DOI: 10.53469/wjimt.2025.08(05).18

HFS-YOLO11 Based Target Detection Algorithm for Thyroid Nodules

Lisha Wang, Fen Liu

School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin, China

Abstract: ***Aiming to address the accuracy bottleneck of the YOLO11 algorithm in thyroid nodule detection, this paper proposes an efficient and accurate haar field spatial-you only look once 11 (HFS-YOLO11) network. The network innovatively incorporates a multi-scale spatial pyramid attention (MSPA) module to enhance feature expression and establish long-range dependencies; introduces a haar wavelet-based downsampling (HWD) module to minimize the information loss during the sampling process; and replaces C3K2 with the receptive field attention convolutional (RFAConv) module at the neck layer to optimize spatial feature extraction and enhance the processing capability for large convolutional kernels. The experimental results show that HFS-YOLO11 significantly improves thyroid nodule detection, with a 3.9% improvement in mAP50, a 1.8% improvement in mAP50-95, a 3.2% improvement in recall, and a 3.6% improvement in precision.***

Keywords: Target detection; Thyroid nodules; HFS-YOLO11.

1. INTRODUCTION

Thyroid nodules are localized hyperplasia or lumps in thyroid tissue [1]. Over the past few years, the prevalence of thyroid nodules has been rising, with global statistics showing that the prevalence ranges from 19% to 68% [2], the vast majority of which are benign. However, 7%-15% of these nodules may still be malignant [3]. Therefore, early identification and assessment of thyroid nodules are crucial to reducing the morbidity and mortality of thyroid cancer. Ultrasound imaging is the most commonly used detection method in clinical diagnosis, offering advantages such as real-time visualization, easy accessibility, and cost-effectiveness. However, the shortcomings of ultrasound images, such as low quality and low signal-to-noise ratio, lead to difficulties in distinguishing thyroid nodules from surrounding tissues and blurring of the boundaries, which increase the challenge of nodule detection. Additionally, thyroid nodules often have characteristics similar to surrounding tissues and irregular morphology, making misjudgments and omissions more common during the detection process. With the continuous evolution of artificial intelligence (AI) technology, deep learning techniques, particularly convolutional neural networks (CNNs), have found extensive application in the medical field. These methods are mainly categorized into two types: the two-stage detection algorithm, which performs target detection by generating a series of candidate regions. Li et al [4] introduced a deep-learning approach for detecting papillary thyroid cancer. By leveraging layer connectivity in faster region-based convolutional neural networks (Faster R-CNN) [5], they extracted more detailed features from low-resolution images. The two-stage detection algorithm achieves superior performance in thyroid nodule detection compared to most networks, though its speed is limited in multi-scale detection tasks. Among one-stage detection algorithms, Song et al. [6] proposed a multi-task cascaded convolutional neural network combining single-shot multi-box detector (SSD) and feature space pyramid techniques for detecting thyroid nodules of varying sizes and shapes. Wang et al. [7] introduced a modified version based on YOLOv2, designing an end-to-end detection network capable of recognizing the position and kind of thyroid nodules. Zhang et al. [8] further improved YOLOv3 by employing a high-resolution network backbone, achieving superior detection accuracy across diverse ultrasound images. Guo et al. [9] proposed an improved version based on YOLOv5, using ResNet18 to replace the backbone network of YOLOv5 to enhance the accuracy of nodule detection. Li et al. [10] used ByteTrack for multi-target tracking based on YOLOv7, enabling accurate real-time tracking of thyroid nodules. While one-stage algorithms are faster, their detection accuracy is usually lower than two-stage algorithms. Therefore, designing a more accurate model for thyroid nodule detection that offers faster detection speeds is crucial. In this paper, we select the YOLO11 in one-stage detection as the base model and construct a haar field spatial-you only look once 11 (HFS-YOLO11) model for thyroid nodule detection. This model addresses the problems of inconsistent size and diverse morphology in thyroid nodule detection. The main contributions of this paper are as follows: 1) We construct an HFS-YOLO11 framework for thyroid nodule detection. 2) By introducing a multi-scale spatial pyramid attention (MSPA) module, multi-scale spatial information features in thyroid nodules can be effectively extracted. The receptive field attention convolutional (RFAConv) module strengthens network performance and processes the details of nodule images. Additionally, replacing the Haar wavelet-based down-sampling (HWD) module minimizes information loss during the sampling

process by reducing the spatial resolution of the feature map. The improved neck layer enhances the network's ability to deal with problems such as blurred edges of thyroid nodules and difficulty in recognizing small targets. 3) Experiments on the thyroid image dataset demonstrate that this approach significantly improves the recognition ability of the you only look once (YOLO) model in thyroid nodule detection.

2. MATERIALS AND METHODS

This paper is broadly divided into two sections: model training and model evaluation. In the model training, all images were resized to 640×640×3 (640 represents the image height and width, and 3 represents the R, G, and B channels). Subsequently, the dataset is expanded by data enhancement operations such as flipping, rotating, and cropping [11]. The model training was performed using a data-enhanced dataset. In the model evaluation, the detection performance and generalization ability of the model are verified by loading test data, using the trained model for forward reasoning, and generating the final detection results by combining with non-maximal value suppression.

2.1 Dataset

Before model training, a sample set of ultrasound images of benign and malignant thyroid nodules from an online open source was collected. These images were de-identified and complied with relevant ethical requirements. Two experienced radiologists independently evaluated the images and labeled the nodule types and locations without prior knowledge of the patient's medical information. Ultrasound images were labeled using the Labelme [12] image annotation tool for network training, with each image in the dataset containing at least one thyroid nodule. To ensure data independence and fairness in the experiments, the dataset was randomly shuffled and divided into training, validation, and test sets. The training set contains 1452 images, the validation set contains 189 images, and the test set contains 241 images, with no overlap between the three. Additionally, during data preprocessing, data augmentation operations such as cropping and rotating were performed on the images to increase the model's generalization ability. The dataset used in this study is a publicly available de-identified dataset that meets the relevant ethical review criteria.

The YOLO11 [13] neural network is currently the one-stage detection network, which has made several improvements to the previous YOLO version to further enhance its performance and flexibility. YOLO11 has become the best choice for target detection tasks due to its fast, accurate, and easy-to-use features. Based on this, we constructed an HFS-YOLO11 network, which shows better robustness in thyroid nodule detection. The network structure of HFS-YOLO11 is shown in Figure 1.

The YOLO11 network is made up of three components: the backbone, neck, and head networks. In the inference process, images of size 640×640×3 are fed into the network. The backbone network is responsible for extracting the texture features of the image and generating three feature maps with varying dimensions to capture texture information at different levels. The neck network uses the feature pyramid network (FPN) [14] structure to generate high-resolution and semantically rich feature maps through feature fusion and enhancement. Finally, the improved feature maps are passed into the YOLO head network for target detection.

Although YOLO11 has a strong capability in texture extraction, there are still some limitations in multi-scale feature extraction, information retention, and detail feature processing. To solve these problems, we add the MSPA module to the network to effectively extract the multi-scale spatial information features of thyroid nodules, and introduce the HWD module to minimize the information loss due to the resolution change. In addition, the RFAConv module is incorporated into the head layer to further enhance the ability of the network to process the details of thyroid nodule images, thus improving the overall detection performance of the network.

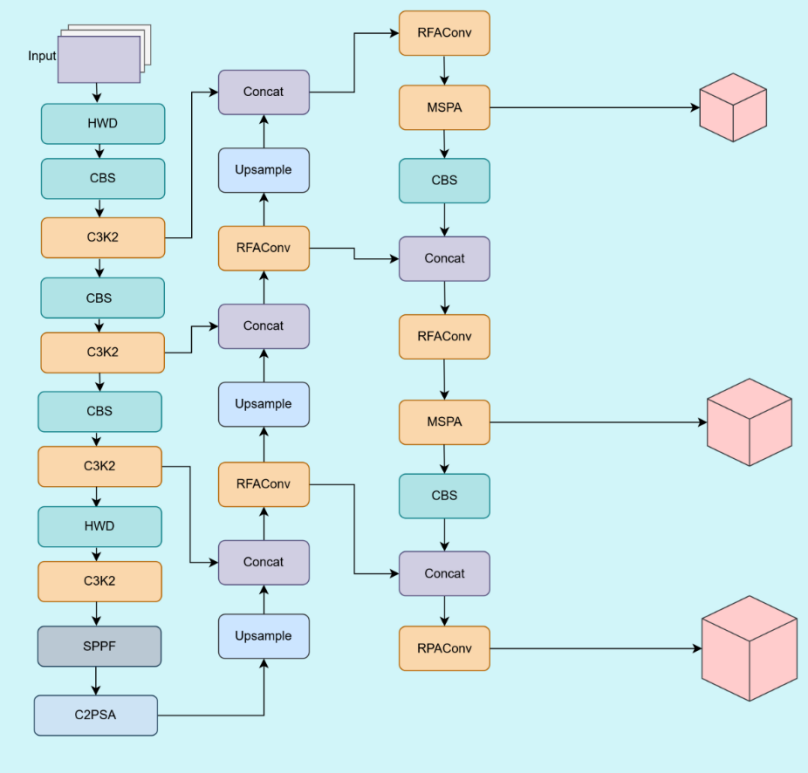

Figure 1: HFS-YOLO11 Network Architecture

2.1.1 MSPA module [15]

The MSPA module consists of three core components. First, unlike traditional CNN extracting multi-scale features in a layer-by-layer manner, MSPA incorporates the hierarchical phantom convolution (HPC) module to effectively capture spatial information across various scales within the input feature maps. Specifically, the HPC module consists of three key steps: split, conv, and concat. The split operation partitions the input feature maps into several subfeature maps along the channel dimension; the conv operation processes these sub-feature maps at multiple scales by layering the residual structure and using multiple convolutional filter banks; and the concat operation recombines the feature maps at different scales in the channel dimension, thus preserving the multi-scale information. Second, the MSPA module learns the channel attention weights of the multi-scale feature maps through the spatial pyramid recalibration (SPR) module to realize cross-dimensional feature interactions. In the SPR module, the spatial pyramid aggregation block is used to dynamically combine both global and local feature responses, combining structural regularization with structural information. In addition, two lightweight point-by-point convolutional layers are used to learn the relationships between channels. Finally, the Softmax function is applied to adjust the channel attention weights, enabling the capture of long-range channel relationships. This design can effectively fuse multi-scale feature information, which improves the sensitivity and expressiveness of the model at different scales in detecting thyroid nodule features, and thus enhances the ability of the model to recognize complex nodule morphology.

Figure 2 illustrates that the input to the MSPA module is the feature map $F \in R^{C\times H\times W}$ is the input to the MSPA module, where C, H, and W represent the number of channels, spatial height, and width. F is then fed into the HPC module, which generates multiple enhanced feature subsets through efficient multi-scale feature decomposition and convolution and other operations, denoted by $[\tilde{F}_1, \tilde{F}_2, \ldots, \tilde{F}_s] \in R^{C\times H\times W}$, the augmented feature maps $\tilde{F}$ with feature maps of various scales passed into the SPR module, the channel attention is efficiently learned, leading to the computation of channel-level attention weights $V = [V_1, V_2, \ldots, V_S] \in R^{C\times 1\times 1}$. To achieve adaptive selection across channels, Softmax is applied to each channel weight Vi to generate calibrated channel-level attention weights $A_i \in R^{C\times 1\times 1}$. The weight expressions are:

$$A_i = \text{Softmax}(V_i) = \exp(V_i) / \sum_{i=1}^{s} \exp(V_i) \quad (1)$$

In this way, long-range channel dependencies across different subsets of feature maps are created. Subsequently, all the calibrated channel-level attentional weights are spliced to generate the final attentional feature matrix A ∈

$R^{C\times1\times1}$, and the attention matrix is denoted as:

$$A = \text{Concat}([A_1, A_2, \ldots, A_S]) \tag{2}$$

Finally, the i subset of augmented feature maps $\tilde{F}_i$ is combined with the corresponding A_i by element-by-element multiplication to generate a refined output feature map $\tilde{F}_i$, which is represented as:

$$\tilde{F}_i = A_i \otimes \tilde{F}_i \tag{3}$$

Similarly, all the refined output feature maps are combined by superposition operation to obtain the final output feature map $\tilde{F}$:

$$\tilde{F} = \text{Concat}([\tilde{F}_1, \tilde{F}_2, \ldots, \tilde{F}_S]) \tag{4}$$

The MSPA multi-scale spatial pyramid attention mechanism not only enhances the interaction between local and global channel attention but also effectively enhances the representation of thyroid nodule image features at different spatial scales, thus improving the accuracy and robustness.

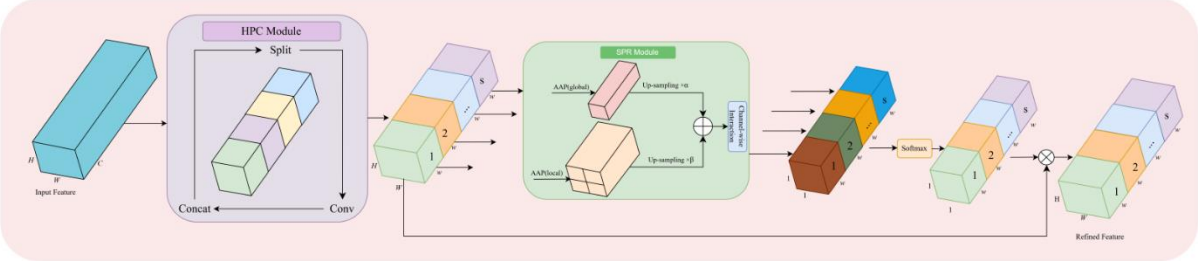

Figure 2: MSPA Module

2.1.2 HWD module [16]

The HWD module is mainly composed of two parts, the lossless feature encoding block and the feature representation learning block, which is designed to solve the difficulties of edge blurring and detail loss in thyroid nodule detection. In the lossless feature encoding block, the thyroid nodule image of resolution H×W undergoes decomposition through the Haar wavelet transform, and the approximate information (low frequency) and detail information (high frequency) of the image are extracted by using the low-pass filter H0 and the high-pass filter H1, respectively, and the image is decomposed into four components by the down-sampling operation (↓2), which includes approximate (low frequency) component and the detail components (high frequency) in the horizontal (H), vertical (V), and diagonal (D) directions. These components not only reduce the spatial resolution of the feature map (H×W→H/2×W/2) but also encode the spatial information efficiently into the channel dimensions (quadrupling the number of channels), realizing lossless information transfer. For the thyroid nodule detection task, the haar wavelet transform can effectively retain the critical detailed texture and edge information in the ultrasound image, which helps accurately localize the nodule boundary, particularly in small-sized nodules and low-contrast regions, thereby significantly enhancing the robustness and accuracy of thyroid nodule detection. The feature representation learning block is designed to adjust the channel count of the feature map, ensuring it aligns with the subsequent layers of the network. This block includes a 1×1 convolution, batch normalization, and a Relu activation function. By removing redundant information, it helps the network efficiently learn the most important features for better performance. The design of the feature representation learning block allows the HWD module to be used interchangeably with different down-sampling techniques, including max pooling or stridden convolution. Compared with traditional maximum pooling, the HWD module not only retains more detailed information in the output feature maps but also provides a good balance between the parameter count and computational cost, which further enhances the discriminative power and representation quality of the features, particularly in the context of thyroid nodule detection.

In summary, the HWD module includes a lossless feature encoding block, which effectively reduces the spatial resolution and preserves the key information through haar wavelet transform, and a feature representation learning block, which filters the redundant information and optimizes the feature representation through the standard 1×1 convolution, batch normalization, and Relu operations. The module is designed to preserve the texture details and edge features of ultrasound images in thyroid nodule detection, and it is especially good at small nodule localization and boundary accurate detection, which provides strong support for improving the robustness and accuracy of detection. The block diagram of the structure is shown in Figure 3.

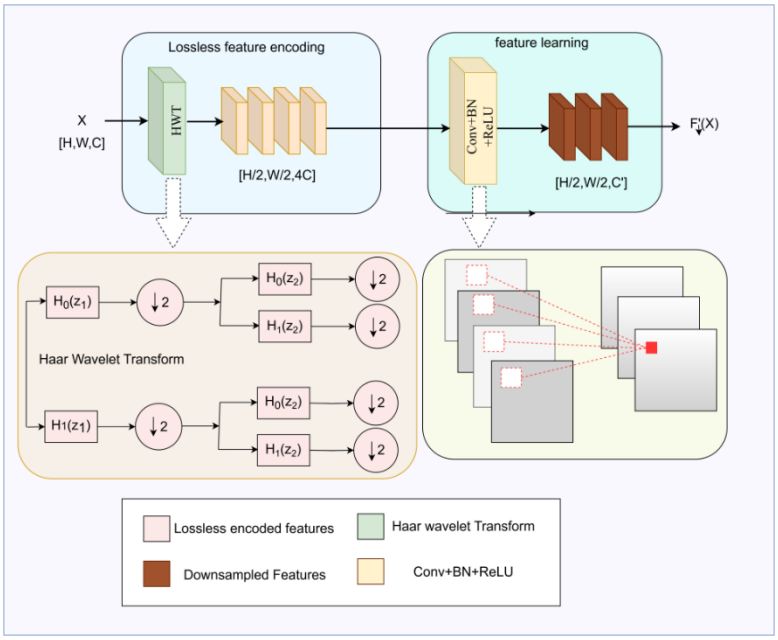

Figure 3: HWD Module

2.1.3 RFAConv module [17]

In thyroid nodule detection, we introduce the receptive field attention (RFA) mechanism, which highlights important features in the receptive field and emphasizes their spatial characteristics, addressing the issue of parameter sharing in convolution kernels. As the spatial features of the receptive field are dynamically generated based on the convolution kernel size, the RFA mechanism is closely integrated with the convolution process, enhancing the overall detection performance. Based on this, we adopt the RFAConv module with a 3×3 convolution kernel as an example, each receptive field slider window represents a spatial feature, and the overall structure is shown in Figure 4. To improve the efficiency of feature extraction, we discard the traditional unfold method in Pytorch, although unfold can extract spatial features parameter-free, the computational speed is slow. In RFAConv, we adopt group convolution instead of unfolding, which significantly accelerates the feature extraction process. Group convolution not only preserves the sensory field spatial features but also improves the operation efficiency.

Meanwhile, compared with the unfolding method, the performance is superior despite the comparable number of parameters. This method performs well in thyroid nodule detection, which can extract nodule details and edge features more accurately, and significantly improves the detection effect, especially in complex backgrounds. RFAConv enhances the ability of the network to perceive the region and details of the thyroid nodules by improving the computational efficiency, which further improves the detection accuracy and robustness, especially in the detection of complex backgrounds and tiny structures.

In thyroid nodule detection, interaction information can significantly strengthen the ability of the network to represent features. RFAConv improves network performance by leveraging the receptive field features to generate an attention map. However, processing each receptive field feature individually may lead to increased computational cost. To mitigate this, Avgpool is applied to aggregate the global information from each receptive field feature, reducing the computation and the number of parameters. Then, a 1×1 grouped convolution operation is used to realize the information interaction between the receptive field features, effectively integrating the global and local feature information. Finally, the Softmax function is applied to highlight the significance of the receptive field features. In summary, the RFA computation can be represented as follows:

$$F = \mathrm{Softmax}(g^{1\times1}(\mathrm{AvgPool}(X))) \times \mathrm{ReLU}(\mathrm{Norm}(g^{k\times k}(X))) = A_{rf} \times F_{rf}, \quad (5)$$

where $g^{i\times i}$represents the grouped convolution of size, k denoting the convolution kernel size, Norm indicates normalization, X refers to the input feature map, and F is obtained by multiplying the attention map A_{rf} with the transformed receptive field space features F_{rf}. Due to the existence of a parameter-sharing mechanism in the convolution operation, the standard convolution is less sensitive to a position change, which restricts the

performance of the convolutional neural network to a certain degree. For this reason, RFAConv effectively addresses the constraints of traditional convolution by dynamically adjusting the spatial feature weights in the receptive field, emphasizing the most critical features within the receptive field slider. RFAConv is not only able to enhance the focus on the thyroid nodule region in complex backgrounds but also accurately capture the nodule edges and fine structures, thus significantly improving accuracy.

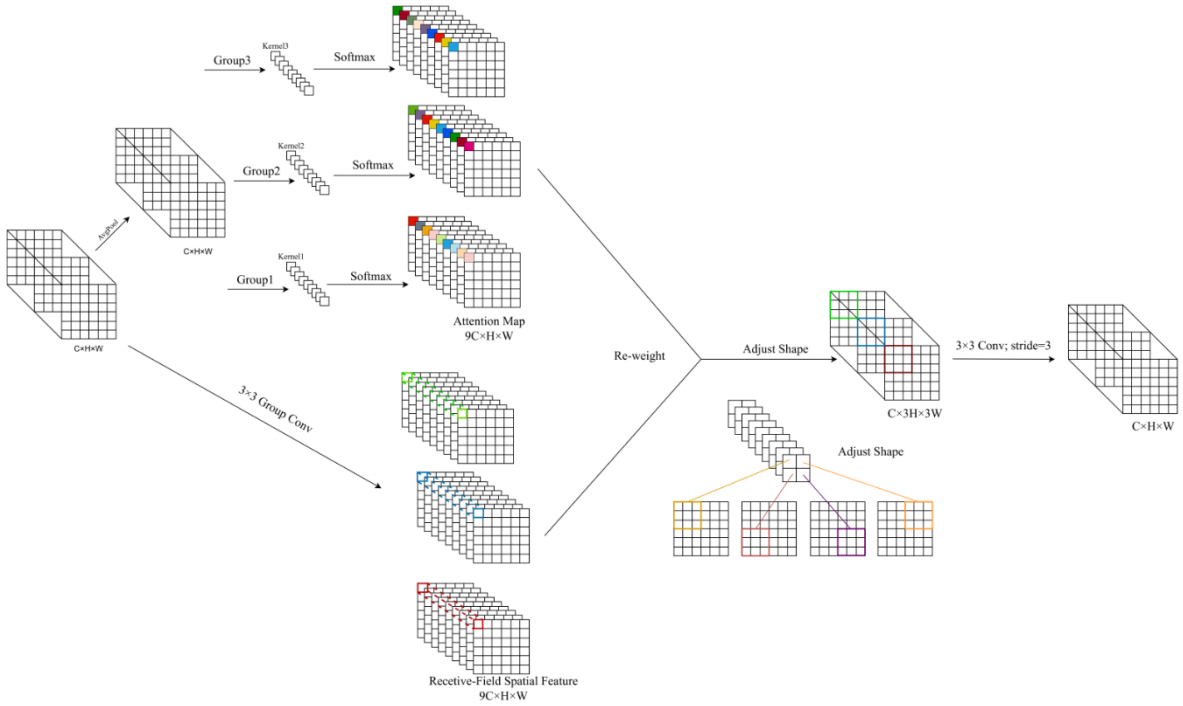

Figure 4: RFAConv Module

2.1.4 Neck layer improvement

In the thyroid nodule detection task, the improved network additionally introduces the upsample, concat, and RFAConv modules at the neck layer once, which further enhances the ability of multi-scale feature fusion and sensory field attention to better cope with the challenges of fine targets and complex backgrounds of thyroid nodules. The upsample operation enables the step-by-step zoom of the feature map, which helps to recover the details of the thyroid nodule edges; the concat module fuses features at different scales, integrating low-level detail features with high-level semantic features, thus improving the detection capability of nodules of different sizes; the introduced RFAConv module utilizes the receptive field attention mechanism to dynamically weight the features in the receptive field, highlighting the key information in the nodule region and suppressing the background noise, further contributes to a significant boost in both detection precision and reliability. The addition of upsample, concat, and RFAConv is more effective in solving the problems of fuzzy edges of the thyroid nodules and the tendency of small targets to be ignored, compared with the original network. The effect is more significant so that the improved neck layer network has a more accurate detection ability in the background of complex ultrasound images.

2.2 Training Setup

The training was performed on an RTX 3090 24GB GPU and a 12vCPU Intel(R) Xeon(R) Platinum 8375C @ 2.90 GHz CPU. The batch size is configured to 16, and data enhancement techniques including random rotation, scaling, cropping, and splicing are applied during the training process, the dataset is augmented to increase its diversity and improve the robustness of the model. In addition, we employed a stochastic gradient descent (SGD) optimizer with cosine annealing for learning rate decay [18]. The starting learning rate was set at 0.01, with a momentum of 0.937, and the network training for 200 epochs.

3. RESULTS

In this study, precision (P), recall (R), and mean average precision (mAP) were used as the assessment metrics for

model performance, which are widely used in various detection tasks and are well suited for evaluating thyroid nodule detection models. These metrics offer a comprehensive assessment of the ability of the model to handle different nodule categories, as well as its reliability and sensitivity in detecting nodules with varying sizes and shapes.

Precision is defined as the proportion of true positive samples among all the samples that the model predicted as positive, which measures how many of the predictions of the model are truly positive, and which is given by:

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{6}$$

Recall is the proportion of actual positive cases correctly identified by the model. Recall measures how many true positive cases the model is able to correctly identify, which is given by:

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{7}$$

TP, FP, and FN stand for true, false positive, and false negative examples, respectively. True and false correspond to whether the inferred results of the model are correct or incorrect compared to the true situation, whereas positive and negative indicate whether the inferred results of the model are in the desired category or some other category.

Another factor influencing precision and recall is the intersection over union (IOU) threshold. The IOU is defined as:

$$\text{IOU} = (\text{A}_\text{p} \cap \text{A}_\text{gt})/(\text{A}_\text{p} \cup \text{A}_\text{gt}) \tag{8}$$

where A_p represents the area of the predicted bounding box and A_gt indicates the area of the real bounding box. IOU measures the overlap between the predicted and the real bounding boxes. We set it to a constant 0.5 and an interval of 0.5-0.95.

We first performed an ablation experiment with the base model the original YOLO11 model, and progressively added the improved neck layer, RFAConv, MSPA, and HWD modules to the network. Table 1. summarizes the performance of each YOLO11 model in thyroid nodule detection, evaluated in terms of precision, recall, and mAP at a IOU threshold of 0.5 (mAP50) and a IOU threshold of 0.5-0.95 (mAP50-95). The experimental results show that by adding the improved neck layer, RFAConv, MSPA, and HWD modules to the original network, the HFS-YOLO11 model demonstrates a notable enhancement in the effectiveness of thyroid nodule detection, with a mAP50 of up to 90.4%. Figure 5. plots the Precision-Recall (PR) curves for each model, the area beneath the PR curve is referred to as average precision (AP), while mAP is the mean of the AP values for all classes.

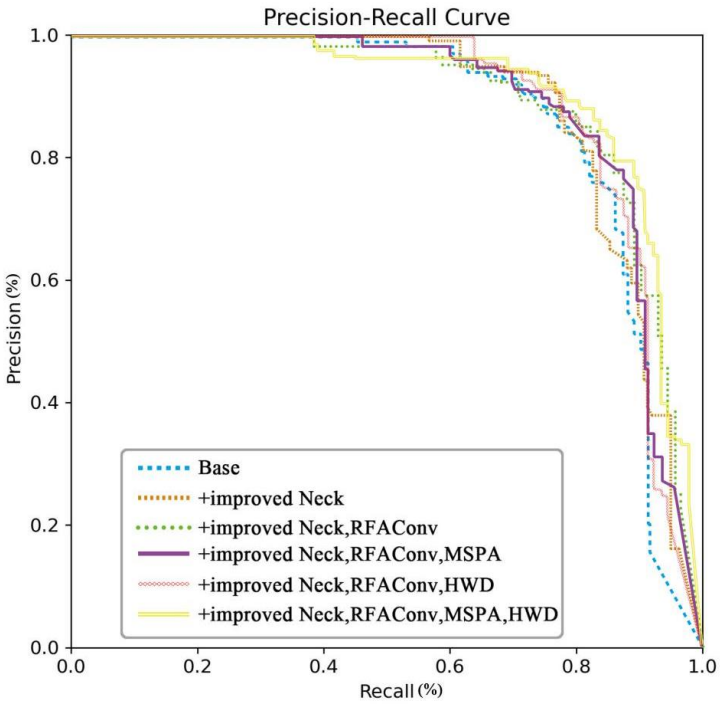

Figure 5: PR-curve of each YOLO11 model in the ablation experiment

Table 1: The Ablation Experiment Result of YOLO11 Network

Model	*Precision(%)*	*Recall(%)*	*mAP50(%)*	*mAP50-95(%)*
YOLO11	83.3	79.5	86.5	65.7
+improved neck	92	76.8	88.2	64.9
+improved neck+RFAConv	84.9	82.2	89.2	66.7
+improved neck+RFAConv+MSPA	83.4	82.2	88.6	67.1
+improved neck+RFAConv+HWD	90.6	77.8	88.7	66.7
+improved neck+RFAConv+MSPA+HWD	86.9	82.7	90.4	67.5

The results of ultrasound image detection of thyroid nodules are shown in Figure 6 and Figure 7. In Figure 6, (a) represents the true label of benign thyroid nodules, and (b)-(g) represent the five results of benign thyroid nodules detected using YOLO11, the network in Table 1. In Figure 7, (a) represents the true labeling of malignant thyroid nodules, and (b)-(g) represents the five outcomes of malignant thyroid nodules detected using YOLO11, the network in Table 1.

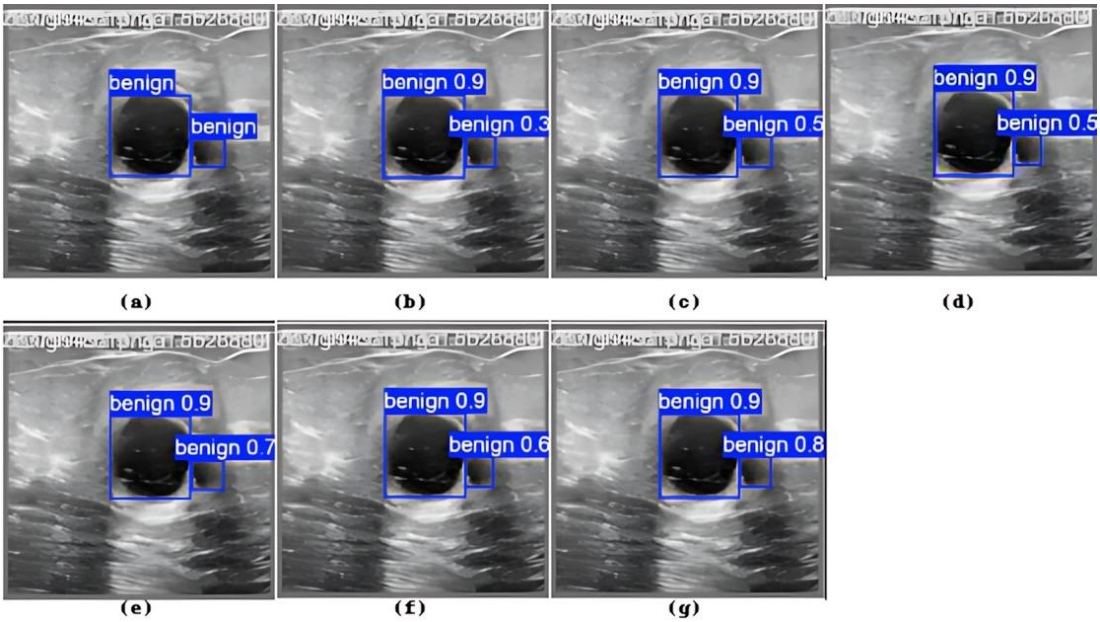

Figure 6: Detection results of different models for benign thyroid nodule images: (a) real situation; (b) base model prediction results; (c) adding improved neck layer; (d) adding improved neck layer and RFAConv; (e) adding improved neck layer, RFAConv, and MSPA; (f) adding improved neck layer, RFAConv, and HWD; and (g) adding the improved neck layer, RFAConv, MSPA, and HWD at the same time. Letters and numbers above the bounding box indicate the type of nodule predicted by the model and the confidence level, respectively.

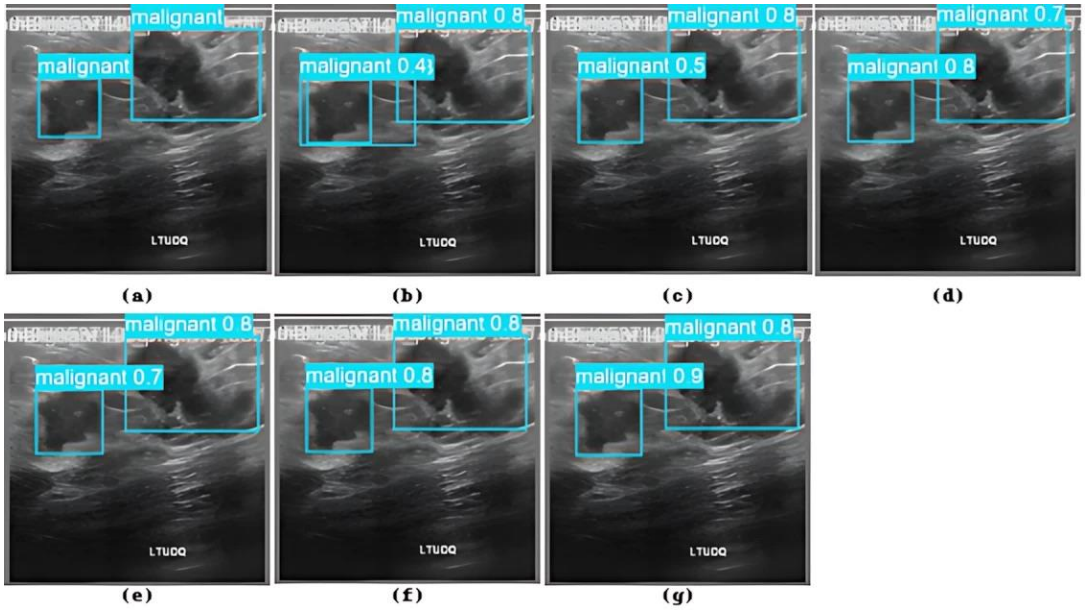

Figure 7: Detection results of different models for malignant thyroid nodule images: (a) real situation; (b) base model prediction results; (c) adding improved neck layer; (d) adding improved neck layer and RFAConv; (e) adding improved neck layer, RFAConv, and MSPA; (f) adding improved neck layer, RFAConv, and HWD; and (g) adding the improved neck layer, RFAConv, MSPA, and HWD at the same time. Letters and numbers above the bounding box indicate the type of nodule predicted by the model and the confidence level, respectively.

To further validate the advantages of the algorithms presented in this paper, we conducted controlled experiments with Faster R-CNN, SSD, YOLOv5, YOLOv8, and YOLO11. This experiment uses the same dataset and evaluation metrics as the original paper, and the comparison results of each algorithm are shown in Table 2.

Table 2: Algorithm control experiment table

Model	*Precision(%)*	*Recall(%)*	*mAP50(%)*	*mAP50-95(%)*
Faster R-CNN	78.2	79.1	84.2	56.7
SSD	80.8	83.9	89	57
YOLOv5	77.2	88	85.8	56
YOLOv8	83.2	84.1	87.1	59.7
YOLO11	83.3	79.5	86.5	65.7
HFS-YOLO11	86.9	82.7	90.4	67.5

4. DISCUSSION

Computer-aided diagnosis (CAD) was first proposed by Lim et al. [19] in 2008 and applied to thyroid nodule detection. As deep learning technology continues to advance, the performance of CAD has been notably strengthened. This study aims to construct an efficient and accurate CAD tool for thyroid nodule detection. In Table 1, We conduct a performance comparison between the original YOLO11 network and the HFS-YOLO11 network in terms of detection, and the experimental findings reveal that HFS-YOLO11 shows a considerable enhancement in mAP50, which verifies the effectiveness of the improved strategy for YOLO11 and enhances the detection of thyroid nodules. The improved neck layer can enhance the multi-scale feature fusion and sensory field attention mechanism to better cope with fine targets and complex backgrounds in thyroid nodules, which enhances the precision of thyroid nodule detection; the RFAConv module boosts the network proficiency in processing the details of thyroid nodule images, which improves the recall rate of thyroid nodule detection; the MSPA fuses the information of multi-scale features, which enhances the capability of the model to capture features at various scales and improves the alignment between the predicted bounding boxes and actual bounding boxes in thyroid nodule detection; the HWD module retains the texture details and edge features in the ultrasound image, and performs well in small nodule localization and boundary precision detection. Through Fig. (6) and Fig. (7), it is found that the HFS-YOLO11 network has an increased confidence level in thyroid nodule detection compared to the original network, and the predicted bounding box is closer to the real situation. This not only lightens the workload for radiologists but also further boosts the detection of thyroid nodules by computer-aided diagnosis (CAD) systems.

Despite the results achieved in this study, there are still some shortcomings: 1) Due to the small size of the dataset, the overfitting problem has not been completely solved despite the introduction of pre-trained models and data enhancement strategies. Future research should prioritize the expansion of the dataset and adopt more efficient technical means to mitigate the overfitting phenomenon. 2) The data sources in this study are relatively homogeneous, which limits the wide applicability of the model. Follow-up work should focus on collecting diverse images from different patients and multiple healthcare organizations to increase the stability and dependability of the model in various demographic groups and healthcare settings.

5. CONCLUSION

In this study, an HFS-YOLO11 network is constructed based on YOLO11, which can better capture the multi-scale spatial information features in thyroid nodules by adding the MSPA module, introducing the HWD module to compensate for the information loss caused by the sampling process as much as possible, and the RFAConv module enhances the performance of the network to process the details of the thyroid nodule images, in addition, the improved neck layer can better cope with fine targets and complex backgrounds in thyroid nodules. The findings indicate that the mAP50 of the HFS-YOLO11 network is 90.4%, which demonstrates a 3.9% improvement over the baseline network, indicating that the HFS-YOLO11 model constructed in this paper effectively improves the detection ability of thyroid nodules, and the model has a broad application prospect in the field of thyroid nodule detection, and it is a highly potential computer-aided tool.

REFERENCES

[1] Yang J, Shi X, Wang B, et al. Ultrasound image classification of thyroid nodules based on deep learning[J]. Frontiers in Oncology, 2022, 12: 905955.
[2] Haugen B R, Alexander E K, Bible K C, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer[J]. Thyroid, 2016, 26(1): 1-133.
[3] Kitahara C M, Sosa J A. Understanding the ever-changing incidence of thyroid cancer[J]. Nature Reviews Endocrinology, 2020, 16(11): 617-618.
[4] Li H, Weng J, Shi Y, et al. An improved deep learning approach for detection of thyroid papillary cancer in ultrasound images[J]. Scientific reports, 2018, 8(1): 6600.
[5] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.
[6] Song W, Li S, Liu J, et al. Multitask cascade convolution neural networks for automatic thyroid nodule detection and recognition[J]. IEEE journal of biomedical and health informatics, 2018, 23(3): 1215-1224.
[7] Wang L, Yang S, Yang S, et al. Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network[J]. World journal of surgical oncology, 2019, 17: 1-9.
[8] Zhang L, Zhuang Y, Hua Z, et al. Automated location of thyroid nodules in ultrasound images with improved YOLOV3 network[J]. Journal of X-ray Science and Technology, 2021, 29(1): 75-90.
[9] Guo R, Xu L, Guo J, et al. Automatic detection and classification algorithm of thyroid nodules in CT images based on YOLOv5s[C]//2023 9th International Conference on Computer and Communications (ICCC). IEEE, 2023: 1701-1706.
[10] Li X, Yin P, Qu Y, et al. Research on Multi-Object Tracking Algorithm for Thyroid Nodules Based on ByteTrack[C]//2024 IEEE 3rd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA). IEEE, 2024: 1589-1592.
[11] Yang D, Xia J, Li R, et al. Automatic thyroid nodule detection in ultrasound imaging with improved yolov5 neural network[J]. IEEE Access, 2024, 12: 22662-22670.
[12] Russell B C, Torralba A, Murphy K P, et al. LabelMe: a database and web-based tool for image annotation[J]. International journal of computer vision, 2008, 77: 157-173.
[13] Khanam R, Hussain M. Yolov11: An overview of the key architectural enhancements[J]. arXiv preprint arXiv:2410.17725, 2024.
[14] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.
[15] Yu Y, Zhang Y, Cheng Z, et al. Multi-scale spatial pyramid attention mechanism for image recognition: An effective approach[J]. Engineering Applications of Artificial Intelligence, 2024, 133: 108261.
[16] G. Xu, W. Liao, X. Zhang, C. Li, X. He, and X. Wu, “Haar wavelet downsampling: A simple but effective downsampling module for semantic segmentation,” Pattern Recogn., vol. 143, Jul. 2023, Art. no. 109819.

[17] Zhang X, Liu C, Yang D, et al. RFAConv: Innovating spatial attention and standard convolutional operation[J]. arXiv preprint arXiv:2304.03198, 2023.
[18] Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts[J]. arXiv preprint arXiv:1608.03983, 2016.
[19] Lim K J, Choi C S, Yoon D Y, et al. Computer-aided diagnosis for the differentiation of malignant from benign thyroid nodules on ultrasonography[J]. Academic radiology, 2008, 15(7): 853-858.