Supplementary materials

1 Watershed Analysis for CEUS branch

1.1 Watershed Analysis within frames

In CEUS videos, contrast agents enter thyroid nodules along blood vessels. Due to pathological differences, the enhanced pattern of contrast agent are different, and it play a key role in the diagnosis of benign and malignant thyroid nodules. To study the distribution characteristics of contrast agent in CEUS video, this paper introduces watershed analysis in remote sensing and combines it with the Lucas-Kanade (LK) Optical Flow method to analyze the enhancement direction (ED) mentioned in CEUS TI-RADS, as shown in Fig.1. Among them, centripetal, and both centrifugal are both considered to be one of the characteristics of malignant tumors. As an exclusive term, watershed analysis originally refers to the closed catchment area formed by the dividing line through which water or other materials flow through the public outlet in different directions. There are similarities between CEUS video and remote sensing basin analysis, in remote sensing, the path of water flow or other substances through the outlet forms a river network, while in CEUS, the region of contrast agent flows, closely related to the distribution of blood vessels, forms a contrast agent network.

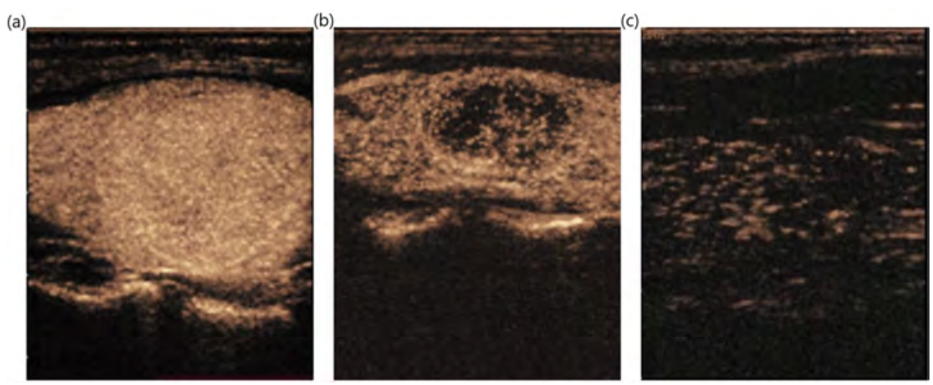

Fig. 1. Enhancement direction of contrast agents in CEUS video (a) scattered, (b) centripetal, and (c)centrifugal.

(i) Determine flow direction within frames: To track the flow relationship between contrast agent units and adjacent contrast agent units in each frame of CEUS video, this paper adapt D8 algorithm to digitally characterize the flow direction of contrast agent. This method assumes that there are only eight possible flow directions for each contrast media unit, of which the final flow direction is determined by the steepest descent method. By calculating the pixel weight drop value θ between the central grid cell and the adjacent 8 grid cells, the grid with the largest pixel weight drop value is identified as the contrast agent flow grid of the central grid.

$$\theta = arctan\left(\frac{h - h_i}{|D|}\right), i = 0, 1, 2, \ldots 7 \tag{1}$$

where $h = \sqrt{r^2 + g^2 + b^2}$ represents the pixel value of the central raster cell. h_i denotes the pixel value of adjacent raster cells, where D is the distance between h and h_i, which is 1 and $\sqrt{2}$ when h is adjacent to h_i horizontally or diagonally, respectively. In a 3×3 contrast agent grid cell, the 8 possible flow directions of each contrast agent cell are represented by the numbers 1, 2, 4, 8, 16, 32, 64 and 128, respectively. When the direction of contrast agent flow cannot be determined in the calculation process, it is the first encountered direction. Then, we generate the flow matrix of contrast agent in the whole basin.

(ii) Confluence analysis: Based on the flow matrix of contrast medium obtained above, confluence analysis is conducted to further determine the flow path. The confluence cumulant is calculated based on the flow direction of contrast agent determined by the steepest descent method, and its basic idea is as follows. Calculate the contrast dose flowing through each contrast unit in the flow matrix above (the total number of all contrast units upstream, which is the confluence value of contrast units in the watershed). It reflects the strength of contrast agent flow convergence ability. The larger the value, the more contrast agent flows to the contrast agent unit, indicating that the more easily contrast runoff is generated in the watershed. Extraction of contrast agent runoff is a significant link in watershed feature extraction.

(iii) Extraction watershed: When the sink flow calculated above reaches the set threshold, the contrast media cell area exceeding the threshold will form a contrast media network, and the contrast media cell just equal to the threshold will be as the contrast media boundary. Based on the given threshold and the confluence accumulation matrix, the contrast agent units greater than and equal to the given threshold are marked as 255, and those less than the threshold are marked as 0 to generate the contrast agent runoff unit matrix. The distribution of contrast agent runoff becomes denser with smaller threshold settings, while with larger thresholds, the distribution becomes more dispersed. Therefore, the selection of threshold value should fully consider the features of CEUS image to avoid affecting the subsequent extraction results of contrast agent ED. To select this threshold adaptively, this paper adopts the minimum value, which is half of the maximum sink traffic and the middle value of the sink traffic distribution, as the threshold.

1.2 Extraction of interframe flow direction

Based on the image sequences obtained from the CEUS film-making network, this study further analyzes the changes of CEUS video in the time domain. The LK Optical Flow method is employed to estimate the motion information of contrast agent between two adjacent frames. It calculates motion information by examining the changes of pixels in the same position in the time domain and the correlation between adjacent frames in the image sequence, the basic idea is based on the following three hypotheses:

1) Constant brightness: The pixel intensity of the target image does not change after the movement of the front and back frames.

2) Time persistence: Changes of time do not cause the drastic change in pixel position, and the displacement between adjacent frames is small.

3) Spatial consistency: Adjacent points on the same surface in the scene exhibit similar movements, and their projection distances to the image plane are relatively close.

For two consecutive image frames $I_0,\ I_1\left(\Omega\epsilon R^2\right)$, the basic optical flow function can be written as:

$$\int_{\Omega}\{\lambda\phi(I_0\left(x\right)-I_1\left(x+u\left(x\right)\right))+\psi(u,\nabla u,\ldots)\} \tag{2}$$

where $u=u(u_1\left(x\right),u_2\left(x\right))^T$ indicates the displacement field, ϕ and ψ are data and smoothness penalty terms respectively, and the balancing parameter λ defines the trade-off between these functions.

2 Evaluation Indicators

To quantitatively evaluate the classification performance of the proposed model, this paper employs five indicators for testing: the area under the receiver operating characteristic curve (AUC) with 95 % Confidence interval (CI), Accuracy, Sensitivity, Specificity, Positive predictive value (PPV), Negative predictive value (NPV), and F_1, as follows:

$$Accuracy=\frac{(TP+TN)}{(TP+FP+TN+FN)} \tag{3}$$

$$Sensitivity=\frac{TP}{(TP+FN)} \tag{4}$$

$$Specificity=\frac{TN}{(TN+FP)} \tag{5}$$

$$PPV=\frac{TP}{(TP+FP)} \tag{6}$$

$$NPV=\frac{TN}{(TN+FN)} \tag{7}$$

$$F_1=\frac{2\times(PPV\times Sensitivity)}{PPV+Sensitivity} \tag{8}$$

where, TP is the number of correctly classified benign nodules, TN is the number of correctly classified malignant nodules, FP is the number of misclassified benign nodules, and FN is the number of misclassified malignant nodules. Then, a P value <0.05 is considered to indicate a statistically significant difference.

3 Datasets

This paper retrospectively collected thyroid data of 986 patients from anonymous Hospital, which the inclusion criteria were as follows: a. CEUS puncture examination was performed in the hospital; b. All have dual-view video data with clear images; c. All patients were confirmed by pathology, and the clinicopathological data were complete. The exclusion criteria were any of the following: a. The thyroid cell pathology report Bethesda system was classified into Bethesda Class I, III or IV with no final pathological results; b. History of fine needle puncture or ablation; c. The patient had significant swallowing during the examination, resulting in poor image quality. All data has been desensitized to protect pathological personal information. Among them, the first frame of the double-view CEUS video was marked by two experienced radiologists who independently labeled the masks of all the nodules, then examined each other's results to reach a consensus, and employed the mask as the area of interest for cropping. Since CEUS and US are dual-view and one-to-one correspondence, the results of US are equally mapped to each frame of CEUS video for clipping, which is the pre-processing work for subsequent classification tasks. Simultaneously extracting CEUS features by two highly qualified sonographers-enhancement directions. This paper intends to employ the three-fold cross-validation method to verify the performance of the subsequent AI model. These data are randomly divided into 3 groups according to the ratio of 6:2:2, which are 590 training sets, 198 verification sets and 198 test sets respectively.

4 Implementation details

The DWFN model proposed in this paper is implemented using the popular deep learning framework PyTorch 1.12.0, and the code runs on a single GPU (i.e., NVIDIA TITAN X 12GB). The network parameters as shown in Table.1. Additionally, random horizontal flipping, random translation along two axes (5%), and random scaling (95%-105%) are used for data augmentation to speed up convergence.

Table 1. Model-parameters

Optimizer	Loss	Larning rate	Batch size	After number epochs (loss does not drop, stop training)
Adam	cross-entropy	1e-4	10	20