



Aided diagnosis of thyroid nodules based on an all-optical diffraction neural network

Lingxiao Zhou^{1#}, Luchen Chang^{2#}, Jie Li¹, Quanzhou Long¹, Junjie Shao¹, Jialin Zhu², Alan Wee-Chung Liew³, Xi Wei², Wanlong Zhang^{1^}, Xiaocong Yuan^{1,4}

¹Nanophotonics Research Center, Institute of Microscale Optoelectronics, Shenzhen University, Shenzhen, China; ²Department of Diagnostic and Therapeutic Ultrasonography, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin, China; ³School of Information and Communication Technology, Griffith University, Queensland, Australia; ⁴Research Center for Humanoid Sensing, Research Institute of Intelligent Sensing, Zhejiang Lab, Hangzhou, China

Contributions: (I) Conception and design: L Zhou, L Chang; (II) Administrative support: X Wei, W Zhang; (III) Provision of study materials or patients: L Zhou, L Chang; (IV) Collection and assembly of data: L Zhou, L Chang, J Li, Q Long, J Shao, J Zhu; (V) Data analysis and interpretation: L Zhou, L Chang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Xi Wei, MD, PhD. Department of Diagnostic and Therapeutic Ultrasonography, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Huanhu West Road, Tianjin 300060, China. Email: weixi@tmu.edu.cn; Wanlong Zhang, PhD. Nanophotonics Research Center, Institute of Microscale Optoelectronics, Shenzhen University, Nanhai Boulevard, Shenzhen 518060, China. Email: zwl@szu.edu.cn.

Background: Thyroid cancer is the most common malignancy in the endocrine system, with its early manifestation being the presence of thyroid nodules. With the advantages of convenience, noninvasiveness, and a lack of radiation, ultrasound is currently the first-line screening tool for the clinical diagnosis of thyroid nodules. The use of artificial intelligence to assist diagnosis is an emerging technology. This paper proposes the use of optical neural networks for potential application in the auxiliary diagnosis of thyroid nodules.

Methods: Ultrasound images obtained from January 2013 to December 2018 at the Institute and Hospital of Oncology, Tianjin Medical University, were included in a dataset. Patients who consecutively underwent thyroid ultrasound diagnosis and follow-up procedures were included. We developed an all-optical diffraction neural network to assist in the diagnosis of thyroid nodules. The network is composed of 5 diffraction layers and 1 detection plane. The input image is placed 10 mm away from the first diffraction layer. The input of the diffractive neural network is light at a wavelength of 632.8 nm, and the output of this network is determined by the amplitude and light intensity obtained from the detection region.

Results: The all-optical neural network was used to assist in the diagnosis of thyroid nodules. In the classification task of benign and malignant thyroid nodules, the accuracy of classification on the test set was 97.79%, with an area under the curve value of 99.8%. In the task of detecting thyroid nodules, we first trained the model to determine whether any nodules were present and achieved an accuracy of 84.92% on the test set.

Conclusions: Our study demonstrates the potential of all-optical neural networks in the field of medical image processing. The performance of the models based on optical neural networks is comparable to other widely used network models in the field of image classification.

[^] ORCID: 0000-0003-4704-9218.

Keywords: Aided diagnosis; thyroid nodules; ultrasound; real-time; all-optical neural networks

Submitted Jan 25, 2023. Accepted for publication Jul 12, 2023. Published online Aug 14, 2023.

doi: 10.21037/qims-23-98

View this article at: <https://dx.doi.org/10.21037/qims-23-98>

Introduction

Thyroid cancer is the most common malignancy in the endocrine system, accounting for 90% of all endocrine tumors. In recent years, the morbidity of thyroid cancer has been on the rise worldwide (1). Although inert thyroid cancer accounts for the majority of the cases, there has been a steady increase in the morbidity and mortality of terminal thyroid cancer and aggressive variants of papillary thyroid cancer. The early manifestation of thyroid cancer consists of the presence of thyroid nodules. Early detection and accurate diagnosis of patients' benign and malignant thyroid nodules is a critical component of clinicians' disease management (2,3). Ultrasound is currently the first-line screening tool for the clinical diagnosis of thyroid nodules by virtue of its convenience, noninvasiveness, and nonradiation (4,5). It is not only used to assess the risk of malignancy of thyroid nodules but is also frequently applied to guiding fine-needle aspiration (FNA) and treatment decisions. In 2017, the American College of Radiology (ACR) published a white paper on the Thyroid Imaging Reporting and Data System (TI-RADS), which is used to help radiologists standardize lesion descriptions (6). Malignant features of thyroid nodules ultrasound as described by the ACR TI-RADS are as follows: solid, extremely hypoechoic echogenicity, microcalcifications, ill-defined borders, and taller-than-wide ≥ 1 . However, the judgment of these features is highly dependent on the clinical experience of physicians. To better achieve the accuracy of identification, physicians usually use auxiliary methods to aid in diagnosis. Since it is time-consuming to analyze the diagnosis of the data collected from ultrasound detection, the use of artificial neural networks has been proposed by some scholars to conduct the initial diagnosis of ultrasound data.

An artificial neural network is a mathematical processing method that simulates neurons in biological neural systems. Since its invention, it has been rapidly developed and applied across numerous fields. In 1989, Cun *et al.* introduced a convolutional neural network model called LeNet-5 based on the error backpropagation

algorithm, which demonstrated promising results in handwritten digit recognition tasks (7). Since then, various models have been proposed, including convolutional neural networks (CNNs) (8), deep belief networks (9), AlexNet (10), generative adversarial networks (11), and neural network models based on self-attention mechanism transformer (12), among others. These models have been widely applied in image recognition and generation (8,11), target detection (7,8), feature extraction, classification, and generation (9), natural language processing (12), and more. The network can act as an expert system due to its parallel processing style, self-learning ability, memory capacity, and ability to predict the development of events. There are several studies in areas such as medical assisted diagnosis (13-16), for example, the use of 2 preoperative medical image modalities for multiclassification of thyroid diseases (i.e., normal, thyroiditis, cystic, multinodular goiter, adenoma, and cancer). Other research has constructed a diagnostic model for thyroid diseases using the currently state-of-the-art deep CNN architecture for differentiating disease types (17-19).

Although artificial neural networks have provided enhanced results in medical assisted diagnosis, improving the network better requires considerable computation time. Cheng *et al.* (20) and Cammarasana *et al.* (21) conducted research on using artificial neural networks for the auxiliary diagnosis of thyroid nodule ultrasound data with an analysis rate of 10 frames per second and 40 frames per second, respectively. With the increasing demand for detailed ultrasound data in clinical practice, high frame rate ultrasound imaging devices have gained attention and have been put into practice. These devices can obtain ultrasound data at a speed of over 50 frames per second (22). However, the processing speed of traditional electronic artificial intelligence networks is slower than the contrast speed of high frame rate ultrasound devices. Hence, these networks are currently unable to analyze high frame rate ultrasound data frame by frame and assist in diagnosis. In-depth research on artificial neural networks has revealed that achieving breakthroughs in processing units and improving the computational speed of the network is challenging

due to the limitations of Moore's law and other related factors (23).

With the development of optical and optoelectronic technology, researchers are increasingly investigating optical neural networks, including on-chip integrated optical neural networks based on Mach-Zendel interferometer topological cascade architecture (24), multimode optical CNNs based on pulse code modulation (25), and diffraction neural networks (26). Among them, diffractive neural networks have received more attention due to their simple structure, low energy consumption, and short processing time. For example, studies have been conducted on improving the energy efficiency of core computational modules and the robustness of diffractive neural networks (27), image classification (28), target detection and recognition (29-31), etc. However, optical diffractive neural networks have rarely been used in medical research.

Therefore, in order to process a large amount of ultrasound data medical images in real time while reducing the resource consumption and waiting time for patients, we aimed to use diffraction neural networks with real-time processing and low resource consumption for the classification and detection tasks of thyroid ultrasound image data. The tasks examined in this study were based on the ultrasound data obtained from the Institute and Hospital of Oncology of Tianjin Medical University. The accuracy of benign-malignant thyroid classification reached 97.69%, and the under the curve (AUC) value reached 99.8%. The detection of thyroid nodules was achieved by the use of sliding detection windows combined with a diffraction neural network model, which yielded with an accuracy of 84.92% for the classification of the presence or absence of nodules in images. The module then framed out the part of the image containing nodules. The study confirms the feasibility of the application of diffraction neural networks in medical image processing. We present this article in accordance with the TRIPOD reporting checklist (available at <https://qims.amegroups.com/article/view/10.21037/qims-23-98/rc>).

Methods

Datasets

All the ultrasound images included in the dataset employed in this study were obtained from the Institute and Hospital of Oncology, Tianjin Medical University, from January 2013 to December 2018. Patients who consecutively underwent

thyroid ultrasound diagnosis and follow-up procedures are included in the study. Two experienced radiologists (Dr. Wei and Dr. Zhang with 15 and 30 years of experience in thyroid cancer ultrasound diagnosis, respectively) examined the ultrasound images of thyroid nodules. The study inclusion criteria were as follows: (I) pathological findings obtained with FNA or surgical resection, (II) ultrasound examination performed within 1 month before surgery or FNA, and (III) complete clinical and ultrasound data. The exclusion criteria were as follows: (I) biopsy or resection prior to ultrasound examination, (II) patients who received preoperative treatment (radiofrequency or microwave ablation, radiotherapy and chemotherapy), and (III) unknown pathological information. This study was approved by the Ethics Committee of Tianjin Medical University Cancer Institute and Hospital (No. bc2020033) and conformed to the provisions of the Declaration of Helsinki (as revised in 2013). The requirements for informed consent were waived due to the retrospective nature of the study.

Research design

The diffraction neural network described in this paper is trained based on the Python 3.10 and PyTorch 1.21.1 environment, and the training procedure is shown in *Figure 1*. First, data are divided into 10 groups evenly and are processed in advance. Nine of the ten data groups are taken as the training set, and the other one as the validation set. Then, the data are input into the diffraction neural network and traditional artificial intelligence (AI) network. The traditional AI neural network used in this study was the fast region-based CNN (fast-RCNN), featuring a simple network module, and Detection Transformer, featuring a complex network module (32). After 1 round of training, the data in the validation set are used for validation. After each round of training, a group of data in the validation set is swapped with one in the training set that has not been used for validation, and then the next round is conducted. The above-described process is repeated 10 times; that is, the 10-fold validation method widely used in machine learning is used. The training is then finished.

All-optical diffraction neural network

The diffraction neural network is composed of 5 diffraction layers and 1 detection plane as shown in *Figure 2A*, where the number indicates the i -th diffraction layer. The input

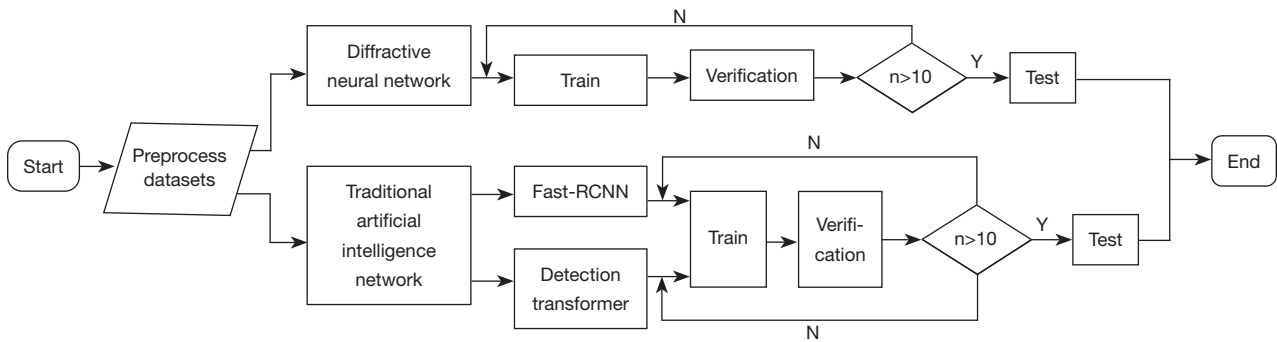


Figure 1 Overall process. Fast-RCNN, fast region-based convolutional neural network; Y, yes; N, no; n, number of cycles.

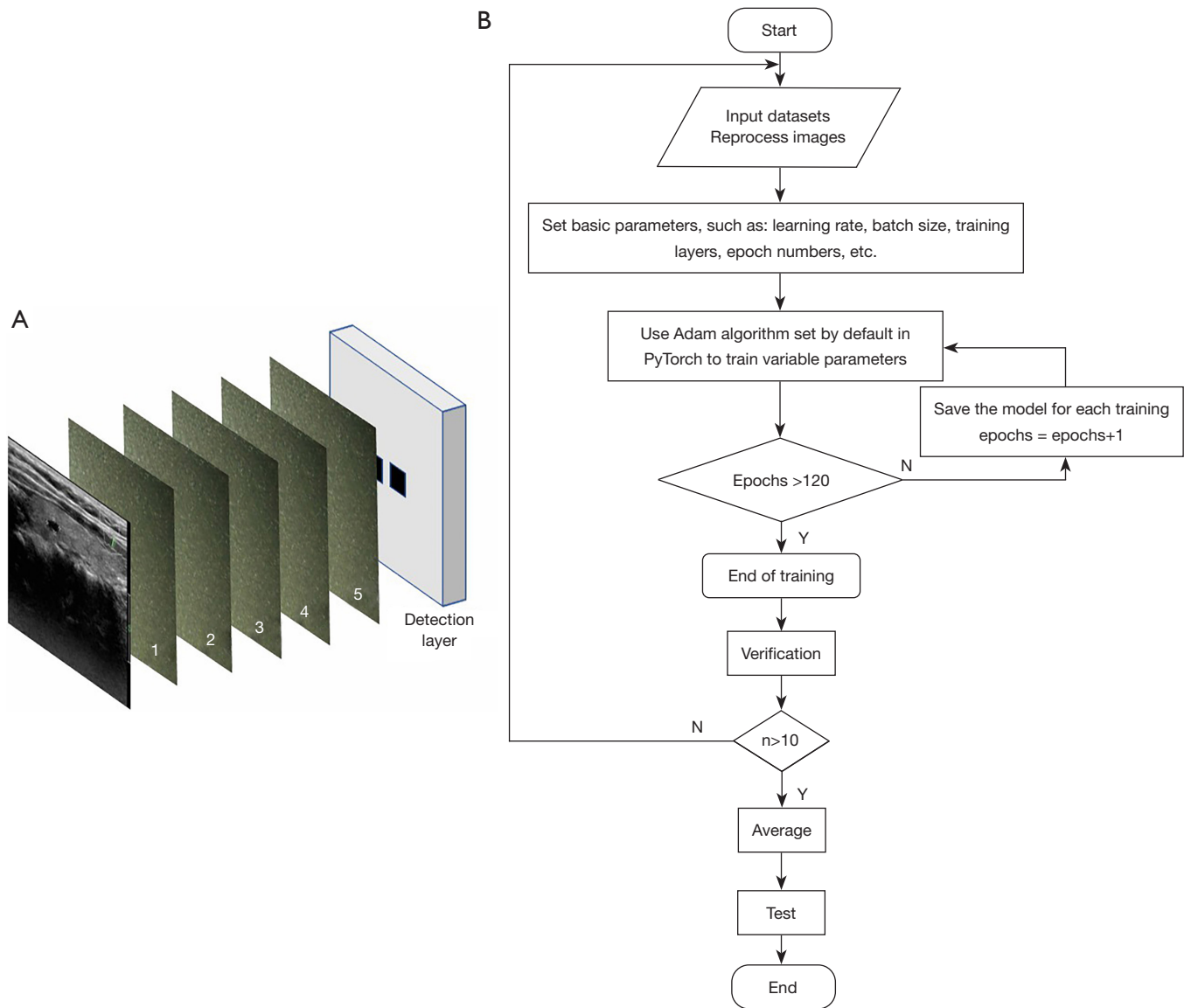


Figure 2 Diffraction neural network structure and training scheme. (A) Composition of the all-optical diffractive neural network. Numbers 1 to 5 represent the layers of diffractive neural networks. (B) Training flow diagram. Y, yes; N, no; n, number of cycles.

image is placed 10 mm away from the first diffraction layer. The diffraction unit size is $400 \mu\text{m} \times 400 \mu\text{m}$, and each diffraction layer is composed of 400×400 diffraction units. The distance between adjacent diffraction layers and the distance between the last diffraction layer and the detection plane are 10 mm. Two detection areas in the size of 50×50 are on the detection plane, which are symmetrically distributed on the detection plane. In this paper, a wave with a frequency of 0.4 THz is used as the input of the diffraction neural network to modulate the amplitude of the input image, and the output of the network is determined by the amplitude and light intensity obtained from the detection area. If the light intensity of the left detection area reaches the threshold while the area on the right does not, the network model determines the result to be malignant. On the contrary if the light intensity of the right detection area reaches the threshold while the area on the left does not, the network model determines the result to be benign. The input image size is cropped to 400×400 . After the diffraction neural network model is built, the data are input for training and testing. The training process is shown in *Figure 2B*; the batch size is set to 64 with a learning rate of 0.0002. After the training is completed, 120 cycles are trained, and a 10-fold validation is performed, with the best model being selected for testing based on the results of the validation. Finally, the amplitude of the images in each diffraction layer is analyzed using the data in the test set, and the inference ability of the model is evaluated.

Diffraction neural networks are physically formed by multiple layers of diffractive surfaces that work in concert to optically perform any function that the network can statistically learn. Analyzed physically, the deduction and prediction mechanisms of the network are all optical while its learning component is performed by a computer. According to the Huygens-Fresnel principle, any point on the wavefront at a given moment in the light propagation process can be viewed as a secondary point source, and the subsequent propagation process can be viewed as advancing with the envelope of these secondary sources while the propagation process can be approximated analytically using scalar diffraction theory. According to the Rayleigh-Sommerfield diffraction equation, each individual neuron of a given diffraction layer can be considered to be a secondary source of a wave consisting of the following optical modes (33):

$$w_i^l(x, y, z) = \frac{z - z_i}{r^2} \left(\frac{1}{2\pi r} + \frac{1}{j\lambda} \right) \exp\left(\frac{j2\pi r}{\lambda}\right) \quad [1]$$

where l represents the l -th layer of the network, and i is the i -th neuron in the l -th layer. The coordinate of the neuron is (x_i, y_i, z_i) , and λ is the working wavelength. $r = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}$ represents the distance between the current neuron and the i -th neuron in the l -th layer and $j = \sqrt{-1}$. The amplitude of the secondary wave and relative phase are confirmed by the multiplication of the input motion of the neuron and its transfer coefficient. The input motion and transfer coefficient are both complex-valued functions. Based on the above theory, the output function of the i -th neuron is as follows:

$$n_i^l(x, y, z) = w_i^l(x, y, z) \cdot t_i^l(x_i, y_i, z_i) \cdot \sum_k n_k^{l-1}(x_i, y_i, z_i) \quad [2]$$

where $\sum_k n_k^{l-1}(x_i, y_i, z_i)$ is the input wave of the i -th neuron on the l -th layer. The transfer coefficient contains the amplitude and phase as follows:

$$t_i^l(x_i, y_i, z_i) = a_i^l(x_i, y_i, z_i) \exp(j\phi_i^l(x_i, y_i, z_i)) \quad [3]$$

For the diffraction neural network modulated only by phase, amplitude $a_i^l(x_i, y_i, z_i)$ (alis the unit constant, so its output light field n_i^l can be expressed as follows:

$$n_i^l(x, y, z) = t_i^l(x_i, y_i, z_i) \cdot \sum_k n_k^{l-1} w_k^{l-1} \quad [4]$$

A plane wave is an analytic solution of Maxwell's set of equations, which are linear differential equations, so that any light wave can be represented as an infinite number of plane waves superimposed on the transmission. The angular spectrum theory starts from the plane wave solution and finally obtains the analytic solution. It is applicable to a situation in which the transmission distance is much larger than the wavelength, only 1 polarization state is considered, the medium is isotropic and linear medium, and most of the substances in life are included. Therefore, in the simulation of diffraction neural network, the angular spectrum method is used to calculate the light propagation process, which can speed up the network training; thus, Eq. [4] can be rewritten as the following equation:

$$n_i^l(x, y, z) = t_i^l(x_i, y_i, z_i) \cdot F^{-1}\left(U^{(l-1)}(u, v) \cdot \exp(j2\pi\gamma\Delta_z)\right) \quad [5]$$

where Δ_z is the distance between the l -th layer and the $(l-1)$ -th layer, and $\gamma = \sqrt{1/\lambda^2 - u^2 - v^2} U^{(l-1)}(u, v)$ is the Fourier transform of the output light field of the $(l-1)$ -th layer; that is

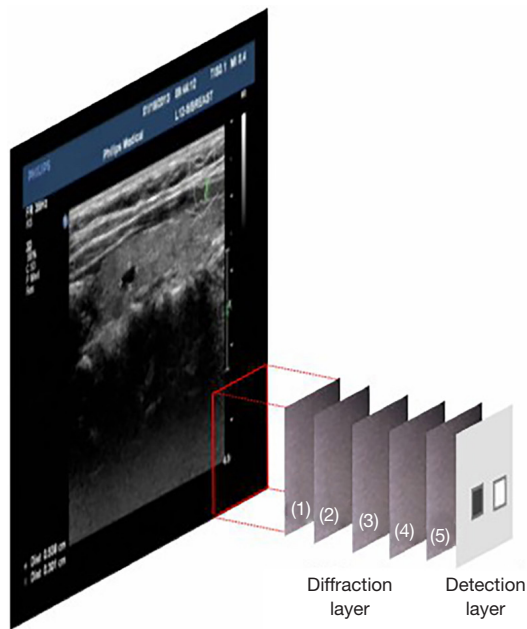


Figure 3 Schematic drawing of thyroid nodules detection. Numbers 1 to 5 represent the layers of diffractive neural networks.

$$U^{-1}(u, v) = F(U^{-1}(x, y, z)) \quad [6]$$

Classification of the malignant or benign thyroid nodules

The datasets contain the annotation of the location of thyroid nodule lesions performed by professional physicians while the thyroid nodule data are classified as benign and malignant and placed in different folders. Next, the diffraction neural network previously described is used for the classification task. Before the classification task, some further data processing is required. In order to fit the network size, we crop the images to size and place them in different folders, with the folder name set to 0 for malignant nodules and 1 for benign nodules, forming a binary classification dataset. The folder name is then used as the label value of the images during training.

During the training process, the light amplitude sum of the diffraction neural network in the 2 detection regions needs to be normalized according to the following normalization setting:

$$A'_i = \frac{A_i}{A_0 + A_i} \quad [7]$$

The training process for diffraction neural networks uses

the cross-entropy loss function (nn.CrossEntropyLoss) set by default in PyTorch for measurement, evaluation, and network optimization. The cross-entropy loss function is shown below:

$$\begin{aligned} \text{loss}(x, \text{class}) &= -\log\left(\frac{\exp(x[\text{class}])}{\sum_i \exp(x[i])}\right) \\ &= -x[\text{class}] + \log\left(\sum_i \exp(x[i])\right) \end{aligned} \quad [8]$$

The process of detecting thyroid nodules is approximately the same as the process of classifying thyroid nodules. First, the images are cropped to a size of 200×200, from which data containing nodules and data without nodules are selected and placed in separate folders to form a dichotomous dataset. After that, this dataset is trained with the same network used for thyroid nodule classification to obtain a model for determining the presence or absence of nodules. Next, we use this model to determine whether the images contain nodules. We place the detector at the edge of the image and gradually move the detector to scan the whole image, determining whether there is a thyroid nodule in the region as we move (the method is shown in *Figure 3*). The detection frame size is set to 75×75, and the sliding step is 5. The numbers in the figure correspond to the *i*-th diffraction plane, and the last plane is the detection plane.

The analysis of medical images through diffraction neural networks has not been extensively researched, and studies on optical network-assisted diagnosis of thyroid nodules are completely lacking. Therefore, it is necessary to compare our study with traditional AI networks. In this paper, 2 widely used AI networks were selected for comparison: faster-RCNN featuring a simple network model and Detection Transformer featuring a more complex network model. Both networks perform the benign and malignant classification task on thyroid nodule datasets.

The basic structure of fast-RCNN is shown in *Figure 4*. The 4 main components and their functions can be summarized as follows:

- (I) Conv layers: the series of convolutional + rectified linear unit (ReLU) + pooling layer is used to extract the feature map of images. The feature map is shared and used for the region proposal network (RPN) layer and fully connected layer.
- (II) RPN: the RPN network is used to produce region proposals. In this process, the network identifies whether anchors are positive or negative through

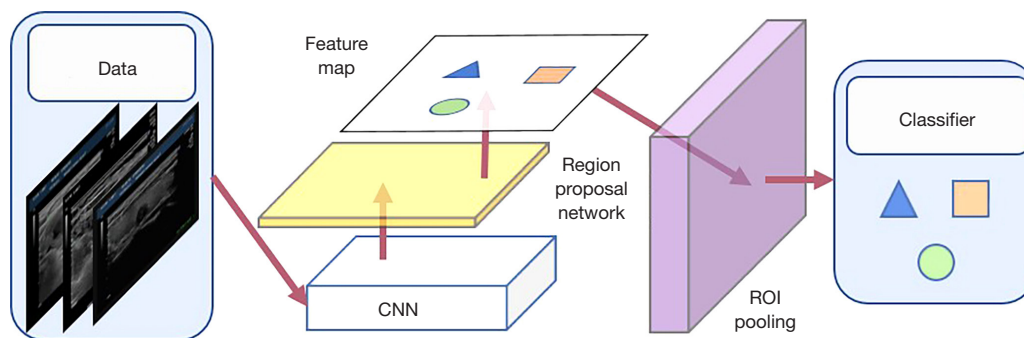


Figure 4 The basic structure of Fast-RCNN. Fast-RCNN, fast region-based convolutional neural network; CNN, convolutional neural network; ROI, region of interest.

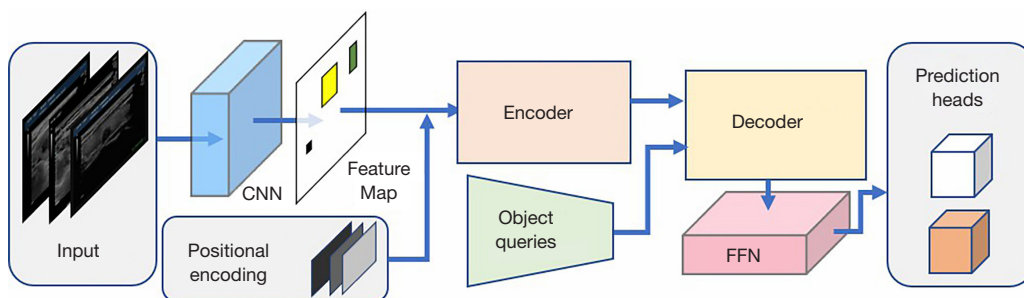


Figure 5 The basic structure of Detection Transformer. FFN, feedforward network; CNN, convolutional neural network.

softmax and correct anchors to obtain precise and accurate proposals through bounding box regression.

- (III) Region of interest (ROI) pooling: this layer collects the input feature map and proposals. After integrating the input information, it extracts proposal feature maps which are input into the fully connected layer for the identification of types.
- (IV) Classification: the type of proposal is determined by a regional feature map, and bounding box regression is used to obtain the exact final position of the detection frame.

The structure of Detection Transformer is shown in *Figure 5* below. It consists of four 4 components: backbone, encoder, decoder and prediction head; their functions described as follows:

- (I) CNN backbone: this part of extraction is used to accelerate network computing.
- (II) Encoder: every encoder layer has a standard framework which comprises a multi-head self-attention module and a feedforward network (FFN). It is used for encoding the input feature map.

- (III) Decoder: all-zero feature vectors are encoded through self-attention, multi-head attention, and a related basic option.

- (IV) Prediction head: the prediction part consists of a ReLU activation function and perceptron with a hidden layer and a linear layer. The center coordinate of FFN prediction frame, height, width, and the input images predict the class labels through softmax, and then the linear layer predicts the class labels through softmax.

Indexes for evaluation

A confusion matrix is a situation analysis table used for summarizing the prediction results of a classification model in machine learning (*Table 1*). It summarizes the records in the datasets in the form of matrix according to 2 criteria: the real category and the category judgment predicted by the classification model. The rows of the matrix represent the real values, and the columns of the matrix represent the predicted values. True positive (TP) indicates positive or benign and the real number of benign cases, false negative

Table 1 Example of binary classification through a confusion matrix

| Confusion matrix | Predicted label | |
|------------------|-----------------|----|
| | 0 | 1 |
| True labels | | |
| 0 | TN | FN |
| 1 | FP | TP |

TN, true negative; FN, false negative; FP, false positive; TP, true positive.

(FN) indicates negative or malignant but the real number of benign cases, false positive (FP) indicates positive or benign but the real number of malignant cases, and true negative (TN) indicates negative or malignant and the real number of malignant cases.

The AUC is defined as the area formed by receiver operating characteristic curve (ROC) curve and coordinate axes. Therefore, it is necessary to make analysis of ROC curve before calculating the AUC.

The abscissa of ROC curve is the FP rate (FPR), and its ordinate is the TP rate (TPR). There are also the TN rate (TNR), the FN rate (FNR), and the quantity average precision (AP), which is commonly used in the field of target detection and also calculated via the confusion matrix. The formulae for calculating these 4 indices are as following:

- (I) $FPR = FP / (FP + TN)$: the probability of positive but not TP (i.e., the probability of positive in true negative);
- (II) $TPR = TP / (TP + FN)$: the probability of positive and also TP [i.e., the probability of positive in TP (positive recall)];
- (III) $FNR = FN / (TP + FN)$: the probability of negative but not true negative (i.e., the probability of negative in TP);
- (IV) $TNR = TN / (FP + TN)$: the probability of negative and also true negative (i.e., the probability of negative in TN);
- (V) $AP = TP / (TP + FP)$.

The ROC curve does not change with the change of the positive and negative samples, so it is of great significance with regard to the reference use of ROC curve as the evaluation standard.

Since the ROC curve requires the calculation of the probability of the output results, we discuss how the network models in this study calculate the probability of the output results. The diffraction neural network identifies

benignancy or malignancy based on the light intensity in the detection area: if the left/right light intensity reaches the threshold, the test thyroid nodule is judged to be malignant/benign. In the network, the light intensity of the detection area of a picture is treated as a tensor with 2 elements (A and B): A indicates the light intensity obtained in the left detection area, and B indicates the light intensity obtained in the right detection area. The threshold value of both A and B is 10. Therefore, we use the equation to calculate the probability value of the output result, with P indicating the probability value: if $P < 0.5$, the result is malignant, and if $P > 0.5$, the result is benign.

At present, clinical data of thyroid nodules are basically obtained from ultrasound. Considering that the volume of ultrasound data is large and the real-time problem of AI-aided diagnosis, we needed us to calculate the time of processing each image by the AI network. The calculation method was devised as follows: single image processing time = total time of processing test set images/the number of test set images. After obtaining the time value, we can determine whether the network meets the real-time requirements needed to analyze ultrasound data.

Results

First, the classification performance of the optical neural network was evaluated. The method described in the previous section was used to train the benign and malignant thyroid nodule data. The change of “loss” during training is shown in *Figure 6A*. As the training process continues, the “loss” decreases rapidly, and the model converges rapidly. The performance of the trained model was evaluated on the test set, and the results of the classification on the test set are shown in *Figure 6B*. The confusion matrix indicated that the diffraction neural network achieved an accuracy of 97.69% in the classification of benign and malignant thyroid nodules. In the confusion matrix, 0 indicates malignant nodules, and 1 indicates benign nodules. The ROC curve is presented in *Figure 6C*, and the AUC value is 99.8%. Next, the data images of the test set were input into the neural network to check the amplitude distribution and the light intensity distribution in the detection region, as shown in *Figure 6D*. The light intensity in the detection region on the left side is stronger, so it is a malignant nodule; meanwhile, the right detection area is stronger, which means it is a benign nodule.

The classification results of thyroid nodules obtained with fast region-based CNN (fast-RCN), Detection

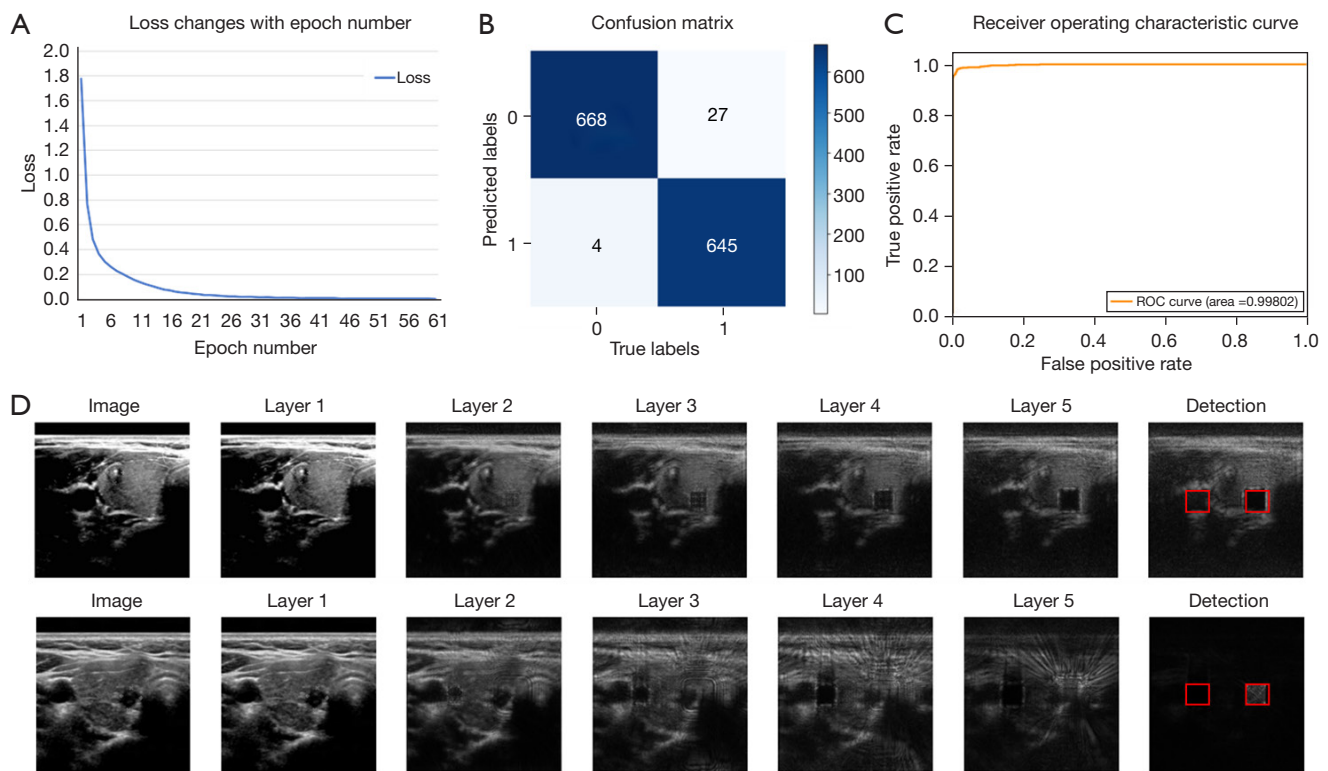


Figure 6 The evaluation of classification performance of the optical neural network. (A) Change of “loss” during the training process. (B) Confusion matrix. (C) ROC curve. (D) The deduction result of the image in the network and the light intensity distribution in the detection area. ROC, receiver operating characteristic.

Transformer, and the proposed model were compared. As shown in *Figure 7A*, the classification accuracy of fast-RCNN is as high as 96%, Detection Transformer has a classification accuracy of up to 94.6%, and the diffractive deep neural network has a classification accuracy of up to 97.69%. To maintain uniformity in the data format, the figure displays the classification accuracy of up to 3 decimal places. We also present the training time consumption for each model in *Figure 7B*. The training time for Detection Transformer is 38 hours, while the training time for our model is only 8 hours.

Figure 7C presents a comparison of the image data processing time for all 3 models. The processing time of fast-RCNN for each image is 0.048 seconds, allowing for approximately 21 images to be processed per second. Similarly, Detection Transformer can process about 26 images per second with a processing time of 0.039 seconds per image. Neither of these models have sufficient throughput to analyze and diagnose the current mainstream ultrasound data of 50 frames per second. On the other hand, our model takes only 0.017 seconds to

process a single image, which implies that it can process approximately 58 images per second. Therefore, the proposed model has the required processing throughput to meet the real-time diagnostic needs of ultrasound data.

We analyzed the results of thyroid nodule detection. The model was trained to determine whether the ultrasound picture contained a nodule had an accuracy of 84.92% on the test set and an AP of 0.831; its confusion matrix is shown in *Figure 8A*. The trained model was then used for the detection task of thyroid nodules using the sliding window method, and its detection results are shown in *Figure 8B*. It can be seen from the figure that the model can better distinguish the regions containing nodules and frame them out.

Discussion

In the classification task of thyroid nodules implemented by diffraction neural network algorithm, in order to evaluate whether diffraction neural networks can be used in the analytical processing work of medical images, this study also

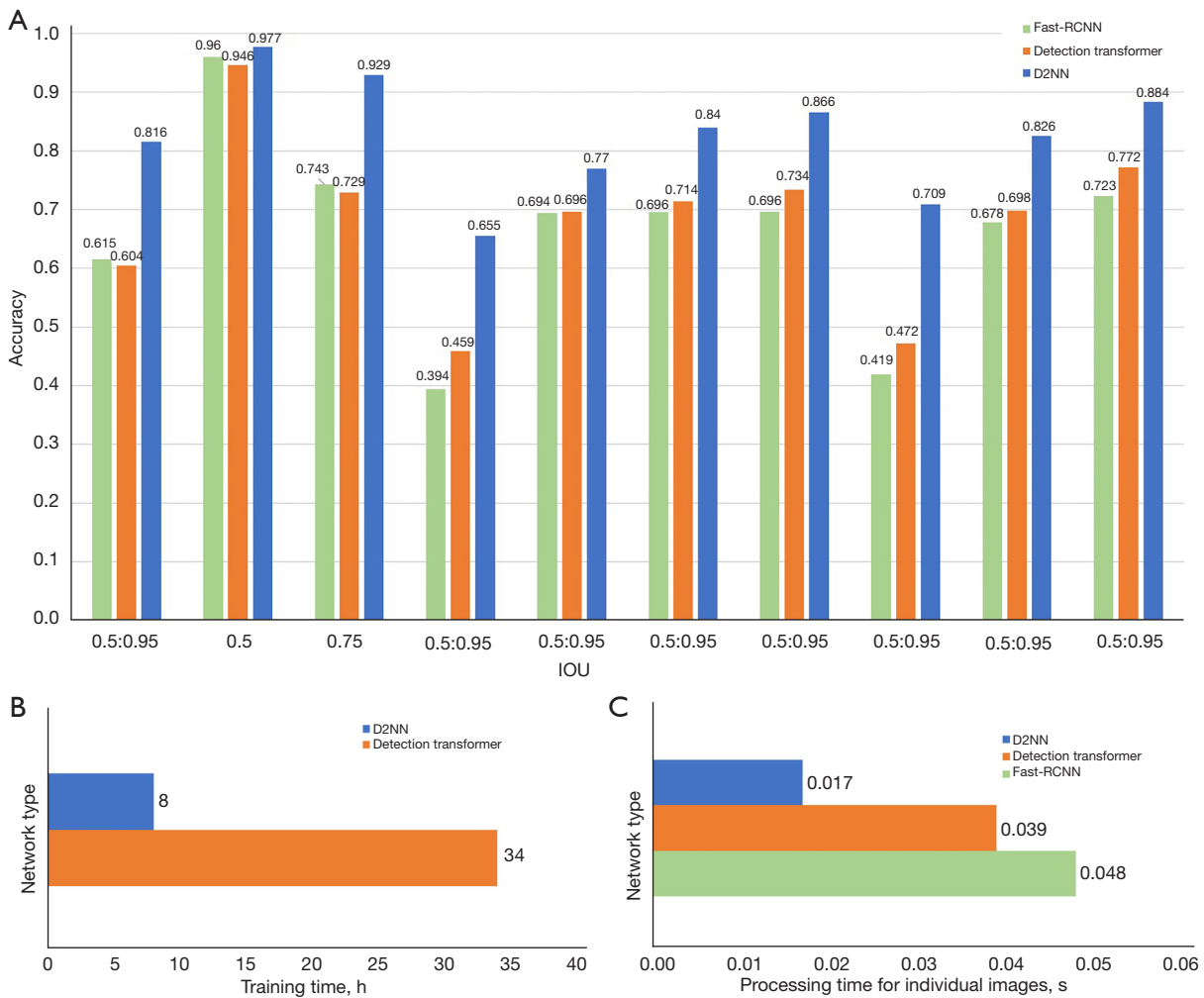


Figure 7 Comparison of the classification performance of different networks. (A) Test results of 3 types of neural networks. (B) Training time for 2 types of neural networks. (C) Processing time of three types of neural networks for a single image. Fast-RCNN, fast region-based convolutional neural network; IOU, intersection over union. D2NN, diffractive deep neural network.

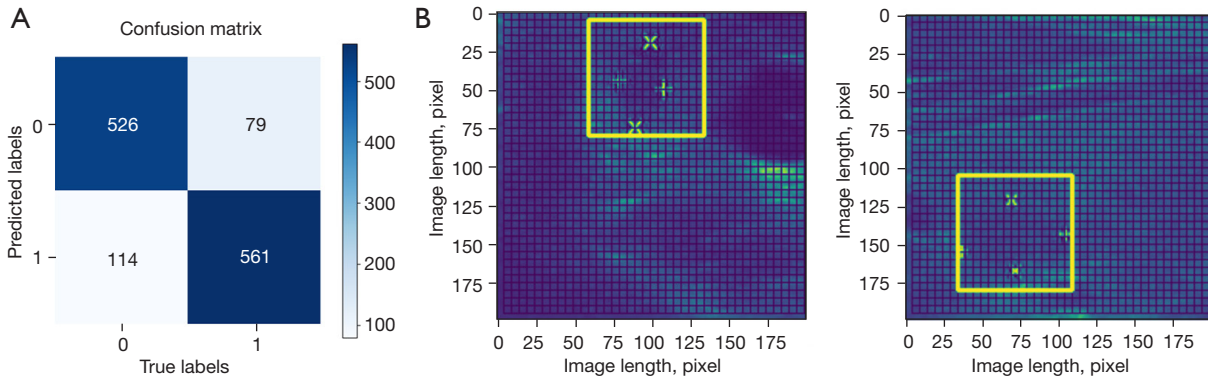


Figure 8 The detection results of optical neural networks. (A) Accuracy of the detection model. (B) Results of detection.

Table 2 The AUC of different network models

| Method | Model | AUC |
|----------------------------|-----------|--------|
| Koundal <i>et al.</i> (34) | SVM | 94.42% |
| Liu <i>et al.</i> (35) | Fast-RCNN | 97.42% |
| Tao <i>et al.</i> (36) | CNN | 90% |
| Proposed model | D2NN | 99.8% |

AUC, area under the curve; SVM, support vector machine; Fast-RCNN, fast region-based convolutional neural network; CNN, convolutional neural networks; D2NN, diffractive deep neural network.

used traditional AI networks for the classification task of the same datasets for comparison. The classification results show that the optical neural network, as an emerging algorithm, has already surpassed the performance of the simpler network fast-RCNN and also reached a relatively higher level than the complex Detection Transformer algorithm. Additionally, our proposed model obtained using the optical diffraction neural network exhibited significant advantages in both training time and image processing speed.

Table 2 shows a comparison of AUC values for several neural networks, and our all-optical neural network model achieved the highest AUC value of 99.8%, demonstrating its excellent performance in classifying thyroid nodules.

As for employing the diffraction neural network to assist in the diagnosis of thyroid ultrasound image data, the traditional CNN network requires considerable time resources due to the large volume of ultrasound data. Chen and Shi *et al.* reported that a diffraction neural network obtained via 3D printing can be used to classify Modified National Institute of Standards and Technology (MNIST) data images after it modulates the terahertz light source (31,37). In the whole process, the 0.4-Hz terahertz light source is the only part that consumes energy, and the whole classification process is performed at the speed of light. In contrast, CNN neural networks consume energy in training models and classifying images. Some researchers have optimized the algorithm of CNN neural network to narrow the gap between CNN design and energy consumption optimization (38), but the energy consumption for large-scale data operations is still much larger than that of diffraction neural networks.

Our proposed model has shown promising results for the classification and diagnosis of thyroid nodules in medical images such as ultrasound data. Moreover, with the help of 3D printing technology, we can physically construct the

diffractive deep neural network (D2NN) model structure to achieve even faster analysis times. Theoretically, the speed of determining benign and malignant thyroid nodules using this model can reach the speed of light, and the analysis time of each image can be reduced to almost 0 seconds. This approach enables us to diagnose ultrasound data at any frame number in real time, making it an efficient and reliable tool for medical professionals.

The comparison between diffraction neural networks and traditional AI networks confirms that diffraction neural networks provide better results in the classification task of thyroid nodules, but detection is not yet possible. To address these problems, we propose using classification to complete detection, but this detection would not yet provide equivalent performance to that of traditional electronic networks.

Conclusions

We propose using diffraction neural networks to classify and detect thyroid nodules. In the task of classifying benign and malignant thyroid nodules, this approach yielded an accuracy of 97.69% and an AUC of 99.8%. In the task of detecting the presence and absence of thyroid nodules, the classification method was used to determine the presence and absence of thyroid nodules through the images, which yielded an accuracy of 84.92%. The model for classifying the presence and absence of nodules is able to frame the areas containing nodules in the images, which confirms the feasibility of using all-optical diffraction neural networks in the field of medical image processing. Compared with fast-RCNN networks and Detection Transformer algorithms, the proposed model has a shorter training time of only 8 hours and a faster image processing speed of 58 frames per second, demonstrating its superior performance compared to traditional neural networks. Even when handling a large number of medical images, the model still maintains a fast image processing speed, proving the efficiency and accuracy of diffraction neural networks in image classification and processing with almost no power consumption.

On the other hand, we acknowledge that the current computational power of the network is limited due to the lack of nonlinear calculations. To address this, we plan to optimize the structure of the network and explore the use of materials with nonlinear optical properties, such as graphene and zinc selenide (39) and photorefractive crystals (40) or by using atomic properties (41). Additionally, we intend to apply the all-optical diffraction

neural network to medical image diagnosis of different parts of the human body and explore the generalization performance of the networks (42). These future directions will deepen our understanding of the potential applications of all-optical neural networks in medicine and improve their computational power, ultimately leading to more accurate and efficient medical diagnoses.

Acknowledgments

Funding: This work was supported by the National Natural Science Foundation of China (No. 62005180 to Wanlong Zhang and No. 82272008 to Xi Wei), the Tianjin Health Research Project (Nos. ZD20018 and QN20018 to Xi Wei), and the Zhejiang Lab Open Research Project (No. K2022MG0AB01 to Wanlong Zhang).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at <https://qims.amegroups.com/article/view/10.21037/qims-23-98/rc>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <https://qims.amegroups.com/article/view/10.21037/qims-23-98/coif>). XW reports grant from the Tianjin Health Research Project (Nos. ZD20018 and QN20018) and the National Natural Science Foundation of China (No. 82272008) during the course of the study; WZ reports grants from the National Natural Science Foundation of China (No. 62005180) and the Zhejiang Lab Open Research Project (No. K2022MG0AB01) during the course of the study. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was approved by the Ethics Committee of Tianjin Medical University Cancer Institute and Hospital (No. bc2020033) and conformed to the provisions of the Declaration of Helsinki (as revised in 2013). The requirements for informed consent were waived due to the retrospective nature of the study.

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons

Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* 2019;69:7-34.
2. Zhang H, Zheng X, Liu J, Gao M, Qian B. Active surveillance as a management strategy for papillary thyroid microcarcinoma. *Cancer Biol Med* 2020;17:543-54.
3. Sun D, Li H, Cao M, He S, Lei L, Peng J, Chen W. Cancer burden in China: trends, risk factors and prevention. *Cancer Biol Med* 2020;17:879-95.
4. Lim H, Devesa SS, Sosa JA, Check D, Kitahara CM. Trends in Thyroid Cancer Incidence and Mortality in the United States, 1974-2013. *JAMA* 2017;317:1338-48.
5. Zhu J, Zhang S, Yu R, Liu Z, Gao H, Yue B, Liu X, Zheng X, Gao M, Wei X. An efficient deep convolutional neural network model for visual localization and automatic diagnosis of thyroid nodules on ultrasound images. *Quant Imaging Med Surg* 2021;11:1368-80.
6. Ho AS, Luu M, Barrios L, Chen I, Melany M, Ali N, Patio C, Chen Y, Bose S, Fan X, Mallen-St Clair J, Braunstein GD, Sacks WL, Zumsteg ZS. Incidence and Mortality Risk Spectrum Across Aggressive Variants of Papillary Thyroid Carcinoma. *JAMA Oncol* 2020;6:706-13.
7. Cun YL, Boser B, Denker J, Henderson D, Howard R, Hubbard W, Jackel L. Handwritten digit recognition with a back-propagation network. In: *Proceedings of the 2nd International Conference on Neural Information Processing Systems (NIPS'89)* 1989;396-404.
8. Cun YL, Huang F, Bottou L. Learning methods for generic object recognition with invariance to pose and lighting. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit* 2004;2:II97-II104.
9. Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. *Neural Comput* 2006;18:1527-54.
10. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12)* 2012;1097-1105.

11. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17) 2017;6000-6010.
12. Barrors C, Mendonca M, Vieira A, Ziviani A. A Survey on Embedding Dynamic Graphs. *ACM Comput Surv* 2021;55:1-37.
13. Bai Z, Chang L, Yu R, Li X, Wei X, Yu M, Liu Z, Gao J, Zhu J, Zhang Y, Wang S, Zhang Z. Thyroid nodules risk stratification through deep learning based on ultrasound images. *Med Phys* 2020;47:6355-65.
14. Li X, Zhang S, Zhang Q, Wei X, Pan Y, Zhao J, et al. Diagnosis of thyroid cancer using deep convolutional neural network models applied to sonographic images: a retrospective, multicohort, diagnostic study. *Lancet Oncol* 2019;20:193-201.
15. Wei X, Gao M, Yu R, Liu Z, Gu Q, Liu X, Zheng Z, Zheng X, Zhu J, Zhang S. Ensemble Deep Learning Model for Multicenter Classification of Thyroid Nodules on Ultrasound Images. *Med Sci Monit* 2020;26:e926096.
16. Wei X, Zhu J, Zhang H, Gao H, Yu R, Liu Z, Zheng X, Gao M, Zhang S. Visual Interpretability in Computer-Assisted Diagnosis of Thyroid Nodules Using Ultrasound Images. *Med Sci Monit* 2020;26:e927007.
17. Jegerlehner S, Bulliard JL, Aujesky D, Rodondi N, Germann S, Konzelmann I, Chiolero A; NICER Working Group. Overdiagnosis and overtreatment of thyroid cancer: A population-based temporal trend study. *PLoS One* 2017;12:e0179387.
18. Yan T, Wu J, Zhou T, Xie H, Xu F, Fan J, Fang L, Lin X, Dai Q. Fourier-space Diffractive Deep Neural Network. *Phys Rev Lett* 2019;123:023901.
19. Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-end object detection with transformers. *European conference on computer vision* 2020;213-229.
20. Cheng JZ, Ni D, Chou YH, Qin J, Tiu CM, Chang YC, Huang CS, Shen D, Chen CM. Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans. *Sci Rep* 2016;6:24454.
21. Cammarasana S, Nicolardi P, Patanè G. Real-time denoising of ultrasound images based on deep learning. *Med Biol Eng Comput* 2022;60:2229-44.
22. Giangregorio F, Garolfi M, Mosconi E, Ricevuti L, Debellis MG, Mendoza M, Esposito C, Vigotti E, Cadei D, Abruzzese D. High frame-rate contrast enhanced ultrasound (HIFR-CEUS) in the characterization of small hepatic lesions in cirrhotic patients. *J Ultrasound* 2023;26:71-9.
23. Durante C, Grani G, Lamartina L, Filetti S, Mandel SJ, Cooper DS. The Diagnosis and Management of Thyroid Nodules: A Review. *JAMA* 2018;319:914-24.
24. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, Cronan JJ, Beland MD, Desser TS, Frates MC, Hammers LW, Hamper UM, Langer JE, Reading CC, Scoutt LM, Stavros AT. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. *J Am Coll Radiol* 2017;14:587-95.
25. Zhang X, Lee VC, Rong J, Lee JC, Liu F. Deep convolutional neural networks in thyroid disease detection: A multi-classification comparison by ultrasonography and computed tomography. *Comput Methods Programs Biomed* 2022;220:106823.
26. Robison RA. Moore's Law: predictor and driver of the silicon era. *World Neurosurg* 2012;78:399-403.
27. Shen Y, Harris NC, Skirlo S, Prabhu M, Baehr-Jones T, Hochberg M, Sun X, Zhao S, Larochelle H, Englund D. Deep learning with coherent nanophotonic circuits. *Nat Photon* 2017;11:441-6.
28. Chang J, Sitzmann V, Dun X, Heidrich W, Wetzstein G. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Sci Rep* 2018;8:12324.
29. Lin X, Rivenson Y, Yardimci NT, Veli M, Luo Y, Jarrahi M, Ozcan A. All-optical machine learning using diffractive deep neural networks. *Science* 2018;361:1004-8.
30. Zhou T, Fang L, Yan T, Wu J, Li Y, Fan J, Wu H, Lin X, Dai Q, editors. Optical backpropagation training method and its applications. In: *Proc of SPIE Vol* 2020;11550:1155002.
31. Chen H, Feng J, Jiang M, Wang Y, Lin J, Tan J, Jin P. Diffractive deep neural networks at visible wavelengths. *Engineering* 2021;7:1483-91.
32. 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* 2017;39:1137-49.
33. Shen F, Wang A. Fast-Fourier-transform based numerical integration method for the Rayleigh-Sommerfeld diffraction formula. *Appl Opt* 2006;45:1102-10.
34. Koundal D, Gupta S, Singh S. Survey of Computer-Aided Diagnosis of Thyroid Nodules in Medical Ultrasound Images. *Adv Intell Syst Comput* 2013;177:459-67.

35. Liu T, Guo Q, Lian C, Ren X, Liang S, Yu J, Niu L, Sun W, Shen D. Automated detection and classification of thyroid nodules in ultrasound images using clinical-knowledge-guided convolutional neural networks. *Med Image Anal* 2019;58:101555.
36. Tao Y, Yu Y, Wu T, Xu X, Dai Q, Kong H, Zhang L, Yu W, Leng X, Qiu W, Tian J. Deep learning for the diagnosis of suspicious thyroid nodules based on multimodal ultrasound images. *Front Oncol* 2022;12:1012724.
37. Shi J, Chen Y, Zhang X. Broad-spectrum diffractive network via ensemble learning. *Opt Lett* 2022;47:605-8.
38. Yang TJ, Chen YH, Sze V. Designing energy-efficient convolutional neural networks using energy-aware pruning. *Proc IEEE Conf Comput Vis Pattern* 2017;5698-5.
39. Sun Y, Dong M, Yu M, Lu L, Liang S, Xia J, Zhu L. Modeling and simulation of all-optical diffractive neural network based on nonlinear optical materials. *Opt Lett* 2022;47:126-9.
40. Luo X, Hu Y, Ou X, Li X, Lai J, Liu N, Cheng X, Pan A, Duan H. Metasurface-enabled on-chip multiplexed diffractive neural networks in the visible. *Light Sci Appl* 2022;11:158.
41. Ryou A, Whitehead J, Zhelyeznyakov M, Anderson P, Keskin C, Bajcsy M, Majumdar A. Free-space optical neural network based on thermal atomic nonlinearity. *Photonics Res* 2021;9:B128.
42. Shao J, Zhou L, Yeung SYF, Lei T, Zhang W, Yuan X. Pulmonary Nodule Detection and Classification Using All-Optical Deep Diffractive Neural Network. *Life (Basel)* 2023;13:1148.

Cite this article as: Zhou L, Chang L, Li J, Long Q, Shao J, Zhu J, Liew AWC, Wei X, Zhang W, Yuan X. Aided diagnosis of thyroid nodules based on an all-optical diffraction neural network. *Quant Imaging Med Surg* 2023;13(9):5713-5726. doi: 10.21037/qims-23-98