technologies

MDPI

Article

Enhancing Thyroid Nodule Detection in Ultrasound Images: A Novel YOLOv8 Architecture with a C2fA Module and Optimized Loss Functions

Shidan Wang [1,†], Zi-An Zhao [2,†], Yuze Chen [3], Ye-Jiao Mao [2] and James Chung-Wai Cheung [2,4,*]

[1] School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China; stacey.os.w@cqu.edu.cn
[2] Department of Biomedical Engineering, Faculty of Engineering, The Hong Kong Polytechnic University, Hong Kong 999077, China; 24057373r@connect.polyu.hk (Z.-A.Z.); yejiao.mao@connect.polyu.hk (Y.-J.M.)
[3] College of Computer Science, Chongqing University, Chongqing 400044, China; yuze.chen@cqu.edu.cn
[4] Research Institute of Smart Ageing, The Hong Kong Polytechnic University, Hong Kong 999077, China
[*] Correspondence: james.chungwai.cheung@polyu.edu.hk; Tel.: +852-2766-7673
[†] These authors contributed equally to this work.

Abstract: Thyroid-related diseases, particularly thyroid cancer, are rising globally, emphasizing the critical need for the early detection and accurate screening of thyroid nodules. Ultrasound imaging has inherent limitations—high noise, low contrast, and blurred boundaries—that make manual interpretation subjective and error-prone. To address these challenges, YOLO-Thyroid, an improved model for the automatic detection of thyroid nodules in ultrasound images, is presented herein. Building upon the YOLOv8 architecture, YOLO-Thyroid introduces the C2fA module—an extension of C2f that incorporates Coordinate Attention (CA)—to enhance feature extraction. Additionally, loss functions were incorporated, including class-weighted binary cross-entropy to alleviate class imbalance and SCYLLA-IoU (SIoU) to improve localization accuracy during boundary regression. A publicly available thyroid ultrasound image dataset was optimized using format conversion and data augmentation. The experimental results demonstrate that YOLO-Thyroid outperforms mainstream object detection models across multiple metrics, achieving a higher detection precision of 54%. The recall, calculated based on the detection of nodules containing at least one feature suspected of being malignant, reaches 58.2%, while the model maintains a lightweight structure. The proposed method significantly advances ultrasound nodule detection, providing an effective and practical solution for enhancing diagnostic accuracy in medical imaging.

Keywords: thyroid nodule detection; ultrasound imaging; YOLO; deep learning; medical image analysis

check for updates

Academic Editors: Juvenal Rodriguez-Resendiz, Gerardo I. Pérez-Soto, Karla Anhel Camarillo-Gómez, Saul Tovar-Arriaga and Pietro Zanuttigh

Received: 21 October 2024
Revised: 27 December 2024
Accepted: 8 January 2025
Published: 9 January 2025

Citation: Wang, S.; Zhao, Z.-A.; Chen, Y.; Mao, Y.-J.; Cheung, J.C.-W. Enhancing Thyroid Nodule Detection in Ultrasound Images: A Novel YOLOv8 Architecture with a C2fA Module and Optimized Loss Functions. *Technologies* **2025**, *13*, 28. https://doi.org/10.3390/technologies13010028

1. Introduction

The thyroid gland is a vital endocrine organ in the human body. The incidence of thyroid-related diseases, particularly thyroid cancer, has been rising rapidly recently and has become a significant global public health concern [1]. Early detection and accurate screening, especially for thyroid nodules—which are often considered early indicators of potential malignancies—play a critical role in mitigating this trend. Clinical studies have confirmed that timely and accurate diagnosis of thyroid nodules can significantly reduce the incidence and mortality rates of thyroid cancer [2]. Various diagnostic examinations are commonly used to evaluate the thyroid gland, including ultrasound (US), computed tomography (CT), magnetic resonance imaging (MRI), thyroid scans, and elastography [3].

While the majority of thyroid nodules are benign [4], a small percentage can be malignant, making accurate diagnosis essential for appropriate treatment. Accurate diagnosis is essential for appropriate treatment, but distinguishing malignant from benign nodules using non-invasive methods remains challenging. Ultrasound imaging is the preferred screening tool due to its convenience, low cost, and absence of radiation exposure. However, ultrasound images often suffer from high noise, low contrast, and blurred boundaries [5], which makes interpretation subjective and dependent on the radiologist's experience. This subjectivity can lead to inconsistent diagnoses and potentially result in unnecessary invasive procedures, such as fine-needle aspiration biopsy (FNAB) [6]. Traditional ultrasonography identifies features associated with malignancy, such as hypoechogenicity, irregular margins, microcalcifications, and a taller-than-wide shape [7]. Although these features aid in assessment, they cannot definitively indicate malignancy, and a biopsy is required for a conclusive diagnosis.

To improve diagnostic accuracy, researchers have explored artificial intelligence (AI) and deep learning techniques to provide a more objective assessment [8,9]. Machine learning models, particularly convolutional neural networks (CNNs), have shown great promise in classifying thyroid nodules and assisting in diagnosis [10]. For instance, Zheng et al. [11] proposed an improved U-Net architecture called DSRU-Net, which enhances the automatic segmentation of thyroid glands and nodules in ultrasound images by incorporating ResNeSt blocks, atrous spatial pyramid pooling, and deformable convolution v3. Similarly, Zhou et al. [12] introduced a thyroid nodule detection model named Thyroid-DETR, utilizing the Transformer architecture along with deformable convolution, multi-head self-attention, and a dual-stream training structure to improve detection accuracy in ultrasound images. Additionally, Chen et al. [13] developed a multi-view learning model called MLMSeg, which integrates CNNs, Transformers, and Graph Convolutional Networks to enhance thyroid nodule segmentation by capturing local, global, and spatial structural features. Consequently, computer-aided detection (CAD) methods for thyroid nodules have become a research hotspot, holding significant potential for future advancements.

Despite advances, existing CAD methods still face challenges when applied to thyroid nodules. One major issue is that datasets often suffer from class imbalance, causing models to be biased toward the majority class during training, which adversely affects performance in detecting and classifying malignant nodules [4,14]. Additionally, many studies employ semantic segmentation techniques for thyroid nodule detection due to their precise pixelwise delineation. However, semantic segmentation networks are often computationally intensive and slower, making them less suitable for real-time clinical applications and large-scale screenings. To address these issues, an improved model named YOLO-Thyroid is proposed for the automatic detection of nodules in thyroid ultrasound images, building upon YOLOv8. YOLO (You Only Look Once) models, with their efficient single-stage detection architecture, have demonstrated strong performance in detecting medical anomalies in pulmonary nodules (chest X-rays), breast masses (mammograms), and brain tumors (MRI scans) [15–19]. YOLO offers a high detection speed suitable for prompt decision-making, which is crucial in clinical settings.

Although newer versions of the YOLO model are available, YOLOv8 [19] was selected due to its optimal balance between speed and accuracy, as well as its optimized performance, flexibility, and extensibility, which make it particularly suitable for the current application scenario. The Coordinate Attention (CA) mechanism is introduced to enhance the extraction of important features. CA embeds positional information into channel attention, allowing the model to focus on the most informative regions of the feature maps, which is crucial for accurately detecting thyroid nodules with variable sizes and blurred boundaries. CA is integrated into the existing C2f module of YOLOv8, resulting in the modified C2fA module. This integration allows YOLO-Thyroid to emphasize important features related to target objects, thereby improving detection performance. Additionally, specialized loss functions are employed: the class-weighted binary cross-entropy (CW-BCE) loss function, used to alleviate the problem of class imbalance, and the SCYLLA-IoU (SIoU) loss function, which comprehensively considers factors such as the target's position, size, and shape during boundary regression to improve localization accuracy. Through these improvements, the YOLO-Thyroid model achieves better detection performance while maintaining a lightweight structure.

1.1. Main Contributions

The main contributions of this paper are as follows:

1. The proposal of an improved model, YOLO-Thyroid, with the introduction of the C2fA module and the CW-BCE and SIoU loss functions, strengthening the model's ability to extract and fuse important features, and improving performance under class imbalance conditions.
2. Extensive experiments and comparisons demonstrating that the YOLO-Thyroid model outperforms current mainstream object detection models across multiple performance metrics, validating its effectiveness.

1.2. Objective

The primary objective of this study is to address the critical challenges in thyroid nodule detection in ultrasound imaging, particularly the limitations of existing methods in terms of feature extraction, class imbalance, and localization accuracy. Specifically, the objectives are:

1. To design an improved YOLOv8-based detection model, YOLO-Thyroid, incorporating a novel C2fA module and optimized loss functions to enhance feature extraction and localization performance.
2. To alleviate the impact of class imbalance in the dataset by introducing a class-weighted binary cross-entropy (CW-BCE) loss function, ensuring the robust detection of both benign and malignant nodules.
3. To integrate advanced attention mechanisms and a lightweight architecture, enabling the model to achieve improved detection accuracy and balanced recall while maintaining computational efficiency for real-time clinical applications.
4. To assess the model's performance in terms of detecting nodules and identifying at least one feature suspected of being malignant (e.g., TIRADS categories 4a, 4b, 4c, and 5), with an emphasis on determining which category exhibits the highest detection sensitivity. Additionally, this objective seeks to evaluate the model's ability to differentiate varying levels of risk and establish a threshold for the number of detected features required to warrant further clinical investigation for potential malignancies.
5. To validate the proposed model on a publicly available thyroid ultrasound dataset and demonstrate its superiority compared to state-of-the-art object detection methods.

This study aims to provide a practical and effective solution for automatic thyroid nodule detection, contributing to improved diagnostic accuracy and efficiency in medical imaging, while offering insights into the clinical implications of the model's ability to differentiate risk levels across classes and establish decision-making thresholds.

The structure of this paper is organized as follows: Section 2 provides a detailed description of the dataset preprocessing methods, as well as the structure and improvements of the YOLO-Thyroid model. Section 3 presents the experimental results and analyses, comparing them with other advanced models. Section 4 discusses the advantages of the model and possible directions for improvement. Section 5 summarizes the findings and suggests future research directions.

2. Materials and Methods

An overview of the methodological workflow is presented in Figure 1. To adapt the dataset for object detection tasks, a comprehensive method was developed to convert ultrasound image labels into the YOLO format. Next, the thyroid ultrasound images were preprocessed to remove irrelevant regions and interfering markers, completing the data preparation for model input. To address the issue of a small sample size, data augmentation techniques were employed. Subsequently, the ultrasound data were trained using the proposed YOLO-Thyroid model, as shown in Figure 1. To improve the model's detection accuracy across different categories of nodules, the C2fA module was proposed, enabling the model to better capture spatial information and important features. Additionally, to enhance the model's robustness against data imbalance and convergence issues, the CW-BCE and SIoU loss functions were incorporated into the object detection loss function, thereby further improving model performance.

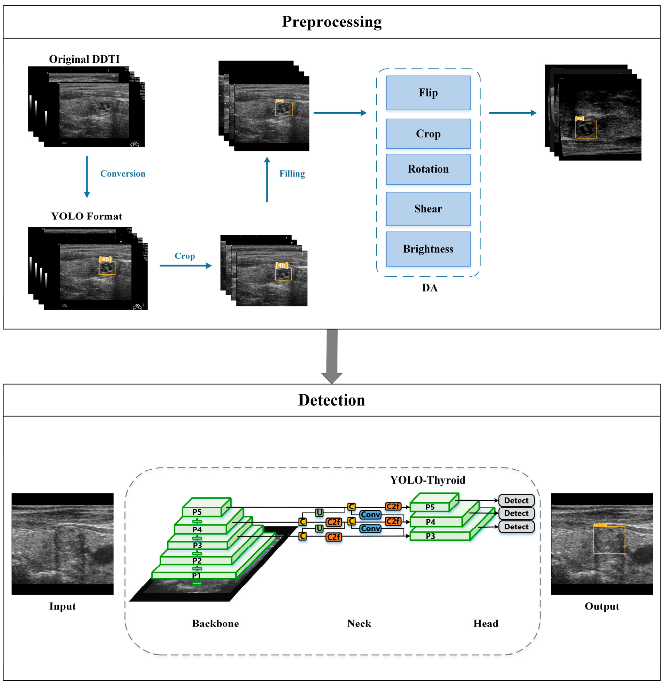


Figure 1. Overall workflow of the proposed method. C and U are concat and upsample layers respectively. The green frame and the number in Preprocessing is the ground truth of nodule location and class. The yellow frame and the number in Detection indicates the prediction of location and class.

2.1. Dataset

2.1.1. Data Description

In this study, the Digital Database of Thyroid Images (DDTI) [20] was utilized. This publicly available dataset provides a comprehensive collection of B-mode ultrasound images, along with detailed descriptions and annotations of suspicious thyroid lesions. The dataset contains 480 images from 400 medical cases, saved in JPG format with a resolution of 560 × 360 pixels. Label information is stored in XML files. The nodule information includes composition, size, echogenicity, edge characteristics, presence or absence of calcification, and Thyroid Imaging Reporting and Data System (TIRADS) scores. Nodule annotations are manually segmented by radiologists and recorded in the form of coordinates. The TIRADS [21] is used to assess the malignancy risk of thyroid nodules by standardizing the evaluation of ultrasound features and classifying nodules into different levels. The TIRADS scores in the dataset include:

- [2] Benign (0% risk of malignancy);
- [3] No suspicious US feature (<5% malignancy);
- [4a] One suspicious US feature (5–10% malignancy);
- [4b] Two suspicious US features (10–80% malignancy);
- [4c] Three or four suspicious US features (10–80% malignancy);
- [5] Five suspicious features (>80% malignancy).

These classifications are illustrated in Figure 2. The dataset was cleaned by removing images with incomplete coordinate annotations and missing TIRADS score labels, ultimately obtaining 339 images. Figure 3 presents a statistical overview of the dataset. According to established guidelines, classes 2 and 3 are categorized as benign, whereas classes 4a, 4b, 4c, and 5 are classified as malignant [20].

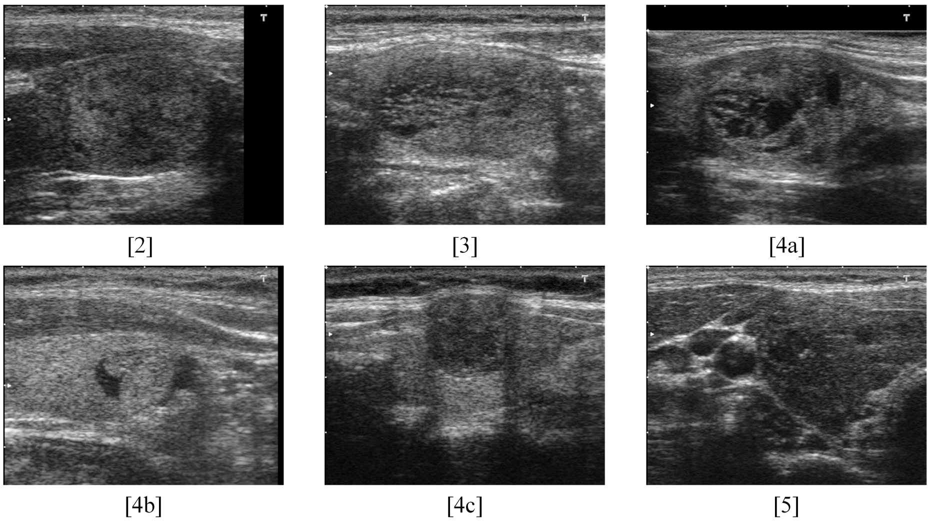


Figure 2. TIRADS classification for assessing malignancy risk in thyroid nodules. [2], [3], [4a], [4b], [4c], [5] are the TIRADS scores to classify nodules. The details are described above.

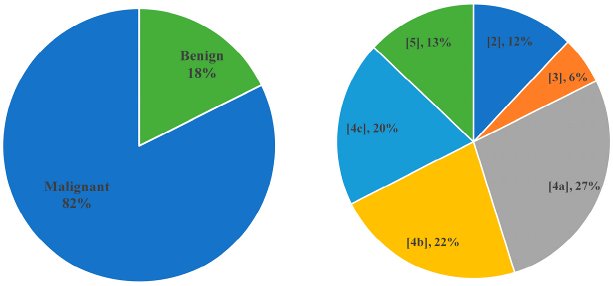
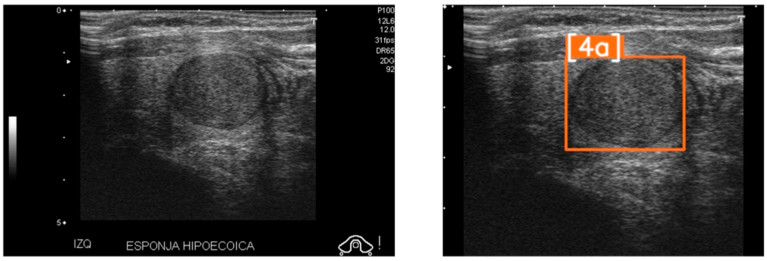


Figure 3. Statistical overview of the DDTI dataset. [2], [3], [4a], [4b], [4c], [5] are the TIRADS scores to classify nodules. The details are described above.

2.1.2. Data Preprocessing

In this paper, the image preprocessing process includes four key steps to convert the dataset with original labels into YOLO data format. Non-essential information had to be removed to ensure that the model focuses only on important features. Next, the images were standardized in size to enhance data consistency and improve model performance. Then, the dataset was divided and augmented to increase data diversity and enhance the model's generalization ability. Examples of the preprocessed images are shown in Figure 4. Detailed descriptions of each step are as follows.



(a) (b)

Figure 4. Examples of preprocessed thyroid ultrasound images. (**a**) Original ultrasound image. (**b**) Preprocessed image. The rectangle in (**b**) represents the detection label in YOLO format, indicating the location of the thyroid nodule. The '4a' label corresponds to the TIRADS score.

Data Format Conversion: The segmentation labels were transformed into the format required by the YOLO model to prepare the data for subsequent object detection model training. For each image, the minimum bounding rectangle of each target nodule was extracted where the nodule is present in the segmentation label. Then, the center coordinates of the bounding box and its width and height were calculated. These values were normalized by dividing by the image's width and height, ensuring they ranged between 0 and 1, as required by the YOLO format. The formatted labels included the class identifier along with the normalized center coordinates, width, and height. This conversion enabled the YOLO model to accurately interpret the bounding boxes for training and detection purposes. The detailed pseudocode is provided in Algorithm A1 in Appendix A.

Removal of Non-essential Information: The original ultrasound images contain additional information besides the target area, such as grayscale scales, parameter information, text labels, and probe icons. This information might interfere with the model's training and detection accuracy. Therefore, during preprocessing, the images were cropped to remove these non-essential elements, retaining only the ultrasound images containing the target area [22,23]. The purpose of this step is to eliminate noise and irrelevant features, enhancing the model's focus on the target.

Image Adjustment: In the cropped images, the edges were padded with black pixels to unify the image size to 640 × 640 pixels. The purpose of this step is to prevent the target shape from being distorted when resizing the images, thus preserving the true characteristics of the target.

Dataset Partitioning and Augmentation: The dataset was divided into three independent subsets: the training set, validation set, and test set, in a ratio of 7:1.5:1.5. Table 1 lists the number of samples in each subset of the DDTI dataset. In the training set, a series of data augmentation techniques were applied to enhance the model's generalization ability and prevent overfitting. Specific augmentation methods included horizontal and vertical flipping to increase sample diversity—acceptable in thyroid imaging due to the gland's bilateral symmetry. Cropping operations were used to simulate different imaging distances and perspective changes, reflecting variations that occur in clinical practice. Rotation within a limited angle range was applied to enhance the model's robustness to varying probe orientations during ultrasound examinations. Shear transformations were employed to mimic slight geometric distortions that may result from probe pressure or angle variations during imaging. Brightness adjustments enabled the model to adapt to variations in image intensity caused by different machine settings or patient characteristics. These augmentation techniques are commonly used in medical image analysis to improve model performance while maintaining clinical validity [24,25].

Table 1. The distribution of samples in the training, validation, and test sets of the DDTI dataset.

Property	Class	Malignancy Risk	Training Set	Validation Set	Testing Set	Total
Recognized Benign	2	0%	30	4	7	41
	3	<5%	12	4	3	19
Suspicious of Malignancy	4a	5–10%	62	14	17	93
	4b	10–80%	58	9	9	76
	4c	10–80%	43	13	10	66
	5	>80%	32	7	5	44
Total			237	51	51	339

In the augmented dataset, the augmentation methods are applied randomly and can be combined, resulting in composite augmentations where multiple techniques are applied together. Figure 5 illustrates the original image alongside the augmented images. Collectively, these augmentation techniques enhanced the diversity of the training data, thereby improving the model's performance and generalization to unseen data. Through these augmentation methods, the number of samples in the training set was expanded to three times the original, and the augmented dataset is shown in Table 2.

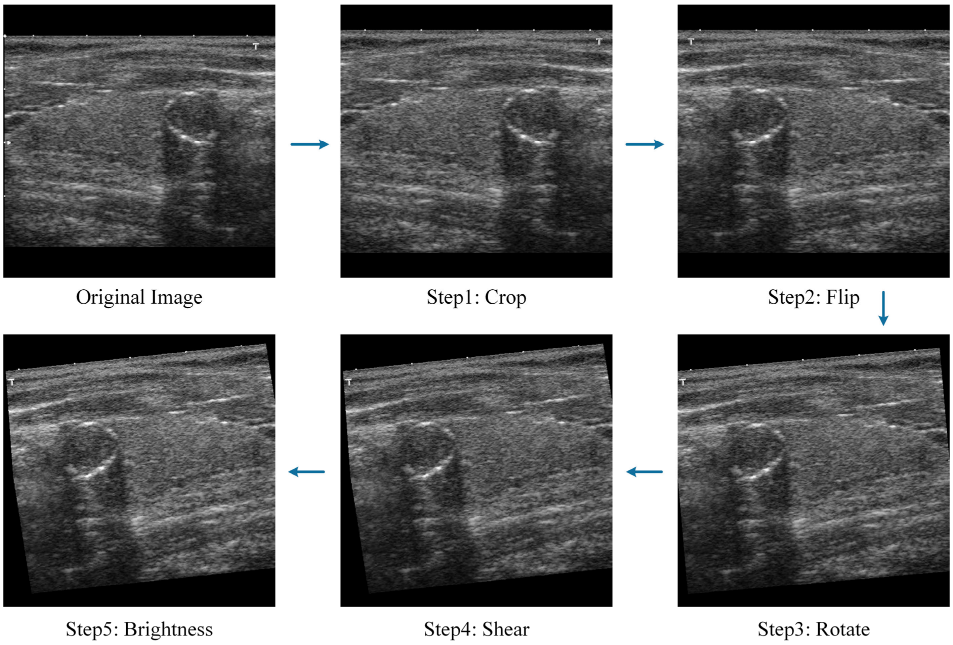


Figure 5. Examples of data augmentation techniques applied to a single image. Step 1: cropping to retain 95% of the image centered at coordinates (7%, 46%). Step 2: horizontal flipping. Step 3: rotation by 5 degrees. Step 4: shear transformations with 9 degrees along the X-axis and −2 degrees along the Y-axis. Step 5: a brightness adjustment of 9%. The final image is the result of the combined effect of these augmentation methods.

Table 2. Expanded sample sizes in the augmented training set.

Property	Class	Malignancy Risk	Training Set	Validation Set	Testing Set	Total
Recognized Benign	2	0%	90	4	7	101
	3	<5%	36	4	3	43
Suspicious of Malignancy	4a	5–10%	186	14	17	217
	4b	10–80%	174	9	9	192
	4c	10–80%	129	13	10	152
	5	>80%	96	7	5	108
Total			711	51	51	813

2.2. Methods

In this study, a detection model based on the YOLOv8 architecture, named YOLO-Thyroid, is proposed. YOLOv8 [19] is an advanced single-stage object detection model with efficient feature extraction capabilities and fast inference speed, as shown in Figure 6. To further enhance performance in thyroid ultrasound nodule detection tasks, two key modules were introduced. First, the feature extraction of the model was optimized by designing a C2fA module to enhance its perception of thyroid nodules, which enhanced detection performance in complex ultrasound image backgrounds. Second, considering the characteristics of ultrasound datasets, its loss function was improved to enhance the model's performance under class imbalance conditions. These two improvements allow YOLO-Thyroid to enhance nodule detection performance while maintaining the original model's efficiency. The design concepts and implementation details will be detailed in the following sections.

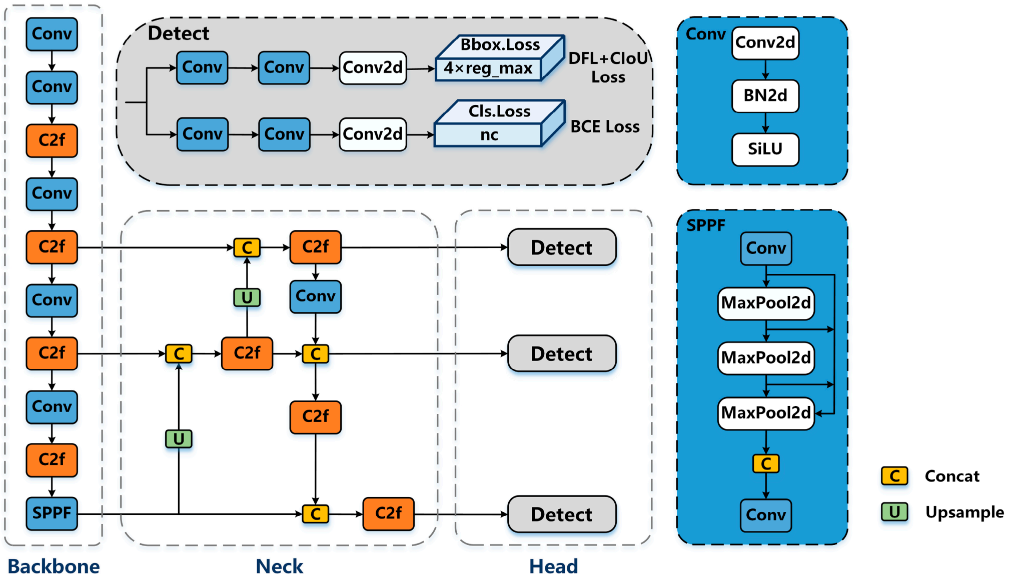


Figure 6. Architecture of YOLOv8 [19].

2.2.1. C2fA Module

When a neural network extends to multiple convolutional layers, its ability to enhance feature representation learning becomes significant. However, increasing deep convolutional layers consumes large memory and computational resources, which is a primary challenge in constructing deep CNNs. To improve model performance without escalating computational complexity, attention mechanisms have emerged as an effective alternative [26]. These mechanisms strengthen the learning of discriminative features and are easily integrated into neural networks due to their flexible structure. The YOLO-Thyroid model incorporates the CA mechanism [27], which embeds positional information into channel attention. By employing one-dimensional pooling operations to capture feature encodings along horizontal and vertical directions, CA effectively integrates spatial coordinate information into attention maps. This enhancement improves the model's ability to perceive target positions and increases the accuracy of feature extraction.

Drawing inspiration from the CA mechanism, the C2fA module was designed and introduced. As illustrated in Figure 7, the C2fA module combines the efficient feature aggregation of the C2f module with the CA mechanism. The architecture of the C2fA module is detailed as follows.

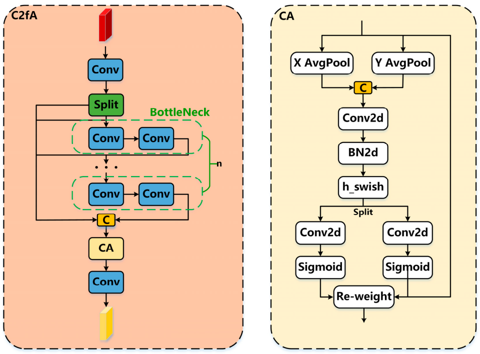


Figure 7. C2fA module design.

The module takes an input feature map $\mathbf{X}$ (a matrix) from the previous layer. First, $\mathbf{X}$ is split into two parts, $\mathbf{X}_1$ and $\mathbf{X}_2$:

$$\mathbf{X}_1, \mathbf{X}_2 = \mathrm{Split}(\mathbf{X}) \tag{1}$$

This splitting operation allows the network to process different portions of the features separately, enabling the diversification of the feature extraction process. The second part $\mathbf{X}_2$ is then passed sequentially through n bottleneck blocks, producing a series of intermediate outputs denoted as $\mathbf{X}_2^{(i)}$ for $i = 1, \ldots, n$. This process extracts higher-level features while reducing computational load. Each bottleneck operation can be described as:

$$\mathbf{X}_2^{(i)} = Bottleneck\left(\mathbf{X}_2^{(i-1)}\right),\ where\ \mathbf{X}_2^{(0)} = \mathbf{X}_2 \tag{2}$$

This bottleneck structure effectively reduces dimensionality and focuses on essential features, thereby enhancing computational efficiency. After the bottleneck transformation, the original and processed feature maps are concatenated to form the output:

$$\mathbf{Y} = \mathrm{Concat}\left(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_2^{(1)}, \ldots, \mathbf{X}_2^{(n)}\right) \tag{3}$$

The concatenation allows the network to combine unprocessed and processed features, providing a richer and more diverse set of features for subsequent layers. It ensures that both the original information and the enhanced features contribute to the learning process.

To enhance the spatial and channel-wise feature representation, a CA mechanism is applied to $\mathbf{Y}$. The output is then passed through another 1 × 1 convolution to adjust the channel size, producing the final output:

$$\mathrm{Output} = \mathrm{Conv}1 \times 1(CA(\mathbf{Y})) \tag{4}$$

By integrating the CA mechanism into the C2f layer, the C2fA module enables the model to identify important features more precisely, thereby improving the detection accuracy of nodules in different categories. Specifically, the last three C2f layers in the backbone network were replaced with C2fA modules. This strategic placement ensures that attention mechanisms are applied to higher-level feature maps. By integrating attention into feature processing, the model can better capture spatial information and important features while maintaining computational efficiency. The experimental results demonstrate that the model incorporating the C2fA module surpasses the original model in both accuracy and efficiency, verifying the effectiveness of the proposed method.

2.2.2. Loss Function

In nodule detection tasks, there is often a significant class imbalance among different categories: some categories have a large number of samples, while others are relatively scarce. This imbalance can cause the model to be biased toward predicting categories with more samples, thereby affecting the overall detection performance. To address this issue, a class-weighted binary cross-entropy (CW-BCE) loss function was introduced. By assigning appropriate weights to each category inversely proportional to its sample frequency, the model focuses more on underrepresented categories during training, thus improving detection performance for these categories.

The weight w_i for each category i is calculated based on its sample frequency:

$$w_i = \frac{N}{C \times n_i} \tag{5}$$

where N is the total number of samples in the dataset, C is the total number of categories, and n_i is the number of samples in category i. This formula ensures that categories with fewer samples receive larger weights, effectively balancing their influence during training. The CW-BCE loss for each sample is then defined as:

$$L_{\text{CW-BCE}} = -w_c[y\ log(p) + (1-y)\ log(1-p)] \tag{6}$$

Here, y is the true label of the sample, set to 1 if the sample belongs to category c and 0 otherwise. p is the predicted probability that the sample belongs to category c, obtained after applying the Sigmoid function to the model's output, and w_c is the weight corresponding to category c.

Consider three categories $[c_1,\ c_2,\ c_3]$ with the following number of samples: [1000, 500, 100] The total number of samples is $N = 1600$. The weights for each category $[w_1,\ w_2,\ w_3]$ are calculated as [0.53, 1.07, 5.33]. Category c_3, which has the fewest samples, receives the highest weight. This higher weight increases the contribution of category c_3 samples to the loss function, encouraging the model to focus more on accurately classifying these underrepresented samples during training. Class weights are incorporated into the loss function to adjust the loss contribution of each sample based on its category weight. This approach ensures that during training, the model pays more attention to underrepresented categories, reducing misclassification rates for these categories and effectively mitigating the impact of class imbalance on overall model performance.

2.2.3. Outcome Measure

To further enhance the localization accuracy of bounding boxes, the SIoU loss function [28] was introduced into the model. SIoU is an improved IoU loss that comprehensively considers the geometric relationship between the predicted box and the ground truth box, including overlap area, center point distance, area ratio, and shape differences. By simultaneously considering these discrepancies, the SIoU loss guides the model to learn more accurately, improving localization precision and regression accuracy. The SIoU loss function is expressed as:

$$L_{box} = 1 - \text{IoU} + \frac{\Delta + \Omega}{2} \tag{7}$$

$$\text{IoU} = \frac{intersection}{union} \tag{8}$$

where Δ represents the distance loss, quantifying the distance between the center points of the predicted box and the ground truth box, and incorporating an angle cost to make the penalty of the distance loss positively correlated with the angle difference; Ω represents the shape loss, penalizing the differences in width and height between the predicted box and the ground truth box; IoU is the Intersection over Union that calculates the ratio of the intersection area over the union area between the predicted box and the ground truth box. mAP (mean average precision) is a metric used to evaluate the performance of object detection models, representing the average precision across all classes at different threshold levels, providing a comprehensive assessment of the model's detection capabilities.

3. Results

3.1. Experimental Setting

To validate the effectiveness of the proposed YOLO-Thyroid model, comprehensive evaluations were conducted. The DDTI dataset [20], professionally labeled and preprocessed, was used in evaluation to ensure the reliability and validity of both the training and testing phases. Nodule diagnoses adhere to the TIRADS scoring system [21], encompassing six nodule categories (2, 3, 4a, 4b, 4c, and 5), including both benign (2 and 3) and malignant (4a, 4b, 4c, and 5) nodules. Table 2 summarizes the sample counts in the training, validation, and test sets of the augmented DDTI dataset. To enhance the dataset, data augmentation techniques were applied to the original dataset, including flipping [29], rotation [29], cropping [29], shearing [30], and brightness [31] adjustment. As a result, the number of samples in the augmented dataset increased from 339 to 813.

In the experiments, all methods were implemented using Ultralytics [19] on an NVIDIA GeForce RTX 3080 GPU (manufactured by NVIDIA Corporation, Santa Clara, California, USA) with 8704 CUDA cores and 10 GB of memory. For thyroid nodule detection, macro-average precision (P), macro-average recall (R), mAP0.5, and mAP0.5:0.95 were utilized to thoroughly evaluate model performance. Precision measures the accuracy of the model's positive predictions, while recall assesses the model's ability to identify all positive samples. mAP0.5 represents the mAP at an IoU threshold of 0.5, and mAP0.5:0.95 represents the mAP across IoU thresholds ranging from 0.5 to 0.95. Additionally, to assess the training and inference efficiency of the model, the training time (Tr, minutes per epoch), testing time (Te, milliseconds per image), parameters (Params), and the number of floating-point operations (FLOPs) were recorded. These metrics provide insights into the computational complexity and resource requirements, contributing to a comprehensive assessment of each model's efficiency and scalability. During model training, the batch size was set to 16, and training was conducted over 300 epochs. Dropout techniques and an early stopping strategy were employed to prevent overfitting.

3.2. Ablation Studies

To evaluate the impact of each module in the proposed model on ultrasound nodule detection performance, ablation experiments were conducted, with the results presented in Table 3. Starting from the baseline model YOLOv8-N, the C2fA module, the CW-BCE loss function, and the SIoU loss function were sequentially added.

As detailed in Table 3, the base YOLOv8-N model achieved a precision of 53.7%, recall of 37.4%, mAP0.5 of 37.4%, and mAP0.5:0.95 of 25.7%. Introducing the C2fA module increased precision to 67.0% but reduced recall to 28.9%, indicating a trade-off between accuracy and nodule detection. Employing the CW-BCE loss function resulted in an improvement of recall to 39.5% and an increase in the mAP metrics. These findings demonstrate that, although CW-BCE is a simple and widely used technique, its application in this manner effectively enhances the model's sensitivity to the minority class. This validates its effectiveness in the specific application of detecting thyroid nodules, where class imbalance poses a significant challenge. The SIoU loss increased recall significantly to 56.6% and improved mAP metrics, though precision decreased to 25.4%, suggesting more target detections but higher false positives due to boundary optimization. The integration of all components yielded the best overall performance, with mAP0.5 reaching 43.6%, mAP0.5:0.95 increasing to 28.7%, average detection precision being 54%, and the detection of nodules containing at least one suspicious feature recall of 58.2%, respectively. This synergy enhances feature extraction, addresses class imbalance, and optimizes boundary regression. In summary, the C2fA module improves feature representation, the weighted binary cross-entropy loss addresses class imbalance, and the SIoU loss enhances boundary localization. Together, they achieve a balance between precision and recall, significantly advancing ultrasound nodule detection performance and highlighting the effectiveness of the proposed method.

Table 3. Results of ablation experiments evaluating module impact.

Base Model	Components			P	R	mAP0.5	mAP0.5:0.95	Tr	Te
	C2fA	Lcwbce	Lsiou	(%)↑	(%)↑	(%)↑	(%)↑	(min/epoch) ↓	(ms/image) ↓
YOLOv8-N				53.7	37.4	37.4	25.7	0.057	9.5
	√			**67.0**	28.9	38.9	25.3	0.058	7.7
		√		35.3	39.5	40.5	28.2	**0.054**	8.6
			√	25.4	**56.6**	39.9	27.4	0.055	8.2
	√		√	55.8	33.5	36.4	25.5	0.059	8.2
	√	√		60.5	36.5	39.2	27.0	0.058	7.8
		√	√	62.1	34.2	39.1	27.3	0.055	**7.2**
	√	√	√	54.0	41.6	**43.6**	**28.7**	0.058	8.1

↑ indicates superior performance with a higher value, while ↓ indicates superior performance with a lower value. √ indicates the components existing. The best performance in the column was bolded.

Figure 8 illustrates the training and validation dynamics of the YOLO-Thyroid model over 300 epochs. The plots depict the evolution of various loss metrics, including box loss, class loss, and distribution-focused loss (DFL loss), alongside performance metrics such as macro-average precision, macro-average recall, mAP0.5, and mAP0.5:0.95. The training losses (top row) consistently decline, indicating effective learning and convergence, while the validation losses (bottom row) also decrease, suggesting good generalization capabilities in this specific dataset. Notably, precision and recall metrics progressively improve, reflecting the model's enhanced ability to accurately detect and classify nodules. The mAP measures show significant growth, underscoring the model's robust performance across varying IoU thresholds. These overall trends confirm the efficacy of the proposed model's improvements in enhancing detection accuracy and reliability.

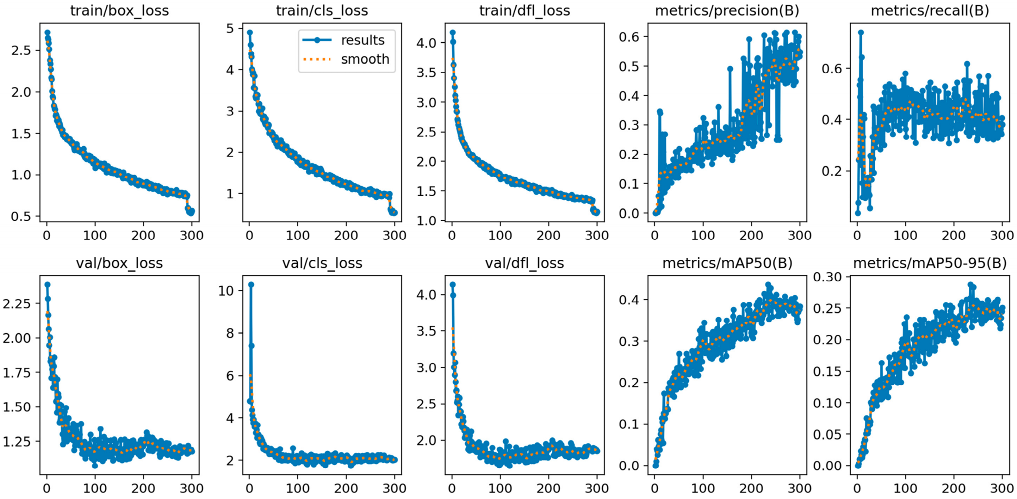


Figure 8. The training and validation dynamics of the YOLO-Thyroid model over 300 epochs. The orange dots indicate the smoothed curves.

3.3. Comparison with State-of-the-Art Methods

In this section, the proposed model is compared with state-of-the-art YOLO series models (YOLOv5 [32], YOLOv6 [33], YOLOv8 [19], YOLOv9 [34], YOLOv10 [35]), and DETR series models (RT-DETR-R50 [36], RT-DETR-L [36]). All models were trained and evaluated on the same dataset and under identical experimental conditions to ensure fairness and reliability in the comparison. The experimental results are presented in Table 4.

Table 4. The comparative performance of the proposed model with state-of-the-art YOLO and DETR series models.

Model Type	Model	mAvg-P (%) ↑	mAvg-R (%) ↑	mAP0.5 (%) ↑	mAP0.5:0.95 (%) ↑	Tr (min/epoch) ↓	Te (ms/image) ↓	Params (M) ↓	FLOPs (G) ↓
DETR	RT-DETR-R50	37.7	24.7	23.8	14.5	0.304	15.4	40.00	125.6
	RT-DETR-L	26.8	27.7	24.5	16.0	0.277	15.2	30.51	103.5
YOLO	YOLOv5-N	**59.9**	33.1	32.4	21.8	**0.055**	9.0	2.39	**7.1**
	YOLOv5-L	59.0	33.6	35.3	22.8	0.257	13.3	50.67	134.7
	YOLOv6-N	57.9	30.5	38.9	26.6	**0.055**	**8.0**	4.04	11.8
	YOLOv6-L	13.3	21.3	17.3	10.3	0.453	19.4	105.73	391.2
	YOLOv8-N	53.7	37.4	37.4	25.7	0.057	9.5	2.87	8.1
	YOLOv8-L	40.4	39.6	31.5	22.3	0.245	14.1	41.59	164.8
	YOLOv9-T	22.1	41.5	34.9	24.1	0.096	9.5	**1.88**	7.6
	YOLOv9-C	40.6	35.5	39.9	23.7	0.211	14.1	24.15	102.3
	YOLOv10-N	51.1	27.9	31.1	20.7	0.07	8.6	2.57	8.2
	YOLO-Thyroid	54.0	41.6	**43.6**	**28.7**	0.058	8.1	2.89	8.1

↑ indicates superior performance with a higher value, while ↓ indicates superior performance with a lower value. The best performance in the column was bolded.

Table 5 presents the performance of precision (P) and recall (R) for different TIRADS classes (4a, 4b, 4c, and 5), along with their weight average values (Avg). Precision indicates the proportion of correctly detected nodules containing at least one relevant feature among all detected nodules, while recall reflects the proportion of true nodules that were successfully detected with at least one suspicious feature.

YOLO-Thyroid demonstrated excellent performance across all metrics. Specifically, it achieved a precision of 54.0%, a recall of 58.2%, an mAP0.5 of 43.6%, and an mAP0.5:0.95 of 28.7%, outperforming the other models overall. This indicated that the proposed model can detect more true positives while reducing false detections, achieving a favorable balance between precision and recall. Additionally, YOLO-Thyroid had a parameter count of 2.89 M, FLOPs of 8.1 G, and an inference time of 8.1 milliseconds per image. It reduced the complexity and computational load while maintaining high accuracy, making it suitable for applications in resource-constrained environments.

Table 5. Performance of Precision (P) and Recall (R) Based on Nodules Containing at Least One Relevant Feature Across Different TIRADS Classes.

	4a	4b	4c	5	Weight Avg
P (%)	61.2	36.9	38.2	58	49.9
R (%)	68.5	57.1	42.9	55.9	58.2

In contrast, other models such as YOLOv5-N, YOLOv6-N, YOLOv8-N, YOLOv9-T, and YOLOv10-N, although they also had smaller model sizes, did not match YOLO-Thyroid in detection performance. Larger models like YOLOv9-C and YOLOv10-L, while showing improvements in some metrics, had significantly increased parameter counts and FLOPs, and their inference speeds decreased noticeably. Moreover, the DETR series models had larger scales and higher computational demands but did not exhibit corresponding advantages in detection performance and inference speed, rendering them inferior to YOLO-Thyroid. This further confirms the advantages of the YOLO-Thyroid model in structural design and optimization, achieving higher detection performance with a smaller model size. These results validate the effectiveness of YOLO-Thyroid for specific medical imaging applications, providing important technical support for diagnosis.

To visually illustrate the superior performance of the proposed method in nodule detection tasks, the actual detection results are presented in Figure 9. This figure compares YOLO-Thyroid with the detection results of other models. As shown in the figure, YOLO-Thyroid can more accurately locate and identify nodules, maintaining high detection accuracy even in complex scenarios and significantly reducing missed detections and false positives. This demonstrates the effectiveness and reliability of YOLO-Thyroid in practical applications.

The generalization ability of the model was assessed by evaluating its performance on both the original and augmented datasets. As shown in Table 6, the YOLO-Thyroid model demonstrated significant improvement with data augmentation: a macro-average recall from 34.5% to 41.6%, and an mAP0.5 from 33.0% to 43.6%. This improvement suggests that data augmentation enhanced the model's ability to learn robust features and generalize to unseen data.

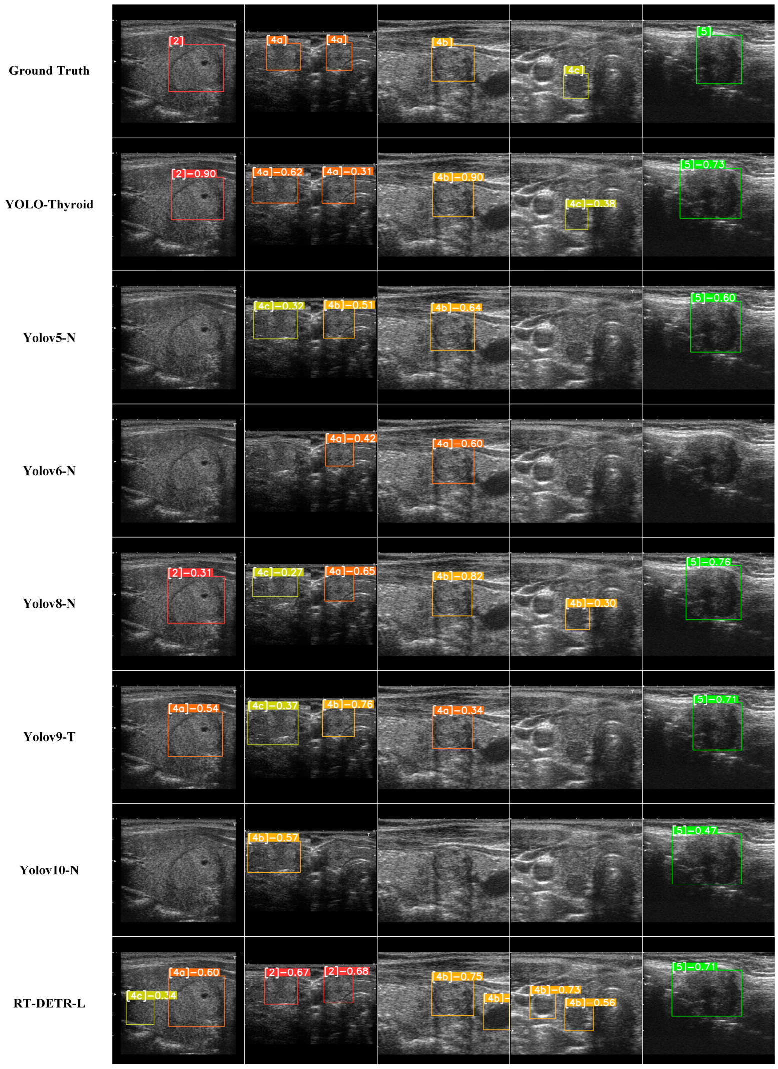


Figure 9. A visual comparison of the nodule detection results across different models. Each detected nodule is highlighted with a bounding box in different colors and labeled with a number in brackets, indicating the class, followed by the confidence score of the prediction.

Table 6. A performance comparison of the YOLO-Thyroid model on original and augmented datasets.

Dataset	mAvg-P (%) ↑	mAvg-R (%) ↑	mAP0.5 (%) ↑	mAP0.5:0.95 (%) ↑
Original	35.8	34.5	33.0	18.5
Augmented	54.0	41.6	43.6	28.7

↑ indicates superior performance with a higher value.

4. Discussion

This article presents the YOLO-Thyroid model, designed for the task of ultrasound nodule detection. YOLO-Thyroid effectively enhances detection performance through the introduction of the C2fA module and improved loss functions. The C2fA module combines spatial and feature information, enabling the model to better capture the features of nodules, particularly improving detection accuracy for small and complex nodules. This enhancement resulted in a precision increase by 18% and an improvement in mAP0.5 by 10.6% compared to the baseline model. The improved loss functions consider the class imbalance issue inherent in the dataset and multiple factors such as the target's position, size, and shape, leading to more accurate localization and reduced detection errors. This is evidenced by a detection macro-average recall increase by 7%. The experimental results demonstrate that this model outperforms current state-of-the-art object detection methods across various performance metrics. Specifically, YOLO-Thyroid achieved an increase in mAP0.5 to 43.6%, compared to 37.4% from the baseline model, indicating a substantial enhancement in overall detection accuracy. This outcome validates the effectiveness of YOLO-Thyroid for specific medical imaging tasks and provides significant technical support for clinical diagnosis.

In object detection tasks, such as thyroid nodule detection, mAP is the standard evaluation metric as it effectively captures both precision and recall across multiple classes and detection thresholds [37]. Precision reflects the model's ability to correctly identify relevant objects, expressed as the percentage of true positive predictions among all positive detections. Recall, on the other hand, measures the model's ability to detect all relevant cases, represented by the percentage of true positive predictions among all ground-truth bounding boxes [37,38]. mAP is widely adopted in major benchmarks and challenges, including PASCAL VOC [39] and COCO [40], because it provides a comprehensive assessment of a model's performance in both classification and localization. This dual capability is crucial in tasks where the accurate localization of thyroid nodules is essential.

Several studies have explored semantic segmentation models like SegNet [41], DeepLab [42], and PSPNet [43], which offer precise pixel-level delineation but come with significant drawbacks. DeepLab, for example, has high GPU utilization due to its complex structure, while models like PSPNet and SegNet are slower because of their intricate architectures, making them less efficient for real-time applications [44–46]. These models are computationally intensive, requiring more processing power and time, which limits their practicality in large-scale clinical screening [47], where speed is crucial. In contrast, YOLO, a one-stage framework, overcomes the shortcomings of two-stage methods like R-CNN [48] by simplifying the detection process. YOLO performs both object localization and classification in a single network, resulting in faster processing and reduced computational overhead [49]. This speed and efficiency make YOLO more suitable for real-time thyroid nodule detection, especially in resource-constrained environments, where rapid decision-making is essential [49,50].

Recent studies have integrated CA mechanisms with YOLOv8 in various applications [51,52]. However, their methods are not directly transferable to ultrasound nodule detection due to the unique challenges of medical ultrasound imaging. The proposed approach introduces a customized C2fA module specifically tailored for ultrasound nodule detection, involving modifications that differ from those in previous studies. Comparing YOLO-Thyroid to the two-stage detection method used by [53], which reports higher precision on the DDTI dataset, highlights important considerations for clinical applications. Two-stage detectors excel in precision due to their sequential proposal and refinement processes but are computationally intensive. Although the one-stage YOLO-Thyroid model has a lower precision, it offers significant advantages in inference speed and computational efficiency, which are critical for real-time ultrasound imaging and prompt clinical decision-making. Additionally, the datasets used in the two studies differ substantially. The DDTI dataset may contain higher-quality images or specific characteristics that favor higher detection precision, whereas the dataset used in this study encompasses a broader spectrum of ultrasound images with varying complexity and noise. These differences emphasize the importance of considering dataset characteristics when evaluating and comparing model performance.

This study further analyzes the model's ability to differentiate and detect nodules across different TIRADS classes to identify categories that require special attention for clinical decision-making. Sensitivity (measured through recall) was used as the primary indicator of the model's effectiveness in detecting nodules within each risk level, particularly for TIRADS 4a, 4b, 4c, and 5 classes.

For TIRADS classes 4a, 4b, and 4c:

- TIRADS 4a demonstrated the highest recall among these three categories, reflecting the model's ability to effectively detect nodules in this lower-risk class. However, nodules in TIRADS 4a generally have fewer malignant features, and their clinical urgency is relatively lower compared to TIRADS 4b and 4c. As a result, while maintaining high recall for 4a is important for ensuring comprehensive screening, it is not the most critical category for guiding clinical decision-making regarding further investigations.
- TIRADS 4b achieved a higher recall compared to TIRADS 4c, highlighting the model's stronger ability to identify nodules in this category. Given its maximum malignancy risk of 80%, TIRADS 4b represents a key decision-making threshold where detecting sufficient diagnostic features is critical for recommending further investigations, such as fine-needle aspiration biopsy (FNAB).
- TIRADS 4c, which shares the same maximum malignancy risk as 4b, exhibited a relatively lower recall. This suggests that further optimization is needed to improve detection for this category to ensure consistent performance across all high-risk classes.
- For TIRADS 5, while recall alone cannot fully evaluate its significance due to its minimum malignancy risk already exceeding 80%, this category remains clinically crucial. Nodules in TIRADS 5 often exhibit obvious malignant features, enabling quicker triage and more immediate clinical action. High detection performance in this category ensures that patients with overtly high-risk nodules are promptly referred for further investigation and treatment, which is vital for improving clinical efficiency.

Based on these findings, TIRADS 4b exhibits the highest detection sensitivity (i.e., two features suspected of being malignant, with a 10% to 80% malignant risk). Therefore, this risk class is recommended as the primary reference for supporting clinical decisions. Meanwhile, maintaining high recall for TIRADS 4a supports comprehensive screening efforts, and consistent performance in TIRADS 5 ensures the rapid identification and triage of overtly malignant nodules.

By establishing TIRADS 4b as the sensitivity benchmark and balancing detection performance across other categories, this study provides a quantitative framework for identifying high-risk nodules that require prompt follow-up. This approach ensures optimal utilization of clinical resources while minimizing missed diagnoses.

Despite the excellent results achieved by the YOLO-Thyroid model, there are still limitations and areas for improvement. First, the research is primarily based on a specific dataset, and the scale and diversity of this dataset may affect the model's generalization capability. In future work, the model's performance will be validated on larger and more diverse datasets to enhance the model's generalizability and robustness. Incorporating other advanced models or technologies, such as the MAMBA [54] and KAN [55] architectures, may further refine the model's performance. Additionally, exploring multimodal data integration, such as combining ultrasound images with patient demographic and clinical data, may improve the model's diagnostic capabilities. Furthermore, combining denoising and image enhancement techniques could reduce the impact of noise and variability in ultrasound images, thereby improving model efficacy. Potential applications of the YOLO-Thyroid model extend beyond thyroid nodule detection. By retraining the model on different medical imaging modalities, such as MRI or CT scans, it could detect other illnesses and abnormalities, thus expanding its diagnostic utility across the healthcare spectrum. Moreover, the model can be adapted for critical applications like fall detection in the elderly [56] by processing visual data from monitoring devices to provide real-time alerts and enhance safety. These adaptations demonstrate the model's versatility and its potential to significantly contribute to patient care and safety in diverse contexts.

5. Conclusions

In this paper, a YOLO-based model, YOLO-Thyroid, is proposed for ultrasound nodule detection. By introducing the C2fA module and improved loss functions, YOLO-Thyroid achieves an optimal balance between detection performance and model complexity. Through a series of ablation experiments, the effectiveness of the C2fA module and the new loss function in enhancing model performance has been verified. These improvements strengthen the model's ability to extract and represent nodule features, increasing detection accuracy for small and complex nodules. Simultaneously, the new loss function enables more precise boundary regression, reducing detection errors. Furthermore, comparative results with state-of-the-art models indicate that the YOLO-Thyroid model achieves superior performance across all evaluation metrics.

The mAP is a primary metric for evaluating detection performance, balancing precision and recall across classes and thresholds. Recall is particularly important in medical contexts to ensure all relevant cases are identified. In this study, the YOLO-Thyroid model was developed and optimized to achieve a high mean average precision (mAP) of 43.6% (mAP@0.5) and a recall of 58.2%. These results indicate superior performance compared to state-of-the-art models.

This research provides an efficient and reliable solution for automatic nodule detection, which is anticipated to play a significant role in clinical diagnosis. In future research, the dataset will be expanded by incorporating more ultrasound images from different devices and patient populations to enhance the model's applicability.

Author Contributions: Conceptualization, J.C.-W.C.; methodology, J.C.-W.C.; software, S.W. and Z.-A.Z.; validation, S.W. and Z.-A.Z.; formal analysis, Z.-A.Z. and Y.C.; investigation, Y.C.; resources, J.C.-W.C.; data curation, S.W. and Z.-A.Z.; writing—original draft preparation, S.W.; writing—review and editing, Y.-J.M. and J.C.-W.C.; supervision, J.C.-W.C.; project administration, J.C.-W.C.; funding acquisition, J.C.-W.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available from the Thyroid Digital Image Database at http://cimalab.unal.edu.co/?lang=en&mod=program&id=5 (accessed on 5 October 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ATA	American Thyroid Association
C2	CSP Bottleneck with Two Convolutions
CA	Coordinate Attention
CNNs	Convolutional Neural Networks
CSP	Cross Stage Partial
CT	Computed Tomography
CW-BCE	Class-Weighted Binary Cross-Entropy
DDTI	Digital Database of Thyroid Images
DFL	Loss Distribution-Focused Loss
FLOPs	Floating-Point Operations
FNAB	Fine-Needle Aspiration Biopsy
IoU	Intersection Over Union
MRI	Magnetic Resonance Imaging
NMS	Non-Maximum Suppression
P	Precision
mAvg-P	Macro-Average Precision
Params	Parameters
R	Recall
mAvg-R	Macro-Average Recall
SIoU	SCYLLA-IoU
Te	Testing Time
TIRADS	Thyroid Imaging Reporting And Data System
Tr	Training Time
US	Ultrasound
YOLO	You Only Look Once

Appendix A

The algorithm outlines the process of converting the original DDTI dataset's segmentation labels into the YOLO format required for object detection. For each image and its corresponding segmentation label, the algorithm extracts the minimum bounding rectangle that encompasses the target nodule by identifying the smallest and largest x and y coordinates where the nodule is present. It then calculates the center coordinates of this bounding box by averaging the minimum and maximum x and y values and determines the width and height by computing the difference between the maximum and minimum coordinates. The normalized center coordinates, width, and height, along with the class identifier of the nodule, are combined to form the YOLO label for each image.

Algorithm A1 Conversion of Original DDTI to YOLO Format

```
Input: Original DDTI with images and their corresponding segmentation labels
Output: YOLO-formatted label dataset
for each image $I_i$ and its segmentation label $L_i$ in Original DDTI do
    Load $I_i$ and $L_i$
        Extract bounding box coordinates:
            $x_{min}, x_{max} = min/max(x \mid L_i(x,y) = 1)$
            $y_{min}, y_{max} = min/max(y \mid L_i(x,y) = 1)$
    Compute YOLO format parameters:
        $x_{center} = (x_{min} + x_{max})/2$
        $y_{center} = (y_{min} + y_{max})/2$
        width $= x_{max} - x_{min}$
        height $= y_{max} - y_{min}$
    Normalize coordinates:
        $x_{center_{norm}} = x_{center}/\text{image\_width}$
        $y_{center_{norm}} = y_{center}/\text{image\_width}$
        $\text{width}_{\text{norm}} = \text{width}/\text{image\_width}$
        $\text{height}_{\text{norm}} = \text{height}/\text{image\_height}$
    Create YOLO label:
        $L_{yolo} = [class\_id, x_{center_{norm}}, y_{center_{norm}}, \text{width}_{\text{norm}}, \text{height}_{\text{norm}}$
    Save $L_{yolo}$ to YOLO label file
end for
```

References

1. Siegel, R.L.; Miller, K.D.; Fuchs, H.E.; Jemal, A. Cancer Statistics, 2022. *CA Cancer J. Clin.* **2022**, *72*, 7–33. [CrossRef] [PubMed]
2. Cancer of the Thyroid—Cancer Stat Facts. Available online: https://seer.cancer.gov/statfacts/html/thyro.html (accessed on 29 September 2024).
3. Mao, Y.-J.; Zha, L.-W.; Tam, A.Y.-C.; Lim, H.-J.; Cheung, A.K.-Y.; Zhang, Y.-Q.; Ni, M.; Cheung, J.C.-W.; Wong, D.W.-C. Endocrine Tumor Classification Via Machine-Learning-Based Elastography: A Systematic Scoping Review. *Cancers* **2023**, *15*, 837. [CrossRef] [PubMed]
4. Zhang, X.-Y.; Wei, Q.; Wu, G.-G.; Tang, Q.; Pan, X.-F.; Chen, G.-Q.; Zhang, D.; Dietrich, C.F.; Cui, X.-W. Artificial Intelligence-Based Ultrasound Elastography for Disease Evaluation-a Narrative Review. *Front. Oncol.* **2023**, *13*, 1197447. [CrossRef] [PubMed]
5. Zheng, Z.; Su, T.; Wang, Y.; Weng, Z.; Chai, J.; Bu, W.; Xu, J.; Chen, J. A Novel Ultrasound Image Diagnostic Method for Thyroid Nodules. *Sci. Rep.* **2023**, *13*, 1654. [CrossRef]
6. Hairu, L.; Yulan, P.; Yan, W.; Hong, A.; Xiaodong, Z.; Lichun, Y.; Kun, Y.; Ying, X.; Lisha, L.; Baoming, L.; et al. Elastography for the Diagnosis of High-Suspicion Thyroid Nodules Based on the 2015 American Thyroid Association Guidelines: A Multicenter Study. *BMC Endocr. Disord.* **2020**, *20*, 43. [CrossRef]
7. Iannuccilli, J.D.; Cronan, J.J.; Monchik, J.M. Risk for Malignancy of Thyroid Nodules as Assessed by Sonographic Criteria: The Need for Biopsy. *J. Ultrasound Med.* **2004**, *23*, 1455–1464. [CrossRef]
8. Sarkar, O.; Islam, R.; Syfullah, K.; Islam, T.; Ahamed, F.; Ahsan, M.; Haider, J. Multi-Scale Cnn: An Explainable Ai-Integrated Unique Deep Learning Framework for Lung-Affected Disease Classification. *Technologies* **2023**, *11*, 134. [CrossRef]
9. Khonina, S.N.; Kazanskiy, N.L.; Oseledets, I.V.; Nikonorov, A.V.; Butt, M.A. Synergy between Artificial Intelligence and Hyperspectral Imagining—A Review. *Technologies* **2024**, *12*, 163. [CrossRef]
10. Kshatri, S.S.; Singh, D. Convolutional Neural Network in Medical Image Analysis: A Review. *Arch. Comput. Methods Eng.* **2023**, *30*, 2793–2810. [CrossRef]
11. Zheng, T.; Qin, H.; Cui, Y.; Wang, R.; Zhao, W.; Zhang, S.; Geng, S.; Zhao, L. Segmentation of Thyroid Glands and Nodules in Ultrasound Images Using the Improved U-Net Architecture. *BMC Med. Imaging* **2023**, *23*, 56. [CrossRef]
12. Zhou, Y.-T.; Yang, T.-Y.; Han, X.-H.; Piao, J.-C. Thyroid-Detr: Thyroid Nodule Detection Model with Transformer in Ultrasound Images. *Biomed. Signal Process. Control* **2024**, *98*, 106762. [CrossRef]
13. Chen, G.; Tan, G.; Duan, M.; Pu, B.; Luo, H.; Li, S.; Li, K. Mlmseg: A Multi-View Learning Model for Ultrasound Thyroid Nodule Segmentation. *Comput. Biol. Med.* **2024**, *169*, 107898. [CrossRef] [PubMed]
14. Ghosh, K.; Bellinger, C.; Corizzo, R.; Branco, P.; Krawczyk, B.; Japkowicz, N. The Class Imbalance Problem in Deep Learning. *Mach. Learn.* **2024**, *113*, 4845–4901. [CrossRef]
15. Montalbo, F.J.P. A Computer-Aided Diagnosis of Brain Tumors Using a Fine-Tuned Yolo-Based Model with Transfer Learning. *KSII Trans. Internet Inf. Syst.* **2020**, *14*, 4816–4834.
16. Al-Antari, M.A.; Han, S.-M.; Kim, T.-S. Evaluation of Deep Learning Detection and Classification Towards Computer-Aided Diagnosis of Breast Lesions in Digital X-Ray Mammograms. *Comput. Methods Programs Biomed.* **2020**, *196*, 105584. [CrossRef] [PubMed]
17. Su, Y.; Liu, Q.; Xie, W.; Hu, P. Yolo-Logo: A Transformer-Based Yolo Segmentation Model for Breast Mass Detection and Segmentation in Digital Mammograms. *Comput. Methods Programs Biomed.* **2022**, *221*, 106903. [CrossRef]

18. Rouzrokh, P.; Ramazanian, T.; Wyles, C.C.; Philbrick, K.A.; Cai, J.C.; Taunton, M.J.; Kremers, H.M.; Lewallen, D.G.; Erickson, B.J. Deep Learning Artificial Intelligence Model for Assessment of Hip Dislocation Risk Following Primary Total Hip Arthroplasty from Postoperative Radiographs. *J. Arthroplast.* **2021**, *36*, 2197–2203.e3. [CrossRef]
19. Jocher, G.; Chaurasia, A.; Qiu, J. Ultralytics YOLOv8. 2023. Available online: https://github.com/ultralytics/ultralytics (accessed on 29 September 2024).
20. Pedraza, L.; Vargas, C.; Narváez, F.; Durán, O.; Muñoz, E.; Romero, E. An Open Access Thyroid Ultrasound Image Database. In Proceedings of the 10th International Symposium on Medical Information Processing and Analysis, Cartagena de Indias, Colombia, 14–16 October 2014.
21. Kwak, J.Y.; Han, K.H.; Yoon, J.H.; Moon, H.J.; Son, E.J.; Park, S.H.; Jung, H.K.; Choi, J.S.; Kim, B.M.; Kim, E.K. Thyroid Imaging Reporting and Data System for Us Features of Nodules: A Step in Establishing Better Stratification of Cancer Risk. *Radiology* **2011**, *260*, 892–899. [CrossRef]
22. Zheng, H.; Dong, Z.; Liu, T.; Zheng, H.; Wan, X.; Bao, J. Enhancing Gastrointestinal Submucosal Tumor Recognition in Endoscopic Ultrasonography: A Novel Multi-Attribute Guided Contextual Attention Network. *Expert Syst. Appl.* **2024**, *242*, 122725. [CrossRef]
23. Ding, X.; Liu, Y.; Zhao, J.; Wang, R.; Li, C.; Luo, Q.; Shen, C. A Novel Wavelet-Transform-Based Convolution Classification Network for Cervical Lymph Node Metastasis of Papillary Thyroid Carcinoma in Ultrasound Images. *Comput. Med. Imaging Graph.* **2023**, *109*, 102298. [CrossRef]
24. Garcea, F.; Serra, A.; Lamberti, F.; Morra, L. Data Augmentation for Medical Imaging: A Systematic Literature Review. *Comput. Biol. Med.* **2023**, *152*, 106391. [CrossRef]
25. Goceri, E. Medical Image Data Augmentation: Techniques, Comparisons and Interpretations. *Artif. Intell. Rev.* **2023**, *56*, 12561–12605. [CrossRef] [PubMed]
26. Zhu, H.; Xie, C.; Fei, Y.; Tao, H. Attention Mechanisms in Cnn-Based Single Image Super-Resolution: A Brief Review and a New Perspective. *Electronics* **2021**, *10*, 1187. [CrossRef]
27. Hou, Q.; Zhou, D.; Feng, J. Coordinate Attention for Efficient Mobile Network Design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2021, Nashville, TN, USA, 20–25 June 2021.
28. Gevorgyan, Z. Siou Loss: More Powerful Learning for Bounding Box Regression. *arXiv* **2022**, arXiv:2205.12740.
29. Maharana, K.; Mondal, S.; Nemade, B. A Review: Data Pre-Processing and Data Augmentation Techniques. *Glob. Transit. Proc.* **2022**, *3*, 91–99. [CrossRef]
30. Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; Yang, Y. Random Erasing Data Augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence 2020, New York, NY, USA, 7–12 February 2020.
31. Alomar, K.; Aysel, H.I.; Cai, X. Data Augmentation in Classification and Segmentation: A Survey and New Strategies. *J. Imaging* **2023**, *9*, 46. [CrossRef]
32. Jocher, G. Ultralytics YOLOv5. 2020. Available online: https://github.com/ultralytics/yolov5 (accessed on 29 September 2024). [CrossRef]
33. Li, C.; Li, L.; Geng, Y.; Jiang, H.; Cheng, M.; Zhang, B.; Ke, Z.; Xu, X.; Chu, X. Yolov6 V3. 0: A Full-Scale Reloading. *arXiv* **2023**, arXiv:2301.05586.
34. Wang, C.-Y.; Yeh, I.-H.; Liao, H.-Y.M. Yolov9: Learning What You Want to Learn Using Programmable Gradient Information. *arXiv* **2024**, arXiv:2402.13616.
35. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. Yolov10: Real-Time End-to-End Object Detection. *arXiv* **2024**, arXiv:2405.14458.
36. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. Detrs Beat Yolos on Real-Time Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024, Seattle, WA, USA, 17–21 June 2024.
37. Padilla, R.; Netto, S.L.; Da Silva, E.A. A Survey on Performance Metrics for Object-Detection Algorithms. In Proceedings of the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP) 2020, Niterói, Brazil, 1–3 July 2020.
38. Padilla, R.; Passos, W.L.; Dias, T.L.B.; Netto, S.L.; Da Silva, E.A.B. A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit. *Electronics* **2021**, *10*, 279. [CrossRef]
39. Everingham, M.; Eslami, S.A.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes Challenge: A Retrospective. *Int. J. Comput. Vis.* **2015**, *111*, 98–136. [CrossRef]
40. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft Coco: Common Objects in Context. In Proceedings of the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, 6–12 September 2014; Part V 13. pp. 740–755.
41. Saood, A.; Hatem, I. COVID-19 Lung Ct Image Segmentation Using Deep Learning Methods: U-Net Versus Segnet. *BMC Med. Imaging* **2021**, *21*, 19. [CrossRef] [PubMed]
42. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected Crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef] [PubMed]

43. Yuan, W.; Wang, J.; Xu, W. Shift Pooling Pspnet: Rethinking Pspnet for Building Extraction in Remote Sensing Images from Entire Local Feature Pooling. *Remote Sens.* **2022**, *14*, 4889. [CrossRef]
44. Yang, R.; Yu, Y. Artificial Convolutional Neural Network in Object Detection and Semantic Segmentation for Medical Imaging Analysis. *Front. Oncol.* **2021**, *11*, 638182. [CrossRef]
45. Cheng, L.; Xiong, R.; Wu, J.; Yan, X.; Yang, C.; Zhang, Y.; He, Y. Fast Segmentation Algorithm of Usv Accessible Area Based on Attention Fast Deeplabv3. *IEEE Sens. J.* **2024**, *24*, 24168–24177. [CrossRef]
46. Guo, Z.; Ma, D.; Luo, X. A Lightweight Semantic Segmentation Algorithm Integrating Ca and Eca-Net Modules. *Optoelectron. Lett.* **2024**, *20*, 568–576. [CrossRef]
47. Zeng, P.; Liu, S.; He, S.; Zheng, Q.; Wu, J.; Liu, Y.; Lyu, G.; Liu, P. Tuspm-Net: A Multi-Task Model for Thyroid Ultrasound Standard Plane Recognition and Detection of Key Anatomical Structures of the Thyroid. *Comput. Biol. Med.* **2023**, *163*, 107069. [CrossRef]
48. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2014, Columbus, OH, USA, 23–28 June 2014.
49. Aldughayfiq, B.; Ashfaq, F.; Jhanjhi, N.Z.; Humayun, M. Yolo-Based Deep Learning Model for Pressure Ulcer Detection and Classification. *Healthcare* **2023**, *11*, 1222. [CrossRef]
50. Ragab, M.G.; Abdulkadir, S.J.; Muneer, A.; Alqushaibi, A.; Sumiea, E.H.; Qureshi, R.; Al-Selwi, S.M.; Alhussian, H. Comprehensive Systematic Review of Yolo for Medical Object Detection (2018 to 2023). *IEEE Access* **2024**, *12*, 57815–57836. [CrossRef]
51. Li, T.; Liu, G.; Tan, S. Superficial Defect Detection for Concrete Bridges Using Yolov8 with Attention Mechanism and Deformation Convolution. *Appl. Sci.* **2024**, *14*, 5497. [CrossRef]
52. Yang, W.; Wu, J.; Zhang, J.; Gao, K.; Du, R.; Wu, Z.; Firkat, E.; Li, D. Deformable Convolution and Coordinate Attention for Fast Cattle Detection. *Comput. Electron. Agric.* **2023**, *211*, 108006. [CrossRef]
53. Gulame, M.B.; Dixit, V.V. Hybrid Deep Learning Assisted Multi Classification: Grading of Malignant Thyroid Nodules. *Int. J. Numer. Methods Biomed. Eng.* **2024**, *40*, e3824. [CrossRef] [PubMed]
54. Gu, A.; Dao, T. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. *arXiv* **2023**, arXiv:2312.00752.
55. Liu, Z.; Wang, Y.; Vaidya, S.; Ruehle, F.; Halverson, J.; Soljačić, M.; Hou, T.Y.; Tegmark, M. Kan: Kolmogorov-Arnold Networks. *arXiv* **2024**, arXiv:2404.19756.
56. Mao, Y.-J.; Tam, A.Y.-C.; Shea, Q.T.-K.; Zheng, Y.-P.; Cheung, J.C.-W. Enighttrack: Restraint-Free Depth-Camera-Based Surveillance and Alarm System for Fall Prevention Using Deep Learning Tracking. *Algorithms* **2023**, *16*, 477. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.