algorithms

MDPI

Article

UCINet: A Multi-Task Network for Umbilical Coiling Index Measurement in Obstetric Ultrasound

Zhuofu Liu [1,*], Lichen Niu [1], Zhixin Di [2] and Meimei Liu [3,*]

[1] The Higher Educational Key Laboratory for Measuring and Control Technology and Instrumentations of Heilongjiang Province, Harbin University of Science and Technology, Harbin 150080, China
[2] Department of Ultrasound Imaging, The Second Affiliated Hospital of Harbin Medical University, Harbin 150086, China
[3] Department of Obstetrics and Gynecology, The Second Affiliated Hospital of Harbin Medical University, Harbin 150086, China
* Correspondence: zhuofu_liu@hrbust.edu.cn (Z.L.); 100184@hrbmu.edu.cn (M.L.); Tel.: +86-139-0451-2205 (Z.L.)

Abstract

The umbilical coiling index (UCI), which quantifies the degree of vascular coiling in the umbilical cord, is a crucial indicator for assessing fetal intrauterine development and predicting perinatal outcomes. However, the existing methods for measuring the UCI primarily rely on manual assessment, which suffers from low efficiency and susceptibility to inter-observer variability. In response to the challenges in measuring the umbilical coiling index during obstetric ultrasound, we propose UCINet, a multi-task neural network engineered explicitly for this purpose. UCINet demonstrates enhanced operational efficiency and significantly improved accuracy in detection, catering to the nuanced requirements of obstetric imaging. Firstly, this paper proposes a Frequency–Spatial Domain Downsampling Module (FSDM) to extract features in both the frequency and spatial domains, thereby reducing the loss of umbilical cord features and enhancing their representational capacity. The proposed Multi-Receptive Field Feature Perception Module (MRPM) employs receptive fields of varying sizes across different stages of the feature maps, enhancing the richness of feature representation. This approach allows the model to capture a more diverse set of spatial information, contributing to improved overall performance in feature extraction. A Multi-Scale Feature Aggregation Module (MSAM) comprehensively leverages multi-scale features via a dynamic fusion mechanism, optimizing the integration of disparate feature scales for enhanced performance. In addition, the UCI dataset, which consisted of 2018 annotated ultrasound images, was constructed, each labeled with the number of vascular coils and keypoints at both ends of the umbilical cord. Compared with state-of-the-art methods, UCINet achieves consistent improvements across two tasks. In object detection, UCINet outperforms Deformable DETR-R50 with an improvement of 1.2% points in mAP@50. In keypoint localization, it further exceeds YOLOv11 with a 3.0% gain in mAP@50, highlighting its effectiveness in both detection accuracy and fine-grained keypoint prediction.

Keywords: umbilical coiling index; object detection; keypoint detection; multi-scale feature

check for updates

Academic Editor: Stefano Mariani

Received: 17 July 2025
Revised: 2 September 2025
Accepted: 17 September 2025
Published: 22 September 2025

Citation: Liu, Z.; Niu, L.; Di, Z.; Liu, M. UCINet: A Multi-Task Network for Umbilical Coiling Index Measurement in Obstetric Ultrasound. *Algorithms* **2025**, *18*, 592. https://doi.org/10.3390/a18090592

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The umbilical coiling index (UCI) is defined as the number of vascular coils per unit length of the umbilical cord. It is an important biophysical parameter for assessing the

tightness of umbilical coiling [1,2]. A regular coiling pattern helps maintain stable fetal blood flow and reduces the risk of umbilical cord compression. However, abnormally high or low UCI values may indicate potential pregnancy complications. These include intrauterine hypoxia, fetal growth restriction, preterm birth, and even fetal death [3,4]. Therefore, accurate assessment of the UCI is clinically valuable for monitoring fetal health during the perinatal period and for identifying pregnancy risks at an early stage.

Currently, the clinical assessment of the UCI primarily relies on manual observation and measurement based on ultrasound images. Experienced sonographers identify the coiling structure of the umbilical cord in 2D or color Doppler ultrasound images and estimate the number of coils per unit length [5,6]. Specifically, they count the number of vascular coils within a selected segment either visually or with image measurement tools. We then manually measure the length of that segment using built-in distance measurement tools, and we subsequently calculate the UCI. However, this manual method suffers from subjectivity, complexity, and low reproducibility. It is prone to errors when image quality is poor or coiling patterns are unclear, limiting its utility for large-scale screening and real-time decision-making. Some studies have explored image processing or traditional machine learning techniques to model umbilical cord structures [7–9]. Nevertheless, due to their limited ability in end-to-end optimization and high-level semantic feature extraction, such methods fail to meet practical clinical demands.

With the widespread adoption of deep learning in medical image analysis, convolutional neural network (CNN)-based methods for object detection and keypoint localization have opened new possibilities for automated and objective measurement of the UCI [10–15]. In recent years, some studies have applied deep learning techniques to the structural identification and quantitative analysis of fetal ultrasound images, including automated measurement of parameters such as fetal head circumference and femur length [16–19]. However, a lack of dedicated methods and datasets designed explicitly for UCI measurement exists.

This study proposes a multi-task network for umbilical coiling index measurement in obstetric ultrasound (UCINet) to address the aforementioned challenges. Using ultrasound images as input, UCINet classifies the number of umbilical coils and precisely localizes key points, enabling automated computation of the umbilical coiling index. Firstly, this paper proposes a Frequency–Spatial Domain Downsampling Module that extracts features jointly in the frequency and spatial domains to reduce feature loss effectively. Secondly, we propose a Multi-Receptive Field Feature Perception Module to enhance the network's adaptability to the complex morphology and scale variations inherent in umbilical cord images. Finally, we propose a Multi-Scale Aggregation Module to fully exploit multi-scale features by dynamically fusing multi-level feature representations. To support the training and evaluation of the method, this study also constructed the first dedicated dataset for measuring the umbilical cord entanglement index, the UCI dataset. Our contributions are summarized as follows:

We propose a Frequency–Spatial Domain Downsampling Module to reduce the loss of critical information in umbilical cord ultrasound images by jointly leveraging frequency and spatial domain features. This paper proposes a Multi-Receptive Field Feature Perception Module to enhance the network's capability in modeling the umbilical cord structure's complex morphology and scale variations. We present a Multi-Scale Feature Aggregation Module to effectively utilize multi-scale contextual information through the dynamic fusion of features across different levels. This paper proposes the UCI Dataset for umbilical coiling index detection, comprising many expert-annotated ultrasound images.

The experimental results demonstrate that the proposed UCINet outperforms the existing methods in both coil count recognition and keypoint localization tasks, exhibiting superior accuracy and robustness.

2. Related Work

2.1. Frequency Domain Features

In recent years, frequency domain features have received increasing attention in medical image analysis. Compared with conventional spatial-domain methods, frequency-domain features are more effective in capturing image edges, textures, and structural details, particularly in regions with repetitive patterns or low contrast. These features provide greater robustness and representational capacity under challenging imaging conditions. DWTNet [18] employs the Discrete Wavelet Transform (DWT) to decompose images into the frequency domain, thereby preserving more edge information and enhancing the network's perceptual ability. FFCNet [19] extracts full-frequency features using frequency filters to improve the model's responsiveness to diverse frequency components. DFANet [20] exploits frequency differences and applies attention mechanisms to filter and enhance feature maps, improving performance in building detection tasks. HWD [21] leverages the Haar wavelet transform to reduce the spatial resolution of feature maps while retaining as much informative content as possible. FeINFN [21] combines frequency-domain analysis with feature fusion techniques for multispectral and hyperspectral image fusion. Hayat has recently proposed lightweight attention mechanisms, such as Attention GhostUNet++, to refine feature representations in CT-based medical segmentation by jointly incorporating channel and spatial cues [22]. Unlike these recalibration-oriented approaches, our Frequency–Spatial Domain Downsampling Module (FSDM) aims to mitigate information loss in ultrasound images by jointly exploiting frequency- and spatial-domain features during downsampling.

2.2. Multi-Scale Feature Fusion

Feature maps at different scales often contain complementary information. Shallow features retain rich texture and edge details, which are beneficial for precise localization. Deeper features capture high-level semantic representations, which are crucial for object discrimination. Therefore, effective fusion of multi-scale features is essential for improving detection accuracy and model robustness. The Feature Pyramid Network (FPN) [23] achieves multi-level feature fusion through a top-down information flow mechanism and has demonstrated significant success. Building upon FPN, PANet [24] introduces a path enhancement module incorporating a bottom-up pathway, facilitating bidirectional communication between semantic and localization features. NAS-FPN [25] further advances this line of work by employing a neural architecture search to automatically identify optimal fusion strategies, achieving competitive performance on multiple object detection benchmarks. HPANet [26] proposes backward multi-scale feature fusion to enhance the processing ability of polyps of different scales. Beyond pyramid-style fusion, recent advances have explored lightweight and boundary-aware strategies for enhancing multi-scale representations in medical imaging. For example, TriConvUNeXt has demonstrated that lightweight multi-scale convolutional blocks—combining dilated, deformable, and depth-wise convolutions with channel shuffle—can significantly improve structural fidelity while reducing computational cost in biomedical segmentation tasks [27]. Similarly, Hayat et al. reviewed and proposed edge-guided super-resolution frameworks integrating channel–spatial attention with boundary-preserving modules, highlighting the clinical importance of multi-scale boundary sharpening in minimally invasive surgery imaging [28].

Our MSAM employs a dynamic weighting mechanism instead of relying on static edge sharpening or predetermined convolutional combinations. This approach allows the network to adaptively prioritize different scales, particularly in speckle noise, blurriness, or varying degrees of cord visibility. Such adaptability is crucial in umbilical cord

ultrasound imaging, where anatomical features can differ significantly across patients and imaging modalities.

2.3. Novelty Statement

Compared with existing frequency-domain or multi-scale architectures, UCINet makes three methodological advances. First, the proposed FSDM performs downsampling by jointly exploiting frequency- and spatial-domain cues, rather than relying on either domain in isolation, as in prior DWT- or filter-based networks. This dual-domain mechanism preserves fine vascular edges while maintaining semantic context, a property particularly critical for umbilical cord ultrasound. Second, the MRPM introduces a lightweight multi-branch pathway design that captures both local coil details and global cord morphology, going beyond generic pyramid pooling or attention schemes. Third, the MSAM departs from static top-down fusion (e.g., FPN, PANet) by employing adaptive weighting across scales. This feature lets the model prioritize the most reliable features in blurred or partially visible images. Collectively, these modules form a synergistic system tailored to the anatomical and clinical challenges of UCI measurement, strengthening the methodological contribution of UCINet relative to conventional frequency- and multi-scale approaches.

3. Method

3.1. Overview

As shown in Figure 1, the umbilical coiling index (UCI) computation based on UCINet involves the following steps. Firstly, the images are manually annotated and divided into training, validation, and test sets. The proposed UCINet is then trained on the constructed dataset to learn the model parameters. Finally, the trained UCINet is used to process the test images, producing outputs such as the number of umbilical coils and the distance between the two ends of the cord. The model can calculate the UCI based on these outputs, providing a quantitative basis for clinical assessment.

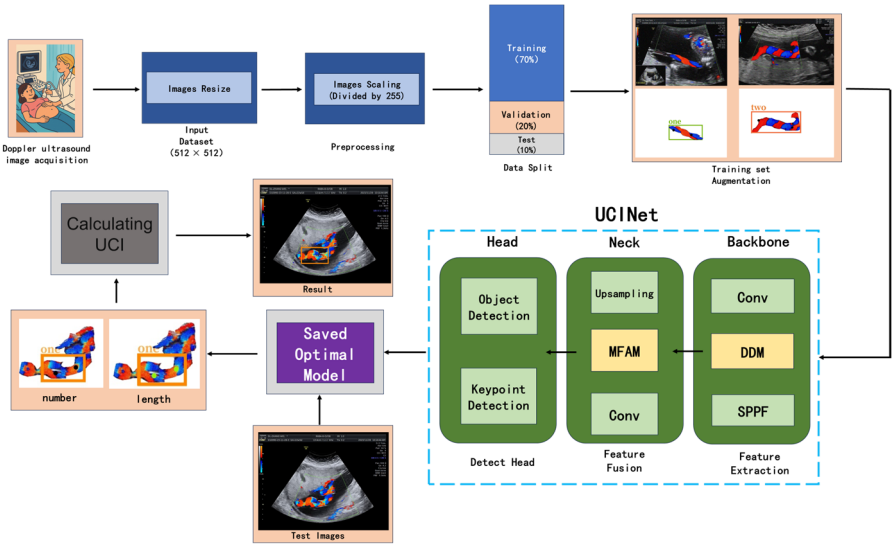


Figure 1. Overview of the UCINet-based umbilical coiling index calculation framework.

The overall architecture of UCINet is illustrated in Figure 2 and consists of three main components: Backbone, Neck, and Head. The Backbone is composed of a cascaded structure integrating the Frequency–Spatial Domain Downsampling Module (FSDM) and the Multi-Receptive Field Perception Module (MRPM), aiming to reduce feature loss and effectively extract multi-scale information from umbilical cord ultrasound images. The Neck incorporates key modules, such as the Multi-Scale Aggregation Module (MSAM)

and FSDM, to further fuse and enhance features across different scales, thereby improving the robustness and discriminability of feature representations. Finally, the Head performs object detection based on the fused features, outputting relevant information, such as the number of coils and the location of key structures, which serves as the foundation for subsequent UCI computation.

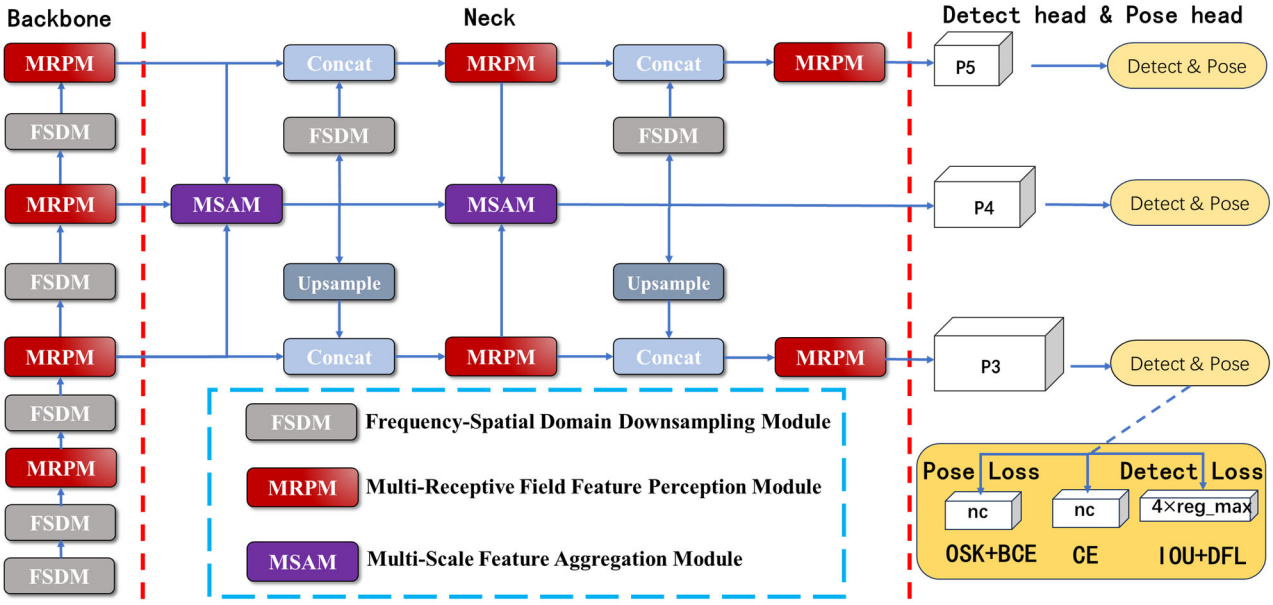


Figure 2. Overview of the proposed UCINet. UCINet consists of three main components: Backbone, Neck, and Head.

3.2. Frequency–Spatial Domain Downsampling Module

During feature extraction, networks typically generate multi-scale feature maps to meet the requirements of detecting objects at different scales. Most existing methods obtain feature maps at varying resolutions through standard convolutional downsampling operations. However, such operations inevitably lead to information redundancy compression and the loss of high-frequency details. Standard convolutions overlook critical diagnostic cues, such as edge structures, and delicate textures in medical images, such as umbilical cord ultrasound scans. These subtle features are essential for accurately determining the number of umbilical coils and localizing the start and end points of the coiling structure. We propose a Frequency–Spatial Domain Downsampling Module (FSDM) to address this issue. This module integrates frequency-domain and spatial-domain information to mitigate the loss of critical features during the downsampling process of umbilical cord images.

As illustrated in Figure 3, the proposed FSDM consists of two main branches: a frequency-domain downsampling branch and a spatial-domain downsampling branch. In the frequency-domain branch, a two-dimensional Haar wavelet transform first decomposes the input feature map $X[m, n]$. A low-pass filter H0(z) and a high-pass filter H1(z) are applied along the column direction, followed by the same filters applied along the row direction. This process yields four frequency sub-band components: C (Approximation), V (Vertical), D (Diagonal), and A (Horizontal), corresponding to the image's approximation content, vertical edge information, diagonal texture features, and horizontal edge information, respectively. This decomposition can be expressed as follows:

$$v_{1,\mathrm{L}}[m,n] = \sum_{k=0}^{K-1} x[m, 2n-k]\mathrm{H0}[k] \quad (1)$$

$$v_{1,\mathrm{H}}[m,n] = \sum_{k=0}^{K-1} x[m,2n-k]\mathrm{H1}[k] \tag{2}$$

$$C{=}\, v_{1,\mathrm{LL}}[m,n] = \sum_{k=0}^{K-1} v_{1,\mathrm{L}}[2m-k,n]\mathrm{H0}[k] \tag{3}$$

$$V{=}\, v_{1,\mathrm{HL}}[m,n] = \sum_{k=0}^{K-1} v_{1,\mathrm{H}}[2m-k,n]\mathrm{H1}[k] \tag{4}$$

$$D{=}\, v_{1,\mathrm{LH}}[m,n] = \sum_{k=0}^{K-1} v_{1,\mathrm{L}}[2m-k,n]\mathrm{H0}[k] \tag{5}$$

$$A = v_{1,\mathrm{HH}}[m,n] = \sum_{k=0}^{K-1} v_{1,\mathrm{H}}[2m-k,n]\mathrm{H1}[k] \tag{6}$$

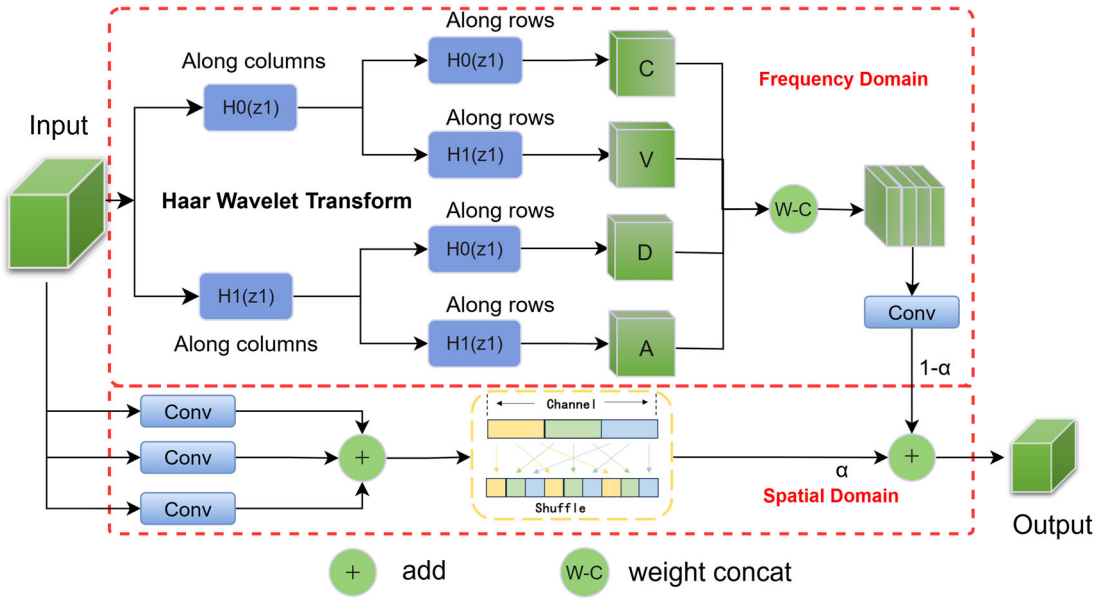


Figure 3. Details of the FSDM. The FSDM mainly consists of two components: frequency-domain downsampling and spatial-domain downsampling.

C, V, D, and A correspond to the low-frequency and three high-frequency features, respectively. Subsequently, a weighted channel concatenation mechanism is employed to integrate the four sub-band features, enhancing the representation capability of frequency-domain features. A convolutional operation is then applied to the fused features to extract the final frequency-domain feature map X_{f}.

In the spatial-domain downsampling branch, convolution operations with kernel sizes of 3, 5, and 7 are applied to the input feature map X to extract local structural features at different receptive fields. The resulting feature maps are then fused through element-wise addition to obtain an intermediate feature map X_{s1}:

$$\begin{aligned} X_{\mathrm{s1}}(i,j) = {} & \left(\textstyle\sum_{m=0}^{2}\sum_{n=0}^{2} X(i-m,j-n)\cdot K^{(3)}(m,n)\right) \\ & + \left(\textstyle\sum_{m=0}^{4}\sum_{n=0}^{4} X(i-m,j-n)\cdot K^{(5)}(m,n)\right) \\ & + \left(\textstyle\sum_{m=0}^{6}\sum_{n=0}^{6} X(i-m,j-n)\cdot K^{(7)}(m,n)\right) \end{aligned} \tag{7}$$

where $X_{\mathrm{s1}}(i,j)$ represents the result of element-wise addition of the results of three convolution operations at position $(i,\ j)$ in X_{s1} and $K^{(k)}$ denotes a kernel size of $k \times k$. A Channel Shuffle operation is applied to reorganize the channels further to enhance cross-channel

feature interaction, thereby strengthening semantic fusion among spatial structures and producing the spatial-domain feature map X_s.

Finally, the FSDM performs a weighted fusion of the frequency-domain and spatial-domain feature maps. A fusion weight parameter $\alpha \in [0,1]$ is introduced to control the relative contribution of each branch to the final output, resulting in the output feature map X_{out}:

$$X_{\text{out}} = \alpha \times X_s + (1 - \alpha) \times X_f \tag{8}$$

3.3. Multi-Receptive Field Feature Perception Module

In umbilical coiling index detection, extracting rich and multi-scale feature representations is crucial for subsequent feature fusion and precise localization, especially under challenging conditions, such as image blur, speckle noise, or background interference. We present a Multi-Receptive Field Feature Perception Module (MRPM) to tackle these challenges. As illustrated in Figure 4, MRPM integrates multi-receptive field convolutions with a lightweight SE-style channel attention mechanism, enabling UCINet to jointly capture local structural cues and global context. This design enhances the accuracy of coil counting and keypoint localization, thereby improving the robustness of UCI computation.

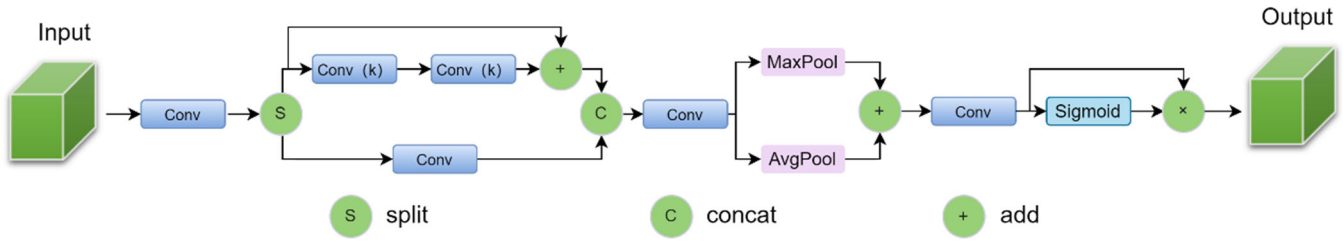


Figure 4. Illustration of the MRPM.

Then, the model evenly divides the compressed feature map along the channel dimension into $\times 1$ and $\times 2$, and each part undergoes feature extraction with different receptive fields.

Two consecutive $k \times k$ convolutions process the first part to enhance its sensitivity to local structures. The second part is transformed via a convolution, facilitating fine-grained feature reconstruction. These two parts are then concatenated along the channel dimension and passed through a fusion convolution to integrate multi-receptive field information, resulting in a richer spatial feature representation. This process can be formally expressed as follows:

$$\begin{array}{l} X_{11}(i,j) = X_1(i,j) + \sum_{m'=0}^{k-1} \sum_{n'=0}^{k-1} K^{(k)}(m',n') \cdot \\ \left(\sum_{m=0}^{k-1} \sum_{n=0}^{k-1} X_1(i-m-m', j-n-n') \cdot K^{(k)}(m,n) \right) \end{array} \tag{9}$$

$$X_{22}(i,j) = \text{Conv}(X_2(i,j)) = \sum_{m=0}^{2} \sum_{n=0}^{2} X_2(i-m, j-n) \cdot K(m,n) \tag{10}$$

$$\text{Concat}[X_{11}, X_{22}] = \begin{cases} X_{11}(i,j,p) & \text{if } 0 \le p \le C_{X11} \\ X_{22}(i,j,p - C_{X11}) & \text{if } C_{X11} \le p \le C_{X11} + C_{X22} \end{cases} \tag{11}$$

$$X_s = \text{Conv}(\text{Concat}[X_{11}, X_{22}]) \tag{12}$$

where the kernel size k takes values of 1, 3, 5, and 7 corresponding to the shallow-to-deep stages of the backbone network. Conv denotes the convolution operation, and Concat$[A, B]$ denotes the concatenation operation between A and B. Among them, C_{X11} and C_{X22} represent the maximum number of channels for X_{11} and X_{22}, respectively.

After obtaining the spatially enhanced feature map X_s, the MRPM further incorporates a channel attention mechanism to amplify the responses of key channels. Specifically, both max pooling and average pooling operations are applied to X_s to capture salient statistical

information, which are then fused (window = full spatial size, stride = 1). We process the fused features through a convolution layer followed by a sigmoid to generate channel attention weights. Finally, these attention weights are multiplied element-wise with the spatial feature map to produce the final output feature X_{out}:

$$P_{max}(X_s(i,j)) = \max_{0 \le m \le k_p, 0 \le n \le k_p} X_s(i \cdot s_p + m, j.s_p + n) \tag{13}$$

$$P_{avg}(X_s(i,j)) = \frac{1}{k_p{}^2} \sum_{m=0}^{k_p-1} \sum_{n=0}^{k_p-1} X_s(i \cdot s_p + m, j.s_p + n) \tag{14}$$

$$X_{c1}(i,j) = \mathrm{Conv}(P_{max}(i,j) + P_{avg}(i,j)) \tag{15}$$

$$X_{out} = X_{c1} \times \sigma \cdot X_{c1} \tag{16}$$

where $P_{max}(\ \ )$ and $P_{avg}(\ \ )$ represent max pooling and average pooling, respectively. k_p represents the window size of the two poolings, and s_p represents the stride. Finally, σ denotes a sigmoid operation.

3.4. Multi-Scale Feature Aggregation Module

Feature maps at different scales often contain complementary semantic information and spatial details. To further enhance the detection performance of object structures in umbilical cord images, particularly the feature representation capability under varying scales and complex deformations, this paper proposes a Multi-Scale Feature Aggregation Module (MSAM). As illustrated in Figure 5, this module effectively integrates contextual information from multiple spatial scales by jointly modeling downsampling, original, and upsampling pathways. Such collaborative modeling strengthens cross-scale feature interactions, thereby improving the perception of coil counts and positional variations of the umbilical cord.

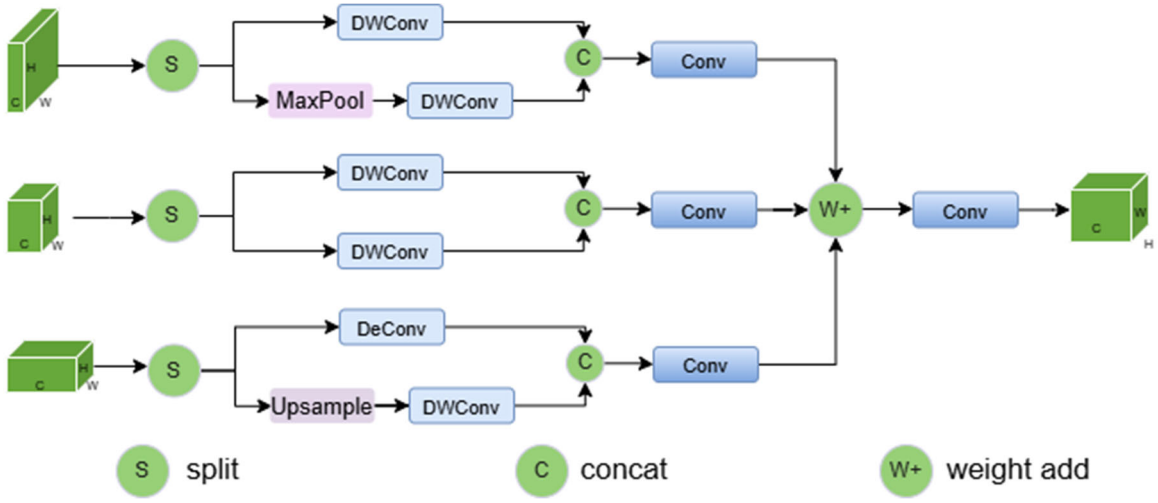


Figure 5. Illustration of the MSAM.

Specifically, the MSFAM comprises three parallel scale modeling pathways corresponding to the downsampling branch, the same-scale branch, and the upsampling branch. Each branch splits the input features evenly along the channel dimension into two parts, which are processed differently to enhance the diversity of channel-wise representations.

One part undergoes local feature extraction via depthwise separable convolution in the downsampling branch. In contrast, the other part is first subjected to max pooling to enlarge the receptive field before being processed by a depthwise separable convolution.

The two processed parts are concatenated and then fused using a standard convolution. In the same-scale branch, the paper separately processes both parts by depthwise separable convolutions, concatenates them, and then integrates them through a standard convolution. The model keeps one part at the original scale and feeds it into a depthwise separable convolution. At the same time, it upsamples the other part to enhance details prior to depthwise separable convolution. The model then combines the fused features through convolution. This process can be formalized as follows:

$$\begin{cases} X_1^t = X^t(i,j,z) & i \in (0,h), j \in (0,w), z \in (0,\frac{c}{2}) \\ X_2^t = X^t(i,j,z) & i \in (0,h), j \in (0,w), z \in (\frac{c}{2},c) \end{cases} \quad t \in [\text{low}, \text{mid}, \text{high}] \tag{17}$$

$$\text{DWConv}(\text{X}(i,j)) = \sum_{\text{m}=0}^{k-1} \sum_{\text{n}=0}^{k-1} X(i-m, j-n) \cdot K^{\text{DW}}(m,n) \tag{18}$$

$$\text{DeConv}(\text{X}(i,j)) = \sum_{\text{m}=0}^{k-1} \sum_{\text{n}=0}^{k-1} X(i-m, j-n) \cdot K^{\text{De}}(m,n) \tag{19}$$

$$X_{\text{low}}(i,j) = \text{Conv}(\text{Concat}(\text{DWConv}(X_1^{\text{low}}(i,j)), \text{DWConv}(\text{P}_{\max}(X_2^{\text{low}}(i,j))))) \tag{20}$$

$$X_{\text{mid}}(i,j) = \text{Conv}(\text{Concat}(\text{DWConv}(X_1^{\text{mid}}(i,j)), \text{DWConv}(X_2^{\text{mid}}(i,j)))) \tag{21}$$

$$X_{\text{high}(i,j)} = \text{Conv}(\text{Concat}(\text{DeConv}(X_1^{\text{high}}(i,j)), \text{DWConv}(\text{Upsample}(X_2^{\text{high}}(i,j))))) \tag{22}$$

Among these, $t \in [\text{low}, \text{mid}, \text{high}]$ represents three feature maps with different spatial resolutions, corresponding to the network's shallow, medium, and deep structures, as shown in Figure 5. DWConv( ) and DeConv( ) represent depth-separable convolution and transposed convolution, respectively, where K^{DW} and K^{De} denote the depth-separable convolution kernel and depth-separable convolution kernel, respectively. Conv() and Concat() are respectively represented as concatenation and convolution operations, as shown in Equations (10) and (11) above.

Finally, the features from the three scales are integrated through a weighted fusion mechanism to enhance semantic complementarity across different scales. The fused features are then passed through a standard convolution operation to produce the final output feature map X_{out}:

$$X_{\text{out}}(i,j) = \text{Conv}\left(X_{\text{low}}(i,j) \oplus X_{\text{mid}}(i,j) \oplus X_{\text{high}}(i,j)\right) \tag{23}$$

where $\oplus$ denotes a weighted fusion operation.

3.5. Loss Function

The loss function of UCINet consists of the following components to achieve multi-task joint optimization for umbilical coiling index detection. For the object detection task, the model employs the Intersection over Union (IoU) loss [29] to measure the overlap between predicted bounding boxes and ground truth boxes, defined as follows:

$$L_1 = 1 - \text{IoU} = 1 - \frac{\text{Area}_{\text{pre}} \cap \text{Area}_{\text{gt}}}{\text{Area}_{\text{pre}} \cup \text{Area}_{\text{gt}}} \tag{24}$$

In keypoint detection, considering the varying impact of different keypoints on the coiling index, a weighted L_1 loss, known as the Object Keypoint Similarity (OKS) loss [30], is employed to measure the spatial deviation between predicted and ground truth keypoints. Its formulation is expressed as follows:

$$L_2 = \frac{1}{K} \sum_{i=1}^{K} w_i \cdot \|p_i - g_i\|_1 \tag{25}$$

where K denotes the number of keypoints, w_i represents the importance weight of the i_{th} keypoint. p_i and g_i correspond to the predicted and ground truth coordinates, respectively.

The classification of coiling categories employs the standard cross-entropy loss function to measure the discrepancy between the predicted probabilities and the true labels, formulated as follows:

$$L_3 = -\sum_{c=1}^{C} y_c \log(p_c) \tag{26}$$

where y_c denotes the one-hot encoding of the ground truth class, and p_c represents the model's predicted probability for class c.

The Distribution Focal Loss (DFL) [31] is introduced to improve the accuracy of keypoint regression further. It evaluates the deviation between the predicted and ground truth distributions by computing the Kullback–Leibler (KL) divergence, defined as follows:

$$L_4 = \sum_{i=1}^{N} D_{\mathrm{KL}}\left(P_i^{\mathrm{pred}} \middle\| P_i^{\mathrm{gt}}\right) \tag{27}$$

In addition, to determine the visibility of the keypoints, a Binary Cross-Entropy (BCE) loss is introduced for visibility classification. The loss function is defined as follows:

$$L_5 = -[y \log(p) + (1-y)\log(1-p)],$$

where y denotes the ground truth visibility of the keypoint (1 for visible, 0 for invisible), and p represents the predicted visibility probability.

In summary, the L_{total} is defined as follows:

$$L_{\mathrm{total}} = \sum_{t=1}^{5} \lambda_t \cdot L_t \tag{28}$$

where λ_t denotes the fixed weight of each loss function.

3.6. Computation of the Umbilical Coiling Index

After completing structural recognition of the umbilical cord in ultrasound images, UCINet can automatically classify the number of vascular coils and accurately localize key anatomical landmarks annotated by clinicians. Based on these outputs, the system can further achieve a quantitative assessment of the umbilical coiling index by combining the predicted structural information with the spatial resolution of the image to compute the degree of coiling [32].

Specifically, after UCINet processes the Doppler ultrasound image of the umbilical cord, the system first predicts the number of vascular coils, where $n \in (1,2,3)$, and localizes key anatomical landmarks within the image. The model denotes the coordinates of the two predicted keypoints as $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$. The pixel distance L between the two points in the image is computed using the Euclidean distance formula, which is given by the following:

$$L = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{29}$$

Considering the variation in imaging scales across different ultrasound devices, a conversion factor α can be obtained based on the specific equipment model to translate pixel distance into real-world measurements. Accordingly, the actual distance between the two keypoints L_{true} can be calculated as follows:

$$L_{\mathrm{true}} = \alpha \times L \tag{30}$$

Since α is directly provided by the ultrasound device after calibration and remains stable across routine imaging settings, its small variability (<±5%) produces only negligible changes in the computed UCI, well below clinical decision thresholds (0.17 and 0.37). Therefore, the robustness of UCI estimation is not materially affected by α, ensuring reliable deployment across heterogeneous scanners (see Appendix B).

It should be emphasized that the coil count n is inherently discrete in clinical practice, as clinicians visually identify and count the number of vascular coils. Thus, modeling n as a discrete variable is sufficient and consistent with the standard clinical definition of UCI, ensuring interpretability and inter-rater consistency.

Using the following formula, we can further compute the umbilical coiling index by combining the predicted number of vascular coils n with the measured distance between keypoints:

$$\mathrm{UCI} = \frac{n}{L_{true}} \tag{31}$$

4. UCI Dataset

No publicly available multi-task ultrasound image dataset is designed explicitly for umbilical coiling index (UCI) computation. A dedicated ultrasound image dataset, UCI-Dataset, has been constructed to advance research in this field for umbilical coiling index detection and calculation.

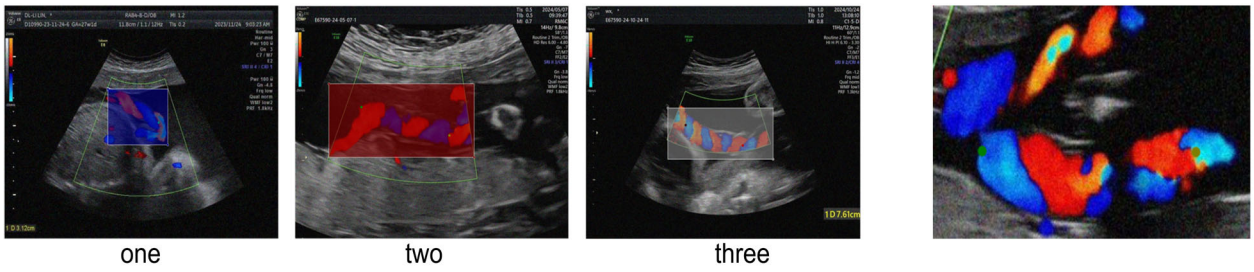
4.1. Data Acquisition

The image acquisition was conducted at the Second Affiliated Hospital of Harbin Medical University using the GE Voluson E10 ultrasound system (GE Healthcare, Chicago, IL, USA), which supports an adjustable operating frequency range of 1–18 MHz. Two thousand eighteen color Doppler ultrasound images of the umbilical cord were collected. The original image resolution was 1129 × 799 pixels.

This study received ethical approval from the Institutional Review Board of the Second Affiliated Hospital of Harbin Medical University (Approval No. KY2023-91) and the Research Ethics Committee of the School of Measurement and Control Engineering, Harbin University of Science and Technology (Approval No. 20241115). All the participants provided written informed consent and authorized the use of their examination data for subsequent scientific research.

4.2. Data Annotation and Splitting

We performed data annotation using the LabelMe tool, encompassing two tasks: object detection and keypoint detection. As illustrated in Figure 6a, for the object detection task, experienced clinicians annotated the number of umbilical cord coils in the ultrasound images, categorized into three classes: “one,” “two,” and “three,” corresponding to 1, 2, and 3 coils, respectively. After annotation, the dataset contained 612, 1004, and 402 instances for each class. For the keypoint detection task, as shown in Figure 6b, ultrasound specialists labeled key anatomical landmarks of the umbilical cord following standardized annotation protocols to facilitate subsequent calculation of the coiling index. For training and validation, the dataset was randomly split at a 7:2:1 ratio into training (1412 images), validation (403 images), and test (203 images) sets. The splitting strictly followed the subject-exclusive principle, ensuring that images from the same subject did not appear across different subsets, thereby preventing data leakage and enhancing external validity.



(**a**) (**b**)

Figure 6. Examples from the UCI Dataset. (**a**) Object detection labels, (**b**) Keypoint detection labels.

5. Experiments

5.1. Experimental Setting

The proposed method was trained utilizing the AdamW optimizer, configured with an initial learning rate of 0.001429, a momentum of 0.9, and a weight decay of 1×10^{-5}. We employed a batch size of 16 and conducted training over 300 epochs, focusing on object and keypoint detection tasks. All the baseline models—SSD, Faster R-CNN, various YOLO architectures, Deformable DETR-R50, HRFormer-S and HRNet—were trained under identical conditions to facilitate a rigorous comparison. We standardized the input resolution at 512×512 pixels and employed a uniform data augmentation strategy. Evaluation metrics, specifically the mean Average Precision (mAP), were computed using the official COCO API implementation to ensure consistency.

5.2. Dataset

Section 4 provides detailed information about the UCI dataset, which is used to evaluate UCINet's effectiveness. The datasets are the proposed UCI and the publicly available Ear210.

Ear210 Dataset: This dataset is primarily used for detecting acupoints in ear images and supports multi-task learning for object detection and keypoint localization. Keypoint annotations cover 21 categories. The dataset contains 210 images, with 168 used for training and 42 for testing. The image resolution is 3712×5568 pixels.

5.3. Evaluation Metrics

Several commonly used metrics are adopted to comprehensively assess performance on the umbilical coiling index (UCI) detection task, including Precision, Recall, mean Average Precision (mAP), Parameter count, and Frames Per Second (FPS).

Precision(R): The proportion of correctly predicted positive samples among all predicted positive ones. It reflects the accuracy of the model's predictions. Higher values indicate a lower false positive rate and more reliable detection results.

Recall: The proportion of actual positive samples that the model correctly detects. It measures the model's ability to avoid false negatives. A higher recall indicates better coverage of true positives.

mAP: A widely used metric in object detection that captures the balance between precision and recall. The mean Average Precision (mAP) is calculated by averaging the Average Precision (AP) scores for all classes. A higher mAP signifies superior overall detection performance across different categories.

Parameters: The model's number of learnable parameters reflects its size and complexity. A smaller parameter count indicates a lighter model that is easier to deploy in resource-constrained environments.

FPS (Frames Per Second): A measure of inference speed, indicating how many images the model can process per second. Higher FPS values denote greater efficiency, making the model more suitable for real-time clinical applications.

5.4. Comparison with the State-of-the-Arts

To validate UCINet's effectiveness, comparative experiments were conducted with existing methods on object detection and keypoint localization tasks, and the computed umbilical coiling index was further compared with manual measurements.

5.4.1. Object Detection Task

In the object detection task, comprehensive comparisons were made between UCINet and several mainstream models, including SSD [33], Faster R-CNN [34], YOLOv5s, YOLOv8s, YOLOv11s [35], and the transformer-based Deformable DETR-R50 [36]. The experiments were conducted on the UCI and public Ear210 datasets to thoroughly evaluate detection accuracy, model complexity, and inference efficiency across different scenarios.

On the UCI dataset, UCINet demonstrated consistent superiority across multiple metrics. As summarized in Table 1, UCINet achieves a mAP@50 of 84.2% and a mAP@50–95 of 61.5%, outperforming all CNN- and Transformer-based baselines. For instance, UCINet surpasses SSD with improvements of 15.3 percentage points in mAP@50 and 18.3% in mAP@50–95, while requiring only 38% of its parameters and achieving more than twice the inference speed. Compared with Faster R-CNN, UCINet reduces the parameter count by 32.2 million, improves mAP@50–95 by 12.6%, and achieves a 5.7-fold increase in FPS, confirming its strong real-time capability.

Table 1. Object detection results of UCINet and other methods on the UCI dataset. (↑ indicates that higher values represent better performance; ↓ indicates that lower values represent better performance. The best result in each column is highlighted in **bold**.)

Method	Precision (%) ↑	Recall (%) ↑	mAP50 (%) ↑	mAP50-95 (%) ↑	Params (M) ↓	FPS ↑
SSD	70.1	65.3	68.9	43.2	24.2	52
Faste R-CNN	74.5	68.0	71.6	48.9	41.4	19
Deformable DETR-R50	75.8	81.5	83.0	60.0	40.5	22
YOLOv5s	72.8	78.5	75.1	52.3	**7.0**	95
YOLOv8s	73.6	80.7	79.6	57.8	11.1	103
YOLOv11s	75.2	82.4	81.5	59.6	9.4	106
UCINet	**77.1**	**83.8**	**84.2**	**61.5**	9.2	**109**

The YOLO family baselines further highlight UCINet's efficiency. Despite having a model size comparable to YOLOv5s and YOLOv11s, UCINet achieves gains of 9.1% and 2.7% in mAP@50, respectively, while maintaining higher inference speed. Even against Deformable DETR-R50, which represents a more advanced transformer-based detector, UCINet provides improvements of 1.2% and 1.5% mAP@50–95 improvements, while reducing parameters from 40.5 M to 9.2 M and boosting FPS from 22 to 109. These results indicate that UCINet achieves more efficient feature representation and superior detection accuracy under clinical imaging conditions.

On the Ear210 dataset, UCINet further validated its generalization ability. As shown in Table 2, UCINet outperforms SSD and Faster R-CNN by 21.0% and 8.8% in mAP@50, and by 22.1% and 11.0% in mAP@50–95, respectively. Compared with YOLOv11s, UCINet achieves additional improvements of 1.1% in mAP@50 and 2.5% in mAP@50–95. The visualization results in Figure 7 further confirm that UCINet provides more accurate localization of

the umbilical cord region with tighter bounding boxes and higher confidence, whereas SSD frequently suffers from false positives. In summary, UCINet effectively balances detection accuracy and inference efficiency, maintaining a lightweight architecture while outperforming CNN- and Transformer-based baselines on both private and public datasets. These results confirm the model's superior generalization ability and highlight its potential for real-world deployment in clinical ultrasound analysis. As illustrated in Figure 8, further demonstrating the effectiveness of the proposed method.

Table 2. Object detection results of UCINet and other methods on the Ear210 dataset. (↑ indicates that a higher value represents better performance; ↓ indicates that lower values represent better performance. The best results in each column are highlighted in **bold**.)

Method	Precision (%) ↑	Recall (%) ↑	mAP50 (%) ↑	mAP50-95 (%) ↑
SSD	85.2	78.7	77.5	50.3
Faster R-CNN	91.2	86.7	89.7	61.4
Deformable DETR-R50	97.8	94.2	97.8	70.5
YOLOv5s	95.6	91.0	95.2	66.3
YOLOv8s	97.1	92.9	97.1	68.6
YOLOv11s	98.2	94.8	97.4	69.9
UCINet	**99.4**	**97.6**	**98.5**	**72.4**

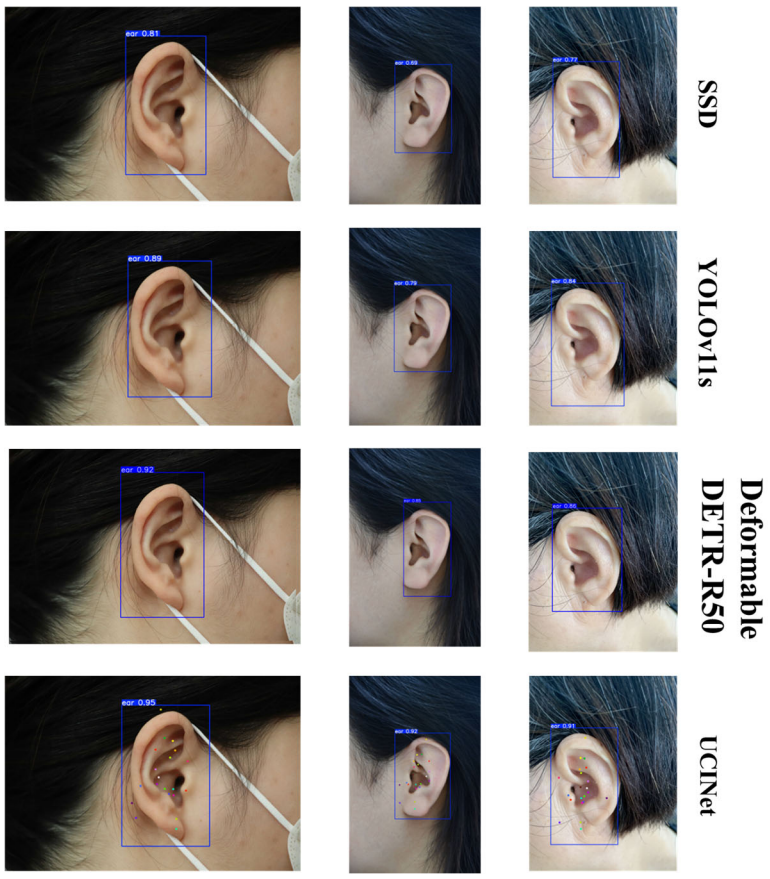


Figure 7. Visualization of object detection results of UCINet and other methods on the Ear210 dataset.

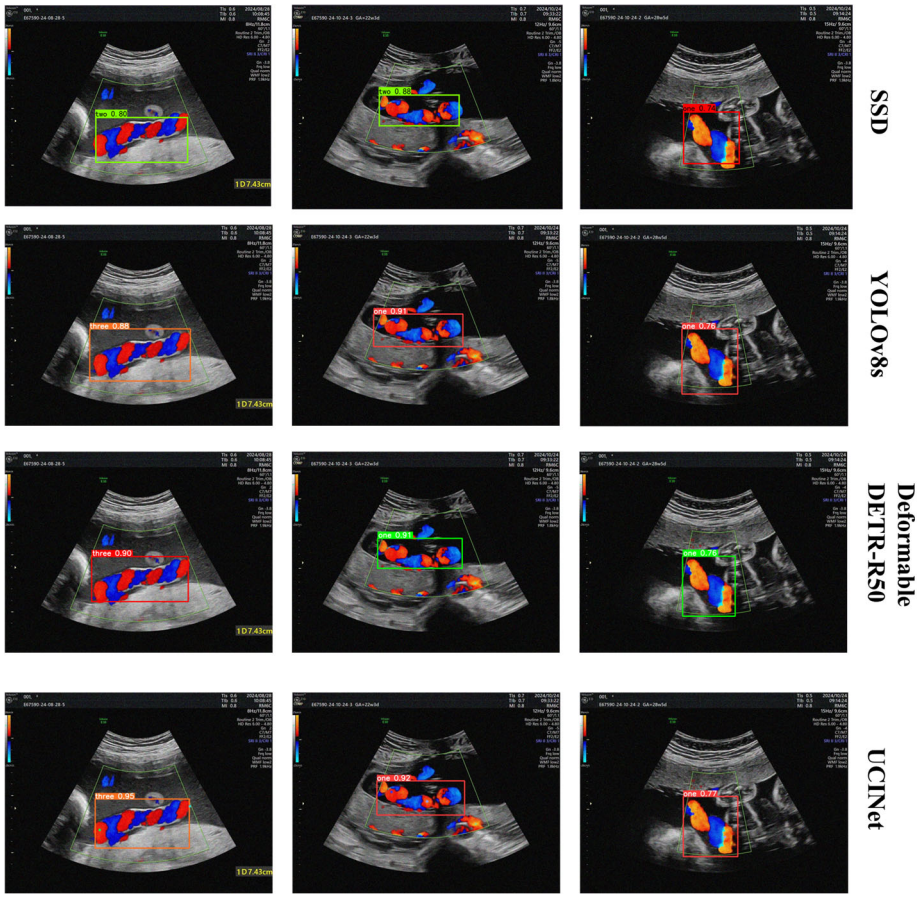


Figure 8. Visualization of object detection results of UCINet and other methods on the UCI dataset.

5.4.2. Keypoint Detection Task

For keypoint detection task, UCINetwas compared with several representative models, including the widely adopted HRNet [37], the transformer-based HRFormer-S [38], and lightweight approaches, such as YOLOv5_Pose and YOLOv11_Pose. We comprehensively evaluated the UCI and Ear210 datasets to analyze the model's performance across diverse domains.

On the UCI dataset, UCINet achieved the highest overall accuracy and efficiency. As summarized in Table 3, UCINet attains a mAP@50 of 88.4% and a mAP@50–95 of 75.4%, outperforming HRNet (+2.5% mAP@50, +0.8% mAP@50–95) and HRFormer-S (+1.3% mAP@50, +0.5% mAP@50–95), despite requiring only one-third of their parameters (9.2 M vs. 28.5–32.1 M). Furthermore, UCINet delivers 109 FPS, 5–6 times faster than HRNet and HRFormer-S, indicating superior real-time capability. Compared with lightweight baselines, UCINet provides notable accuracy gains of 4.9% and 5.3% in mAP@50 and mAP@50–95 over YOLOv5_Pose, and 1.6% and 2.6% over YOLOv11_Pose, while maintaining a comparable parameter budget, highlighting its enhanced localization precision and robustness. The visualization results in Figure 9 further illustrate that UCINet identifies umbilical cord endpoints with sharper localization and higher confidence, whereas YOLO-based baselines occasionally suffer from drift or partial misalignment.

Table 3. Keypoint detection results of UCINet and other methods on the UCI dataset. (↑ indicates that a higher value represents better performance; ↓ indicates that lower values represent better performance. The best results in each column are highlighted in **bold**.)

Method	Precision (%) ↑	Recall (%) ↑	mAP50 (%) ↑	mAP50-95 (%) ↑	Params (M) ↓	FPS ↑
HRNet	78.5	84.1	85.9	74.6	28.5	18
HRFormer-S	80.2	85.3	87.1	74.9	32.1	20
YO LOv5_Pose	79.6	79.2	83.5	70.1	**7.0**	92
YOLOv11_Pose	82.3	85.0	86.8	72.8	9.4	95
UCINet	**84.6**	**87.2**	**88.4**	**75.4**	9.2	**109**

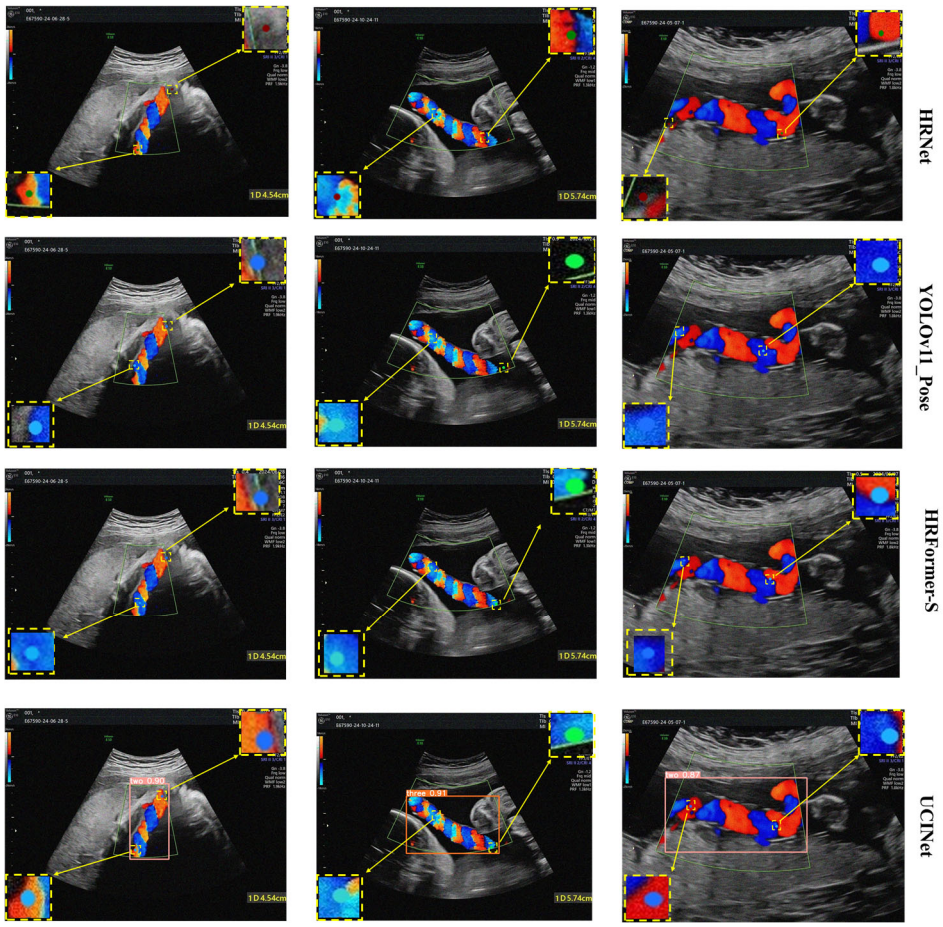


Figure 9. Visualization of keypoint detection results of UCINet and other methods on the UCI dataset.

On the Ear210 dataset, UCINet also demonstrates strong generalization ability. As shown in Table 4, UCINet achieves a mAP@50 of 86.2% and a mAP@50–95 of 50.5%, surpassing HRNetby 7.7% and 7.3%, HRFormer-S by 5.0% and 4.0%, and YOLOv5_Pose by 6.1% and 4.8%, respectively. Compared with the stronger YOLOv11_Pose baseline, UCINet achieves improvements of 3.0% in mAP@50 and 2.3% in mAP@50–95. Figure 10 illustrates that UCINet produces more reliable endpoint localization under challenging conditions, such as blurred edges or complex ear anatomy, where competing models show apparent deviations. In summary, UCINet effectively balances accuracy, efficiency, and model compactness in keypoint detection. It consistently outperforms CNN-, Transformer-, and YOLO-based baselines across private and public datasets, confirming its superior robustness and strong potential for deployment in clinical applications.

Table 4. Keypoint detection results of UCINet and other methods on the Ear210 dataset. (↑ indicates that a higher value represents better performance. The best results in each column are highlighted in **bold**.)

Method	Precision (%) ↑	Recall (%) ↑	mAP50 (%) ↑	mAP50-95 (%) ↑
HRNet	79.3	79.7	78.5	43.2
HRFormer-S	82.0	81.5	81.2	46.5
YO LOv5_Pose	81.1	80.2	80.1	45.7
YOLOv11_Pose	84.5	82.4	83.2	48.2
UCINet	**88.1**	**84.9**	**86.2**	**50.5**



Figure 10. Visualization of keypoint detection results of UCINet and other methods on the Ear210 dataset.

5.5. Calculation of the Umbilical Coiling Index

To evaluate the accuracy and reliability of the proposed UCINet model in calculating the umbilical cord entanglement index (UCI), descriptive and inferential statistical analyses were performed on a dataset containing 343 annotated ultrasound images. The actual UCI values, obtained by manual measurement and averaging by two experienced ultrasound physicians, were compared with the predicted UCI values generated by the model.

5.5.1. Descriptive Statistics

Table 5 and Figure 11 summarize the central tendency and dispersion metrics for actual and predicted UCI values. The observed UCI values demonstrated a mean of 0.4499

and a standard deviation of 0.0861, while the predicted values yielded a mean of 0.4474 with a standard deviation of 0.0925. The metrics indicate a close alignment between actual and predicted outcomes, with standard errors measuring 0.0046 for the observed values and 0.0050 for the predicted values. Both groups 95% confidence intervals showed a high degree of overlap, indicating strong consistency.

Table 5. Descriptive statistics of actual and predicted UCI values.

Variable	Sample Size (n)	Mean	SD	SEM	95% CI
Actual UCI	343	0.4499	0.0861	0.0046	[0.4408, 0.4590]
Predicted UCI	343	0.4474	0.0925	0.0050	[0.4376, 0.4572]

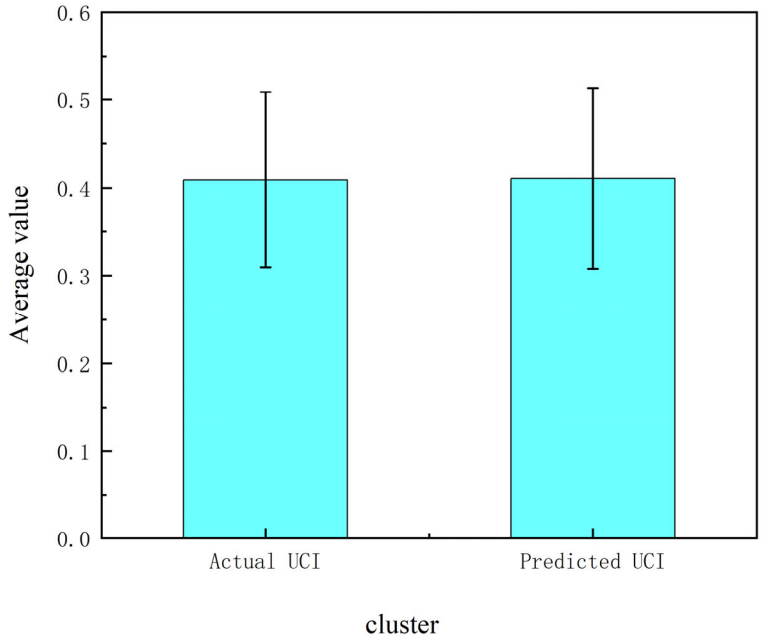


Figure 11. Plotting and comparing the actual value to the mean of the measured value. Error bars indicate standard error.

5.5.2. Paired-Samples *t*-Test

To assess whether the differences between predicted and actual UCI values were statistically significant, a paired-samples *t*-test was conducted under the following hypotheses:

Null Hypothesis (H_0): $\mu_{difference} = 0$ (no significant difference between actual and predicted UCI values);

Alternative Hypothesis (H_1): $\mu_{difference} \neq 0$.

The analysis results are summarized in Table 6. The computed t-statistic is 1.5601, based on a sample size of 342 degrees of freedom. The associated *p*-value of 0.300 confirms that the observed difference does not reach statistical significance. The calculated Cohen's d was 0.0842, suggesting a negligible effect size.

Table 6. Paired-samples *t*-test results.

Metric	Value
t-statistic	1.5601
Degrees of freedom (df)	342
p-value	0.300
Cohen's d	0.0842
Conclusion	No significant difference

5.5.3. Independent-Samples *t*-Test

To further validate the consistency between the two groups, an independent-samples *t*-test was performed. Although the paired-samples test is generally more appropriate for matched data, this analysis provides additional robustness.

Null Hypothesis (H_0): $\mu_{\text{actual}} = \mu_{\text{predicted}}$;

Alternative Hypothesis (H_1): $\mu_{\text{actual}} \neq \mu_{\text{predicted}}$.

As shown in Table 7, the *t*-statistic was 0.3690, with a *p*-value of 0.700 and an effect size η^2 of 0.0002, again supporting the absence of a significant difference.

Table 7. Independent-samples *t*-test results.

Metric	Value
t-statistic	0.3690
Degrees of freedom (df)	684
p-value	0.700
Effect size (η^2)	0.0002
Conclusion	No significant difference

5.5.4. IndependentBland–Altman Analysis

We performed a Bland–Altman analysis to assess the agreement between the automated predictive model outputs and the corresponding clinical reference values. The mean values of the model-predicted upper confidence interval (UCI) and the manual UCI reference—derived from the average of two independent observers—were plotted on the x-axis. The y-axis represents the difference between these two measurements. We computed the mean bias along with the 95% limits of agreement (LoA) to assess both systematic error and measurement variability. As shown in Figure 12, the analysis yielded a mean bias of -0.0007, indicating that the model predictions neither systematically overestimate nor underestimate UCI compared with manual assessment. The 95% LoA ranged from-0.096 to +0.095, with nearly all paired differences falling within this interval. This narrow margin of disagreement demonstrates that the automated approach achieves close concordance with manual measurements.

Given that the clinically relevant UCI range typically spans from 0.17 to 0.37, the observed discrepancy is relatively minor and unlikely to affect clinical decision-making. These findings confirm that UCINet provides clinically interchangeable UCI estimates and can be a reliable adjunct to manual evaluation in routine obstetric ultrasound practice.

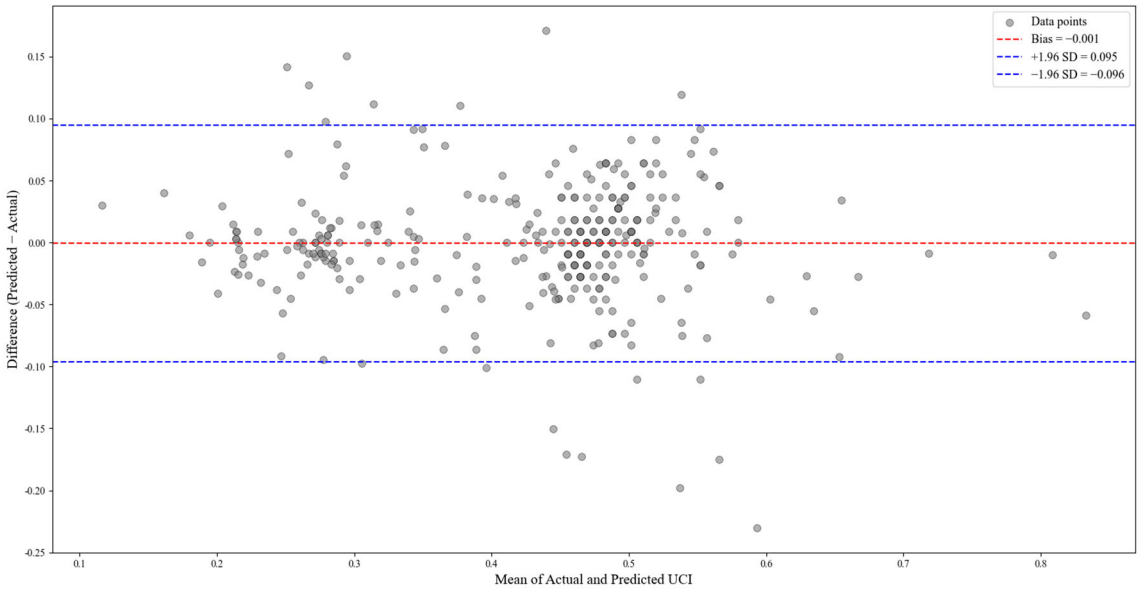


Figure 12. Bland–Altman analysis of UCI measurements.

5.5.5. Boundary-Focused Error Distribution

To further evaluate the clinical precision of UCINet beyond overlap-style detection metrics, we analyzed the absolute error of the umbilical coiling index (UCI) on a per-case basis. Figure 13 illustrates the distribution of absolute errors across all the test cases. The histogram (left) shows that the majority of predictions deviated by less than 0.05 from the manual reference, indicating tight clustering around the ground truth. The boxplot (right) further confirms the narrow dispersion, with only a few outliers approaching 0.10.

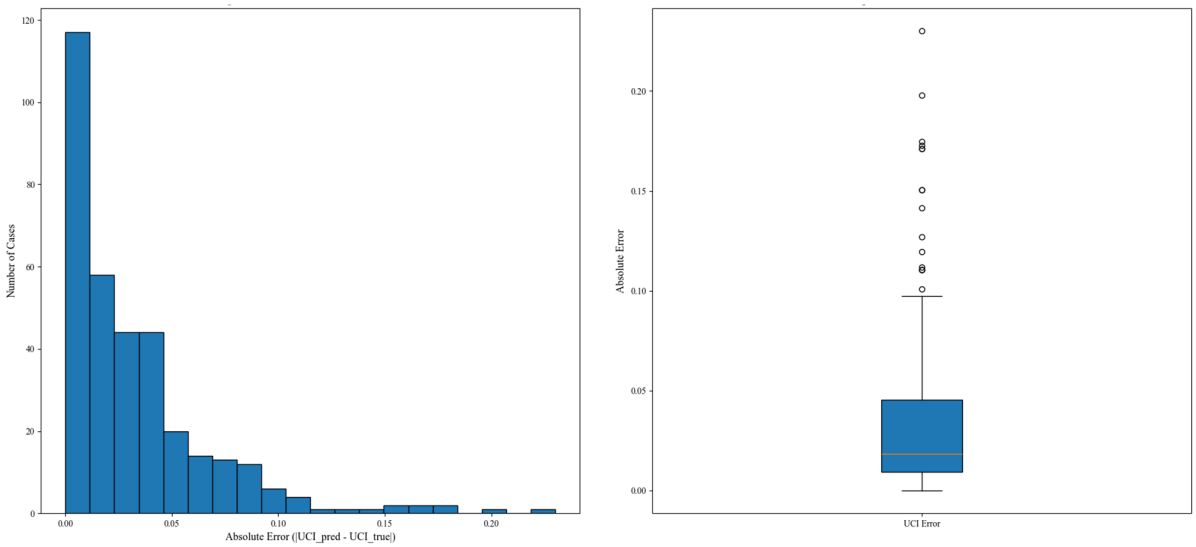


Figure 13. Distribution of per-case absolute errors in UCI estimation. **Left**: histogram of absolute errors across 343 test cases. **Right**: boxplot summarizing the error distribution.

Quantitatively, UCINet achieved a mean absolute error (MAE) of 0.0332, a median absolute error (MedAE) of 0.0184, and a 95th percentile absolute error of 0.0971. These results indicate improved overall accuracy and robustness compared with the previous version, particularly with a substantial reduction in median error, reflecting higher consistency across typical cases. Importantly, all the error values remain substantially below the clinically meaningful thresholds used to distinguish hypocoiled (<0.17), normocoiled (0.17–0.37),

and hypercoiled (>0.37) cords. These findings demonstrate that UCINetachieves clinically precise UCI estimation, with prediction errors unlikely to alter clinical classification.

5.6. Ablation Study

To evaluate the proposed modules' effectiveness, we conducted strengthened ablation experiments on the UCI dataset for both object detection and keypoint detection tasks, in accordance with the reviewers' recommendations. Specifically, we examined all single removals (eliminating FSDM, MRPM, or MSAM individually) and all pairwise combinations (retaining only one module), in addition to the complete model. To ensure statistical robustness, each configuration was trained and evaluated with five random seeds, and the results are reported as mean ± 95% confidence intervals. Moreover, parameter-matched controls were constructed by substituting each removed module with lightweight convolutional blocks of equivalent parameter count, thereby ruling out the possibility that performance gains were merely attributable to increased model capacity.

As shown in Table 8, the complete model (FSDM + MRPM + MSAM) yielded the best performance for the object detection task with mAP@50 = 84.2% ± 0.4 and mAP@50–95 = 61.5% ± 0.3. Removing FSDM caused a 7% drop in mAP@50 and 1.6% drop in mAP@50–95, indicating its critical role in preserving fine-grained features in umbilical cord ultrasound images. The joint extraction of frequency- and spatial-domain features is especially beneficial in enhancing model robustness in low-contrast regions. Excluding the MRPM caused declines of 2.0% in mAP@50 and 1.8% in mAP@50-95, demonstrating that multi-receptive field design effectively captures structural information at different scales and improves the model's adaptability to complex morphological variations of the umbilical cord. The MSAM also plays a key role in guiding cross-scale feature fusion. Its removal resulted in a 1.6% decrease in mAP@50 and a 1.4% decrease in mAP@50-95, confirming its contribution to multi-level contextual integration. Pairwise removals resulted in further degradation, with performance approaching that of the parameter-matched baseline, thereby confirming the complementary contributions of the three modules. Notably, the 95% confidence intervals of the improvements did not overlap with those of the baselines, verifying that the 1–2 point mAP gains are statistically significant rather than capacity-driven.

Table 8. Effectiveness of each module in the object detection task on the UCI dataset (mean ± 95% CI). (↑ indicates that a higher value represents better performance. The best results in each column are highlighted in **bold**).

FSDM	MRPM	MSAM	Precision (%) ↑	Recall (%) ↑	mAP50 (%) ↑	mAP50-95(%) ↑
√	√	√	**77.1 ± 0.3**	**83.8 ± 0.4**	**84.2 ± 0.4**	**61.5 ± 0.3**
	√	√	75.6 ± 0.3	82.8 ± 0.4	82.5 ± 0.4	59.9 ± 0.4
√		√	75.8 ± 0.4	82.3 ± 0.3	82.2 ± 0.4	59.7 ± 0.3
√	√		76.0 ± 0.4	82.6 ± 0.3	82.6 ± 0.4	60.1 ± 0.3
√			74.8 ± 0.4	81.5 ± 0.5	81.6 ± 0.3	58.9 ± 0.4
	√		74.5 ± 0.5	81.8 ± 0.4	81.4 ± 0.4	58.6 ± 0.5
		√	74.9 ± 0.4	82.0 ± 0.4	81.7 ± 0.3	59.0 ± 0.4

Similarly, as shown in Table 9, in the keypoint detection task, removing the FSDM caused a 1.3% drop in mAP@50 and 1.8% in mAP@50:95. When the MRPM was excluded, mAP@50 and mAP@50-95 decreased by 1.5% and 2.1%, respectively. Removing the MSAM led to a 1.2% drop in mAP@50 and 1.6% in mAP@50:95. Pairwise removals again led to further reductions, while the parameter-matched controls confirmed that the observed improvements were attributable to the proposed mechanisms rather than parameter scaling.

Table 9. Effectiveness of each module in the keypoint detection task on the UCI dataset (mean ± 95% CI). (↑ indicates that a higher value represents better performance. The best results in each column are highlighted in **bold**.)

FSDM	MRPM	MSAM	Precision (%) ↑	Recall (%) ↑	mAP50 (%) ↑	mAP50-95 (%) ↑
√	√	√	84.6 ± 0.3	87.2 ± 0.4	88.4 ± 0.3	75.4 ± 0.4
	√	√	83.1 ± 0.3	86.1 ± 0.3	87.1 ± 0.4	73.6 ± 0.4
√		√	83.3 ± 0.3	85.6 ± 0.4	86.9 ± 0.3	73.3 ± 0.3
√	√		83.7 ± 0.3	86.0 ± 0.3	87.2 ± 0.4	73.8 ± 0.4
√			82.4 ± 0.4	85.0 ± 0.3	86.0 ± 0.4	72.4 ± 0.4
	√		82.2 ± 0.4	85.2 ± 0.4	86.1 ± 0.3	72.2 ± 0.4
		√	82.6 ± 0.4	85.4 ± 0.3	86.3 ± 0.4	72.7 ± 0.3

Visual comparisons in Figure 14 further corroborate these findings: without FSDM, the model blurred vessel edges; without MRPM, it incompletely captured coiling structures; without MSAM, it produced inconsistent cross-scale feature fusion. These results demonstrate that all three modules are indispensable and mutually complementary, and their joint inclusion delivers statistically robust and clinically meaningful improvements in detection and keypoint localization.

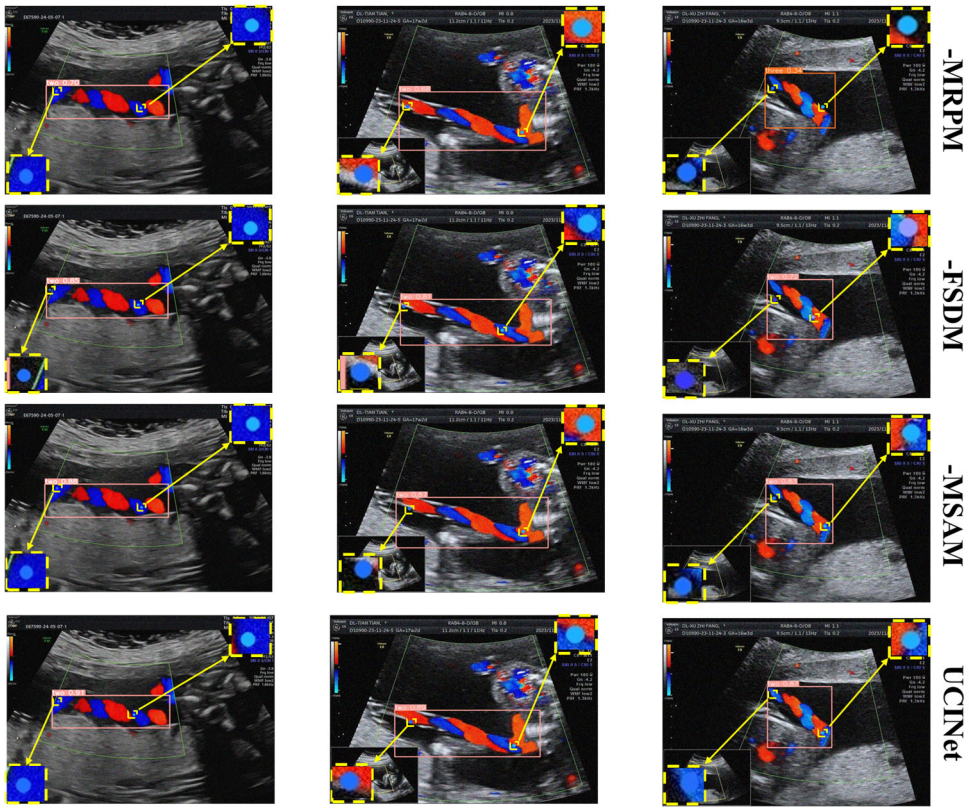


Figure 14. Visualization of the effects after removing each module.

6. Discussion

6.1. Comparative Analysis

The proposed UCINet demonstrates consistent and notable improvements over existing state-of-the-art methods in object detection and keypoint localization. Compared with classical CNN-based detectors, such as SSD and Faster R-CNN, UCINet achieves significantly higher detection accuracy—showing 15.3% and 12.6%increases in mAP@50,

respectively—while reducing parameter counts and substantially increasing inference speed. Compared with recent YOLO variants, UCINet surpasses YOLOv5s and YOLOv11s by 9.1% and 2.7% in mAP@50, respectively, while maintaining superior real-time efficiency. Furthermore, UCINet outperforms the transformer-based Deformable DETR-R50 by 1.2% in mAP@50 and by1.5% in mAP@50–95, using fewer than one-quarter of its parameters and achieving nearly fivefold higher FPS.

In the keypoint detection task, UCINet surpasses HRNet and HRFormer-S by 2.5% and 1.3% points in mAP@50, while requiring only one-third of their parameters and delivering a 5–6-fold increase in inference speed. Compared with lightweight baselines, such as YOLOv5_Pose and YOLOv11_Pose, UCINet achieves improvements of 4.9% and 1.6% in mAP@50, respectively, while maintaining a comparable model size. Experiments on the external Ear210 dataset further confirm the generalization ability of UCINet, demonstrating consistent advantages across datasets with different anatomical structures and imaging conditions.

The statistical evaluation of UCI computation further validated these findings. Both paired and independent samples *t*-tests confirmed no significant difference between UCINet-predicted and expert-measured UCI values ($p > 0.05$). Theeffect sizes (Cohen's d = 0.0842; $\eta^2 = 0.0002$) were negligible, while the Bland–Altman analysis indicated a mean bias of -0.0007 and 95% limits of agreement within ± 0.10. Moreover, the boundary-focused error analysis showed that most predictions deviated by less than 0.05 from the ground truth, with MAE = 0.0332 and MedAE = 0.0184—values well below clinical decision thresholds. These results confirm that UCINet delivers both statistical and clinical reliability.

6.2. Strengths, Limitations, and Future Work

The main strength of this study is the application of deep learning to automate the computation of the umbilical coiling index (UCI), replacing manual, error-prone measurement with an objective and reproducible process. UCINet integrates three tailored modules: the Frequency–Spatial Domain Downsampling Module (FSDM) for preserving fine textures and vessel edges, the Multi-Receptive Field Perception Module (MRPM) for capturing diverse cord morphologies, and the Multi-Scale Aggregation Module (MSAM) for adaptively balancing cross-scale features under noisy or partially visible conditions. Together, these components enhance detection accuracy, keypoint localization, and clinical reliability in obstetric ultrasound.

Several limitations of this study should be acknowledged. First, the error conversion factor α was derived through simulations rather than empirical measurements across different ultrasound scanners, zoom levels, and image overlays, which may affect the precision of the results. Second, the external validity is limited by the absence of cross-center and cross-device evaluations. Finally, the dataset provides insufficient coverage of rare scenarios, such as severe speckle noise, partial visibility, and structural abnormalities, which may reduce the robustness of the model in real-world settings. Future work will address these limitations by empirically calibrating α across multiple devices, implementing subject-exclusive and cross-center validations to enhance generalizability, and expanding the dataset to include challenging cases, such as severe artifacts and rare structural variations, thereby ensuring the clinical robustness and practical applicability of UCINet.

6.3. Clinical Significance

The umbilical coiling index (UCI) is a critical indicator of fetal health, with both low and high coiling patterns associated with hypoxia, intrauterine growth restriction, and other adverse perinatal outcomes. Manual measurement, however, is subjective and variable. UCINet provides an automated and reproducible solution, achieving close agreement with

expert annotations and mean absolute errors well below clinically relevant thresholds. With real-time performance and low variability, UCINet enables standardized UCI assessment, reduces observer dependence, and supports timely identification of high-risk pregnancies, offering strong potential for routine clinical integration.

7. Conclusions

We introduce UCINet, a multi-task neural network that measures the umbilical coiling index (UCI) from obstetric ultrasound data. UCINet executes object and keypoint detection to quantify the number of vascular coils and the necessary inter-point distances for UCI calculation.

To mitigate information loss caused by downsampling in ultrasound imaging, we developed a Frequency–Spatial Domain Downsampling Module (FSDM). This module integrates frequency-domain decomposition with spatial-domain compensation to maintain essential structural details effectively.

Additionally, we incorporate a Multi-Receptive Field Feature Perception Module (MRPM) that enhances feature diversity and representational capacity by leveraging convolutions across a range of receptive fields in the spatial domain.

To further improve the model's capability in capturing multi-scale information, we designed a Multi-Scale Aggregation Module (MSAM) that enables dynamic feature fusion across different scales, thus facilitating richer representation learning.

Author Contributions: Conceptualization, Z.L., Z.D., M.L., and L.N.; data collection and analysis, L.N., M.L., and Z.D.; methodology, Z.L. and L.N. original draft preparation, L.N.; review and editing, Z.L. and L.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and was approved by the Faculty Research Committee of Measuring and Control Technology and Instrumentation (No. 20241115) and the Institutional Review Board of the Second Affiliated Hospital of Harbin Medical University (No. KY2023-91).

Informed Consent Statement: Written informed consent was obtained from the subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors would like to express their gratitude for the participation of the volunteers. AI was merely used as a translation aid.

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

Appendix A

Note that training was conducted with a batch size of 16 to stabilize optimization and accelerate convergence, whereas runtime profiling was performed with a batch size of **1**. This choice reflects the real-time clinical deployment scenario, where ultrasound images are typically processed sequentially, and thus the latency per frame is the most relevant metric.

See Table A1 for details. To provide clarity for future deployment, we conducted detailed runtime and memory profiling of UCINet on an NVIDIA RTX 3060Ti (8 GB) with CUDA 12.1 and PyTorch 2.2 (automatic mixed precision enabled). The input resolution was fixed at 512×512. The latency was measured using CUDA events averaged over 1000 runs, while GPU memory consumption was recorded using torch.cuda.max_memory_allocated().

Table A1. Runtime breakdown and memory usage of UCINet on NVIDIA RTX 3060Ti (8 GB). ↓ indicates that lower values are better.

Stage/Module	Latency (ms) ↓	Share (%)	Notes
Pre-processing	0.4	4.3	Image resizing and normalization
Backbone (standard conv blocks)	2.4	26.1	Feature extraction layers
FSDM	1.1	12.0	Frequency–spatial dual-branch fusion
MRPM	0.9	9.8	Multi-receptive branches with channel attention
MSAM	1.5	16.3	Three-path adaptive multi-scale aggregation
Detection Head	0.8	8.7	Classification and box regression
Keypoint Decoding	1.4	15.2	Heatmap decoding and offset refinement
Post-processing (NMS, formatting)	0.7	7.6	NMS and tensor packing
Total	9.2	100	≈108.7 FPS

Appendix B

Figures A1 and A2 illustrate the decomposition results of a typical umbilical ultrasound image using a fixed Haar basis and a learned spectral filter, respectively.

As shown in Figure A1, the Haar basis decomposition retains most of the image energy in the low-frequency LL band while allocating a small proportion to the LH, HL, and HH sub-bands. Despite their lower energy, these high-frequency bands capture vertical, horizontal, and diagonal edge textures, which are crucial for automatic coil measurement. Moreover, the decomposition results are deterministic and consistent across the runs.

In contrast, Figure A2 shows that the learned spectral filter also achieves a frequency decomposition, but its energy allocation across sub-bands is more variable and sensitive to initialization and training. This instability reduces its interpretability and reproducibility compared to Haar decomposition.

Therefore, adopting a fixed Haar basis in FSDM is more justified, as it stably extracts diagnostically relevant edge and texture cues without requiring additional parameters or training, thereby improving interpretability and robustness.

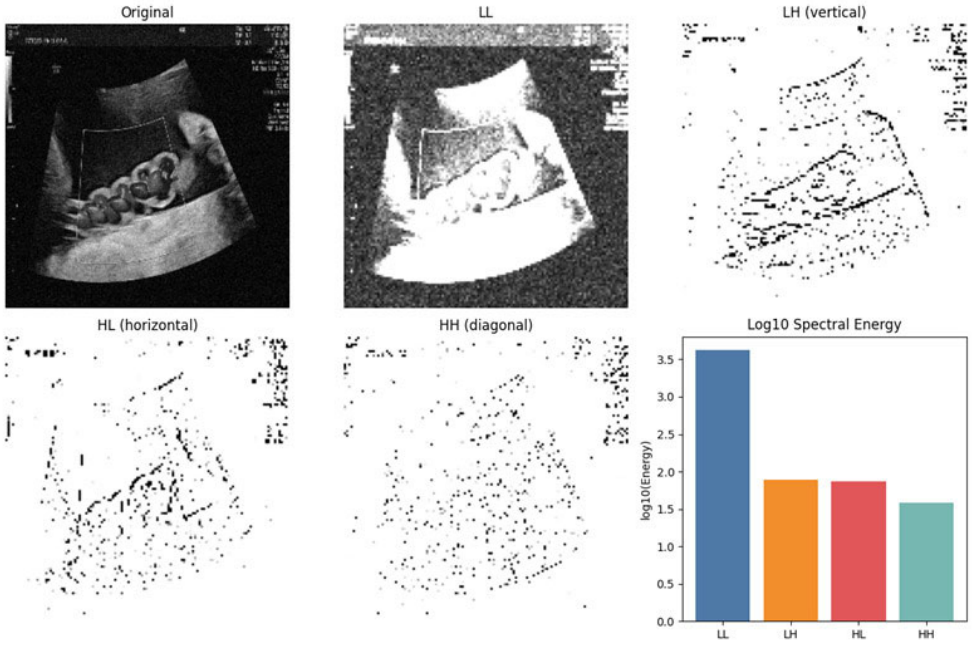


Figure A1. Decomposition of an umbilical ultrasound image using a fixed Haar basis. The LL sub-band preserves overall grayscale and structural information, while the LH, HL, and HH sub-bands, despite their lower energy, effectively highlight vertical, horizontal, and diagonal edge and texture features.

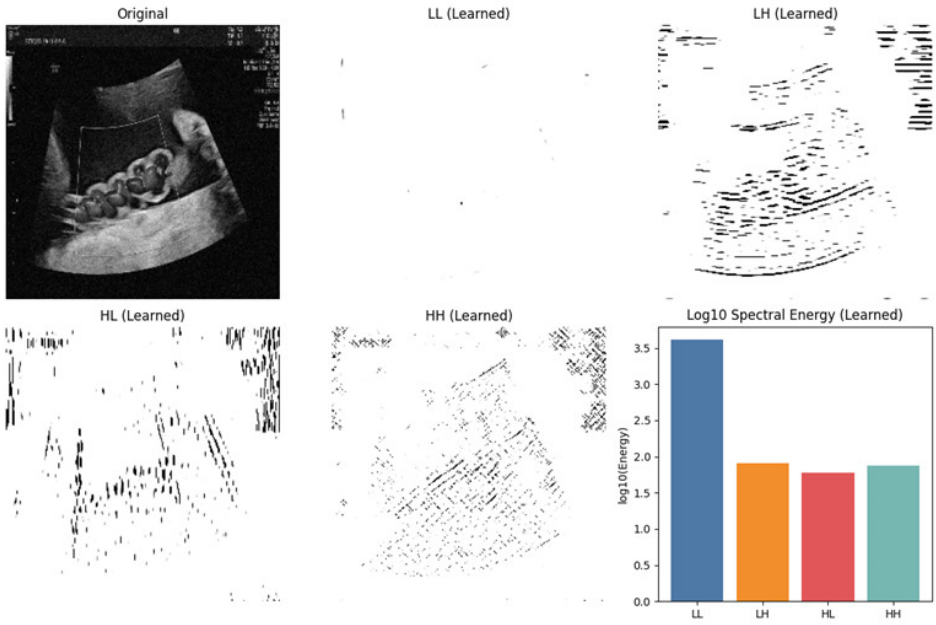


Figure A2. Decomposition of an umbilical ultrasound image using a learned spectral filter. Although it also separates low- and high-frequency information, its energy allocation across sub-bands is less stable and depends on initialization and training, making its directional interpretability and stability weaker than the fixed Haar basis.

Appendix C

To assess robustness against scanner variability, we perturbed the conversion factor α within $\pm 5\%$ across 343 test cases. As shown in Figure A3, both the mean absolute error (MAE) and the 95thpercentile absolute error remained stable, with the MAE ranging from 0.0377 to 0.0442 and the 95thpercentile error ranging from 0.1099 to 0.1269. These deviations are well below the clinical thresholds (0.17, 0.37), confirming that UCINet provides reliable UCI estimation even under realistic variations in α.

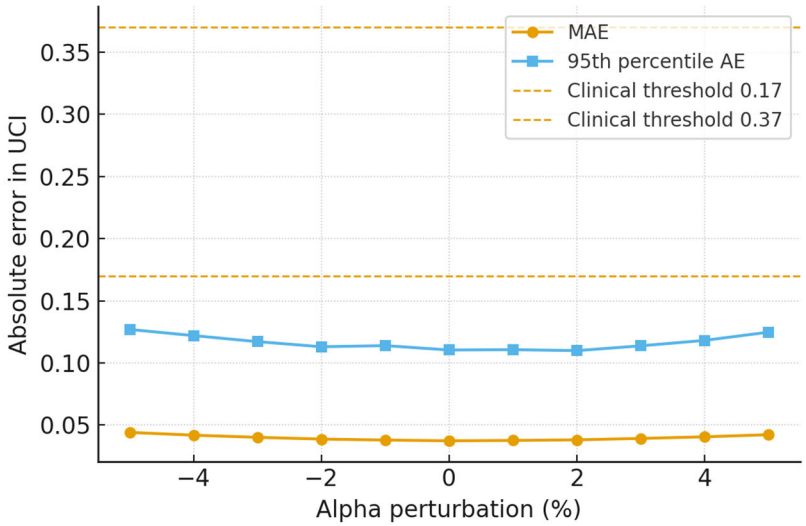


Figure A3. UCI error under $\pm 5\%$ perturbation of α (MAE and 95thpercentile AE curves).

References

1. Mittal, A.; Nanda, S.; Sen, J. Antenatal umbilical coiling index as a predictor of perinatal outcome. *Arch. Gynecol. Obstet.* **2015**, *291*, 763–768. [CrossRef]
2. Singh, S.; Pai, S.; Sahu, B. Study of umbilical coiling index and perinatal outcome. *Int. J. Reprod. Contracept. Obstet. Gynecol.* **2020**, *9*, 3977–3982. [CrossRef]

3. Kalluru, P.K.R.; Kalluru, H.R.; Allagadda, T.R.; Talur, M.; Gonepogu, M.C.; Gupta, S. Abnormal umbilical cord coiling and association with pregnancy factors. *J. Turk. Ger. Gynecol. Assoc.* **2024**, *25*, 44–52. [CrossRef] [PubMed]
4. Singireddy, N.; Chugh, A.; Bal, H.; Jadhav, S. Re-evaluation of umbilical cord coiling index in adverse pregnancy outcome–Does it have role in obstetric management? *Eur. J. Obstet. Gynecol. Reprod. Biol. X* **2024**, *21*, 100265. [CrossRef] [PubMed]
5. Ladella, S.; Wu, J. EP17. 04: Abnormal umbilical cord coiling index diagnosed prenatally is a predictor for adverse perinatal outcomes. *Ultrasound Obstet. Gynecol.* **2023**, *62*, 204–205. [CrossRef]
6. Kothari, A.; Gupta, S.; Gupta, V.K.; Shekhawat, U.; Shoaib, M. Umbilical Cord Coiling Index as a Prognostic Marker of Perinatal Outcome. *Eur. J. Mol. Clin. Med.* **2022**, *9*, 10141–10146.
7. Pradipta, G.A.; Wardoyo, R.; Musdholifah, A.; Sanjaya, I.N.H. Machine learning model for umbilical cord classification using combination coiling index and texture feature based on 2-D Doppler ultrasound images. *Health Inform. J.* **2022**, *28*, 1–19. [CrossRef]
8. Zope, A.M.; Zende, U.M.; Kumar, S.; Gautam, A. Optimizing fetal health assessment with ai-driven umbilical cord classification via 2-d doppler ultrasound imaging. *Obstet. Gynaecol. Forum* **2024**, *34*, 139–142.
9. Yousefpour Shahrivar, R.; Karami, F.; Karami, E. Enhancing fetal anomaly detection in ultrasonography images: A review of machine learning-based approaches. *Biomimetics* **2023**, *8*, 519. [CrossRef]
10. Li, X.; Chen, H.; Qi, X.; Dou, Q.; Fu, C.-W.; Heng, P.-A. H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans. Med. Imaging* **2018**, *37*, 2663–2674. [CrossRef]
11. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep high-resolution representation learning for human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–19 June 2019; pp. 5693–5703.
12. Tyagi, A.K.; Nair, M.M. Deep learning for clinical and health informatics. In *Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications*; Scrivener Publishing LLC: Beverly, MA, USA, 2021; pp. 107–129.
13. Im, S. Analysis of Trends of Medical Image Processing based on Deep Learning. *Int. J. Adv. Cult. Technol.* **2023**, *11*, 283–289.
14. Rajpurkar, P.; Lungren, M.P. The current and future state of AI interpretation of medical images. *N. Engl. J. Med.* **2023**, *388*, 1981–1990. [CrossRef] [PubMed]
15. Apostolopoulos, D.J.; Apostolopoulos, I.D.; Papathanasiou, N.D.; Spyridonidis, T.; Panayiotakis, G.S. Detection and Localisation of Abnormal Parathyroid Glands: An Explainable Deep Learning Approach. *Algorithms* **2022**, *15*, 455. [CrossRef]
16. Parvathavarthini, S.; Sharvanthika, K.; Sindhu, S.; Kaviya, K. Fetal head circumference measurement from ultrasound images using attention U-net. In Proceedings of the 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT), Erode, India, 22–24 February 2023; pp. 1–5.
17. D'Alberti, E.; Patey, O.; Smith, C.; Šalović, B.; Hernandez-Cruz, N.; Noble, J.A.; Papageorghiou, A.T. Artificial intelligence-enabled prenatal ultrasound for the detection of fetal cardiac abnormalities: A systematic review and meta-analysis. *eClinicalMedicine* **2025**, *84*, 103250. [CrossRef]
18. Chen, L.; Wu, X.; Ma, J.; Li, S.; Shi, Y.; Huang, M. DWT-Net: A Medical Image Segmentation Model Incorporating Frequency Domain Information. In Proceedings of the 2024 4th International Conference on Communication Technology and Information Technology (ICCTIT), Guangzhou, China, 27–29 December 2024; pp. 589–594.
19. Wang, K.-N.; He, Y.; Zhuang, S.; Miao, J.; He, X.; Zhou, P.; Yang, G.; Zhou, G.-Q.; Li, S. Ffcnet: Fourier transform-based frequency learning and complex convolutional network for colon disease classification. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Singapore, 18–22 September 2022; pp. 78–87.
20. Lu, L.; Liu, T.; Jiang, F.; Han, B.; Zhao, P.; Wang, G. DFANet: Denoising Frequency Attention Network for Building Footprint Extraction in Very-High-Resolution Remote Sensing Images. *Electronics* **2023**, *12*, 4592. [CrossRef]
21. Liang, Y.; Cao, Z.; Deng, S.; Dou, H.-X.; Deng, L.-J. Fourier-enhanced implicit neural fusion network for multispectral and hyperspectral image fusion. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 63441–63465.
22. Hayat, M.; Aramvith, S.; Bhattacharjee, S.; Ahmad, N. Attention ghostunet++: Enhanced segmentation of adipose tissue and liver in ct images. *arXiv* **2025**, arXiv:2504.11491.
23. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
24. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
25. Ghiasi, G.; Lin, T.-Y.; Le, Q.V. Nas-fpn: Learning scalable feature pyramid architecture for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–19 June 2019; pp. 7036–7045.
26. Ying, Y.; Li, H.; Zhong, Y.; Lin, M. HPANet: Hierarchical Path Aggregation Network with Pyramid Vision Transformers for Colorectal Polyp Segmentation. *Algorithms* **2025**, *18*, 281. [CrossRef]
27. Ma, C.; Gu, Y.; Wang, Z. TriConvUNeXt: A pure CNN-Based lightweight symmetrical network for biomedical image segmentation. *J. Imaging Inform. Med.* **2024**, *37*, 2311–2323. [CrossRef]

28. Hayat, M.; Gupta, M.; Suanpang, P.; Nanthaamornphong, A. Super-resolution methods for endoscopic imaging: A review. In Proceedings of the 2024 12th International Conference on Internet of Everything, Microwave, Embedded, Communication and Networks (IEMECON), Jaipur, India, 24–26 October 2024; pp. 1–6.
29. Yu, J.; Jiang, Y.; Wang, Z.; Cao, Z.; Huang, T. Unitbox: An advanced object detection network. In Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016; pp. 516–520.
30. 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.
31. Li, X.; Wang, W.; Wu, L.; Chen, S.; Hu, X.; Li, J.; Tang, J.; Yang, J. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 21002–21012.
32. Strong Jr, T.H.; Jarles, D.L.; Vega, J.S.; Feldman, D.B. The umbilical coiling index. *Am. J. Obstet. Gynecol.* **1994**, *170*, 29–32. [CrossRef]
33. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016.
34. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [CrossRef]
35. Khanam, R.; Hussain, M. Yolov11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725. [CrossRef]
36. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv* **2020**, arXiv:2010.04159.
37. Lu, B.; Sun, Y.; Yang, Z.; Song, R.; Jiang, H.; Liu, Y. HRNet: 3D object detection network for point cloud with hierarchical refinement. *Pattern Recognit.* **2024**, *149*, 110254. [CrossRef]
38. Yuan, Y.; Fu, R.; Huang, L.; Lin, W.; Zhang, C.; Chen, X.; Wang, J. Hrformer: High-resolution transformer for dense prediction. *arXiv* **2021**, arXiv:2110.09408. [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.