraction of candidate regions. The RPN shares identical feature maps with the detection network, leading to a substantial improvement in processing speed [16]. YOLO introduces the YOLOX-DA data enhancement technique and an expedited model initialization method, thereby endowing the model with enhanced generalization capabilities and expediting convergence [17]. Moreover, RT-DETR [18] combines deformable convolution and a transformer structure. It utilizes the self-attention mechanism to enable the model to globally model the image, transcending reliance solely on local information. This architectural choice contributes to a deeper understanding of contextual relationships within the image, ultimately leading to an improvement in target detection accuracy.

With the use of digital images in pathology, deep learning networks are beginning to provide solutions for object recognition in medical images. During training, pathologists typically label mitotic cells (enclosed in bounding boxes) in pathological slides as labels. Previous studies on mitosis detection have mainly focused on pathological slides of breast cancer and neuroendocrine tumor tissues. Li et al. [19] first proposed the use of deep detection networks to solve the mitosis detection problem. They designed a multi-stage deep learning framework in the ICPR 2014 MITOSIS dataset that only provided mitosis centroid positions. They used a segmentation network to estimate the annotations of bounding boxes and simultaneously applied a verification network to eliminate some false positives to improve the overall detection performance of the model. Alom et al. [20] proposed an end-to-end multi-task learning model called MitosisNet for breast cancer, which consists of segmentation, detection and classification models for confirming mitotic regions to improve the overall detection performance during testing. Wang et al. [21] developed a Fourier-based algorithm (FMDet) that converts the mitosis detection task into a semantic segmentation task based on the MIDOG2021 dataset and uses the attention mechanism to generate channel-level multi-scale features, achieving accurate mitosis detection in multi-center breast tissue pathology images. In another study, Topuz et al. [22] used the YOLO models (YOLOv3, YOLOv5, YOLOv7 and YOLOv8) on the MIDOG 2022 dataset, which contains 5 different cancer types to perform the mitotic cell detection task on images and found that the YOLOv8 architecture provided the more robust results, achieving a recall value of 89.1%. Yücel et al. [23] used the YOLOv5 framework combined with the transformer module to achieve a precision of 0.89 and a recall of 0.68 in the task of detecting mitosis in neuroendocrine tumor tissue. Since the mitotic figure morphology of different tumor types is similar [24], the above discussion shows that similar deep learning algorithms can be applied to GIST.

In this study, we selected YOLOv8 as the foundational framework. As shown in Figure 1, the network architecture of YOLOv8 is comprised of three components: backbone, neck, and head. The backbone network used in YOLOv8 is CSP-Darknet [25], which consists of CBS, C2f and SPPF modules. Inspired from the VGG [26] architecture, it incorporates a number of 3×3 convolutions, effectively improving the receptive field of the network through each pooling operation. Other backbone networks that are commonly used in the field of target detection include ResNet [27], EfficientNet [28], SwinTransformer [29], and others. The neck of YOLOv8 uses the PA-FPN strategy approach, where FPN is a top-down passing down of strong semantic features from higher levels to augment the whole feature pyramid, but this method only enhances the semantic information and does not pass on the localization information, whereas Path Aggregation Network (PAN) adds a bottom-up pyramid to the back of the FPN, which complements the FPN with the strong localization features from the lower levels and further enhances the expression of the multiscale features. The head of YOLOv8 has been changed from a coupled head to a decoupled head, which consists of two CBS convolution modules and one Conv2d for object recognition and classification respectively, and finally the Bbox loss and classification loss are calculated respectively.

III. PROPOSED METHOD

As shown in Figure 2, the input image data will be processed through three parts of the improved framework we propose. Initially, the backbone network VanillaNet-6 conducts feature extraction on the images. Concurrently, the AFPN module adaptively fuses features extracted at different stages to fully leverage contextual information. Subsequently, we utilize the decoupled head from the original YOLOv8 framework to optimize predictions for object categories and bounding boxes.

A. UTILIZATION OF THE VANILLANET MODULAR

Current deep learning networks typically rely on a large number of complex layers to extract high-level features

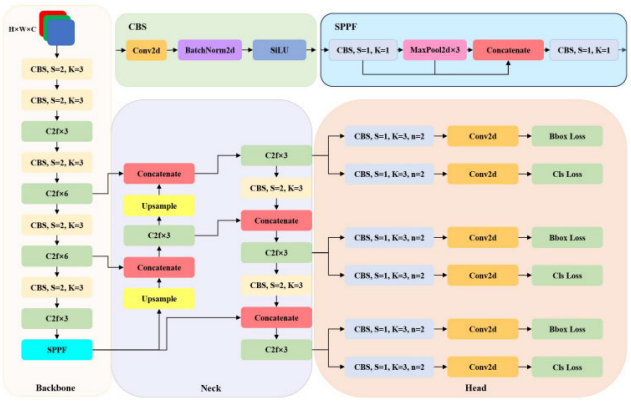

FIGURE 1. YOLOv8 network structure.

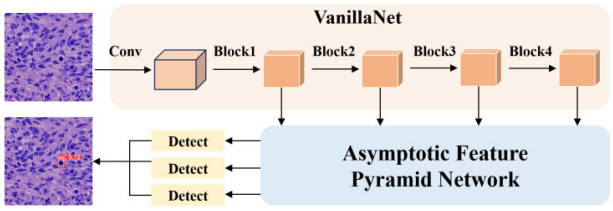

FIGURE 2. Proposed framework for GIST mitosis detection.

in order to achieve good performance. Convolutional neural networks, based on shortcuts, remain the mainstream approach for neural network architecture design. However, VanillaNet demonstrates that a simple structure, similar to LeNet and AlexNet, can achieve even higher accuracy and faster performance than traditional deep learning networks. Therefore, we replaced the backbone network of YOLOv8 with VanillaNet-6 to more effectively extract image features and reduce computational complexity in this study.

As shown in Table 2, VanillaNet-6 incorporates a total of six layers. To begin, the stem utilizes a $4 \times 4 \times 3 \times C$ convolutional layer with a stride of 4, which is consistent with the commonly used configurations in ResNet. This layer serves the purpose of transforming the input images with 3 channels into feature maps with C channels. In stages 1, 2, and 3, a max-pooling layer with a stride of 2 is used to reduce the size of the feature maps and increase the number of channels by 2 simultaneously. In stage 4, the number of channels remains unchanged, as it follows an average pooling layer. The final layer consists of a fully connected layer that produces the classification results. The kernel size for each convolutional layer is set to 1×1, as this choice aims to minimize computational costs while preserving the information contained within the feature maps. An activation function is applied after each 1×1 convolutional layer. To facilitate the training process of the network, batch normalization is incorporated after each layer. For VanillaNet with varying numbers of layers, additional blocks are introduced at each stage.

TABLE 2. The architecture of the VanillaNet-6.

	Input	Operator
Stem	224×224	4×4, 512, stride 4
Stage1	56×56	[1×1, 1024]×1 MaxPool 2×2
Stage2	28×28	[1×1, 2048]×1 MaxPool 2×2
Stage3	14×14	[1×1, 4096] ×1 MaxPool 2×2
Stage4	7×7	[1×1, 4096]×1

B. ASYMPTOTIC FEATURE PYRAMID NETWORK APPLICATION

A common strategy for multiscale feature extraction involves utilizing classical top-down and bottom-up feature pyramid networks. However, these methods suffer from the loss or degradation of feature information, which weakens the fusion effect of non-neighboring levels. The AFPN is initiated by fusing two neighboring low-level features and progressively incorporates high-level features into the fusion process [30], which helps to reduce the semantic gap between features of different levels and improves the feature fusion effect [31].

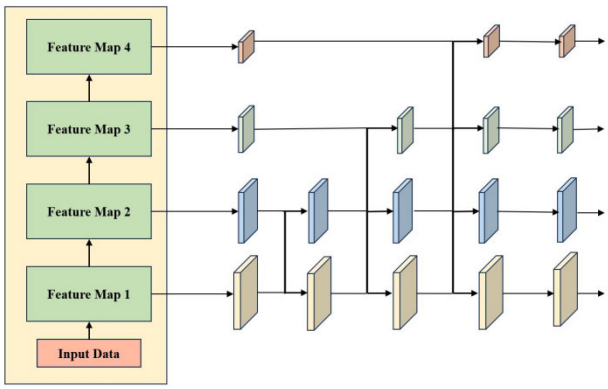

FIGURE 3. The feature fusion process of the AFPN.

As shown in Figure 3, after extracting features from the backbone, the AFPN can obtain various layers of features for fusion. Specifically, (1) Bottom-Level Feature Fusion: AFPN introduces the gradual fusion of bottom-level features, which firstly fuses bottom-level features, then deep-level features, and finally integrates top-level features. (2) Adaptive Spatial Fusion: Adaptive Spatial Fusion Mechanism (ASFF) is introduced, which introduces the changing spatial weights during multi-level feature fusion to strengthen the importance of the key level, and at the same time, suppresses the influence of the contradictory information from different objects. For features from different levels, corresponding up-sampling or down-sampling operations are usually performed on the feature layers to ensure that they have the same spatial dimensions. Next, the weights of each feature layer are computed using a convolution operation, normalized by SoftMax, and finally,

the feature layers are weighted and fused according to the weights. (3) Alignment of Underlying Features: AFPN incorporates the concept of asymptotic fusion, which facilitates the gradual convergence of features at various levels during the fusion process, thereby minimizing the semantic gap between them. Through the implementation of this module, the network's feature fusion effect is enhanced, allowing the model to better understand and utilize information across different levels.

C. LOSS FUNCTION

For the loss function selection in object recognition, we utilize the default CIOU and DFL loss functions in the YOLOv8 model, allowing the network to prioritize the target location and the distribution of the neighborhood more efficiently.

$$LossCIoU = 1 - IoU + \frac{\rho^2\left(b, b^{gt}\right)}{c^2} + \alpha v \quad (1)$$

$$v = \frac{4}{\pi^2}(arctan\frac{w^{gt}}{h^{gt}} - arctan\frac{w}{h})^2 \quad (2)$$

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (3)$$

$$DFL(S_i, S_{i+1}) = -((y_{i+1} - y)\, log\,(S_i) + (y - y_i)\, log\,(S_{i+1})) \quad (4)$$

where IoU denotes the intersection over union of the true and forecasted bounding boxes; b, w, h and b^{gt}, w^{gt}, h^{gt} represent the center points, width and height of the forecasted and true boxes, respectively, and ρ indicates the Euclidean distance between the two center points; c stands for the diagonal distance of the smallest enclosing area capable of containing both the forecasted and true boxes. S_I and S_{i+1} represent the predicted probabilities of categories i and i+1, while y and y+1 are their corresponding true label values.

Since the task is the detection of mitotic cells, we use binary cross entropy as the classification loss function:

$$BCE = \frac{1}{N}\sum_i L_i = \frac{1}{N}\sum_i -[y_i \cdot log\,(p_i) + (1 - y_i) \cdot log\,(1 - p_i)] \quad (5)$$

where N is the total number of samples, y_I represents the true label of sample i; p_I represents the probability of the model predicting sample i.

IV. EXPERIMENTS

A. DATASET ACQUISITION, PREPROCESSING, AND LABELING

This study presents a comprehensive dataset from Zhujiang Hospital of Southern Medical University, which was collected under the supervision of experienced medical professionals. The study sample comprised 105 patients who met the inclusion and exclusion criteria at Pearl River Hospital of Southern Medical University between January 2019 and August 2023.

Adhering to the standardized protocol, the inclusion criteria included: (1) patients undergoing surgery with a pathological diagnosis of GIST; (2) immunohistochemistry indicating positive staining of CD117 and Dog-1 in pathological slides; (3) availability of comprehensive clinical and pathological data, requiring complete information such as age, gender, tumor volume, number of mitotic figures, primary tumor site, and risk classification to ensure dataset accuracy.

Exclusion criteria were defined as follows: (1) patients with an insufficient amount of tumor tissue on pathology slides that hindered proper marking for analysis; (2) patients with discolored or contaminated pathology slides unsuitable for accurate assessment; (3) patients with torn or defective pathology slides that could not be fully scanned; (4) patients with other systemic tumor diseases or conditions of primary etiology.

Importantly, all pathological assessments were conducted on H&E stained tissue slides, utilizing standardized histological staining techniques to facilitate precise analysis. The study underwent thorough review and approval by the Medical Ethics Committee of Zhujiang Hospital, Southern Medical University (approval number: 2024SL0042). Informed consent was waived due to the non-invasive and retrospective nature of our study.

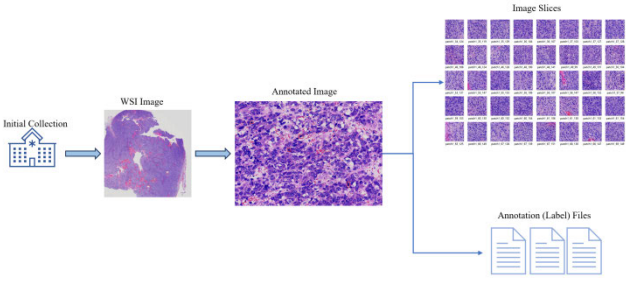

FIGURE 4. Dataset construction for GIST mitotic detection.

Figure 4 illustrates the process of data preparation. After acquiring pathological slides of GIST, the pathologist initiated the process by scanning the slides into WSIs using the Intemedic (Neo-5x, China), a pathology slide image scanning system. Subsequently, for the purpose of annotation, pathologists manually annotated the images using QuPath (v0.4.2, England). To ensure the accuracy of the annotation results, pathologists also independently reviewed the images and their corresponding annotations. It is noteworthy that any discrepancies between the pathologists regarding the aforementioned results led to a collaborative decision to discard the annotation. Following this, we performed a slicing operation on pathological slides, obtaining 512 × 512 JPG images. From these, we selected images containing mitotic cells, computed, and transformed the annotation coordinates of the mitotic cells within. Ultimately, we selected 914 images and corresponding label files for each image. Each file contains the positions of each annotation within the image, represented by the vector [x, y, height, width]. Here, the coordinate (x, y) represents the normalized values of the center point of the

bounding box. The height represents the normalized height of the bounding box, and the width represents its normalized width (in pixels).

As depicted in Figure 5, there are some positive signals regarding the mitosis detection task that can be observed in the label distribution of this dataset. Firstly, the positions of the targets to be detected in the image are diverse, which facilitates the training of a robust detection model. Secondly, the bounding boxes of most mitotic events have similar sizes and shapes, making it easier for the model to learn these dimensional features. Therefore, this dataset can provide rich and balanced information for the task of mitosis detection. From the dataset, we divided it into training, validation, and testing sets in an 8:1:1 ratio, respectively comprising 731, 91, and 92 images. The testing set was completely isolated from the training and validation sets and was utilized solely for the final evaluation of the model's performance. The validation set was employed as a tool for model tuning, facilitating the identification and optimization of the model's hyperparameters and weights. This approach allows for rigorous assessment and fine-tuning of the model prior to its final evaluation. For data augmentation, we implemented the online data augmentation technique used in the YOLOv8 network to make the number of images in training set reach 3838, which could improve the detection of smaller targets and enhance the model's generalization ability. These techniques include random cropping, which introduces variability in object sizes and positions within the images. We also utilized the Mosaic augmentation method, which combines four different images into one composite image. Notably, we deactivated the Mosaic augmentation in the last 10 epochs of training to stabilize learning as we fine-tuned the model. Additionally, we incorporated image flipping, which horizontally mirrors the images.

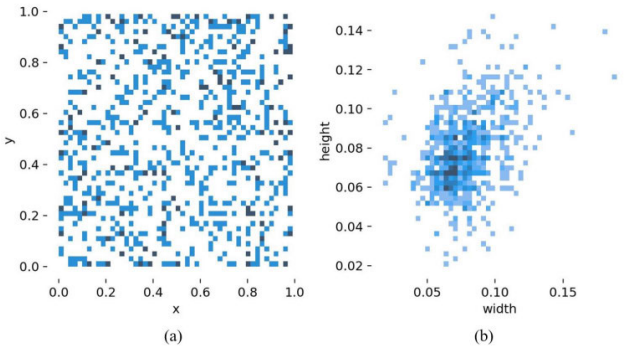

FIGURE 5. Distribution of annotated values for the mitosis dataset. (a) Indicates the position distribution of the centers of the detection boxes. The center points of the detection boxes are relatively evenly distributed in the image, and there is no obvious central tendency. (b) shows the height and width distribution of detection boxes in the data. We can find that most of the label box widths are concentrated in a small range, and there is an obvious dense area of height and width, which may indicate that many mitotic events have similar sizes and shapes.

B. EXPERIMENTAL CONFIGURATION

This paper's experiments were implemented with PyTorch version 2.1.2. These experiments are conducted on the Intel(R) Core(TM) i9-13900 with NVIDIA GeForce RTX 4090 as its hardware configuration. The learning rate is 0.01. We used the Stochastic Gradient Descent (SGD) optimizer, set the batch size to 32, and conducted 200 epochs for training.

C. EVALUATION METRICS OF DETECTION MODEL

There are four possible outcomes in the test set when objectives are identified based on the relationship between actual and expected values. To be more precise, we define True Positive (TP) as a target that has a mitotic event and is successfully identified by the improved YOLOv8 model. On the other hand, False Positive (FP) refers to situations in which the model mistakenly detects a target as having a mitotic event even while there isn't one. Furthermore, False Negative (FN) describes circumstances where a mitotic event occurs but the model incorrectly classifies it as negative. Table 3 presents these delineations directly.

TABLE 3. Four possible classification results.

		Actual Label	
		Positive	Negative
Predicted Label	Positive	TP	FP
	Negative	FN	TN

Three critical criteria were used in our tests to assess the accuracy of mitotic image detection: recall, precision, and F1 score. As stated technically by equation (6), recall measures the model's capacity to identify all relevant instances or true positives. A higher recall is suggestive of the model's competence in identifying positive examples and reducing false negatives. Equation (7) expresses accuracy as the ratio of true positive predictions to all positive predictions, where a greater precision denotes a lower number of false positives. The F1 score is a useful metric when dealing with scenarios with uneven class distributions or when both false positives and false negatives are significant. It is calculated as the harmonic mean of accuracy and recall, and it finds a balance between the two metrics. The F1 score ranges from 0 to 1, with elevated values denoting superior model performance [9], as precisely delineated in equation (8).

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$F1Score = \frac{2 * Recall * Precision}{Recall + Precision} \tag{8}$$

V. RESULTS

A. BASIC INFORMATION OF RESEARCH POPULATION

This study included 105 eligible patients with GIST, based on the exclusion criteria. Among these patients, 52 (49.5%) were male and 53 (50.5%) were female, with a median (IQR) age of 60 (46.3-71.6) years. The tumor volume ranged from 0.2 cm to 16 cm. Among the cases, 43 (40.9%) had a tumor size of $\leqslant$ 2.0 cm, 29 (27.6%) had a tumor size of 2.1-5.0 cm, 30 (28.6%) had a tumor size of 5.1-10.0 cm, and 3 (2.9%) had a tumor size of $>$10.0 cm. The count of mitotic figures revealed that 85 cases (81%) had $\leqslant$ 5/50 HPF, 16 cases (15.2%) had 6-10/50 HPF, and 4 cases (3.8%) had $>$10/50 HPF. The primary tumor sites were distributed as follows: 84 cases (80%) in the stomach, 3 cases (2.9%) in the duodenum, 8 cases (7.6%) in the jejunum and ileum, 2 cases (1.9%) in the colon, and 8 cases (7.6%) in other areas. The postoperative pathological results revealed that 39 cases (37.1%) had an extremely low risk, 21 cases (20%) had a low risk, 20 cases (19.1%) had a medium risk, and 25 cases (23.8%) had a high risk. Table 4 displays the clinical and pathological characteristics of all patients with GIST in this study.

TABLE 4. Brief summary of clinical characteristics of patients with GIST.

Patient characteristics	Training set	Validation set	Testing set	Total
Number of patients	84	10	11	105
Age range (years)				
21-30	4	0	0	4
31-40	3	1	1	5
41 - 50	13	2	0	15
51 - 60	24	3	5	32
61 - 70	29	1	3	33
71 - 80	7	2	2	11
81 - 90	4	1	0	5
Gender				
Male	39	3	10	52
Female	45	7	1	53
Tumor Size (cm)				
$\leqslant$2.0	36	3	4	43
2.1~5.0	26	1	2	29
5.1~10.0	19	6	5	30
>10.0	3	0	0	3
Mitotic Count / 50HPF				
$\leqslant$5	68	9	8	85
6~10	14	1	1	16
>10	2	0	2	4
Primary Tumor Site				
Gastric	65	9	10	84
Duodenum	2	0	1	3
Jejunum/ileum	7	1	0	8
Rectum	2	0	0	2
Others*	8	0	0	8
Rupture	0	0	0	0
Risk Classification				
Very Low Risk	32	3	4	39
Low Risk	16	2	3	21
Intermediate Risk	17	3	0	20
High Risk	19	2	4	25

Others*: Abdominal cavity, pelvic cavity, small intestine, and other sites.

B. PERFORMANCE EVALUATION OF DETECTION MODEL

In the application case of mitotic figure detection, the two measures we chose indicate the accuracy of the model and its ability to prevent missed detections. A high recall rate indicates that the model can minimise the problem of missing cells moving through mitosis when identifying mitotic figures, minimising the loss of important data and providing support for further analysis and categorisation. High recall is therefore essential to ensure that the majority of positive classes (mitotic cells) are accurately identified. Higher accuracy reduces the chance of the model incorrectly classifying cells that are not undergoing mitosis as mitotic cells, resulting in fewer false positives. Given these two factors, the F1 score becomes a good measure as it is the harmonic mean of precision and recall, balancing the performance of both and representing the overall performance of the model.

As shown in Figure 6, we implemented an early stopping strategy during our validation experiments. Training was halted if there was no significant decrease in loss, indicating no notable improvement in model performance, within a predetermined number of epochs. We observed that while the loss on the training set continued to decrease, the loss on the validation set had already converged. Continuing training beyond this point would likely lead to overfitting. The final loss values on the validation sets signify that the model had reached its optimal performance without overfitting. Figure 7 illustrates the performance of our model on the validation set, demonstrating good reliability. We also observed that the model detected some false positives, which bear significant similarities to real mitotic cells in color and shape. This suggests that during the learning process, the model's strong representative features include these characteristics, indicating its learning preference to some extent. This insight into the model's decision-making process enhances our understanding of its interpretability. Based on this, we further validated the model on the testing set and compared its performance with other models.

In the comparative analysis of model performance, as shown in Table 5, our proposed model shows a robust balance across different metrics, demonstrating its effectiveness in the detection task. From the comparison of experimental results, the region proposal-based object detection algorithm Faster R-CNN achieved a precision of 0.820, slightly behind the highest result of 0.827 from YOLOv8-m, but its recall

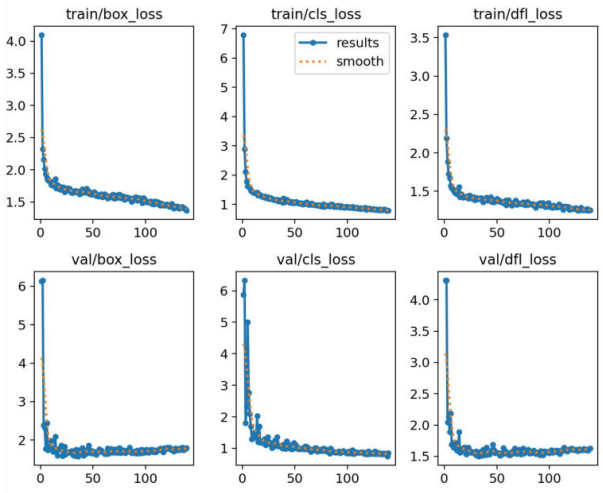

FIGURE 6. Training and validation loss over epochs. This figure shows that as the number of epochs increases, both training and validation loss gradually decrease and stabilize, which indicates that the model is learning effectively and improving its performance on both the training and validation datasets, demonstrating good convergence.

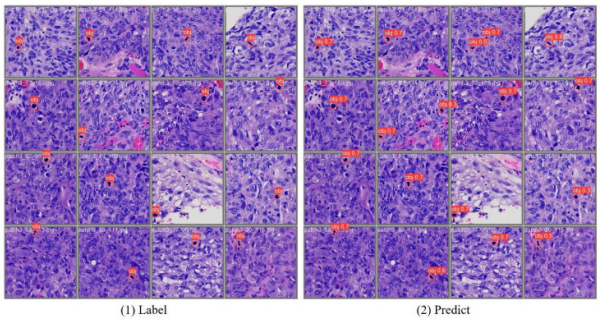

FIGURE 7. Detection performance of proposed model on the validation set.

rate is significantly lower than other models; the series of YOLOv5 models, except for YOLOv5-s and YOLOv5-l, which achieved recalls of 0.857 and 0.888 respectively, have generally lower precision, resulting in slightly poorer overall performance; RT-DETR performed similarly to YOLOv5-n, lower than other models; while the YOLOv8 series of models were competitive, showing good recall but overall lower precision, with YOLOv8-m achieving the highest precision of 0.827 and a decent recall of 0.816, with an F1 score of 0.822; our model achieved a precision of 0.816 and a recall of 0.858, as well as the highest F1 score of 0.837, with precision only behind YOLOv8-m and Faster R-CNN, and recall only behind YOLOv8-n, YOLOv5-s and YOLOv8-l. Our model outperforms the standard YOLOv5 series of n, s, m, RT-DETR and YOLOv8-s in terms of precision and recall, demonstrating its excellent detection capability. This is particularly evident when considering the F1 score, where our model not only outperforms individual YOLOv8 variants, but also exceeds the performance of the original YOLOv8 framework, indicating that it can identify relevant objects with high accuracy while maintaining a low miss rate, which is crucial for practical applications.

This paper uses the original YOLOv8-m, YOLOv8-n models and our proposed model to test a selection of image data from the test set to better verify the detection performance of the proposed algorithm in practical applications of mitosis. As shown in Figure 8, we find that detection performance of our model is better than that of YOLOv8-m and YOLOv8-n in the cases that are easily overlooked or difficult to detect. Both YOLOv8-m and YOLOv8-n have false detections with low confidence, unable to classify correctly, while our algorithm can accurately detect and classify the mitotic targets. This is mainly due to the use of a new backbone network and the AFPN strategy in the model, improving detection precision and expanding the receptive field, thus enhancing the overall performance.

TABLE 5. Comparison results with other networks.

Models	Precision	Recall	F1-Scores
Faster R-CNN	0.820	0.802	0.811
YOLOv5-n	0.779	0.796	0.787
YOLOv5-s	0.741	0.857	0.795
YOLOv5-m	0.799	0.771	0.785
YOLOv5-l	0.751	0.888	0.792
RT-DETR	0.773	0.809	0.790
YOLOv8-n	0.711	0.908	0.798
YOLOv8-s	0.777	0.837	0.806
YOLOv8-m	0.827	0.816	0.822
YOLOv8-l	0.767	0.874	0.817
Ours	0.816	0.858	0.837

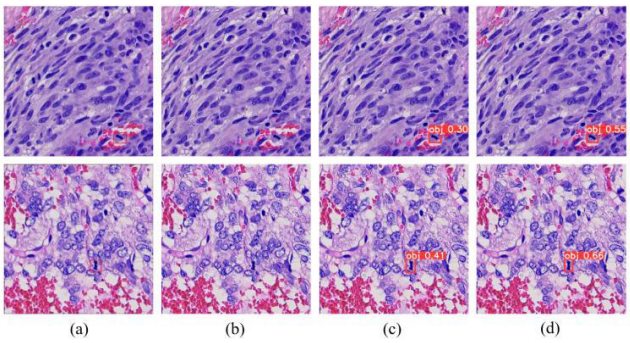

FIGURE 8. Comparison of the mitosis detection results. (a) Original labels. (b) YOLOv8-m detection results, no detection boxes indicate that the confidence is below the set threshold of 0.2. (c) YOLOv8-n detection results (d) Proposed model detection results.

VI. DISCUSSION

In the clinical diagnosis and management of GIST patients, the NIH risk classification is frequently used to assess the likelihood of recurrence and metastasis, as well as to precisely guide postoperative adjuvant therapy [4]. Unfortunately, the widespread use of mitotic figure identification and counting in clinical practice is constrained by a number of factors, including a scarcity of experienced pathologists, the high subjectivity and lack of reproducibility of diagnostic results,

and the difficulty of applying it to large-scale data analysis [12]. To our knowledge, this is the first attempt to apply a deep learning model to investigate GIST mitotic figure recognition. In this study, we have selected YOLOv8 as our foundational framework. Based on this, we used VanillaNet in place of the backbone network and enhanced the model by using the multi-scale feature extraction and fusion approach known as AFPN. It is pleasing that our suggested model outperformed the others in mitotic figure detection and demonstrated the best overall performance.

It is noteworthy that the YOLO series of models are designed to achieve fast detection while maintaining high accuracy, which treats the object detection task as a single regression problem, mapping directly from image pixels to bounding box coordinates and class probabilities [17]. This approach reduces error propagation and increases efficiency compared to traditional methods that decompose the task into multiple independent steps, such as region proposal extraction, feature extraction, and classification. The architecture of YOLOv8 allows for relatively easy customization and extension, offering flexibility to adapt the model structure according to specific detection needs, making it suitable for particular biomedical image analysis tasks [32]. Tasks such as mitotic detection often involve processing a large volume of image data, and YOLOv8 is capable of providing real-time or near-real-time detection speeds, which is particularly crucial for applications that demand rapid processing and analysis of a vast amount of biomedical images.

VanillaNet has made numerous innovations, such as refining the network architecture for higher computing performance, to increase the efficiency and precision of feature extraction compared to the original backbone network used in YOLOv8 [33]. For instance, VanillaNet uses lightweight convolution modules, which can lower computational costs and parameter counts while preserving or enhancing the representational capability of the network [34], [35]. It also incorporates a new deep training strategy, which can facilitate the training of deeper networks and allow the model to focus more on relevant features for the detection task [36]. Therefore, it might be designed to work more efficiently with the specific object classes that are commonly encountered in targeted detection tasks. The bidirectional path strategy of AFPN ensures that the model can obtain rich semantic and detailed information at each scale [31], which can exploit wider contextual information [37]. It also can help the model capture key features related to mitotic states, thereby improving the ability to identify different stages. And this scale adaptability enhances robustness of the model, providing more accurate and stable detection results. These improvements enable the network to process input images more quickly, allowing for relatively fast inference times. Additionally, the network can more accurately localize and classify objects in the image, improving the overall performance of the YOLOv8 model.

Based on the experimental results and the refined approach, this model is anticipated to serve as a robust assisting tool for medical practitioners, thereby enhancing the efficiency of diagnosing mitotic figures in GIST. By integrating this model into the risk classification of GIST, it can function as a secondary diagnostic aid for physicians [38]. This comprehensive and accurate identification offers more precise suggestions for clinical diagnosis and treatment [39], [40], [41]. In addition, the model exhibits automatic identification capabilities, expediting the mitosis detection process and subsequently reducing diagnosis time. This acceleration contributes to swift generation of pathological results, providing robust support for clinical large-scale repetitive and tedious tasks. It is noteworthy that the rapid advancement of deep learning in medical image analysis has brought about a revolutionary shift in disease diagnosis and treatment. Beyond replicating human subjective evaluation, deep learning holds immense potential in identifying image features that are challenging for the human naked eye to detect. In due course, deep learning may surpass existing diagnostic standards, adapting more accurate and detailed evaluation models and standards.

Furthermore, while our implementation focuses on a specific dataset, it's important to recognize that recent advancements in mitotic figure detection have successfully leveraged deep learning models for extensive data analysis and real-time applications. For instance, Wang et al. [21] proposed FMDet, which uses Fourier-based data augmentation, pixel-level annotation, and segmentation-based detection, achieving top ranking in the MIDOG 2021 challenge and demonstrating generalization across four datasets. Similarly, Jahanifar et al. [42] developed a two-stage framework for mitotic detection, excelling in both MIDOG21 and MIDOG22 challenges by applying their model to the largest publicly available datasets. In actual clinical, pathologists typically spend an average of several minutes to tens of minutes reviewing a patient's pathology images. However, by using our model in an efficient hardware and optimized software environment, the inference speed for a single 512×512 image can be reduced to approximately ten milliseconds, and processing an entire WSI can be shortened to about 10 minutes. This significantly reduces the workload in clinical settings. Nevertheless, there are still some limitations in comparing accuracy. The degree of model optimization and the specific software environment configuration need to be considered. Additionally, it requires the averaged results of multiple pathology experts, which necessitates more time to complete.

In addition, robust data encryption and access control mechanisms should be implemented to safeguard patient data throughout the processing pipeline. Furthermore, to address potential biases in model predictions, diverse and representative datasets should be established and conduct comprehensive bias assessments to detect and rectify any discrepancies

in model performance among various demographic groups promptly.

There are several limitations to this study. Firstly, the source of the obtained image set is too limited. In future studies, it is imperative to conduct multi-center investigations encompassing a broader spectrum of image data from more diverse patient populations and healthcare systems. We need to increase the size of the dataset to verify the universality of the model in different populations and environments, which aims to enhance the generalizability of the model. Secondly, the reliability and accuracy of the deep learning model significantly depend on the establishment of the GIST mitotic image dataset in the training set. The current dataset has not yet reached the desired goal. Thirdly, our WSIs digital pathology images are derived from pathology specimens transformed by an image scanning system. This process, on one hand, impacts image quality due to pathology specimen acquisition, filming, and staining. On the other hand, the image quality may also be compromised due to the introduction of optical aberrations or digital noise during the image transformation process. In addition, current research focuses on the accuracy of model detection and lacks exploration of real-time performance, which is crucial for practical deployment in clinical environments. Future work should also enhance and evaluate the real-time capability and computational efficiency of models, considering the integration with existing clinical systems to provide efficient support for doctors. Future work should also focus on enhancing the interpretability of models. This may involve developing visualization tools to explain the decision-making process of the model, highlighting the most influential features for prediction through techniques such as attention maps in the attention mechanism, and conducting research to evaluate the clinical utility of these explanations. These factors have a detrimental effect on the recognition and analysis of images by automatic image analysis algorithms.

In future work, we will focus on the following aspects: Firstly, the comparison of detection speed performance is crucial, as current work primarily emphasizes improving accuracy while neglecting the speed of detection. Secondly, enhancing model interpretability is essential. Our study revealed that the model may bias towards certain features in target detection, indicating a need for further exploration to develop interpretability. Machine learning methods (such as radiomics feature extraction) or visualization of attention maps within attention mechanisms can be considered for use. Lastly, addressing data security and privacy protection in deep learning applications is imperative. Ethical and security issues arising from data collection and training could potentially be resolved through new technologies such as federated learning. We plan to conduct a multi-center study to amass a larger dataset, enhance image quality, and optimize the deep learning model to rectify these problems.

VII. CONCLUSION

In this paper, we have improved the state-of-the-art YOLOv8 detection model by employing a backbone network modification method and a multi-scale feature fusion strategy to detect mitosis from pathological slides. The model can be used to detect mitotic figures at different stages. The model proposed in this paper uses VanillaNet to optimise the network structure of the original YOLOv8 model for customised detection tasks, thereby improving the network performance. Experimental results show that the model proposed in this paper has better performance compared to YOLOv8 and other deep learning models, and adopts the AFPN feature pyramid network, which is currently a more competitive scale feature fusion module, leading to further improving the detector's performance. Our proposed model is expected to be used for real-time detection of mitotic figures in pathological slides. Future work will focus on applying the existing model to actual clinical tasks. The characteristics and features of different mitotic stages not covered in this work will also be analysed. In addition, data augmentation methods and detection models will be optimised to further improve detection accuracy.

ACKNOWLEDGMENT

The authors appreciate the editors and reviewers for their diligent efforts and valuable feedback, and extend their gratitude to everyone who contributed to the writing and revision of this manuscript.

(Haoxin Liang and Zhichun Li are co-first authors.)

REFERENCES

[1] K. Søreide, O. M. Sandvik, J. A. Søreide, V. Giljaca, A. Jureckova, and V. R. Bulusu, "Global epidemiology of gastrointestinal stromal tumours (GIST): A systematic review of population-based cohort studies," *Cancer Epidemiology*, vol. 40, pp. 39–46, Feb. 2016, doi: 10.1016/j.canep.2015.10.031.

[2] P. G. Casali et al., "Gastrointestinal stromal tumours: ESMO–EURACAN clinical practice guidelines for diagnosis, treatment and follow-up," *Ann. Oncol.*, vol. 29, pp. iv68–iv78, Oct. 2018, doi: 10.1093/annonc/mdy095.

[3] H. Joensuu, M. Eriksson, K. S. Hall, A. Reichardt, B. Hermes, J. Schütte, S. Cameron, P. Hohenberger, P. J. Jost, S.-E. Al-Batran, L. H. Lindner, S. Bauer, E. Wardelmann, B. Nilsson, R. Kallio, P. Jaakkola, J. Junnila, T. Alvegård, and P. Reichardt, "Survival outcomes associated with 3 years vs 1 year of adjuvant imatinib for patients with high-risk gastrointestinal stromal tumors: An analysis of a randomized clinical trial after 10-year follow-up," *J. Amer. Med. Assoc. Oncol.*, vol. 6, no. 8, p. 1241, Aug. 2020, doi: 10.1001/jamaoncol.2020.2091.

[4] C. M. Kelly, L. G. Sainz, and P. Chi, "The management of metastatic GIST: Current standard and investigational therapeutics," *J. Hematol. Oncol.*, vol. 14, no. 1, p. 2, Jan. 2021, doi: 10.1186/s13045-020-01026-6.

[5] L. R. Klug, H. M. Khosroyani, J. D. Kent, and M. C. Heinrich, "New treatment strategies for advanced-stage gastrointestinal stromal tumours," *Nature Rev. Clin. Oncol.*, vol. 19, no. 5, pp. 328–341, May 2022, doi: 10.1038/s41571-022-00606-4.

[6] D. M. Metter, T. J. Colgan, S. T. Leung, C. F. Timmons, and J. Y. Park, "Trends in the U.S. and Canadian pathologist workforces from 2007 to 2017," *J. Amer. Med. Assoc. Netw. Open*, vol. 2, no. 5, May 2019, Art. no. e194337, doi: 10.1001/jamanetworkopen.2019.4337.

[7] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[8] B. Sahiner, A. Pezeshk, L. M. Hadjiiski, X. Wang, K. Drukker, K. H. Cha, R. M. Summers, and M. L. Giger, "Deep learning in medical imaging and radiation therapy," *Med. Phys.*, vol. 46, no. 1, pp. e1–e36, Jan. 2019, doi: 10.1002/mp.13264.

[9] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022, doi: 10.1109/TNNLS.2021.3084827.

[10] A. Janowczyk and A. Madabhushi, "Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases," *J. Pathol. Informat.*, vol. 7, no. 1, p. 29, Jan. 2016, doi: 10.4103/2153-3539.186902.

[11] L. Pantanowitz, D. Hartman, Y. Qi, E. Y. Cho, B. Suh, K. Paeng, R. Dhir, P. Michelow, S. Hazelhurst, S. Y. Song, and S. Y. Cho, "Accuracy and efficiency of an artificial intelligence tool when counting breast mitoses," *Diagnostic Pathol.*, vol. 15, no. 1, p. 80, Jul. 2020, doi: 10.1186/s13000-020-00995-z.

[12] G. Campanella, M. G. Hanna, L. Geneslaw, A. Miraflor, V. W. K. Silva, K. J. Busam, E. Brogi, V. E. Reuter, D. S. Klimstra, and T. J. Fuchs, "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature Med.*, vol. 25, no. 8, pp. 1301–1309, Aug. 2019, doi: 10.1038/s41591-019-0508-1.

[13] Y. Xiao, X. Wang, P. Zhang, F. Meng, and F. Shao, "Object detection based on faster R-CNN algorithm with skip pooling and fusion of contextual information," *Sensors*, vol. 20, no. 19, p. 5490, Sep. 2020, doi: 10.3390/s20195490.

[14] G. H. Aly, M. Marey, S. A. El-Sayed, and M. F. Tolba, "YOLO based breast masses detection and classification in full-field digital mammograms," *Comput. Methods Programs Biomed.*, vol. 200, Mar. 2021, Art. no. 105823, doi: 10.1016/j.cmpb.2020.105823.

[15] E. Guemas, B. Routier, T. Ghelfenstein-Ferreira, C. Cordier, S. Hartuis, B. Marion, S. Bertout, E. Varlet-Marie, D. Costa, and G. Pasquier, "Automatic patient-level recognition of four plasmodium species on thin blood smear by a real-time detection transformer (RT-DETR) object detection algorithm: A proof-of-concept and evaluation," *Microbiol. Spectr.*, vol. 12, no. 2, Feb. 2024, Art. no. e0144023, doi: 10.1128/spectrum.01440-23.

[16] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[17] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," *Multimedia Tools Appl.*, vol. 82, no. 6, pp. 9243–9275, Aug. 2022, doi: 10.1007/s11042-022-13644-y.

[18] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "DETRs Beat YOLOs on real-time object detection," Apr. 2023, *arXiv:2304.08069*.

[19] C. Li, X. Wang, W. Liu, and L. J. Latecki, "DeepMitosis: Mitosis detection via deep detection, verification and segmentation networks," *Med. Image Anal.*, vol. 45, pp. 121–133, Apr. 2018, doi: 10.1016/j.media.2017.12.002.

[20] M. Z. Alom, T. Aspiras, T. M. Taha, T. Bowen, and V. K. Asari, "MitosisNet: End-to-end mitotic cell detection by multi-task learning," *IEEE Access*, vol. 8, pp. 68695–68710, 2020, doi: 10.1109/ACCESS.2020.2983995.

[21] X. Wang, J. Zhang, S. Yang, J. Xiang, F. Luo, M. Wang, J. Zhang, W. Yang, J. Huang, and X. Han, "A generalizable and robust deep learning algorithm for mitosis detection in multicenter breast histopathological images," *Med. Image Anal.*, vol. 84, Feb. 2023, Art. no. 102703, doi: 10.1016/j.media.2022.102703.

[22] Y. Topuz, S. Yıldız, and S. Varlı, "Performance analysis of the YOLO series for object detection: Detection of mitosis cells in histopathology images," in *Proc. Med. Technol. Congr. (TIPTEKNO)*, Nov. 2023, pp. 1–4.

[23] Z. Yücel, F. Akal, and P. Oltulu, "Mitotic cell detection in histopathological images of neuroendocrine tumors using improved YOLOv5 by transformer mechanism," *Signal, Image Video Process.*, vol. 17, no. 8, pp. 4107–4114, Jun. 2023, doi: 10.1007/s11760-023-02642-8.

[24] S. Choi and S. Kim, "Artificial intelligence in the pathology of gastric cancer," *J. Gastric Cancer*, vol. 23, no. 3, p. 410, Jul. 2023, doi: 10.5230/jgc.2023.23.e25.

[25] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[27] S. Wu, S. Zhong, and Y. Liu, "Deep residual learning for image steganalysis," *Multimedia Tools Appl.*, vol. 77, no. 9, pp. 10437–10453, May 2018, doi: 10.1007/s11042-017-4440-4.

[28] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, May 2019, pp. 6105–6114.

[29] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.

[30] G. Yang, J. Lei, Z. Zhu, S. Cheng, Z. Feng, and R. Liang, "AFPN: Asymptotic feature pyramid network for object detection," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2023, pp. 2184–2189.

[31] K. Min, G.-H. Lee, and S.-W. Lee, "Attentional feature pyramid network for small object detection," *Neural Netw.*, vol. 155, pp. 439–450, Nov. 2022, doi: 10.1016/j.neunet.2022.08.029.

[32] X. Xiao and X. Feng, "Multi-object pedestrian tracking using improved YOLOv8 and OC-SORT," *Sensors*, vol. 23, no. 20, p. 8439, Oct. 2023, doi: 10.3390/s23208439.

[33] X. Xu, T. Sanford, B. Turkbey, S. Xu, B. J. Wood, and P. Yan, "Shadow-consistent semi-supervised learning for prostate ultrasound segmentation," *IEEE Trans. Med. Imag.*, vol. 41, no. 6, pp. 1331–1345, Jun. 2022, doi: 10.1109/TMI.2021.3139999.

[34] M. Mubashar, H. Ali, C. Grönlund, and S. Azmat, "R2U++: A multiscale recurrent residual U-Net with dense skip connections for medical image segmentation," *Neural Comput. Appl.*, vol. 34, no. 20, pp. 17723–17739, Oct. 2022, doi: 10.1007/s00521-022-07419-7.

[35] J. Kugelman, J. Allman, S. A. Read, S. J. Vincent, J. Tong, M. Kalloniatis, F. K. Chen, M. J. Collins, and D. Alonso-Caneiro, "A comparison of deep learning U-Net architectures for posterior segment OCT retinal layer segmentation," *Sci. Rep.*, vol. 12, no. 1, p. 14888, Sep. 2022, doi: 10.1038/s41598-022-18646-2.

[36] D. Abdelhafiz, J. Bi, R. Ammar, C. Yang, and S. Nabavi, "Convolutional neural network for automated mass segmentation in mammography," *BMC Bioinf.*, vol. 21, no. S1, p. 192, Dec. 2020, doi: 10.1186/s12859-020-3521-y.

[37] S. Li, H. Huang, X. Meng, M. Wang, Y. Li, and L. Xie, "A glove-wearing detection algorithm based on improved YOLOv8," *Sensors*, vol. 23, no. 24, p. 9906, Dec. 2023, doi: 10.3390/s23249906.

[38] C. Hassan, M. Spadaccini, A. Iannone, R. Maselli, M. Jovani, V. T. Chandrasekar, G. Antonelli, H. Yu, M. Areia, M. Dinis-Ribeiro, P. Bhandari, P. Sharma, D. K. Rex, T. Rösch, M. Wallace, and A. Repici, "Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: A systematic review and meta-analysis," *Gastrointestinal Endoscopy*, vol. 93, no. 1, pp. 77–85, Jan. 2021, doi: 10.1016/j.gie.2020.06.059.

[39] S. Kuntz, E. Krieghoff-Henning, J. N. Kather, T. Jutzi, J. Höhn, L. Kiehl, A. Hekler, E. Alwers, C. von Kalle, S. Fröhling, J. S. Utikal, H. Brenner, M. Hoffmeister, and T. J. Brinker, "Gastrointestinal cancer classification and prognostication from histology using deep learning: Systematic review," *Eur. J. Cancer*, vol. 155, pp. 200–215, Sep. 2021, doi: 10.1016/j.ejca.2021.07.012.

[40] H.-P. Chan, R. K. Samala, L. M. Hadjiiski, and C. Zhou, "Deep learning in medical image analysis," *Adv. Exp. Med. Biol.*, vol. 1213, pp. 3–21, Feb. 2020, doi: 10.1007/978-3-030-33128-3_1.

[41] B. J. Erickson, "Basic artificial intelligence techniques," *Radiologic Clinics North Amer.*, vol. 59, no. 6, pp. 933–940, Nov. 2021, doi: 10.1016/j.rcl.2021.06.004.

[42] M. Jahanifar, A. Shephard, N. Zamanitajeddin, S. Graham, S. E. A. Raza, F. Minhas, and N. Rajpoot, "Mitosis detection, fast and slow: Robust and efficient detection of mitotic figures," *Med. Image Anal.*, vol. 94, May 2024, Art. no. 103132, doi: 10.1016/j.media.2024.103132.

HAOXIN LIANG was born in Foshan, Guangdong, China, in 2000. She is currently pursuing the bachelor's degree in clinical medicine with Southern Medical University.

She is working on two articles and one patent. Her research interests include the application of medical-industrial deep learning in pathology diagnosis and basic research on gastrointestinal tumors. She has received several school-level scholarships from Southern Medical University, including the First Prize, in 2018, the First Prize, in 2019, the Second Prize, in 2020, the Second Prize, in 2021, and the Second Prize, in 2022. She was awarded the Silver Prize in the 9th China International "Internet Plus" Student Innovation and Entrepreneurship Competition.

ZHICHUN LI received the B.S. degree in computer science and technology from Hong Kong Baptist University, in 2022. He is currently pursuing the M.Phil. degree with the Department of Health Technology and Informatics, The Hong Kong Polytechnic University. His research interests include medical image processing, image-guided radiation therapy, and deep learning applications.

WEIJIE LIN was born in Shantou, Guangdong, China, in 2002. He is currently pursuing the bachelor's degree in clinical medicine with Southern Medical University.

He diligently pursued his passion for clinical medicine with Southern Medical University. Throughout his undergraduate journey, he actively engaged in practical learning opportunities, collaborated with peers on research projects, and eagerly participated in extracurricular activities related to his field of study.

Mr. Lin was awarded the Silver Prize in the 9th China International "Internet Plus" Student Innovation and Entrepreneurship Competition.

YUHENG XIE was born in Weinan, Shanxi, China, in 2004. He is currently pursuing the bachelor's degree in clinical medicine with Southern Medical University.

He hasn't won many awards due to his lower grades, but he actively participates in various school activities and makes efforts to engage in scientific research with elder students.

Mr. Xie was awarded the Silver Prize in the 9th China International "Internet Plus" Student Innovation and Entrepreneurship Competition.

SHUO ZHANG received the B.S. degree in biomedical engineering from Southern Medical University, China, in 2017, and the M.S. degree in biomedical engineering from China Medical University, China, in 2020. He is currently pursuing the Ph.D. degree with the School of Biological Science and Medical Engineering, Southeast University, China. His main research interests include machine learning and inaccurate supervision, especially in learning from health data.

ZHOU LI has presided over two projects each for Guangdong Natural Science Foundation, Science and Technology Plan, and major industry-university-research projects in Guangzhou; one sub-project for Guangdong Strategic Emerging Industry Core Technology Research Project; three sub-projects for major industry-university-research projects in Guangzhou; and one project each for the Southern Medical University Science and Technology Development Plan and Clinical Research Project, with over one million in ongoing funding. In recent years, he has published 11 articles as the first author or the corresponding author, with nine articles having an impact factor of more than 3.0 and two articles more than 5.0, totaling an impact factor of 42.595. He has also published several case reports in domestic and international journals and holds three invention patents.

Dr. Li appeared on CCTV's Special Program "Thumbs Up China" for the 19th National Congress of the Communist Party of China as the "Warmest Doctor." In 2017, he was awarded the first "Norman Bethune-Style Good Doctor" Nomination Award from the Chinese Medical Humanities Conference. He was honored with the "Improving Medical Service" Advanced Individual title by the National Health Commission, in 2021.

HONGYU LUO the bachelor's degree in clinical medicine from Hengyang Medical College, in 1994.

He is currently the Chief Physician of General Surgery with The Sixth People's Hospital of Huizhou City. His research interests include tension-free hernia repair and recurrent hernia surgery; radical surgery for gastric and colorectal cancer; extrahepatic cholelithiasis surgery: cholecystectomy, choledochotomy with internal and external choledochal duct exploration and drainage; splenectomy; and portage dissection for cirrhosis.

Mr. Luo is a member of the Hospital's Party Committee and the Secretary of the First Party Branch of Surgery. As the first person in charge, he was awarded the "New Technology, New Project Application Award" of The Sixth People's Hospital of Huizhou City, the First Famous Doctor Award of The Sixth People's Hospital of Huizhou City, and the Advanced Science and Technology Worker of The Sixth People's Hospital of Huizhou City, in 2020.

TIAN LI received the B.S. degree in chemical biology from Peking University, Beijing, China, in 2014, the M.S. degree in medical physics from Duke University, Durham, NC, USA, in 2016, and the Ph.D. degree in health technology and informatics from The Hong Kong Polytechnic University, Hong Kong, China, in 2020.

From 2016 to 2018, he was a Medical Physicist with Beijing Cancer Hospital. From 2021 to 2023, he was a Research Assistant Professor with the Department of Health Technology and Informatics, The Hong Kong Polytechnic University. Since 2023, he has been an Assistant Professor with the Department of Health Technology and Informatics, The Hong Kong Polytechnic University. He is the author of more than 30 journal articles. His research interests include image analysis, image enhancement, image synthesis, image reconstruction, four-dimensional imaging techniques, magnetic resonance fingerprinting, radiotherapy, and radiomics.

Dr. Li received several awards and honors, including Hong Kong Ph.D. Fellowship and the Faculty Distinguished Thesis Award. He is an Associate Editor of *Medical Physics*.

SHUAI HAN received the Bachelor of Medicine degree from Henan University School of Medicine, in 2010, and the Ph.D. degree in surgery from Southern Medical University, in 2015.

He is currently with the Department of General Surgery, Zhujiang Hospital, Southern Medical University, as an Attending Physician. He has authored more than 20 articles and holds more than ten patents. His research interests include basic research on gastrointestinal tumors, medical-industrial integration of deep learning in the diagnosis of gastrointestinal tumors, laparoscopic minimally invasive treatment for gastrointestinal tumors, multidisciplinary comprehensive treatment of intermediate- and advanced-stage gastrointestinal tumors, and minimally invasive surgeries, such as laser intracavernous closure of saphenous varicose vein. He also performs minimally invasive laparoscopic repair of inguinal hernia, incisional hernia, and fistula hernia.

Dr. Han was awarded the 2023 Third Individual Third Prize of "Chengzhu Young Research Advancement Award" in Obesity and Metabolic Surgery and the Winner of the 1st Chinese Electronic Journal of Obesity and Metabolic Diseases "Precision Cup" Weight Loss Surgery Video Contest.