scientific reports

OPEN

Application of dual branch and bidirectional feedback feature extraction networks for real time accurate positioning of stents

Shixiao Wu[1,6], Rui Hu[2], Chengcheng Guo[3,4]✉, Xingyuan Lu[5,6]✉, Peng Leng[1] & Zhiwei Wang[2]

The installation of arterial stents refers to the use of stents (also known as vascular stents) to maintain the patency of arteries during the treatment of arterial stenosis or blockage. Arterial stents are typically made of metal or polymer materials and are structured as a mesh that provides support within the blood vessel, preventing it from collapsing again after interventional treatment. The installation of arterial stents is an effective interventional therapy that can significantly improve symptoms caused by arterial stenosis or blockage and enhance the quality of life for patients. Endovascular therapy has become increasingly important for treating both thoracic and abdominal aortic diseases. A critical aspect of this procedure is the precise positioning of stents and complete isolation of the pathology. To enhance stent placement accuracy, we propose a deep learning model called the Double Branch Medical Image Detector (DBMedDet), which offers real-time guidance for stent placement during implantation surgeries. The DBMedDet model features a parallel dual-branch edge feature extraction network, a bidirectional feedback feature fusion neck sub-network, as well as a position detection head and a classification head specifically designed for thoracic and abdominal aortic stents. The model has achieved a detection Mean Average Precision (mAP) of 0.841 (mAP@0.5) and a real-time detection speed of 127 Frames Per Second (FPS). For mAP@0.5, when employing 5-fold cross-validation, DBMedDet demonstrates superior performance compared to several YOLO models, achieving improvements of 4.88% over YOLOv8l, 4.61% over YOLOv8m, 3.20% over YOLOv8s, 6.23% over YOLOv8n, 6.09% over YOLOv10s, 3.92% over YOLOv9s, 3.20% over YOLOv8s, 3.00% over YOLOv7tiny, and 5.01% over YOLOv5s. This study presents a precise and easily implementable method for the automatic detection of stent placement limits in the thoracic and abdominal aorta. The model can be applied in various areas such as coronary intervention therapy, peripheral vascular intervention therapy, cerebrovascular intervention therapy, postoperative monitoring and follow-up, and medical training and education. By utilizing real-time imaging guidance and deep learning models (such as DBMedDet), stent placement procedures in these application areas can be performed with greater precision and safety, thereby enhancing patient treatment outcomes and quality of life.

Keywords DBMedDet, Stent installation area limitations, Real-time, Object detection, Deep learning

Endovascular therapy is a minimally invasive procedure used to treat various vascular diseases and conditions through the vascular system. This approach involves accessing the blood vessels via small incisions, typically in the groin or wrist, and employing specialized tools such as catheters, stents, and balloons to diagnose and treat conditions like aneurysms, arterial blockages, and vascular malformations. One of the key advantages of endovascular therapy is its reduced recovery time compared to traditional open surgeries, as it minimizes tissue trauma and decreases the risk of complications. Additionally, advancements in imaging technologies, such as intravascular ultrasound and fluoroscopy, have significantly improved the precision and efficacy of these procedures, enabling clinicians to navigate complex vascular anatomy more effectively.

[1]Scholl of Information Engineering, Wuhan Business University, Wuhan 430056, Hubei, China. [2]Cardiovascular Department, Renmin Hospital of Wuhan University, Wuhan 430071, Hubei, China. [3]Scholl of Information Egineering, Wuhan College, Wuhan 430212, Hubei, China. [4]Scholl of Electronic Information, Wuhan University, Wuhan 430072, Hubei, China. [5]School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300382, China. [6]Shixiao Wu and Xingyuan Lu contributed equally to this work. ✉email: netccg@whu.edu.cn; lxy1466792674@stud.tjut.edu.cn

Recent studies have shown that endovascular techniques can lead to favorable outcomes in various patient populations. For instance, the use of covered stents in the treatment of aortoiliac occlusive disease has demonstrated significant improvements in both patient symptoms and quality of life[1]. Moreover, the integration of advanced imaging modalities has enhanced the success rates of procedures like endovascular aneurysm repair (EVAR), providing better visualization of the vascular structures involved[2]. As technology continues to evolve, endovascular therapy is likely to expand in scope, offering new possibilities for the treatment of challenging vascular conditions.

Lately, the use of endovascular therapy for aortic diseases, including both thoracic and abdominal conditions, has seen a significant increase. Precise positioning of the stent and complete isolation of the pathology are crucial. The aorta has numerous branches that transport oxygenated blood to multiple organs, which might be covered during endovascular surgery due to incorrect judgment of intraoperative images, with reported incidences ranging from 0.1–3.5%. Coverage of the left subclavian artery in thoracic endovascular repair can lead to left arm ischemia, subclavian steal syndrome, or even stroke, while coverage of the renal artery in abdominal endovascular repair may result in renal ischemia or necrosis. Intraoperative images often have lower resolution due to equipment limitations or the need for real-time processing. This makes it difficult to clearly identify small structures, such as stents or vascular lesions. Additionally, intraoperative images are often subject to various interferences (such as electromagnetic interference, motion artifacts, etc.), leading to higher noise levels, which in turn affects the quality of the images and the accuracy of subsequent analyses. Endovascular surgery involves complex vascular networks and surrounding tissue structures, requiring the imaging processing system to accurately understand and distinguish these structures for effective object detection. At the same time, during the surgery, the image analysis system must be able to provide feedback in almost real-time to ensure the safety and effectiveness of the procedure.

In recent years, significant advancements have been made in the field of stent detection and visualization, highlighting the importance of accurate assessment of stent positioning in relation to the vessel wall. Canero introduced an innovative automatic technique for visualizing and quantifying the mutual positioning of stents and vessel walls through three-dimensional reconstructions[3]. Building on this, J. Dijkstra proposed a knowledge- and model-guided system for the semi-automatic detection of stent borders, focusing on the analysis of transverse slices within specified segments[4]. Complementing these efforts, David Rotger presented a novel approach that employs a cascade of GentleBoost classifiers to detect stent struts, utilizing structural features to extract information from various subregions of the struts[5]. Furthermore, R. Hua introduced an automated method that integrates both local and contextual information about strut appearance. This method begins with the extraction of local features from a diverse filter bank, which are then used for pixel-wise classification to identify candidate pixels corresponding to stent struts. Recognizing the role of contextual surroundings in differentiating artifacts that may mimic stent struts, the initial detection map is refined by incorporating contextual features from the image patch centered around the candidate pixel[6]. Ciompi et al. developed a fully automatic computer-aided detection framework to identify and analyze stented coronary anomalies from 3D CT images[7]. This system utilizes a machine learning approach to achieve precise segmentation of arteries, accurate localization of blockages, and proper placement of stent struts and shaping. Additionally, Wang et al. employed Adaboost and Support Vector Machines (SVM) techniques to facilitate the automatic detection and annotation of stent struts in intravascular ultrasound images[8–10].

The machine learning models discussed earlier demonstrate strong capabilities in establishing correlations between metrics (images or datasets) and underlying attributes. However, these models face significant challenges in detecting the placement location of aortic stents during surgery. These challenges stem from several factors, including reliance on manual measurements, variations in image quality, and limited availability of measurement data. Adaboost typically requires multiple iterations to train several weak learners. Each additional weak learner increases the model’s training time and computational complexity. In real-time systems, there is a demand for quick response times, and the training process of Adaboost may lead to processing delays. SVM and certain decision tree algorithms can also experience significant increases in computational complexity when handling large-scale data, resulting in extended response times. Deep learning may serve as a means to provide intraoperative guidance for surgical teams, reducing errors in visual perception and enhancing surgical decision-making. In a noteworthy contribution to the dataset available in this domain, Yao et al. introduced Image Type-B Aortic Dissection (ImageTBAD), the first 3D computed tomography angiography (CTA) image dataset of thoracic aortic dissection (TBAD), which is uniquely annotated with true lumen (TL), false lumen (FL), and false lumen thrombus (FLT)[11,12]. Segmentation models like Fully Convolutional Networks (FCN) and U-Net require a substantial amount of computation during the forward propagation process. These models typically involve multiple convolutional layers and complex feature extraction processes, leading to high computational loads and slower processing speeds. Nils Gessert developed a convolutional neural network designed to predict the type of stent typically observed in Intravascular Optical Coherence Tomography (IVOCT) images[13,14]. “Training and Saliency Map Construction” two-stage detection method affects real-time performance.

Although deep learning is continually advancing, the aforementioned articles have hardly achieved real-time detection. Sindhu Ramachandran S utilizes a YOLO-based deep learning network for the real-time detection and localization of lung nodules in low-dose CT scans, achieves a performance of 20 images or frames per second (20 fps) on an Intel i5 system equipped with an NVIDIA GeForce GTX 1080[15]. But in general, FPS over 30 is considered to have achieved real-time detection[16]. Shuai Hao proposed YOLO-based model for the real-time detection of natural disaster victims, achieving an image detection speed of 42 frames per second at a resolution of 640×640, which demonstrates strong real-time performance[17]. The main contribution of this article is as follows:

(1) A parallel dual-branch edge feature extraction network is proposed for the initial extraction of backbone features. The dual-branch network typically has different branch structures that can process information at different scales in parallel. This enables the network to better capture the diversity and details of the objects, especially in medical imaging, where objects (such as stents) may vary in size and shape. The dual-branch structure allows the network to handle multiple features simultaneously, reducing processing time. This helps improve the overall inference speed of the network, meeting the demands of real-time detection.
(2) A bidirectional feedback feature fusion neck sub-network is developed for the secondary extraction of features. The bidirectional feedback mechanism allows the network to continuously fuse and adjust feature information during both forward and backward propagation. This dynamic feature fusion can better capture the complex characteristics of medical images, enhancing the model's ability to understand the objects. Bidirectional feedback can facilitate faster information transfer between feature layers, reducing redundant computations and improving processing efficiency. This is especially important for real-time applications, as it can accelerate the inference process.
(3) An auxiliary classification detection head is introduced for the classification of thoracic or abdominal aortic stents.
(4) Finally, a deep learning model named DBMedDet is presented, which provides real-time constraint cues for the placement location of stents during stent implantation surgery.

The remainder of this paper is organized as follows: In section "Materials and methods", we present the methodology used in this study, describing the components of the model, including a parallel dual-branch edge feature extraction backbone network, basic modules C2f and C2, an improved Neck sub-network, and a classification detection head. Section "Results" presents the results of the article, including comparisons of mAP/loss curves, ablation experiments (different convolutional kernels, different feature pyramid networks, different edge algorithms, different modules, pre-training, auxiliary classification heads, data augmentation, multi-scale training), and comparisons of detection results. Section "Discussion" summarizes the methods and results of the article, while section "Conclusion" discusses the details and limitations of the experiments.

Materials and methods

The dataset

The data was collected from the Department of Cardiology at Renmin Hospital of Wuhan University and consists of 2,014 thoracic/abdominal aortic images annotated by experienced physicians, as shown in Fig. 1. The green sections in the first and second images from the top left indicate the placement constraints for the thoracic aortic stent. The green sections in the remaining four images indicate the placement constraints for the abdominal aortic stent. This dataset is unique for detecting stents in the thoracic and abdominal aorta, as there are no other existing datasets available. The training set consists of 1612 samples, with 813 chest images and 799 abdominal images. The test set consists of 402 samples, with 203 chest images and 199 abdominal images. The ratio of training set to test set is 8:2. The above are images of the thoracic aorta, and the below are images of the abdominal aorta. We employed a 5-fold cross-validation method to evaluate the model's performance.

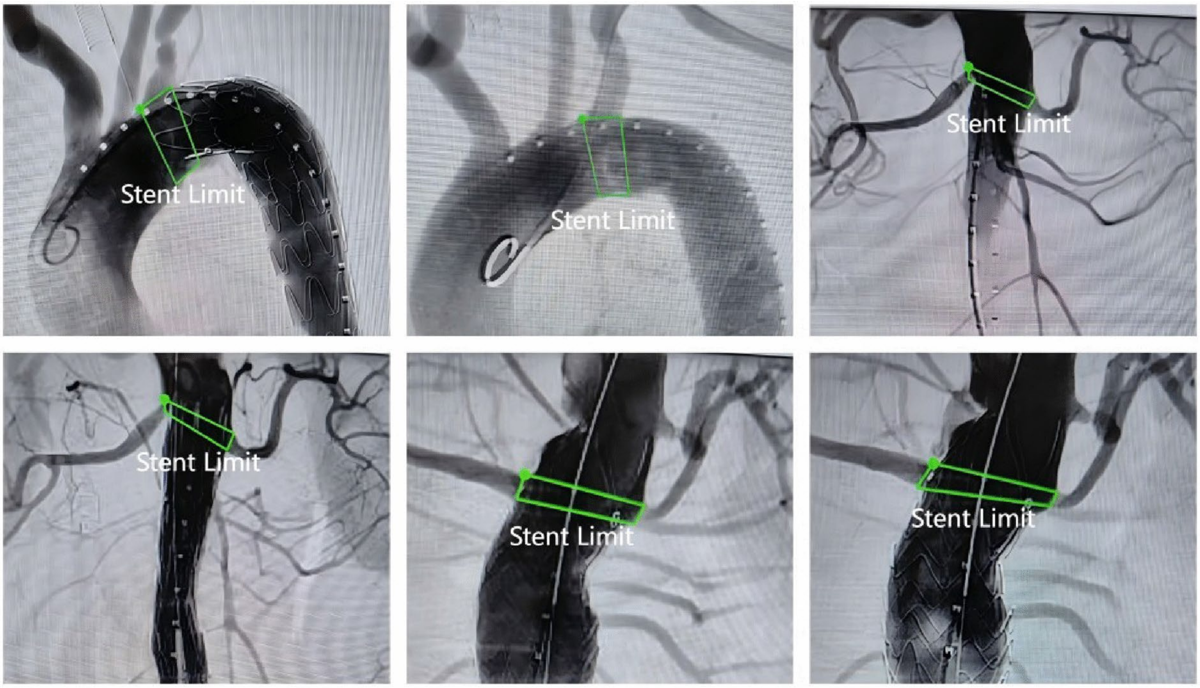

Fig. 1. The dataset.

The mAP@0.5, parameter count, FLOPs, and FPS for each model can be found in Table 1. All methods were performed in accordance with the relevant guidelines and regulations.

DBMedDet

The proposed DBMedDet model is composed of the previously mentioned dual-branch edge feature extraction backbone network, a neck section that utilizes bidirectional feedback feature fusion, and a head section that incorporates an auxiliary classification head, as illustrated in Fig. 2.

The model employs a dual-branch feature extraction network. This dual-branch structure enables the simultaneous extraction of different types of features, such as local and global features, enhancing the model's ability to capture diverse information. By processing different feature streams separately, the dual-branch network improves the model's robustness when faced with noise, occlusion, or variations in input. The two branches can complementarily extract information, helping the model to achieve a more comprehensive understanding of the input data, thereby increasing its representational power. Furthermore, the dual-branch structure allows for efficient feature fusion between the branches, enhancing the expressiveness of the features and ultimately improving the model's performance. It also enables the model to learn features at a deeper level, capturing more complex patterns and relationships.

The features of DBMedDet can flow in all four directions. This four-directional flow allows the model to better retain and utilize local features, reducing information loss during transmission. The model can comprehensively capture contextual information from different directions, enhancing its ability to recognize complex structures and boundaries. The fusion of features from different directions helps the model achieve more accurate detection and classification of targets at various scales, thus improving overall performance. When facing objects with different poses, directions, or occlusions, the four-directional feature flow can increase the model's robustness and adaptability. This flow enables the model to generate more detailed feature representations, allowing for a more accurate capture of subtle differences in objects.

Backbone

DBMedDet's backbone network consists of two branches (Fig. 3). The main branch is the original YOLOv8's backbone, while the other branch is dedicated to processing edge information[18]. The initial output of the edge branch undergoes edge information extraction using the Sobel operator, resulting in edge information features of batch × 1 × w × h. This edge information is then downsampled to 1/4 using the stem and processed in parallel with the main branch. At the same scale, before each passage through C2, feature fusion is performed with the features processed by the edge information. This allows the network to learn feature information from the main branch, effectively adding additional edge information to the original feature set. By combining the original feature information with the edge information, the network becomes more suitable for medical image detection.

In this backbone network, the innovative C2 module is introduced. This module consists of split operation, Conv1 × 1, and Conv3 × 3. The use of C2 aims to reduce Flops, as adding edge information does not require the model to learn excessive parameters for features.

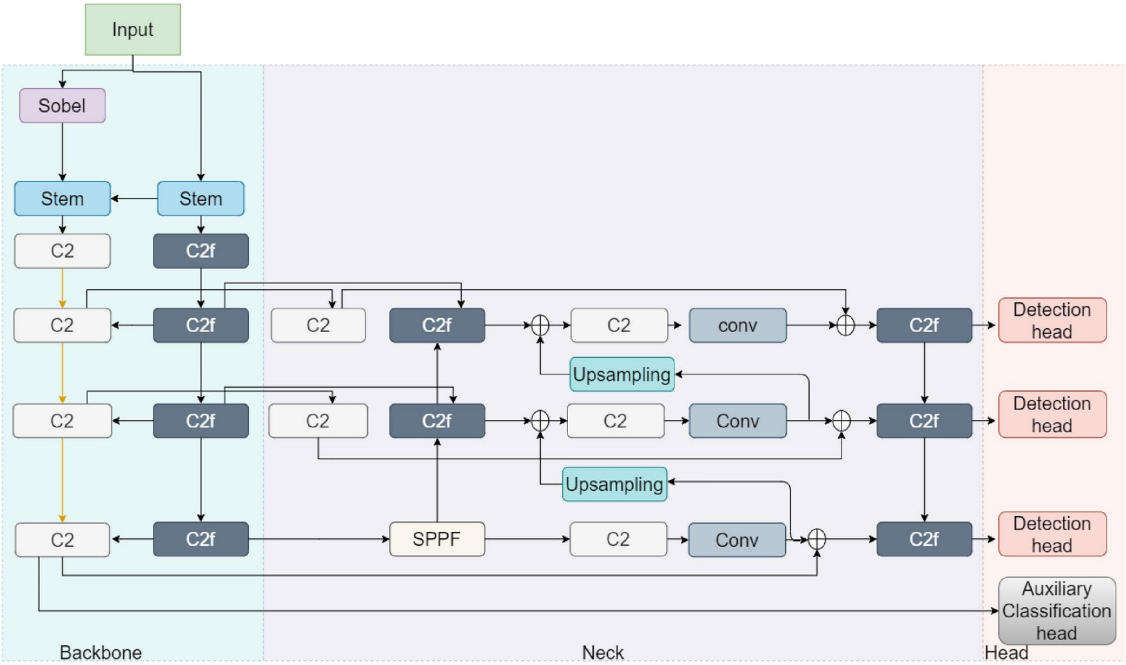

Fig. 2. DBMedDet (The auxiliary classification head of DBMedDet is a multilayer perceptron).

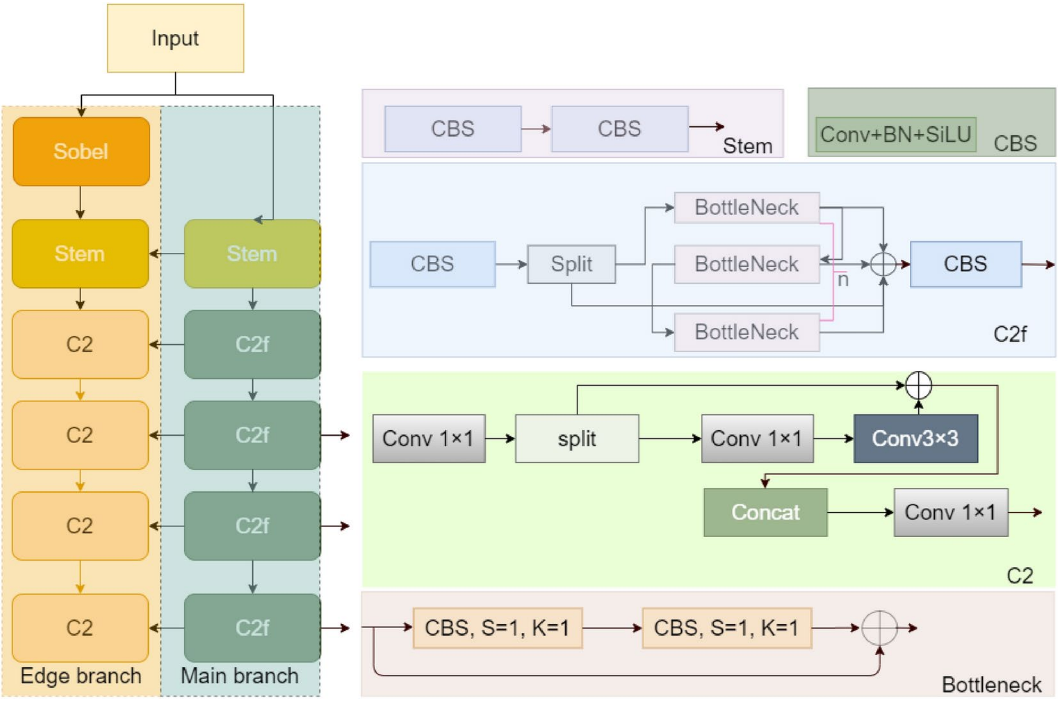

Fig. 3. The backbone network (DBMedDet's backbone network includes YOLOv8's backbone network and an edge branch feature extraction network. In the backbone, there are four C2f modules from top to bottom, where the values of n are 1, 2, 2, and 1, respectively. CBS's parameter K and S means convolution's kenerl and stride. Stem refers to downsampling).

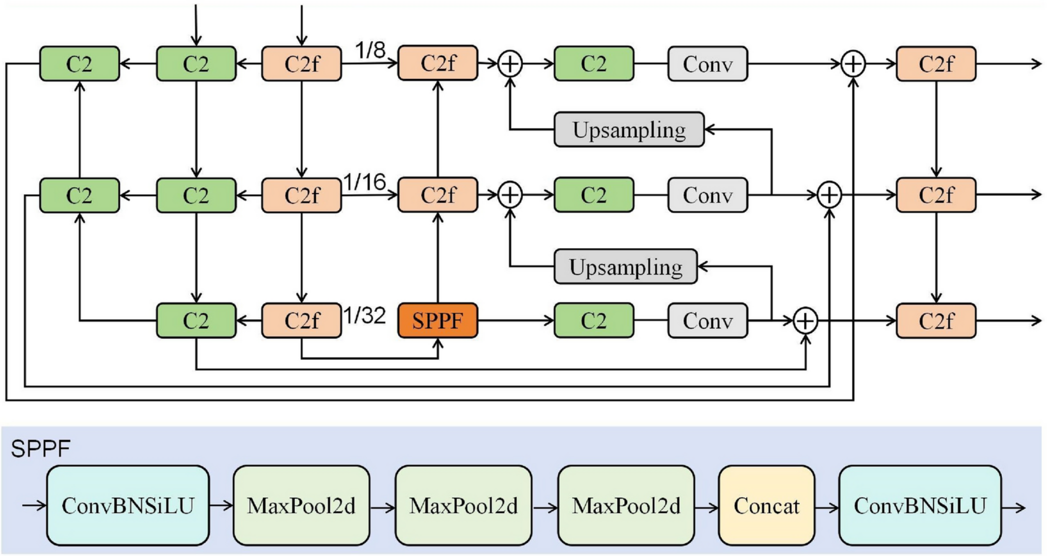

Fig. 4. Recurrent Multi-Direction Network (RMNet), the neck.

Neck

In the Neck sub-network, at the same scale, before each pass through C2, the feature information processed by the main branch is fused with the edge information. This effectively adds additional edge information to the original feature set. The network combines the raw feature information with the edge information to learn features simultaneously, better adapting to medical image detection. The use of the C2 module helps to reduce computational load, as adding edge information alone does not require the model to learn features with excessive parameters. Finally, the fused information is combined with the outputs of the main branch at three scales for a final fusion, then input through a block to the detection head (Fig. 4).

Feature Pyramid Network (FPN) is a framework aimed at enhancing object detection through the utilization of multi-scale feature maps[19]. It tackles the difficulty of identifying objects of varying sizes by constructing a feature pyramid that provides a more effective representation of objects across different scales. Bidirectional Feature Pyramid Network (BiFPN) builds on this idea by incorporating bidirectional feature fusion, which improves feature representation and boosts the efficiency of the model[20]. Meanwhile, Path Aggregation Network (PANet) takes this a step further by concentrating on the aggregation of features across various scales and levels, thereby further refining the feature representation in object detection models[21]. As shown in Fig. 5, RMNet is a powerful architecture that leverages recurrent connections and multi-directional feature flow to improve feature extraction and representation in computer vision tasks. Its focus on contextual awareness, dynamic feature aggregation, and enhanced representation makes it particularly effective for complex object detection and segmentation challenges. By combining the strengths of recurrent and multi-directional approaches, RMNet aims to push the boundaries of performance in various visual recognition applications.

RMNet is a powerful architecture that leverages recurrent connections and multi-directional feature flow to improve feature extraction and representation in computer vision tasks. Its focus on contextual awareness, dynamic feature aggregation, and enhanced representation makes it particularly effective for complex object detection and segmentation challenges. By combining the strengths of recurrent and multi-directional approaches, RMNet aims to push the boundaries of performance in various visual recognition applications.

Head

The detection task in this study involves detecting the abdominal aorta and thoracic aorta (see Fig. 6). To enhance YOLOv8 for this purpose, an auxiliary classification head has been incorporated. This additional classification head consists of global average pooling followed by a softmax layer. When predicting two categories, the last convolutional layer of the feature extraction module generates two feature maps. These feature maps are then processed through global average pooling to produce two 1×1 feature maps. These 1×1 feature maps are fed into the softmax layer, where each output indicates the probability (or confidence) associated with the two categories.

"The Classification Loss (Cls loss)" means Binary Cross-Entropy Loss (BCE Loss). BCE loss is a commonly used loss function for binary classification tasks. The formula is defined as:

$$\text{BCE Loss} = -\frac{1}{N}\sum_{i=1}^{N}\left(y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)\right) \tag{1}$$

In this equation:

- N is the number of samples.
- y_i represents the true label for sample i (0 or 1).
- $\hat{y}_i$ is the predicted probability of the positive class for sample i.

"Bounding Box Loss (Bbox Loss)" is a composite loss function used in object detection tasks. It combines different loss components, including Complete Intersection over Union (CIoU) Loss and Distance Focal Loss (DFU Loss), to enhance the performance of bounding box regression.

Complete Intersection over Union (CIoU) Loss improves traditional IoU metrics by considering additional factors such as the distance between the center points of the predicted and ground truth boxes, the aspect ratio, and the size.

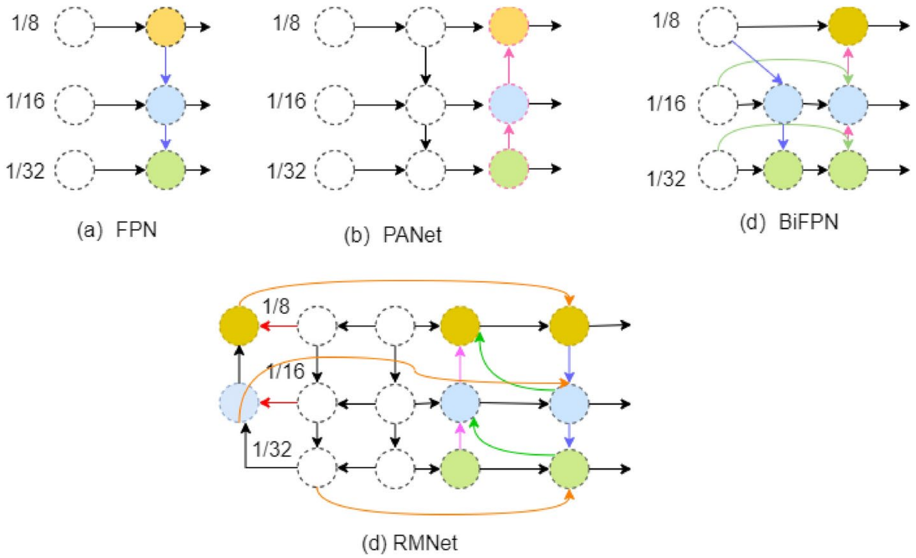

Fig. 5. RMNet and other feature pyramid network.

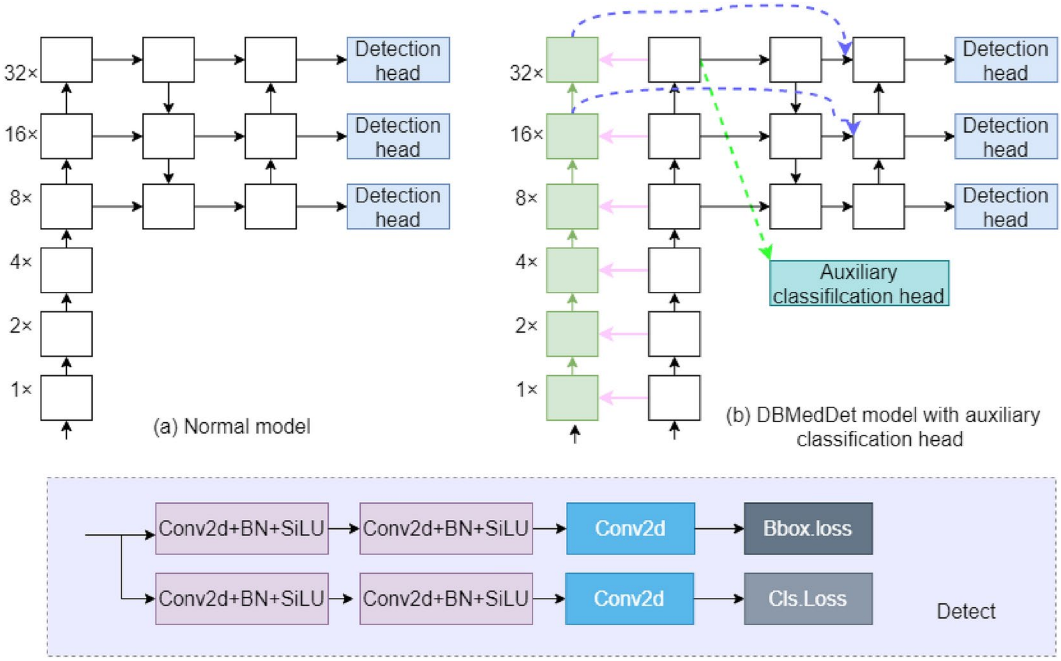

Fig. 6. The head network(The auxiliary classification head of DBMedDet is a multilayer perceptron. General models rely on unidirectional networks for feature extraction, while DBMedDet employs a bidirectional parallel network to extract more enriched features. The auxiliary detection head is added to the last layer of the main backbone network. In DBMedDet, the feature information can flow in four directions?up, down, left, and right?creating favorable conditions for information supplementation and capture).

Intersection over Union (IoU) is a metric used to evaluate the performance of object detection algorithms. It quantifies the accuracy of an object detector by calculating the overlap between the predicted bounding box and the ground truth bounding box.

The formula for IoU is defined as:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{2}$$

Where the Area of Union is calculated as:

$$\text{Area of Union} = \text{Area of A} + \text{Area of B} - \text{Area of Overlap} \tag{3}$$

Thus, the complete IoU formula can be expressed as:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of A} + \text{Area of B} - \text{Area of Overlap}} \tag{4}$$

The CIoU Loss is defined as:

$$\text{CIoU Loss} = 1 - \text{IoU} + \frac{d^2}{c^2} + \alpha v \tag{5}$$

Where:

- IoU is the Intersection over Union.
- d is the Euclidean distance between the centers of the predicted and ground truth boxes.
- c is the diagonal length of the smallest enclosing box.
- v is the aspect ratio consistency.
- α is a positive trade-off parameter.

"Distance Focal Loss (dfl Loss)" is designed to focus on hard-to-regress bounding boxes by down-weighting easy examples. The DFU Loss can be expressed as:

$$\text{dfu Loss} = -\lambda(1-\hat{p})^{\gamma}\log(\hat{p}) \tag{6}$$

Where:

- $\hat{p}$ is the predicted probability of the positive class.
- λ is a balancing factor.
- γ is a focusing parameter that adjusts the rate at which easy examples are down-weighted.

The Bounding Box Loss (Bbox Loss) combines CIoU Loss and dfu Loss as follows:

$$\text{Bblox Loss} = \text{CIoU Loss} + \text{dfu Loss} \tag{7}$$

This combined loss function helps to improve the accuracy of bounding box predictions in object detection tasks by leveraging the strengths of both CIoU and DFU Loss.

Incorporating CIoU Loss and dfu Loss into the Bounding Box Loss allows for more robust training of object detection models, effectively addressing the challenges associated with bounding box regression.

Results

Training details

In DBMedDet training, the batch size was 4, the epoch was 100, the image size was 640×640, the initial learning rate was 0.01, the floating learning rate was 0.01, the momentum was 0.937, the optimizer weight decay was 5e-4, the warmup epochs was 3.0, the warmup initial momentum was 0.8, and the warmup initial bias learning rate was 0.1. In actual deployment, it is recommended to configure the hardware environment with an NVIDIA RTX 4090, 16 vCPUs from the Intel(R) Xeon(R) Platinum 8352V CPU @ 2.10GHz, and 120GB of memory.

Evation metrics

Precision assesses the ratio of true positive samples to the total number of samples predicted as positive by the model. The formula for precision is:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{8}$$

In this formula, True Positives (TPs) denote the number of samples that are accurately identified as positive, while False Positives (FPs) refer to samples that are mistakenly classified as positive despite being negative. A high precision level indicates the model's accuracy and reliability in distinguishing positive samples.

Recall, on the other hand, measures the proportion of actual positive samples that the model successfully identifies, reflecting its ability to recognize true positives. The formula for calculating recall is:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

In object detection, the F1 score is an important performance evaluation metric that effectively measures the model's ability to recognize samples across different categories. By analyzing the F1 score, the model can be optimized to enhance its performance in practical applications. The formula is as follows:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

The mean Average Precision (mAP) is calculated as follows:

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^{N} \text{AP}_i \tag{11}$$

Where N is the number of classes, and AP_i is the Average Precision for class i, calculated as:

$$\text{AP}_i = \int_0^1 P(i)\, dR(i) \tag{12}$$

Where $P(i)$ is the precision at a given recall level, and $R(i)$ is the recall level. The Floating Point Operations (FLOPs) typically represents the number of floating-point operations required by the model during inference. For a convolutional layer. Frames Per Second (FPS) represents the number of frames processed per second and is commonly used to measure the speed of video processing or real-time inference. FLOPs and FPS can be defined as follows:

$$\text{FLOPs} = 2 \times C_{\text{in}} \times H_{\text{out}} \times W_{\text{out}} \times K_h \times K_w \times C_{\text{out}} \tag{13}$$

Where:

- C_{in}: Number of input channels.
- H_{out}: Height of the output feature map.
- W_{out}: Width of the output feature map.
- K_h: Height of the convolution kernel.
- K_w: Width of the convolution kernel.

- C_{out}: Number of output channels.

$$\text{FPS} = \frac{\text{Total Frames}}{\text{Total Time (seconds)}} \tag{14}$$

Where:

- Total Frames: Total number of frames processed.
- Total Time: Total time taken to process these frames (in seconds).

Comparison of state-of-the-art methods

We compared DBMedDet with advanced algorithms such as Oriented-RCNN, R3Det, ROI Transformer, Rotated Faster R-CNN, Rotated RepPoints, R3Det-Tiny, and found that the proposed algorithm achieved the highest mAP(Table 1, the fourth validation).

DBMedDet has 3.6M fewer parameters than YOLOv8m, which means that DBMedDet is a more lightweight model compared to YOLOv8m, helping to reduce storage requirements and inference time. The computational complexity (FLOPs) of DBMedDet is 8.3M lower than that of YOLOv8m. The lower FLOPs indicate reduced computational resource consumption during inference, making DBMedDet more suitable for operation on resource-constrained devices. In terms of detection accuracy, DBMedDet achieves an mAP@0.5 that is 4.6% higher than YOLOv8m. This indicates that DBMedDet is better at detecting objects, enhancing the model's effectiveness and reliability.

Overall, DBMedDet outperforms YOLOv8m in terms of parameter count, computational complexity, and detection accuracy. Although it falls slightly short in frame rate, it still maintains real-time performance. This result indicates that DBMedDet is an effective and efficient object detection model suitable for various applications.

Ablation experiments and analysis

We investigated the impacts of multi-scale detection, data augmentation, pre-trained weights, and various convolutional kernels on DBMedDet. Implementing multi-scale detection enhances the model's ability to identify objects more effectively. Additionally, both the box loss and classification loss are lower than those of the other compared models.

Multi-scale impact

Using multi-scale training in DBMedDet experiments can significantly improve the model's detection accuracy and generalization ability, although it may increase the demand for computational resources. Conversely, not using multi-scale training may limit the model's performance in practical applications. Therefore, the decision to use multi-scale training should be based on the specific application scenarios and resource conditions.

As shown in Table 2, after using multi-scale training, the mAP@0.5 increased by 12.3%, the mAP@0.5:0.95 improved by 11.6%, and the F1 score increased by 1.9%.

	map@0.5					
Method	**avg**	**min**	**max**	**Param**	**FLOPs**	**FPS**
YOlOv8l[18]	0.792	0.704	0.930	44.5	169.1	133
YOLOv8m[18]	0.795	0.719	0.923	26.4	81.2	154
YOLOv8s[18]	0.809	0.735	0.915	11.4	29.6	196
YOLOv8n[18]	0.779	0.706	0.958	3.1	8.4	204
YOLOv10s	0.780	0.711	0.919	8.8	27.8	169
YOLOv9s	0.802	0.725	0.876	7.5	28.3	84
YOLOv7tiny[22]	0.811	0.742	0.942	8.4	22.3	189
YOLOv5s[23]	0.791	0.733	0.932	9.4	25.0	189
Oriented R-CNN[24]	0.433	0.252	0.635	41.1	211.4	11
Rotated Reppoints[25]	0.329	0.110	0.702	36.6	194.2	9
R3Det-tiny[26]	0.400	0.158	0.906	36.9	225.6	11
ROI Transformer[27]	0.518	0.363	0.699	55.0	225.2	9
Rotated Faster R-CNN[28]	0.444	0.297	0.672	41.1	211.3	11
R3Det[26]	0.269	0.171	0.387	41.6	328.7	9
DBMedDet (Ours)	0.841	0.762	0.967	22.8	72.9	127

Table 1. Performance comparison of object detection models. [a]mAP@0.5 (mean Average Precision at IoU = 0.5) is a metric used to evaluate the performance of object detection. It represents the average precision (AP) of the model across different classes when the IoU threshold is set to 0.5. [b]Rotated RepPoints, R3Det, R3Det-tiny, ROI Transformer, and Rotated Faster R-CNN have all not achieved real-time performance. DBMedDet can realize real-time detection

Method	map@0.5	map@0.5:0.95	F1
Multi-scale	0.841	0.469	76.0
No multi-scale	0.718	0.353	74.1

Table 2. Comparative experiment using multi-scale input training. The term “mAP@0.5:0.95” refers to the mAP calculated over multiple IoU thresholds ranging from 0.5 to 0.95, typically in increments of 0.05. This metric is commonly used to evaluate the performance of object detection models, providing a more comprehensive measure of performance across varying levels of detection precision

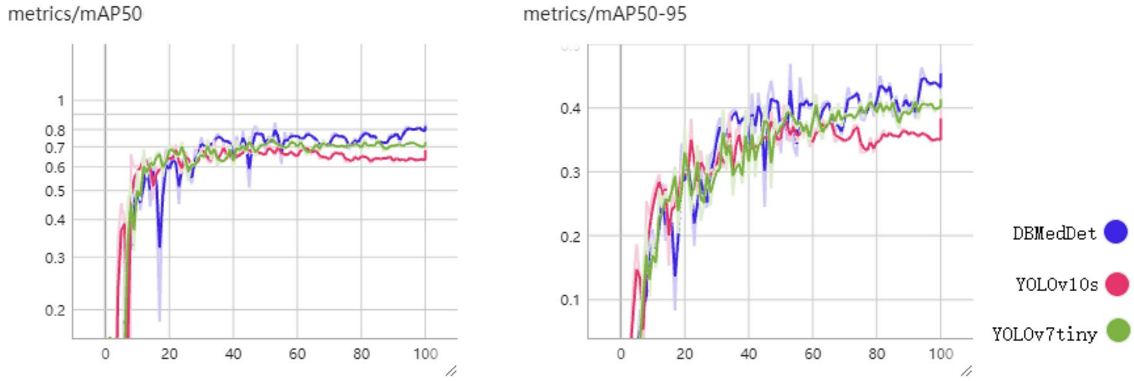

Fig. 7. The visualization of multi-scale impact (mAP50 means mAP@0.5 and mAP50-95 means mAP@0.5:0.95).

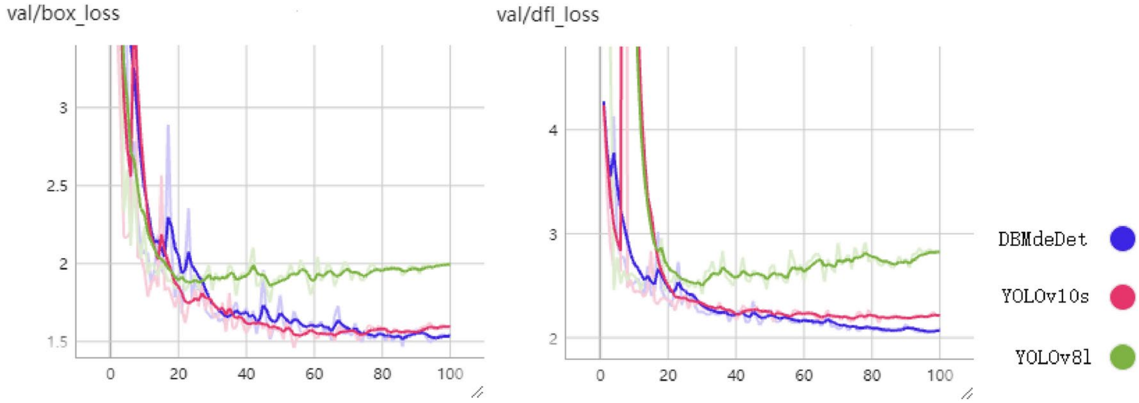

Fig. 8. The visualization of multi-scale impact (Bbox loss and Cls loss).

Additionally, we provided visualizations of the mAP curves during training, offering a multi-dimensional comparison of the results when using multi-scale training versus not using it(Fig. 7).

From the training results of various models, it can be observed that under multi-scale training conditions, the proposed model achieves higher values for mAP@0.5 and mAP@0.5:0.95 compared to the baseline models. This indicates that multi-scale training contributes positively to improving the model’s prediction accuracy. Here mAP50 means mAP@0.5, mAP50-95 means mAP0.5:0.95.

Figure 8 provides visualizations of the loss curves during the fourth validation, specifically for “Bbox loss (box_loss)” and “dfl loss (dfl_loss)”. The results show that DBMedDet exhibits lower regression loss (Bbox loss) compared to other models for the last few iterations, and its dfl loss curve also demonstrates relatively low values.

The edge detection operator

This paper selects the Sobel operator for edge feature extraction. In the experiments, three operators such as Sobel, Canny, and Laplacian were compared, with Sobel ultimately being chosen. The Sobel operator is a convolution-based edge detection algorithm that identifies edges by calculating the gradient of the image. It performs well on images with low noise due to its straightforward calculations and yielded favorable results in the experiments.

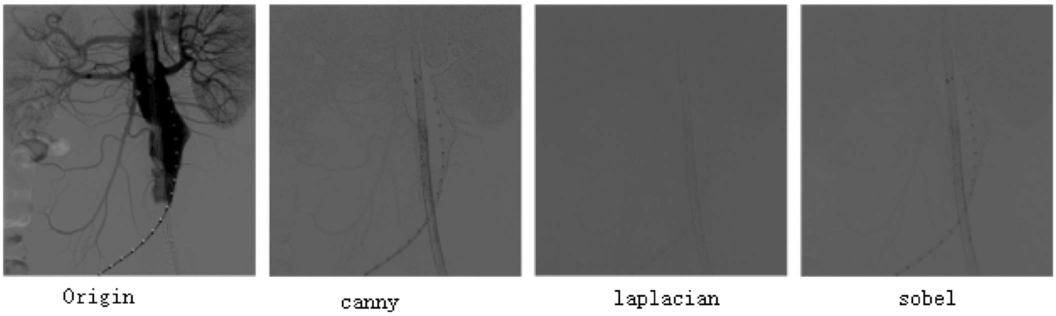

Fig. 9. The visualization of multi-scale impact.

Method	map@0.5	map@0.5:0.95	F1	FPS
Canny	0.843	0.469	76.7	103
Laplacian	0.723	0.401	80.5	130
Sobel (Ours)	0.841	0.469	76.0	127

Table 3. Performance comparison of different methods. When using the Canny edge detection algorithm, the performance metrics mAP@0.5 and mAP@0.5:0.95 were comparable to those of the Sobel operator; however, the processing speed decreased. Therefore, we opted to use the Sobel operator

Method		map@0.5	map@0.5:0.95	F1	Param	FLOPs	FPS
✓		0.729	0.417	72.9	13.8	34.7	193
	✓	0.724	0.418	74.7	13.8	34.7	190
✓	✓	0.841	0.469	76.0	22.8	72.9	127

Table 4. Performance comparison of different branch configurations. The main branch refers to the backbone network included in YOLOv8, while the edge branch refers to the network added in the dual-branch network with C2 and Sobel modules. The parallel dual-branch edge network performs better

The Canny edge detection algorithm is a multi-stage process that includes Gaussian filtering, gradient calculation, non-maximum suppression, and double threshold detection. It accurately detects edges in images and exhibits a degree of robustness to noise. Although Canny outperforms Sobel slightly, its calculations are more complex because of operations like non-maximum suppression.

The Laplacian operator detects edges by calculating the second derivative of the image and is effective for highlighting detailed edges. However, in the experiments, the Laplacian primarily accentuated certain edges and struggled to extract comprehensive edge information. The edge detection results of the various operators are illustrated in Fig. 9 and Table 3.

The dual branches

For the proposed model, we compared the impact of using different branches on the model(see Table 4). Compared to the single branch backbone network, the proposed DBMedDet network with its dual branch architecture achieves the highest mAP, improving performance by at least 11%. However, it should be noted that the FPS is lower than that of the single branch network.

From Fig. 10, it can be seen that when the dual-branch structure is not used, the attention of the neural network is focused on the first half of the detection (main branch). The edge branch enhances the details of the image. After the main branch and edge branch are fused, the network's attention becomes highly concentrated on the detection part. This can be observed in the "fusion" column, where the detection area receives focused attention in the dual-branch network.

The different convolutional kernels

Testing various convolutional kernel sizes (3×3, 5×5, 7×7, and 9×9) and types can affect the model's capacity to capture different features at various levels of abstraction, which in turn impacts the overall detection performance of DBMedDet (see Table 5). Ultimately, the model chose to use 3×3 convolutional kernels.

When assessing metrics such as mAP@50, mAP@0.5:0.95, the 3×3 kernel exhibits outstanding performance (see Fig. 11). Moreover, the loss curves indicate that the 3×3 kernel leads to lower loss values (refer to Fig. 12).

When the kernel size is 3×3, the values of Bbox loss, Cls loss, and dfl loss are all at their lowest, indicating that this configuration is well-suited for the model.

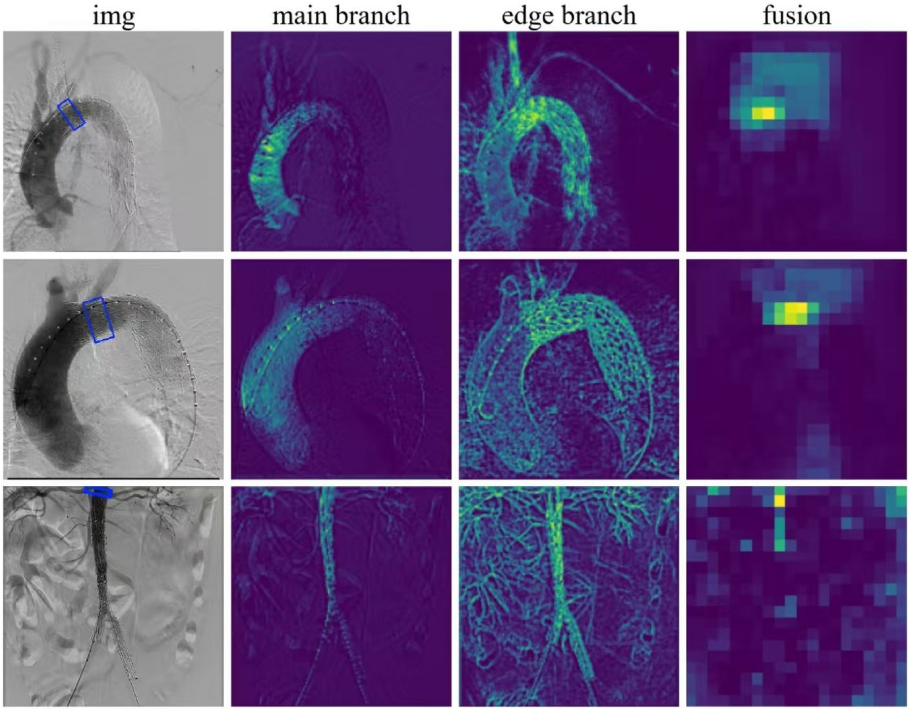

Fig. 10. The dual branch visulization.

Kernel size	map@0.5	map@0.5:0.95	F1	Param	FLOPs	FPS
3×3 (Ours)	0.822	0.443	75.3	22.8	72.9	127
5×5	0.792	0.359	75.2	33.4	106.5	95
7×7	0.709	0.349	73.7	49.5	156.8	95
9×9	0.719	0.357	73.8	70.9	223.9	121

Table 5. Performance comparison of different kernel sizes. *k* refers to the parameters within the convolutional kernels. As the size of the convolutional kernel increases, the F1 score gradually decreases, while the number of parameters and FLOPs also increase, resulting in slower processing speeds. When the convolutional kernel is 3×3, the model achieves the highest mAP, the fastest speed, the maximum F1 score, and the lowest number of parameters. Here, the mAP is not 0.841 because pre-trained weights were not used in this training

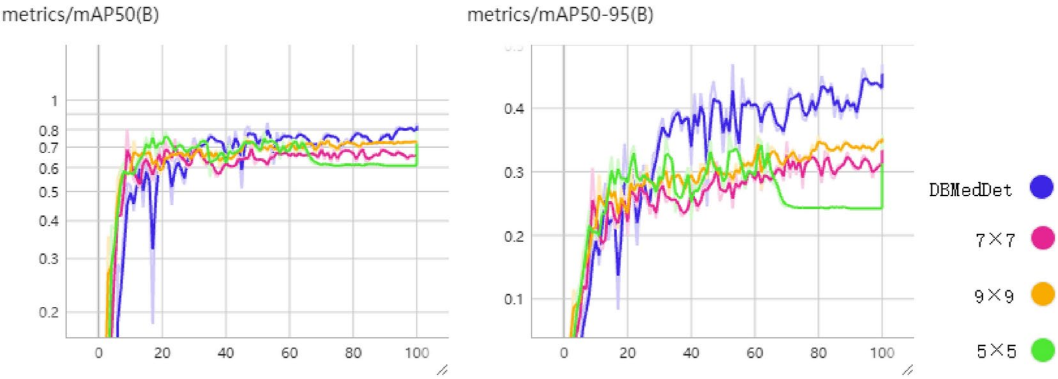

Fig. 11. The visulization of different kernel size impact(mAP@0.5,mAP@0.5:0.95).

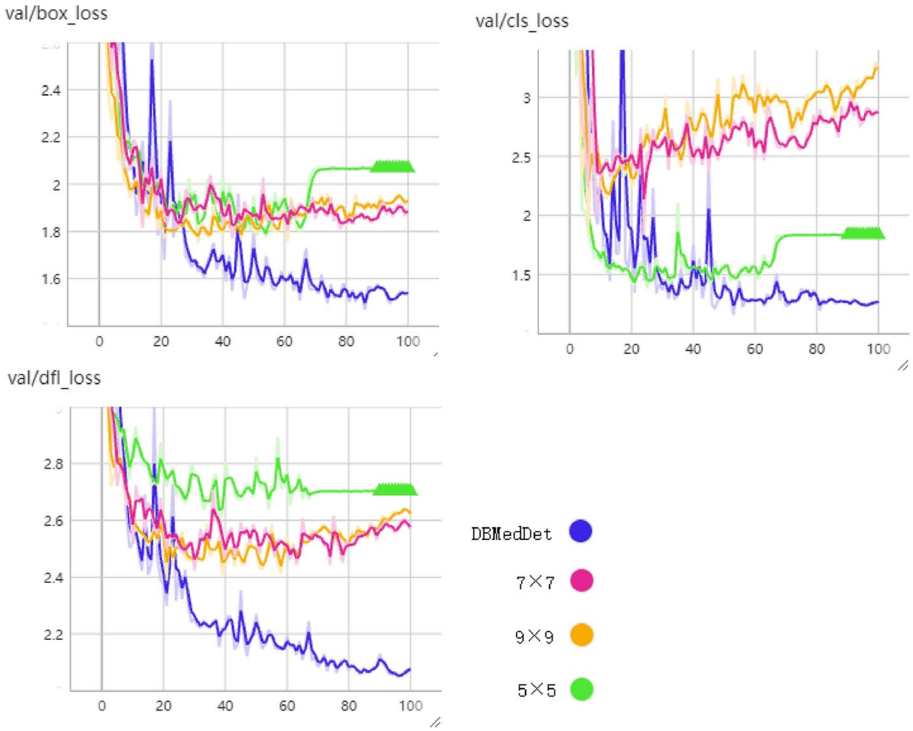

Fig. 12. The visulization of different kernel size impact(Bbox loss, Cls loss and dfl loss).

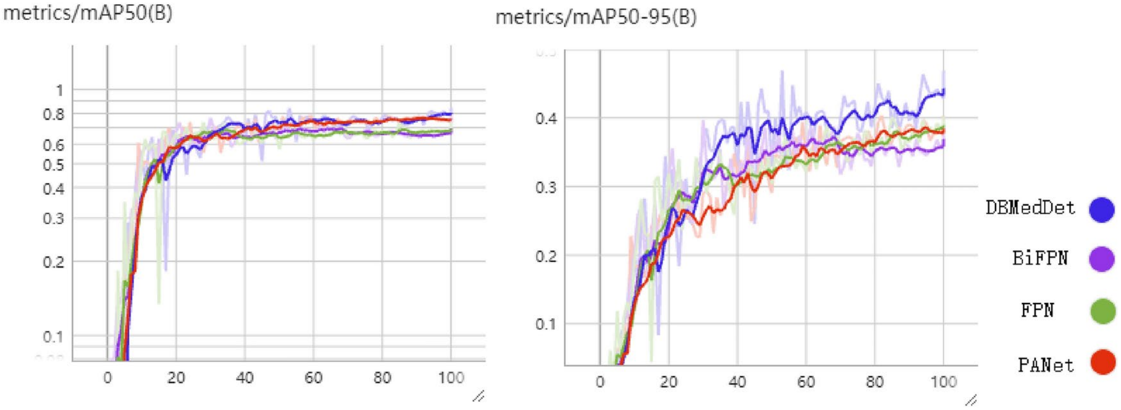

Fig. 13. The visulization curves of different feature pyramid networks impact(mAP@0.5,mAP@0.5:0.95).

The different feature pyramid neworks

We introduced a new feature pyramid network called RMNet and compared its performance against FPN, BiFPN, and PAN(PANet), focusing on mAP changes during training. Our results show that ReMPN outperforms the other feature pyramid models in terms of mAP@0.5, mAP@0.5:0.95(Fig. 13). Consequently, DBMedDet achieved the highest mAP and F1 score.

From Table 6, it can be seen that when the proposed RMNet is used as the neck network for DBMedDet, both mAP and F1 score reach their highest values. However, this improvement comes at the cost of increased parameter count, with a corresponding decrease in FLOPs and FPS, resulting in slower processing speed.

In the mAP@50 curve, DBMedDet performed exceptionally well in the last few epochs. In the mAP@50:95 curve, DBMedDet showed outstanding performance after the first 20 epochs, outperforming other pyramid networks.

The auxiliary classification head

In the supportive surgery, the areas designated for stent installation encompass the thoracic aorta and abdominal aorta. The experiment compared the effects of incorporating an auxiliary classification head versus not including it, as illustrated in Table 7. M1 refers to the basic YOLOv8s model, while M2 denotes YOLOv8s enhanced with

Method	map@0.5	map@0.5:0.95	F1	Param	FLOPs	FPS
FPN	0.734	0.402	75.3	13.1	39.6	166
PANet	0.785	0.404	77.7	20.4	67.8	160
BiFPN	0.777	0.398	73.8	20.4	67.8	159
RMNet (Ours)	0.841	0.469	76.0	22.8	72.9	127

Table 6. Performance comparison of different methods.

Method	map@0.5	map@0.5:0.95	F1	Param	FLOPs	FPS
M1	0.756	0.449	76.1	11.4	29.6	196
M2	0.770	0.449	72.1	13.8	34.7	193
M3	0.832	0.465	75.3	25.2	78.1	127
M4	0.841	0.469	76.0	22.8	72.9	127

Table 7. Step-by-step ablation study.

Mosaic	map@0.5	map@0.5:0.95	F1
✓	0.578	0.255	67.2
×	0.841	0.469	76.0

Table 8. Comparative experiment using data augmentation.

a dual-branch parallel network. M3 builds upon M2 by improving the neck using the proposed ReMPN, and M4 incorporates an auxiliary classification head on top of M3, also known as DBMedDet. The proposed model achieved the highest mAP, reaching 84.1% after adding the auxiliary classification head, enhancing the backbone network, and refining the neck. The introduction of the classification head contributed to a 1.2% increase in the model's detection mAP.

Data augmentation

Data augmentation techniques can improve a model's generalization and robustness. However, our experiments revealed that applying mosaic data augmentation did not enhance the model's performance, leading us to forgo its implementation (see Table 8).

Data augmentation is a commonly used technique aimed at improving the generalization ability of machine learning models, especially when training samples are insufficient. Although data augmentation can effectively enhance model performance in many cases, it can sometimes lead to worse results. This may be due to several reasons: certain augmentation methods (such as random cropping, blurring, etc.) may introduce noise, causing the model to misjudge critical features. Additionally, if the model is tasked with recognizing small objects, excessive rotation or cropping may result in the loss of important features in the images, thus affecting the model's recognition capability.

In all curves(Fig. 14), including mAP, precision, and recall values decreased after data augmentation. We speculate that this is due to the detection targets being too small, leading to target loss caused by excessive cropping ,blurring or rotation.

The pre-trained weight

Leveraging pre-trained weights from models trained on large datasets can serve as an effective foundation for training on specific tasks, such as DBMedDet, potentially accelerating convergence and enhancing performance. Utilizing a COCO pre-trained model facilitated faster convergence and led to improved detection performance (see Table 9).

The detection result

In Fig. 15, we compare the detection results of the proposed algorithm DBMedDET with those of YOLOv7tiny and YOLOv8s. The results indicate that DBMedDET outperforms the other algorithms, which tend to produce more false positives (incorrectly detecting bracket installations) and false negatives (failing to detect installations that exceed bracket limits).

For the first image, YOLOv8s fails to detect the position of the support bracket. For the second image, both YOLOv8s and YOLOv7tiny have inaccuracies in their detected positions, which are slightly skewed. In the third image, YOLOv8s and YOLOv7tiny detect positions that are too low. In the fourth image, there is not much difference in the detection results among the three algorithms.

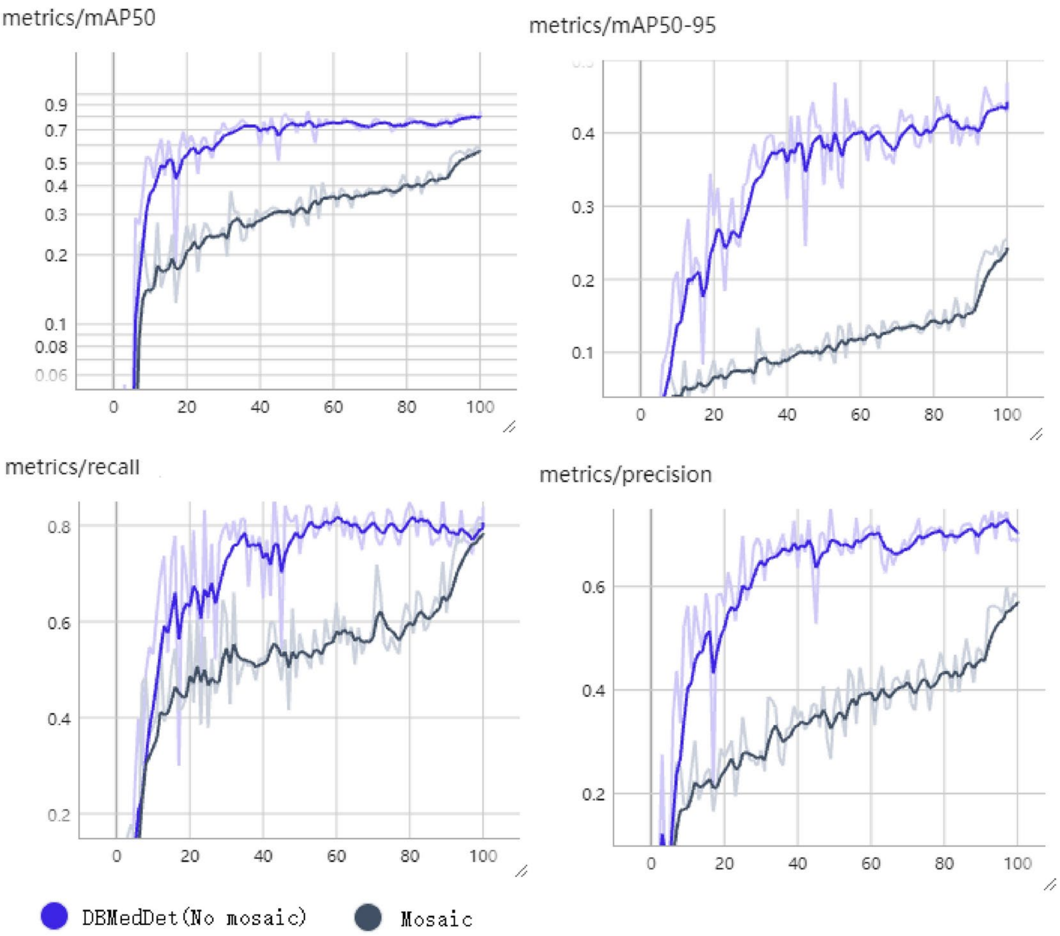

Fig. 14. The visulization of data augmentation impact(mAP50, mAP50-95, recall and precision).

Method	map@0.5	map@0.5:0.95	F1
✓	0.841	0.469	76.0
×	0.822	0.443	75.3

Table 9. Comparison experiment on using the COCO2017 pre-trained model.

Discussion

In recent years, the endovascular therapy has been increasingly utilized in aortic disease treatment, both thoracic and abdominal, precise positioning of the stent and fully isolation of the lesions are key points. Aorta possess numerous branches which transport oxygenated blood to multiple organs, however some of them are susceptible to be covered during the surgery due to limited landing zone (refer to the distance between lesions and branches). A partial or full branch occlusion, with reported incidence ranges from 0.1–3.5%, could be detrimental. Such as left subclavian artery (LSA) coverage in thoracic endovascular repair (TEVAR) may lead to left arm ischemia, subclavian steal syndrome or stroke, while coverage of the renal artery or internal iliac artery in abdominal endovascular repair (EVAR) may result in renal ischemia or lower extremity claudication.

Preoperative 3-dimensional computed tomography angiography (CTA) is always helpful to identify the branches and lesions. However, intraoperative images are 2-dimentional, which are misleading on some occasions due to the vessel tortuosity or angulation. Artificial intelligence (AI) and machine learning (ML) has shown promising result in preoperative diagnosis and prognostication of aortic disease, while intraoperative guidance is less investigated. An AI tool, which helps to identify real-time position of the branches and lower the risk of unintentional coverage will greatly benefit the patients. By quick idenfication and shortening exposure time, the tool protects the surgeon as well. DBMedDet offers several advantages, including a high mean Average Precision (mAP) of 0.841 (mAP@0.5) and impressive real-time performance with a frame rate of 127 FPS. By integrating original feature information with edge data, the network simultaneously learns two types of features, enhancing its suitability for medical image detection. In the Neck subnetwork, feature information processed by the main branch is fused before each pass through C2 at the same scale. This approach effectively incorporates

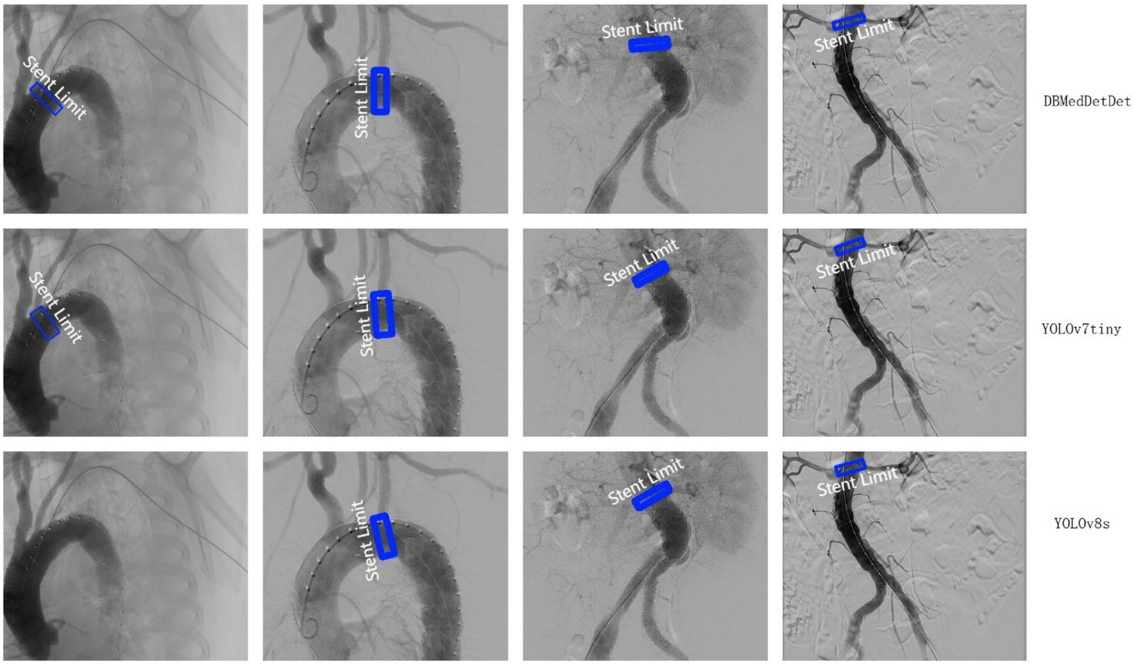

Fig. 15. The detection result(DBMedDet, YOLOv7tiny and YOLOv8s.

additional edge information into the original features. Utilizing C2 helps to minimize computational load, as integrating edge information does not necessitate the model learning an excessive number of parameters.

The fused information is then combined with the outputs from the main branch at three different scales, which is subsequently passed through a block to the detection head. Additionally, we introduced an auxiliary classification head to the model, resulting in a 1.2% increase in detection mAP.

DBMedDet extracts features from different scales and angles through two independent branches, enabling the network to capture the diversity of various anatomical structures and lesions. At the same time, using different training datasets for each branch enhances the network's robustness to various anatomical variations and artifacts. One branch of the network focuses on classification tasks (such as predicting the type of stent), while the other branch concentrates on regression tasks (such as locating the position of the stent), which helps the model to comprehensively consider different features and information. DBMedDet combines features from different levels and branches through bidirectional fusion, enhancing the model's ability to understand complex scenarios. During the feature fusion process, the introduction of contextual information can improve the model's performance in the presence of anatomical variations and image artifacts.

In contrast, algorithms such as Oriented-RCNN, R3Det, ROI Transformer, Rotated Faster RCNN, Rotated RepPoints, and R3Det-Tiny all have frame rates below 15 FPS. This limitation is primarily due to the challenges presented by small objects, which often have fewer pixels and less distinct features, making them difficult for detection models to identify accurately.

Conclusion

The primary contribution of this article is the introduction of DBMedDet, a deep learning model designed to provide real-time constraint cues for the optimal placement of stents during implantation surgeries. This model features a parallel dual-branch edge feature extraction network, a bidirectional feedback feature fusion Neck sub-network, along with a position detection head and a classification head specifically tailored for thoracic and abdominal aortic stents.

For edge feature extraction, we employed the Sobel operator. We conducted comparisons of mean Average Precision (mAP at various thresholds: mAP@0.5 and mAP@0.5:0.95), precision, and recall between several models, including YOLOv8s_obb, YOLOv8m_obb, YOLOv8l_obb, YOLOv8n_obb, and a single edge branch algorithm, which integrates our single-branch network with YOLOv8's backbone. Additionally, we introduced a novel feature pyramid network, RMNet, and evaluated its performance against FPN, BiFPN, and PAN in terms of mAP and loss curve trends during training. Our findings indicate that RMNet surpasses other feature pyramid models in mAP@0.5, mAP@0.5:0.95, precision, and recall.

Moreover, we enhanced the proposed model by incorporating an auxiliary classification head, refining the backbone network, and improving the Neck structure, resulting in a detection accuracy of 84.1%. A comparison of DBMedDet with advanced algorithms such as Oriented-RCNN, R3Det, ROI Transformer, Rotated Faster RCNN, Rotated RepPoints, and R3Det-tiny revealed that our proposed algorithm achieved the highest mAP among these models. In the various experiments conducted with mmrotate, taking R3Det-tiny as an example, the best mAP@0.5 achieved in the 5-fold cross-validation experiment reached 0.906, while the worst result was only 0.158. This discrepancy is quite significant and is related to the data split operation in mmrotate. The dataset for the 5-fold cross-validation is established after the data split, which contributes to the differences in results. In our future work, we will focus on improving the variability in the dataset. It is essential to ensure that the training dataset encompasses patients of varying ages, genders, races, and health conditions, enabling the model to adapt to a wide range of patient characteristics. Additionally, the dataset should be modified to include samples from different anatomical structures and lesion types to strengthen the model's ability to handle anatomical variations. When considering imaging conditions, it is important to account for varying levels of noise, image quality, and different types of artifacts to ensure the model performs effectively in real-world clinical settings. To address model complexity and overfitting, it is crucial to monitor the complexity of the model while applying regularization techniques and data augmentation methods to enhance its generalization capabilities. Furthermore, gathering clinical feedback during actual deployment is vital, allowing for fine-tuning and improvements based on new data and user input to boost the model's adaptability and accuracy.

Small objects typically occupy a smaller area in images, and feature extraction networks in algorithms like Oriented-RCNN and R3Det may struggle to capture sufficient detail, leading to reduced detection accuracy for small targets. Scale invariance is an important issue in object detection tasks. The design of multi-scale feature fusion in algorithms like Oriented-RCNN and R3Det may not fully enhance the representation capability for small objects, especially when processing small targets at different scales, which can result in information loss. Additionally, algorithms like Oriented-RCNN and R3Det may perform worse in bounding box regression for small objects compared to larger ones, particularly in cases where the objects have complex shapes or overlap. Additionally, the results using the smaller network outperform those obtained with a larger network as the backbone. Therefore, we will concentrate on studying the detection performance of smaller networks for small objects.

Data availability

For any questions regarding the study or to request access to the data, please contact the corresponding author: [Chengcheng Guo, netccg@whu.eud.cn].

Received: 13 September 2024; Accepted: 9 January 2025
Published online: 28 March 2025

References

1. Zhao, J., Zhang, J. & Liu, Y. Outcomes of covered stenting for aortoiliac occlusive disease: A systematic review and meta-analysis. *J. Vasc. Surg.* **75**(3), 1026–1036 (2022).
2. Wang, H., Li, X. & Chen, G. Advances in imaging techniques for endovascular aneurysm repair: A review. *Eur. J. Vasc. Endovasc. Surg.* **65**(1), 15–24 (2023).
3. Cañero, C. et al. Optimal stent implantation: Three-dimensional evaluation of the mutual position of stent and vessel via intracoronary echocardiography, 261–264. https://doi.org/10.1109/CIC.1999.825956 (1999).
4. Dijkstra, J., Koning, G., Tuinenburg, J.C., Oemrawsingh, P.V., & Reiber, J. Automatic border detection in intravascular iltrasound images for quantitative measurements of the vessel, lumen and stent parameters. In *Computers in Cardiology 2001*. Vol. 28 (Cat. No. 01CH37287), 25–28 (IEEE, 2001).
5. Rotger, D., Radeva, P. & Bruining, N. Automatic detection of bioabsorbable coronary stents in ivus images using a cascade of classifiers. *IEEE Trans. Inf. Technol. Biomed.* **14**(2), 535–537 (2009).
6. Hua, R. et al. Stent strut detection by classifying a wide set of ivus features. In *MICCAI Workshop on Computer Assisted Stenting*, 130–137 (2012)
7. Ciompi, F. et al. Computer-aided detection of intracoronary stent in intravascular ultrasound sequences. *Med. Phys.* **43**(10), 5616–5625 (2016).
8. Wang, A. et al. Automatic stent strut detection in intravascular optical coherence tomographic pullback runs. *Int. J. Cardiovasc. Imaging* **29**, 29–38 (2013).
9. Lu, H. *et al.* Automatic stent strut detection in intravascular oct images using image processing and classification technique. In *Medical Imaging 2013: Computer-Aided Diagnosis*, Vol. 8670, 295–302 (SPIE, 2013).
10. Jiang, X. et al. Automatic detection of coronary metallic stent struts based on yolov3 and r-fcn. *Comput. Math. Methods Med.* **2020**(1), 1793517 (2020).
11. Yao, Z. et al. Imagetbad: A 3d computed tomography angiography image dataset for automatic segmentation of type-b aortic dissection. *Front. Physiol.* **12**, 732711 (2021).
12. Shalhub, S. et al. Characterization of syndromic, nonsyndromic familial, and sporadic type b aortic dissection. *J. Vasc. Surg.* **73**(6), 1906–1914 (2021).
13. Gessert, N. et al. Bioresorbable scaffold visualization in ivoct images using cnns and weakly supervised localization. *Image Process.* **10949**, 1094 (2019).
14. Zubatiuk, T. & Isayev, O. Development of multimodal machine learning potentials: Toward a physics-aware artificial intelligence. *Acc. Chem. Res.* **54**(7), 1575–1585 (2021).
15. Ramachandran, S., George, J., Skaria, S. & Varun, V. V. Using yolo based deep learning network for real time detection and localization of lung nodules from low dose ct scans **10575**, 1–9 (2018).
16. Tan, L., Huangfu, T., Wu, L. & Chen, W. Comparison of retinanet, ssd, and yolo v3 for real-time pill identification. *BMC Med. Inform. Decis. Mak.* **21**(1), 1–11 (2021).
17. Hao, S., Zhao, Q. & He, Y. T. Yolo-msfr: Real-time natural disaster victim detection based on improved yolov5 network. *J. Real-Time Image Process.* **21**(1), 7–1712 (2024).
18. Wang, C.-Y. et al. Yolov8: The next generation of object detection. arXiv preprint arXiv:2307.00747 (2023)
19. Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2117–2125 (2017).
20. Tan, M., Pang, R., & Le, Q.V. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 13029–13038 (2020)
21. Liu, S., Qi, L., Hu, H., Wang, Z., & Zhang, Z. Path aggregation network for instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 8759–8768 (2018).
22. Wang, Z. et al. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696 (2022)
23. Jocher, G. D. A., et al.: Yolov5: A high-performance object detection model. GitHub repository (2020)

24. Yang, Y., Li, Z., Huo, Z., Yang, J., Zhang, Y., & Wang, Q. Oriented r-cnn for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11244–11253 (IEEE, 2021).
25. Yang, Z., Zhang, P., Liu, W., Wang, J., & Zhang, Z. Rotated reppoints for arbitrarily-oriented object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 14301–14310 (2021).
26. Yang, Z., Zhang, P., Liu, W., Wang, J., & Zhang, Z. R3det: Refined rotated region proposal network for arbitrarily-oriented object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10089–10098 (2020).
27. Dai, J., Li, Y., He, K., & Sun, J. Roi transformer for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1020–1029 (IEEE, 2021).
28. Li, Z., Zhang, W., Wang, J., & Zhang, Z. Rotated faster r-cnn for arbitrarily-oriented object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32 (2018).

Acknowledgements

This research was funded by the establishment of 5G+ Artificial intelligence remote treatment and diagnosis platform for major aortic diseases (Grant Number: 2022BCA035) and the doctoral foundation of Wuhan Business University (Grant Number:2023KB001).

Author contributions

Conceptualization, Sxw and Xyl.; methodology, Xyl.; validation, Sxw and Xyl.; formal analysis, Sxw and Ccg.; investigation, Sxw and Xyl.; data curation, Sxw, Pl and Xyl.; writing—original draft preparation, Sxw.; writing—review and editing, Sxw and Xyl.; visualization, Xyl.; supervision, Ccg.; project administration, Ccg.; funding acquisition, Ccg and Zww. All authors have read and agreed to the published version of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Ethics committee approval

In this study, the publication of all identifying information and images involving human participants has been approved with written informed consent from the participants and their legal guardians. The study has been approved by the Ethics Committee of Renmin Hospital of Wuhan University.

Additional information

Correspondence and requests for materials should be addressed to C.G. or X.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.