Ovarian Tumors Detection and Classification from Ultrasound Images Based on YOLOv8

Thi-Loan Pham [1,2] and Van-Hung Le [3,*]

[1] Faculty of Information Technology, Hai Duong University, Hai Duong, Vietnam
[2] School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam
[3] Faculty of Basic Sciences, Tan Trao University, Tuyen Quang, Vietnam
Email: phamthiloan2011@gmail.com (T.-L.P.); van-hung.le@mica.edu.vn (V.-H.L.)
*Corresponding author

***Abstract*—Ovarian cancer is the 7th most common malignant tumor and the 8th leading cause of death in women. Therefore, ovarian cancer detection early and on the image data as ultrasound images is an issue that needs to be studied. YOLO is a highly accurate CNN, especially with very low processing time, and can calculate on the CPU for object detection problems in computer vision. In this paper, we perform a comparative study on the latest version of the YOLO family, YOLOv8 (YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x) for the detection and classification of ovarian tumors on the *OTU 2D-OS*. This study performed a fine-tuned model to detect and classify the *OTU 2D-OS*: (1) 1 label (with or without ovarian tumor); (2) 8 labels; (3) 2 labels (benign ovarian tumor and malignant). The precision of YOLOv8X is the best and higher than YOLOv7 is 19% for detecting and classifying 8 ovarian tumor classes on the *OTU 2D-OS* subset. The calculation time of YOLOv8 is also shown, and the processing time of YOLOv8x is slower than YOLOv7 (YOLOv8x is 186fps on GPU, 1.84fps on CPU). However, these results are still low compared to the requirements for the actual diagnosis and detection of ovarian tumors. It can be accepted that "it is better to catch a mistake than to miss it". Therefore, the problem of detecting ovarian tumors on ultrasound images brings many challenges and needs further research.**

***Keywords*—ovarian tumor detection, ovarian tumor classification, benign and malignant, YOLOv8, ultrasound images, Convolutional Neural Networks (CNNs)**

I. Introduction

According to worldwide statistics, ovarian cancer is the 8th most common cancer in women [1]. In 2020 alone, more than 313,000 new cases were detected. In developing countries such as Vietnam, Indonesia, and Malaysia, cancer screening is often low, so when the disease is detected, it is mainly in the late stages. In Vietnam, more than 70% of cancer cases are detected at a late stage [2]. Therefore, the treatment of cancer is very difficult. The detection and diagnosis of cancer increases the chance of survival by up to 50% for patients [3]. However, the ongoing cancer screening will consume large financial sums. Since this problem carries a great burden, the development of simple techniques that can detect cancer is very useful research for human life. Ultrasound images are now clear information that can be collected and disseminated from the medical examination and treatment of female patients. At the same time, doctors to detect ovarian cancer on ovarian ultrasound images, they must be experienced doctors and also have many errors. Currently, the strong development of artificial intelligence, especially the introduction of deep learning networks, has brought impressive results when solving computer vision problems such as object detection, estimation, and recognition of computer vision. Recently, there have also been several studies on the detection, classification, and segmentation of ovarian tumors on ultrasound images of the ovaries [4, 5]. However, these studies performed the detection, semantic segmentation, and classification of ovarian tumors on ultrasound images of the MMOTU dataset [4]. They still have low results and need to be performed on computer systems with GPUs, so the cost is high. At the same time, the research of Liu *et al.* [6] also presented the challenges when applying deep learning to perform computer vision problems on ultrasound images as the quality of ultrasound images is very low, the number of samples for training the models is limited, so overfitting can easily occur. In this study, the authors also present two ways to limit this problem: optimizing the model and increasing the training data. YOLOv8 [7] is an improved and current version of the YOLO family (from YOLOv1 to YOLOv8) [8] for object detection and classification. To see the challenges of detecting and classifying ovarian tumors on ultrasound images and the current results obtained with the latest deep learning on object detection and classification. Unlike the approach of Qi *et al.* [4], to build an approach capable of quickly handling the problem of detecting and classifying ovarian tumors on ultrasound images. We use CNNs for ovarian tumor detection and classification without performing semantic segmentation to reduce computation time and make it possible to perform on the CPU. In this paper, we perform fine-tuning of the YOLOv8 model and its variations (YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, YOLOv8x) for the detection and classification of ovarian tumors on ultrasound images of the ovaries on the OTU 2D subset

Manuscript received August 18, 2023; revised September 18, 2023; accepted October 12, 2023; published February 24, 2024.

(*OTU 2D original set—OTU 2D-OS*) of the MMOTU dataset [4]. Fine-tuning and testing are performed on three configurations: (1) detect with or without ovarian tumors in the image; (2) detect and classify benign or malignant tumors in the image; (3) detect and classify of 8 classes ovarian tumors in the image.

The structure of the paper is organized as follows. In Section II, we briefly present some related works on ovarian tumor detection and classification. In Section III, we describe our framework, YOLOv8 model, and fine-tuning model for ovarian tumor detection and classification. The dataset and experimental results will be reported in Section IV. We finally conclude and give some ideas for future works.

II. Related Works

Early detection and diagnosis of ovarian cancer is a very important problem in medical examination and treatment. Early detection of ovarian tumors as benign or malignant greatly increases the chances of survival for patients. The problem of detecting, classifying, and segmenting ovarian tumors on ultrasound images is not new anymore. In 2014 and 2015, Khazendar *et al.* [9, 10] used the traditional approach of using Support Vector Machine (SVM) to train a classification model based on Local Binary Patterns (LBP) and histogram features to classify ovarian tumors on ultrasound images. The classification results on the accuracy measurement are 90%, 81%, and 69% based on classification decisions of high, medium, and low confidence, respectively. In this study, the labels: teratoma, endometrioma, serous cystadenofibroma, fibroma, serous cystadenoma, mucinous scystadenoma, functional cyst, and endometriosis are benign ovarian tumors. Labels: borderline, invasive malignant, and metastatic malignant are ovarian malignancies. The database for model training and testing the classification model includes 187 ultrasound images. This database is completely different from the *OTU 2D-OS* of the MMOTU dataset [4].

Tegnoor *et al.* [11] detected and classified ovarian cysts by pre-processing the ultrasound images using edge detection, histogram equalization, active contours without edges method, and morphological operations applied for image segmentation. Seven geometric features were extracted, and an SVM classifier was used to classify the image into follicle or non-follicle with a False Acceptance Rate (FAR) of 2.00% and False Rejection Rate (FRR) of 0.32% on the 35 images for training and 35 images for testing.

Srivastava *et al.* [12] used a fine-tuned VGG-16 network to detect an ovarian cyst with or without the ultrasound image. In fine-tuning, an already existing pre-trained model was modified with a similar type of task that it had earlier performed. Ovarian cyst detection results were evaluated on the accuracy measure, which is 92.11%.

Qi *et al.* [4] proposed an MMOTU dataset with a set of OTU 2D and OTU CEUS consisting of 1469 2D ultrasound images and 170 contrast-enhanced ultrasonography. In particular, the alignment feature was also proposed to the DS2Net network for semantic segmentation of ovarian tumor data on ultrasound images. Experiments have compared semantic segmentation of ovarian tumor data on ultrasound images between methods: PSPNet [13, 14] DANet [15], SegFormer [16], U-Net [17], TransUNet [18], BiseNetV2 [19]. The best binary semantic segmentation result on the OTU 2D is *IoU* = 82.46%, *mIoU* = 80% of SegFormer [16].

We recently performed a comparative study on ovarian tumor segmentation and classification on ultrasound imaging of the ovaries. This study evaluated the problem of ovarian tumor classification and segmentation on 8 labels of the OTU 2D dataset with some CNNs with the following average results: PSPNet (IoU = 56.87%, accuracy = 67.37%), U-net (IoU = 49.70%, accuracy = 61.27%), DANet (IoU = 58.65%, accuracy = 71.56%), Deeplabv3 (IoU = 59.75%, accuracy = 69.99%), PSANet (IoU = 58.66%, accuracy = 70.06%).

By YOLO categories, Terven *et al.* [8] performed a full survey of the architecture, new on each version of YOLO (from YOLOv1 to YOLOv8 and YOLO-NAS). Especially, the comparison results of object detection on Pascal VOC 2007 and COCO2017 datasets were also shown. At the same time, the future development directions of YOLO were also predicted.

III. Ovarian Tumors Detection and Classification

A. Background of YOLO

Before presenting YOLOv8, we reintroduce the general architecture of YOLO, which is shown in Fig. 1.

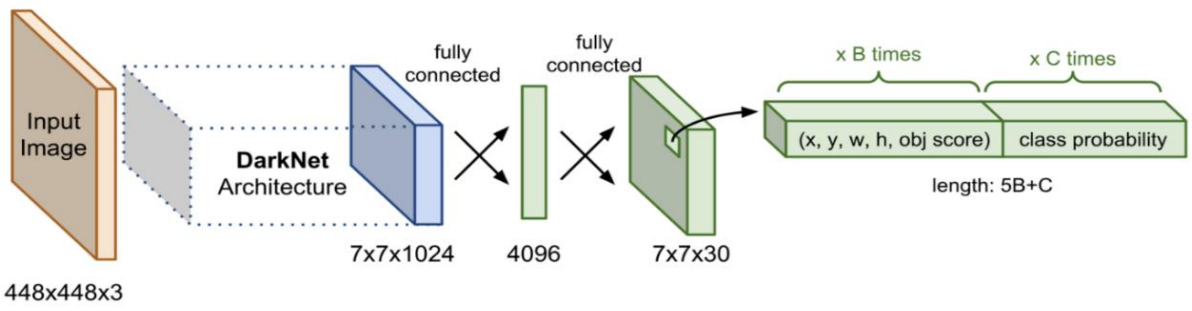

Fig. 1. The general architecture of YOLO [20].

YOLO architecture includes: base networks are convolution networks that perform the task of extracting features from images. The back part is the Extra Layers applied to detect objects on the feature map of the base

network. The base network is the Darknet. YOLO architectures are also quite diverse and can be customized into versions for many different input shapes. As shown in Fig. 1, the output of the base network is a feature map with dimensions of 7×7×1024 that will be used as input for Extra layers that predict the label and bounding box coordinates of the object. The output of the YOLO model is a vector that will include the components as Eq. (1).

$$y^T = [p_0, (t_x, t_y, t_w, t_h), (p1, p2, \ldots, pc)] \quad (1)$$

where p_0 is the predicted probability of an object appearing in the bounding box. (t_x, t_y, t_w, t_h) helps define the bounding box. (t_x, t_y) is the coordinate of the center and (t_w, t_h) is the width and height of the bounding box. $(p_1, p_2, \ldots, p_c)$ is the vector of prediction probability distributions of the classes.

- The size of the output will be determined by the number of classes. To find the bounding box for an object, YOLO will need anchor boxes as a basis for estimation. These anchor boxes will be predetermined and will surround the object relatively accurately. After that, the regression bounding box algorithm will refine the anchor box to create a predicted bounding box for the object. In a YOLO model include:
- Each object in the training image is distributed to an anchor box. In case there are 2 or more anchor boxes surrounding the object, we will determine the anchor box that has the highes IoU with a ground truth bounding box, as illustrated in Fig. 2.
- Each object in the training image is distributed to a cell on the feature map that contains the midpoint of the object.

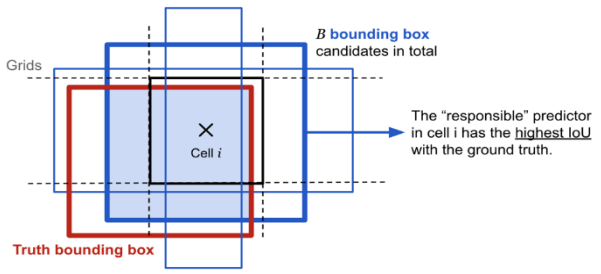

Fig. 2. Illustration of defining an anchor box for an object [21].

From Celli, we can identify 3 anchor boxes with blue borders. All three of these anchor boxes intersect with the object's bounding box. However, only the anchor box with the thickest blue border is selected as the anchor box for the object because it has the highest *IoU* compared to the ground truth bounding box.

To evaluate the error of the bounding box, YOLO uses used Loss function. YOLO uses the Sum-Squared Error (SSE) function. The values *x, y, w, h, C* are the values of the ground truth box, and the values $\tilde{x}, \tilde{y}, \tilde{w}, \tilde{h}, \tilde{C}$ are the predicted bounding box.

$$SSE = E_1 + E_2 + E_3 + E_4 + E_5 \quad (2)$$

where

$$E_1 = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} LF_{ij}^{obj} (x_i - \tilde{x}_i)^2 (y_i - \tilde{y}_i)^2 \quad (3)$$

$$E_2 = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} LF_{ij}^{obj} \left(\sqrt{w_i} - \sqrt{\tilde{w}_i}\right)^2 \left(\sqrt{h_i} - \sqrt{\tilde{h}_i}\right)^2 \quad (4)$$

$$E_3 = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} LF_{ij}^{obj} \left(C_i - \tilde{C}_i\right)^2 \quad (5)$$

$$E_4 = \lambda_{no\_obj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} LF_{ij}^{no\_obj} \left(C_i - \tilde{C}_i\right)^2 \quad (6)$$

$$E_5 = \sum_{i=0}^{S^2} \sum_{j=0}^{B} LF_{ij}^{obj} \sum_{c \in classes} \left(p_i(c) - \tilde{p}_i(c)\right)^2 \quad (7)$$

where E_1 is *xy loss* when the object exists at box_j in $cell_i$;
E_2 is *wh loss* when the object exists at box_j in $cell_i$;
E_3 is *confidence loss* when *the* object exists at box_j in box_j;
E_4 is *confidence loss* when objects do not exist in the boxes;
E_5 is *class probability loss* in the cell where the object exists.
Furthermore, $LF_{ij}^{obj} = 1$ if in the i^{th} cell, there is a j^{th} box containing an object;
$LF_{ij}^{no\_obj}$ is the opposite of LF_{ij}^{obj};
$LF_{ij}^{obj} = 1$ if the i^{th} the cell contains an object (otherwise, it is 0);
$\lambda_{coord}, \lambda_{no\_obj}$ is the component weight.

However, even a good model still has a case: predicting multiple bounding boxes for the same object. To solve this problem, YOLO filters out redundant bounding boxes (duplicate and same class) by non-maximum suppression. The steps of non-max suppression are as follows:

Step 1: First, we will find a way to reduce the number of bounding boxes by filtering out all bounding boxes whose probability of containing objects is less than a certain threshold, usually 0.5.

Step 2: For intersecting bounding boxes, non-max suppression will select a bounding box with the highest probability of containing an object. Then calculate the *IoU* interference index with the remaining bounding boxes.

B. YOLOv8

Since 2015 there have been many versions of YOLO developed to perform object detection, the versions of YOLO are shown in Table I.

TABLE I. TIMELINE VERSIONS OF YOLO

YOLO series	Year	Framework	Backbone
YOLOv1 [20]	2015	Darknet	Darknet-19
YOLOv2 [21]	2016	Darknet	Darknet-19
YOLOv9000 [21]	2016	Darknet	Darknet-19
YOLOv3 [22]	2018	Darknet	Darknet-53
PP-YOLO [23]	2020	ResNet50-vd	ResNet50-vd
YOLOv4 [24]	2020	Darknet	CSPDarknet-53
YOLOv5 [25]	2020	PyTorch	Modified CSPv7
YOLOS [26]	2021	PyTorch	Transformer block
PP-YOLOV2 [27]	2021	PyTorch	ResNet50-vd-dcn
YOLOv6 [28]	2022	PyTorch	EfficientRep
YOLOv7 [29]	2022	PyTorch	RepConvN
YOLOv8 [7]	2023	PyTorch	YOLOv8

YOLOv8 is the current latest version of the YOLO family to be published. Like other versions of YOLO, YOLOv8 also includes two main components: Backbone and Head. The architecture of YOLOv8 is shown in Fig. 3. YOLOv8 is developed by Ultralytics, who also created the YOLOv5 model. YOLOv8 includes many changes and improvements in architecture and developer experience compared to YOLOv5. The YOLOv8 architecture features several new enhancements and combinations introduced by Ultralytics.

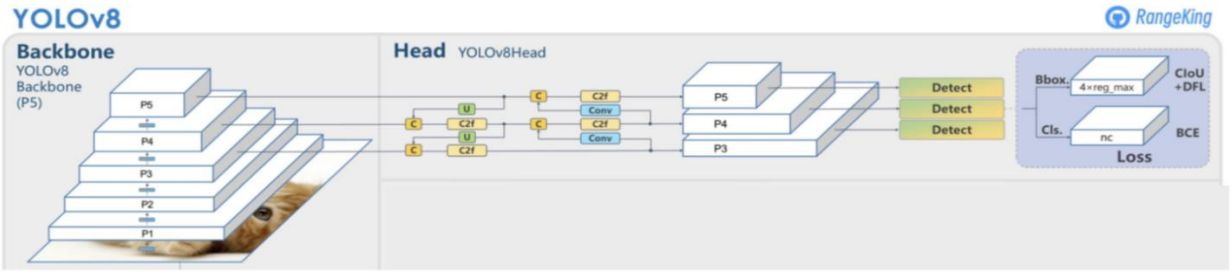

Fig. 3. YOLOv8 architecture [7].

First, the backbone of YOLOv8 is the same as YOLOv5, as it uses the CSPDarknet53 feature extractor. It has some changes like $C2f$ replacing $C3$ to combine high-level features with contextual information to improve detection accuracy. The first 6×6 convolution in the body is converted to the 3×3 convolution. In $C2f$, the output from the bottleneck (which is a combination of two 3×3 transitions with the remaining connections) is combined, as illustrated in Fig. 4, where "f" is the number of features, "e" is the expansion rate and CBS is a block composed of a Conv, a BatchNorm and a SiLU. While in $C3$, only the output from the last bottleneck is used.

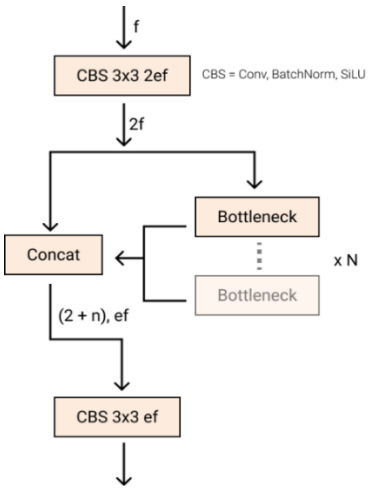

Fig. 4. Illustrating of the C2f module in YOLOv8.

Second, YOLOv8 uses an anchor-free model with a detached head to handle object, classification, and regression tasks independently. This design allows each branch to focus on its task and improves the overall accuracy of the model. The sigmoid function as the activation function for the feature score is used in the output layer of YOLOv8. From this, the probability that the bounding box contains an object is represented. It uses the SoftMax function for the class probability, which represents the probability of objects belonging to each possible class.

Third, two convolutions (#10 and #14 in the YOLOv5 config) were removed. YOLOv8 provides a semantic segmentation model called the YOLOv8-Seg model to achieve state-of-the-art results on various object detection and semantic segmentation benchmarks while maintaining high speed and efficiency. The loss functions for bounding box loss and binary cross-entropy for classification loss and to improve the predicted results on small objects.

Fourth, the bottleneck in YOLOv8 is still the same as in YOLOv5, except that the kernel size of the first convolution has been changed from 1×1 to 3×3. This change represents a change to the ResNet block, as illustrated in Fig. 5.

Currently, YOLOv8 offers five types of pre-trained models for object detection: YOLOv8n (Nano), YOLOv8s (Small), YOLOv8m (Medium), YOLOv8l (Large), and YOLOv8x (Extra Large). Among them, YOLOv8n is the fastest and smallest, and YOLOv8x is the most accurate but slowest.

C. *Fine-Tuning YOLOv8 for Ovarian Tumors Detection and Classification*

YOLOv8 allows fine-tuning of ovarian tumor detection models on custom datasets. In this paper, the input data is the *OTU 2D-OS* of the MMOTU dataset [4]. The ultrasound images of the *OTU 2D-OS* normalized to the size of 640×480 and the bounding box is normalized in the format of the COCO2017 dataset [30], and the labels of ovarian tumors on the images. We perform finetuning YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x for three types of output predictions: (1) with or without ovarian tumor on the image; (2) detection and classification of 8 types of ovarian tumors; (3) classification of benign ovarian tumor and malignant ovarian tumor on the image. The implementation process is shown in Fig. 6.

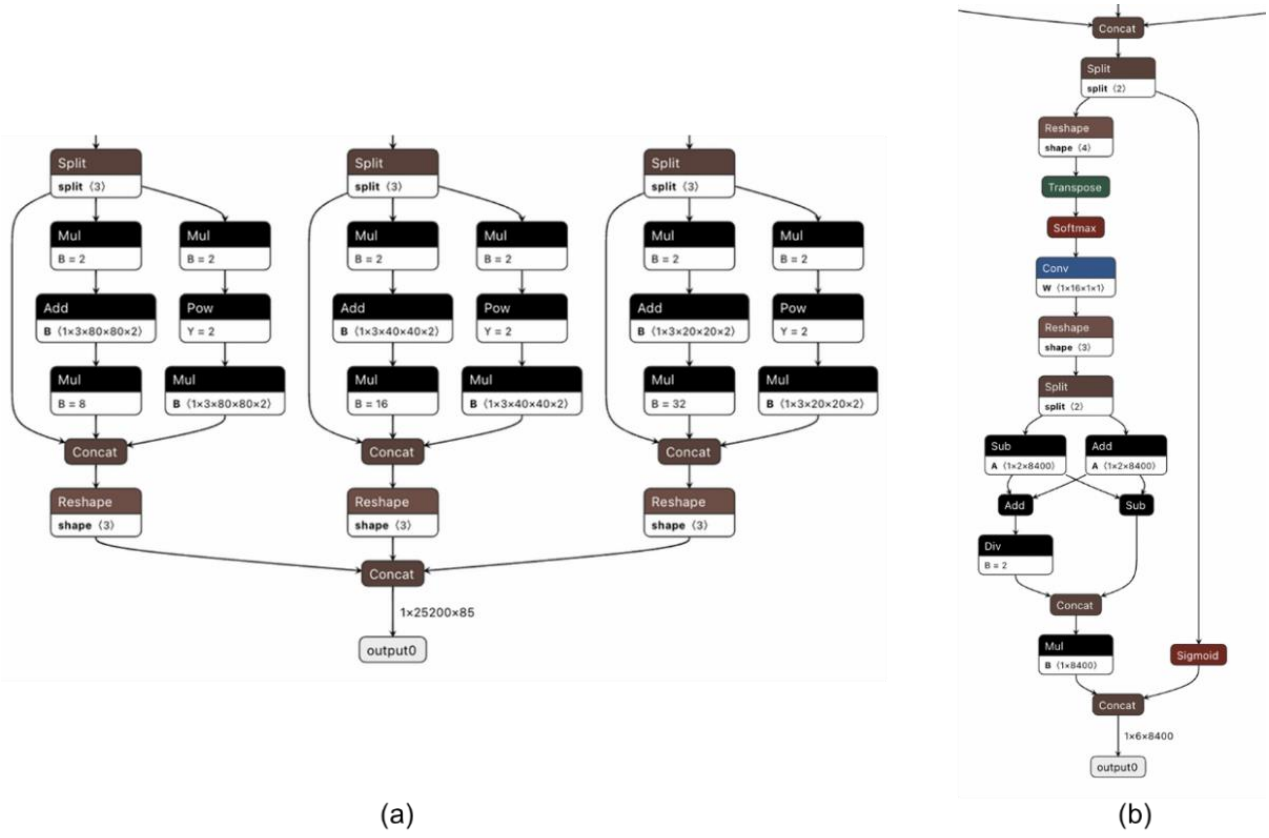

Fig. 5. Architecture of the head part of YOLOv5 and YOLOv8, (a) the detection head of YOLOv5, (b) The detection head for YOLOv8.

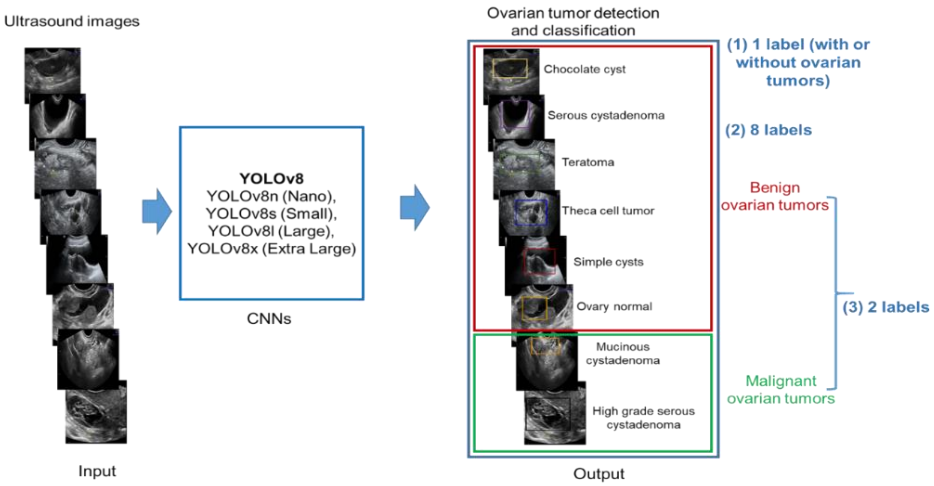

Fig. 6. The model of performing fine-tuning of the ovarian tumor detection and classification model on the *OTU 2D-OS*.

In this paper, we use the source code of YOLOv8 in [31]. Before performing the training and evaluation of the right-hand detection model on the image, we normalize the bounding box annotation (as illustrated in Fig. 7) of the hand to YOLO's bounding box format with the COCO2017 dataset [30], as presented in Eq. (8).

$$\alpha = \frac{x\_max+x\_min}{\frac{2}{w\_b}}; b = \frac{x\_max+x\_min}{\frac{2}{h\_b}}$$

$$c = \frac{w\_b}{w\_im}; d = \frac{h\_b}{h\_im} \quad (8)$$

where $(x\ max, y\ max)$ are the coordinates of the top right corner of the hand bounding box on the image, $(x\ max, y\ max)$ is the coordinates of the bottom left corner of the hand bounding box on the image, w_b is the width size of the bounding box of the hand on the image, h_b is the height size of the hand bounding box on the image, w_im is the width size of the image, h_im is the height size of the image. The data structure of the bounding box in the format of COCO2017 used for training and evaluation has the form (l, a, b, c, d), where l is the label of the object.

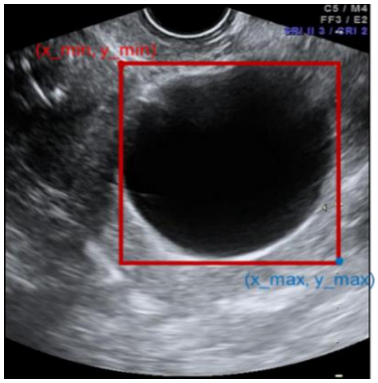

Fig. 7. Illustration of the bounding box annotation of the ovarian tumor.

IV. Experimental Result

A. Data Collection

In this paper, we use the MMOTU dataset [4]. This is a dataset of ultrasound images of the ovaries. Ultrasound is one of the techniques used in diagnosing ovarian cancer. It is a technique that possesses advantages such as simplicity, ease of implementation, low cost, and high effectiveness. In this technique, doctors can perform an ultrasound through the vaginal route by inserting an ultrasound probe into the vagina, using sound waves to detect tumors in the female reproductive organs, including the ovaries.

The MMOTU dataset, the pixel-wise semantic annotation, and the global-wise category annotation are provided by 27 experts in the Obstetrics and Gynecology department. Each image is first annotated by one expert and then checked by another expert, which guarantees the annotating quality. During annotating, experts refer to

Pathhological reports, which makes the annotations accurate and convincing.

In this paper, we perform a follow-up study based on the *OTU 2D-OS* of the MMOTU dataset [4]. The *OTU 2D-OS* includes 1469 2D ultrasound images and eight labels, as illustrated in Fig. 8.

At the same time, we perform data preparation according to three configurations of datasets, which are singlelabel data with or without ovarian tumors (1-label), (8-label) data on ovarian tumor types as shown in of MMOTU dataset [4], and (2-label), the first six labels (labels 1 to 6) of the eight labels of *OTU 2D-OS* are benign ovarian tumors and the remaining two labels (labels 7th and 8th) of the eight labels of *OTU 2D-OS* are malignant ovarian tumors. This problem is explained as follows: From the 8 classes/labels of the *OTU 2D-OS*, considering the opinions of experts, these 8 classes can be divided into two categories [9, 10]: benign and malignant. Benign ovarian tummor includes Chocolate cyst, Serous cystadenoma, Teratoma, Theca cell tumor, Simple cysts, and Ovary normal. Maligant ovarian tumor includes Mucinous cystadenoma, Hight-grade serous cystadenoma. The number of samples to train and validate the YOLOv8 model is shown in Table II.

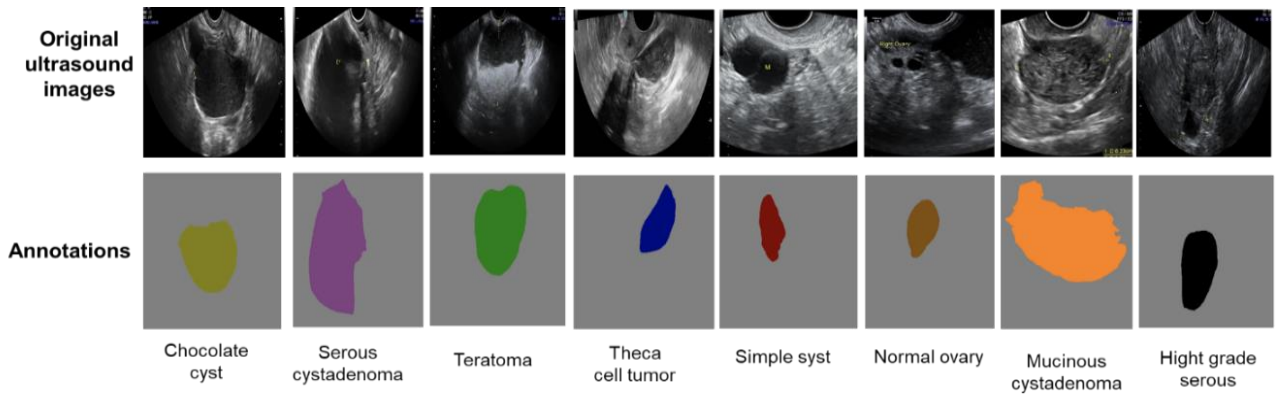

Fig. 8. Illustration of some *OTU 2D-OS* images with eight labels of ovarian tumor in the MMOTU dataset. The top row is an ultrasound image. The bottom row is the annotation data, each label has a different color.

TABLE II. The Number of Image Samples for the Training Set and Validation Set of *OTU 2D-OS*

Classification type	Labels of ovarian tumors	Training set	Validation set	Training set	Validation set
Begin	Chocolate cyst	226	110	891	421
	Serous cystadenoma	153	66		
	Teratoma	228	108		
	Theca cell tumor	57	31		
	Simple syst	47	19		
	Normal ovary	180	87		
Malignant	Mucinous cystadenoma	71	33	109	48
	Hight grade serous	38	15		

B. Metrics

To evaluate the performance of detection of ovarian tumors, we utilize the *IoU* (Intersection over Union) measure, as presented in Fig. 9. Fig. 9 shows the intersection area (blue area on the numerator) of the bounding box annotation and the detected bounding box divided by the union area (blue area under the denominator) of the bounding box annotation and the detected bounding box. The Precision (*P*), Recall (R), and mAP (mean

Average Precision) are also used to calculate and compare results.

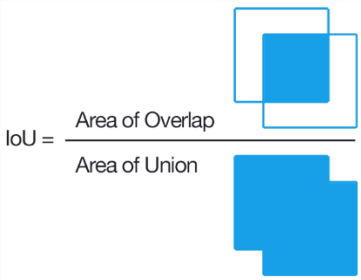

Fig. 9. The formula to calculate the Intersection over Union (IoU).

$$P = \frac{TP}{TP + FP};\ R = \frac{TP}{TP + FN} \quad (9)$$

$$mAP = \frac{1}{|Classes|} \sum_{c\ \epsilon\ Classes} \frac{TP(c)}{TP(c) + FP(c)} \quad (10)$$

where TP is the correct prediction, FP is the false prediction, TN is a false prediction (the label is positive but the prediction is negative), and FN is the unpredicted ground truth data area. The threshold of *IoU* we use is 0.5.

In this paper, we deployed the studied models on a server with NVIDIA GeForce RTX 2080 Ti, 12GB GPU for fine-tuning, training, and testing. The programs were written in the Python language (≥3.9 version) with the support of the CUDA 11.2/cuDNN 8.1.0 libraries. In addition, there are several libraries such as OpenCV, matplotlib, ultralytics, numpy, Pillow, tensorboard, Torch, etc. In this paper, we fine-tune the YOLOv7, and YOLOv8 models with 100 epochs, and the size of the image is normalized to 640.

C. Results and Discussions

The results of detection and classification with or without ovarian tumor on *OTU 2D-OS* when fine-tuning by YOLOv7 and YOLOv8 are shown in Table III. The results show that YOLOv8 is about 19% more accurate (precision-*P*) than YOLOv7. The detection result is greater than 90%. *P* is a very meaningful measure because it is calculated based on the number of samples correctly predicted to be an ovarian tumor divided by the number of samples correctly predicted to be an ovarian tumor plus the number of samples incorrectly predicted to be an ovarian tumor (false positive). This means that an image region of a normal ovary is mistakenly predicted to be an ovarian tumor region. The wrong prediction result of YOLO8x is 8.74%, this result is still very high. Therefore, this model can not be applied to practical applications. The *R* result of YOLOv8x is 82.3%, which is not a good result. *R* is also a very important measure, as it is calculated based on the total number of correct prediction samples divided by the total number of correct prediction samples plus the number of wrong prediction samples (the image area containing the ovarian tumor is predicted to be without an ovarian tumor). This problem is very dangerous in detecting and diagnosing ovarian tumors since cases of patients with ovarian tumors are diagnosed as not containing ovarian tumors. In a practical application such as detecting and diagnosing cancer, it is often accepted that "a mistake is better than a miss", that is, it is possible to accept an image area that does not contain an ovarian tumor but can be detected mistaken for an ovarian tumor. This happens in measure *P*. *mAP*50 (Mean Average Precision) is the average precision of predictions with an *IoU* threshold of 50%, which means that an ultrasound image region whose ground-truth data is an ovarian tumor and the predicted data region is an ovarian tumor and has an area greater than or equal to 50% of the area of the ground-truth ovarian tumor data area is considered a correct detection. This is a bad result and can not be applied to help doctors diagnose and detect ovarian tumors based on ultrasound images.

TABLE III. RESULTS OF DETECTION AND CLASSIFICATION WITH OR WITHOUT OVARIAN TUMOR ON THE IMAGE OF *OTU 2D-OS* (EPOCH=100)

Label	Methods/ Measurement	Precision (*P*-%)	Recall	mAP50 (%)
With/without Ovarian tumors	**YOLOv7**	86	87	91
	YOLOv7-w6	83	82	86
	YOLOv7-d6	84	88	92
	YOLOv7-e6	90	87	92
	YOLOv8n	89.2	87.7	93
	YOLOv8s	89.3	88.9	92.9
	YOLOv8m	89.6	86.1	93.3
	YOLOv8l	88.8	87.7	94.1
	YOLOv8x	**91.26**	**83.3**	**92.67**

The results of the detection and classification of 8 ovarian tumor classes on the *OTU 2D-OS* are shown in Table IV. The results in Table IV show that YOLOv8X has the highest accuracy and is 5% higher than YOLOv7. The results of detecting and classifying ovarian tumors on ultrasound images of YOLOv8X are the best (*P* = 71.5 %, *R* = 68.7%, *mAP*50=70.6%). Similar to the above presentation about *P*, *R*, and *mAP*50, the rate of ultrasound image data areas that are not ovarian tumors that are mistakenly detected as ovarian tumors with one of the 8 labels of YOLOv8x is 28.5%. The proportion of ultrasound image data regions labeled as one of the 8 ovarian tumor labels that were detected as not ovarian tumors or wrong classified to another label was 31.3%. It still has a huge error. Therefore, this result has not been applied in practice to diagnose ovarian tumors on ovarian ultrasound images to help doctors perform because each doctor's diagnosis has a great psychological impact on the patient.

The results of the detection and classification of benign and malignant ovarian tumors on the *OTU 2D-OS* are shown in Table V. In Table V, the best ovarian tumor detection and classification is (*P* = 74.9%) of YOLOv8x. Just as shown above, the rate of ultrasound image data areas that are not ovarian tumors that are mistakenly detected as benign or malignant ovarian tumors by YOLOv8x is 25.1%. The rate of ultrasound image data regions labeled as benign or malignant ovarian tumors that were detected as not ovarian tumors or wrong classified to other labels was 41%. These results need to be improved so that they can be applied in practice to support doctors in diagnosing ovarian tumors as benign or malignant. Since then there has been no need for further tests for the patient

to save on medical examination and treatment costs. Fig. 10 shows the confusion matrix of ovarian tumor detection and classification with 8 classes of ovarian tumors of the *OTU 2D-OS* when using the fine-tuned YOLOv8x model for prediction. Fig. 10 also shows the prediction result of the 3rd label (Teratoma) having the highest result of 0.81, and the 5th label (Simple syst) having the lowest result of 53%. As shown in Fig. 8, the geometric structure of ovarian tumors on ultrasound and ground-truth images is the simplest geometric structure. The 5th label is detected and mistakenly classified with the 2nd label (Serous cystadenoma) and the background. Another label that also has a high rate of incorrect classification is the 4th label, which is often mistakenly classified with the 5th label.

TABLE IV. OVARIAN TUMORS RESULTS OF DETECTION AND CLASSIFICATION OF 8 CLASSES ON THE *OTU 2D-OS*

Methods	Labels	Measurements		
		Precision (*P*-%)	Recall (R-%)	mAP50 (%)
YOLOv7	**Average**	52	53	49
YOLOv7-w6		39	51	46
YOLOv7-d6		38	64	46
YOLOv7-e6		32	61	45
YOLOv8n	Chocolate cyst	74.6	70.9	75.2
	Serous cystadenoma	74.7	71.2	76.8
	Teratoma	76.4	81.5	86.3
	Theca cell tumor	71.9	58.1	58.3
	Simple syst	69.9	52.6	60.7
	Normal ovary	71.2	68.0	73.3
	Mucinous cystadenoma	78.0	57.6	69.4
	Hight grade serous	36.3	46.7	46.2
	Average	**69.1**	**63.3**	**68.3**
YOLOv8s	Chocolate cyst	73.5	68.2	73.7
	Serous cystadenoma	67.4	63.6	70.1
	Teratoma	77.7	78.7	82.9
	Theca cell tumor	54.5	51.6	54.5
	Simple syst	26.6	47.4	43.2
	Normal ovary	68.7	64.4	72.2
	Mucinous cystadenoma	45.7	56.2	47.8
	Hight grade serous	22.5	46.7	40.1
	Average	**54.6**	**59.6**	**60.6**
YOLOv8m	Chocolate cyst	69.2	71.8	76.8
	Serous cystadenoma	63.7	71.2	73.1
	Teratoma	85.2	79.8	86.6
	Theca cell tumor	72.5	59.5	62.6
	Simple syst	66.5	52.2	53.9
	Normal ovary	89.4	69.0	77.1
	Mucinous cystadenoma	68.8	63.3	61.6
	Hight grade serous	51.8	60.0	51.5
	Average	**70.9**	**65.9**	**67.9**
YOLOv8l	Chocolate cyst	80.0	62.7	75.1
	Serous cystadenoma	72.0	66.2	73.9
	Teratoma	85.8	72.2	81.8
	Theca cell tumor	75.3	58.9	67.5
	Simple syst	39.0	53.8	46.1
	Normal ovary	77.6	59.8	72.6
	Mucinous cystadenoma	58.9	57.6	61.5
	Hight grade serous	43.2	46.7	34.5
	Average	**66.5**	**59.7**	**64.1**
YOLOv8x	Chocolate cyst	80.0	69.0	74.3
	Serous cystadenoma	71.4	75.8	80.6
	Teratoma	87.1	82.4	86.3
	Theca cell tumor	69.7	58.1	69.6
	Simple syst	61.8	68.4	61.7
	Normal ovary	82.0	71.3	82.5
	Mucinous cystadenoma	63.1	72.7	63.3
	Hight grade serous	56.6	52.3	46.4
	Average	**71.5**	**68.7**	**70.6**

TABLE IV. THE RESULTS OF DETECTION AND CLASSIFICATION OF BENIGN AND MALIGNANT OVARIAN TUMORS ON THE *OTU 2D-OS*

Methods	Labels	Measurements		
		Precision (*P*-%)	Recall (R-%)	mAP50 (%)
YOLOv8n	Benign	82.7	89.5	89.7
	Malignant	55.3	45.8	54.4
	Average	69	67.7	72
YOLOv8s	Benign	86.1	81.5	87.9
	Malignant	60.4	44.4	54.5
	Average	73.2	63	71.2

YOLOv8m	Benign	82.2	86.5	89.3
	Malignant	56.3	62.5	56.7
	Average	69.3	74.5	73
YOLOv8l	Benign	81.2	86.2	88.6
	Malignant	48.5	54.2	45.4
	Average	64.9	70.2	67
YOLOv8x	Benign	86.3	78.4	89.3
	Malignant	63.4	39.6	44.4
	Average	74.9	59	66.8

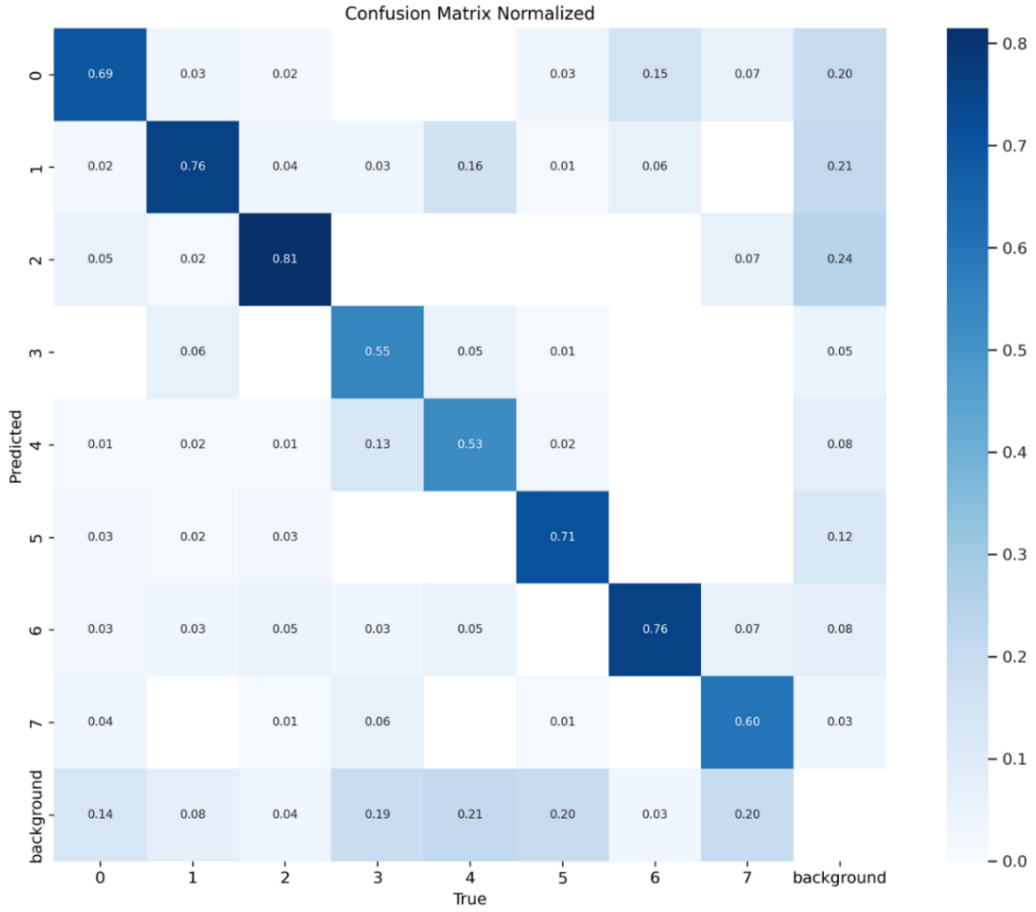

Fig. 10. Illustrate the confusion matrix normalized for detection and classification of 8 classes of ovarian tumors of the *OTU 2D-OS* using the fine-tuned YOLOv8x model.

Fig. 11 illustrates the ground truth and predictive results for 8 classes of ovarian tumors and the bounding box of ovarian tumors on the *OTU 2D-OS*. The top (four rows of images above) is the ground truth boundingbox and tumor label data and the bottom (four rows of images below) is the bounding box and label prediction result of ovarian tumor. At the top, the first image is named *"1002.jpg"*, labeled *"5"* and the ovarian tumor is represented by a green bounding box. Prediction results at the bottom, the first image is named *"1002 .jpg"*, the ovarian tumor label prediction is *"6"* and the confidence score is 0.3, represented by a lighter blue bounding box labeled *"5"*. This is a case where the label of an ovarian tumor was wrongly predicted. In the second image of row 1 at the top, the image is named *"548.jpg"*, the label of the ovarian tumor is *"1"* and is represented by a bounding box with the IndianRed color. The prediction and classification results are shown in the second image row 1 at the bottom, the image is named *"548.jpg"*, the prediction result is labeled *"1"*, the confidence score is 0.5, and is represented by a bounding box equals bounding box with color IndianRed. This result is the result of the correct classification of ovarian tumors. The results in Tables III–V also show that as the number of ovarian tumor classes increases, the precision and recall decrease. Currently, the rate of detection and wrong classification is high. In particular, the results of R are much lower than P , so the pre-trained models have a high rate of wrong predictions and classifications (labeled as a type of ovarian tumor but wrongly predicted as no ovarian tumor or wrongly classified to another label).

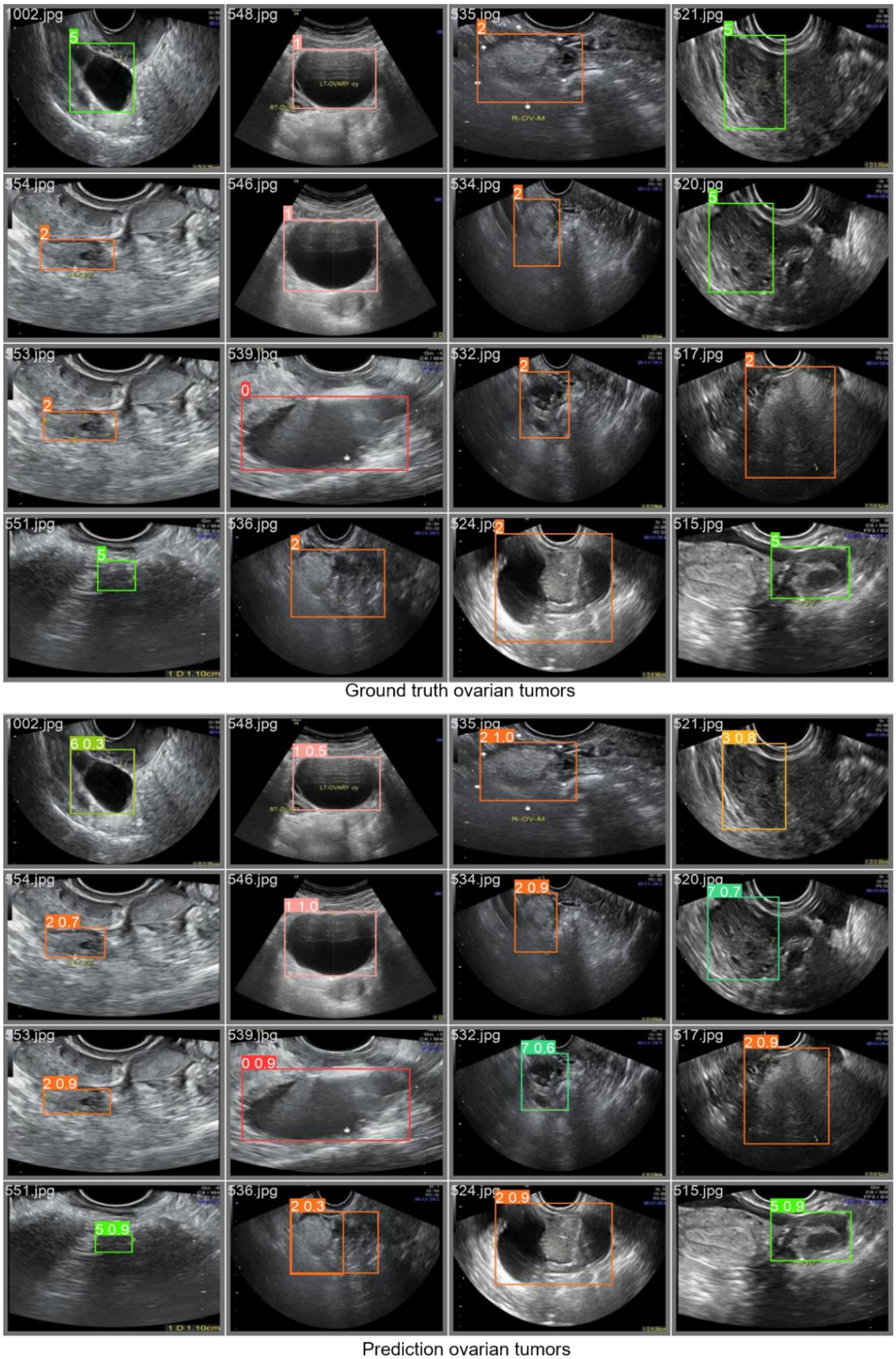

Fig. 11. Illustrating the ground truth data and prediction of 8 classes of ovarian tumors by YOLOv8x. The top is the ground truth data about the label and bounding box of the ovarian tumor area, the bottom is the prediction result of the label, confidence score, and bounding box of the ovarian tumor.

Table VI shows the processing time of YOLOv7 and YOLOv8x on the GPU and CPU.

TABLE VI. THE PROCESSING TIME FOR DETECTING AND CLASSIFYING 8 OVARIAN TUMOR CLASSES ON THE *OTU 2D-OS* WHEN PERFORMING CALCULATIONS ON GPU AND CPU

Measurements	Computing Device	YOLOv7 (fps)	YOLOv8x (fps)
Processing time	GPU	204	186
Processing time	CPU	2.04	1.84

Based on the results, analysis, and discussion presented above, the main contributions of this paper include:

- Normalizing the *OTU 2D-OS* with the data in the box format of the YOLO model; Normalizing the data when divided into one class (with or without ovarian tumor), two classes (benign and malignant), and eight classes (split in the MMOTU dataset).
- Performing fine-tuning the YOLOv8 model with variations (YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, YOLOv8x) for ovarian tumor detection and classification on three configurations: one class (with or without of ovarian tumor); two classes (benign or malignant); eight classes (chocolate cyst, serous cystadenoma, teratoma, theca cell tumor, simple syst, normal ovary, mucinous cystadenoma, high grade serous).
- Testing and comparing the results of the YOLOv8 model with its variants compared with YOLOv7 when performing ovarian tumor detection and classification on the *OTU 2D-OS*.

V. CONCLUSION AND FUTURE WORKS

The problem of detecting and classifying ovarian tumors on ultrasound images is very important to research in medical examination and treatment as well as diagnosing female patients. In this paper, we perform a comparative study based on fine-tuning YOLOv8 with variants (YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x) to detect and classify ovarian tumors on OTU ultrasound images. Ovarian tumor detection and classification model with three models: detection and classification model with or without ovarian tumor on the image; the model for detecting and classifying 8 classes of ovarian tumors; the model for detecting and classifying two classes of ovarian tumors as benign and malignant. The best average results of ovarian tumor detection and classification on ovarian ultrasound images are as follows: 1 label (P = 91.26%, R = 83.3%, mAP50 = 92.67%), 8 labels (P = 71.5 %, R = 68.7%, mAP50 = 70.6%), 2 labels (P = 74.9%, R = 59%, mAP50 = 66.8%). The results showing the presence or absence of ovarian tumors on ultrasound images of the ovaries can be applied in practice to help doctors spend less time reviewing the results, and further tests can be performed to classify and diagnose ovarian tumors. The results on the classification of ovarian tumors with 8 classes and two classes of benign and malignant ovarian tumors require further research and improvement of results soon.

CONFLIC OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Thi-Loan Pham did methods, writing, and editing; Van-Hung Le did methods, experimenting, writing, and editing. Both authors had approved the final version.

FUNDING

This research is funded by Tan Trao University in Tuyen Quang province, Vietnam.

REFERENCES

[1] WCRF International. (2020). Ovarian cancer statistics. [Online]. Available: https://www.wcrf.org/cancer-trends/ovarian-cancer-statistics/

[2] Baoson hospital. (2020). More than 70late stage. [Online]. Available: https://www.baosonhospital.com/hon-70-nguoi-benh-ung-thu-phat-hien-benh-o-giai-doan-muon

[3] D. Crosby, S. Bhatia, K. M. Brindle, L. M. Coussens, C. Dive, M. Emberton, S. Esener, R. C. Fitzgerald, S. S. Gambhir, P. Kuhn, T. R. Rebbeck, and S. Balasubramanian, "Early detection of cancer," *Science*, vol. 375, 6586, 2022.

[4] Q. Zhao, S. Lyu, W. Bai, L. Cai, B. Liu, M. Wu, X. Sang, M. Yang, and L. Chen, "A multi-modality ovarian tumor ultrasound image dataset for unsupervised cross-domain semantic segmentation," arXiv preprint, arXiv:2207.06799, 2022.

[5] T.-L. Pham, V.-H. L. B, T.-H. Tran, and D. H. Vu, "Comprehensive study on semantic segmentation of ovarian tumors," *LNNS 734*, vol. 3, pp. 262–273, 2023.

[6] S. Liu, Y. Wang, X. Yang, B. Lei, L. Liu, S. X. Li, D. Ni, and T. Wang, "Deep learning in medical ultrasound analysis: A review," *Engineering*, vol. 5, no. 2, pp. 261–275, 2019. https://doi.org/10.1016/j.eng.2018.11.020

[7] D. Reis, J. Kupec, J. Hong, and A. Daoudi, "Real-time flying object detection with YOLOv8," arXiv preprint, arXiv:2305.09972, 2023.

[8] J. Terven and D. Cordova-Esparza, "A comprehensive review of YOLO: From YOLOv1 and beyond," arXiv preprint, arXiv:2304.00501, 2023.

[9] S. Khazendar, H. Al-Assam, H. Du, S. Jassim, A. Sayasneh, T. Bourne, J. Kaijser, and D. Timmerman, "Automated classification of static ultrasound images of ovarian tumours based on decision level fusion," in *Proc. 2014 6th Computer Science and Electronic Engineering Conference, CEEC 2014*, 2014, pp. 148–153.

[10] S. Khazendar, A. Sayasneh, H. Al-Assam, H. Du, J. Kaijser, L. Ferrara, D. Timmerman, S. Jassim, and T. Bourne, "Automated characterisation of ultrasound images of ovarian tumours: The diagnostic accuracy of a support vector machine and image processing with a local binary pattern operator." *Facts, Views Vision in ObGyn*, vol. 7, no. 1, pp. 7–15, 2015.

[11] J. R. Tegnoor, "Automated ovarian classification in digital ultrasound images using SVM," *International Journal of Engineering Research & Technology*, vol. 1, no. 6, pp. 1–17, 2012.

[12] S. Srivastava, P. Kumar, V. Chaudhry, and A. Singh, "Detection of ovarian cyst in ultrasound images using fine-tuned VGG-16 deep learning network," *SN Computer Science*, vol. 1, pp. 1–8, 2020.

[13] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[14] R. Wightman, H. Touvron, and H. Jegou, "Resnet strikes back: An improved training procedure in TIMM," arXiv preprint, arXiv:2110.00476, 2021.

[15] J. Fu *et al*., "Dual attention network for scene segmentation," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[16] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," arXiv preprint, arXiv:2105.15203, 2021.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2015, pp. 234–241.
[18] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," arXiv preprint, arXiv:2102.04306, 2021.
[19] C. Yu, C. Gao, J. Wang, G. Yu, C. Shen, and N. Sang, "Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation," *International Journal of Computer Vision*, pp. 1–18, 2021.
[20] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.
[21] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc.30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, pp. 6517– 6525.
[22] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint, arXiv:1804.02767, 2018.
[23] X. Long, K. Deng, G. Wang, Y. Zhang, Q. Dang, Y. Gao, H. Shen, J. Ren, S. Han, E. Ding, and S. Wen, "PP-YOLO: An effective and efficient implementation of object detector," arXiv preprint, arXiv:2007.12099, 2020.
[24] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv Preprint, arXiv:2004.10934, 2020.
[25] R. Couturier, H. N. Noura, O. Salman, and A. Sider, "A deep learning object detection method for an efficient clusters initialization," arXiv preprint, arXiv:2104.13634, 2021.
[26] Y. Fang, B. Liao, X. Wang, J. Fang, J. Qi, R. Wu, J. Niu, and W. Liu, "You only look at one sequence: Rethinking transformer in vision through object detection," *Advances in Neural Information Processing Systems*, vol. 31, pp. 26183–26197, 2021.
[27] X. Huang, X. Wang, W. Lv, X. Bai, X. Long, K. Deng, Q. Dang, S. Han, Q. Liu, X. Hu, D. Yu, Y. Ma, and O. Yoshie, "PP-YOLOv2: A practical object detector," arXiv preprint, arXiv:2104.10419, 2021.
[28] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, Y. Li, B. Zhang, Y. Liang, L. Zhou, X. Xu, X. Chu, X. Wei, and X. Wei, "YOLOv6: A single stage object detection framework for industrial applications," arXiv preprint, arXiv:2209.02976, 2022.
[29] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," arXiv preprint, arXiv:2207.02696, 2022.
[30] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8693 LNCS, no. PART 5, 2014, pp. 740–755.
[31] S. Rath. (2023). YOLOv8 Ultralytics: State-of-the-art YOLO models. [Online]. Available: https://learnopencv.com/ultralytics-yolov8/

Copyright © 2024 by the authors. This is an open access article distributed under the Creative Commons Attribution License (CC BY-NC-ND 4.0), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.