Empirical Study of Ovarian Tumors Detection and Classification in Ultrasound Images

Thi-Loan Pham [1,3], Van-Hung Le [2], Thi-Lan Le [3], Hai Vu [3], Duy-Hai Vu [3], Thanh-Hai Tran [3]

[1] Hai Duong University, Vietnam

[2] Tan Trao University, Vietnam

[3] School of Electrical and Electronic Engineering (SEEE), Hanoi University of Science and Technology, Vietnam

Correspondence: Van-Hung Le, email: van-hung.le@mica.edu.vn

Communication: received 11 Oct 2023, revised 09 Feb 2024, accepted 16 May 2024

Digital Object Identifier: 10.32913/mic-ict-research.v2024.n2.1237

Abstract: Ovarian cancer is one of the leading causes of death in women. Ultrasound images often provide initial findings and diagnosis before further tests are performed. Despite some progress in medical image analysis, the detection and recognition of ovarian cancers remains still quite limited. Most existing works focused on detecting or classifying two tumor categories, namely benign and malignant lesions. This paper presents an initial comprehensive study of ovarian tumor detection, and at the same time classifies eight classes of ovarian tumors in ultrasound images using deep-learning models. Due to the lack of training data, we implemented a data augmentation process and examined the improvement of performance with and without data augmentation. Experiments carried out on the OTU-2D set of MMOTU dataset with YOLOv5, YOLOv7, and YOLOv7 variants show promising detection and classification results.

Keywords: *Ovarian tumor, Classification, Data augmentation, Ultrasound images, YOLOv5 and YOLOv7.*

I. INTRODUCTION

According to statistics, the new incidence of ovarian cancer is 10.3 per 100,000 women per year, and the mortality rate is 6.3 per 100,000 women per year [1]. Currently, overcrowding in hospitals, particularly in developing countries such as Vietnam [2], leads to missed or misdiagnosed diseases, causing unintended consequences for patients and an economic burden on families and society. In addition, the diagnosis of the disease often depends on the doctor's experience. Therefore, to tackle these issues, it is necessary to develop tools to encourage the automatic detection and determination of ovarian tumors from ovarian ultrasound images.

Today, with the increased development of artificial intelligence, many applications of deep learning in medical image analysis has achieved expressive results. For instance, automatic detection of lung tumors from x-ray images [3], automatic detection and segmentation of lesions in gastrointestinal endoscopy images [4–6], breast and cervical cancer images analysis [7], [8], MRI images analysis [9], etc. However, the studies dedicated to ovary ultrasound analysis are limited and few datasets are publicly available for the research community. Most existing works focus on classifying images as normal or abnormal, or detecting abnormalities [10]. Further analysis such as classifying tumors into different tumor classes is still not thoroughly investigated.

Regarding ovarian ultrasound image analysis, recently Pham et al. [11] performed a study on the semantic segmentation of ovarian tumors on the OTU_2D subset of the MMOTU dataset. However, as the analysis is carried out at the pixel level, the computation requires GPU and is quite consuming time. The current best accuracy results for segmentation of 8 classes of ovarian tumors when using the accuracy measure is 71.56%, because the ovarian tumors on ultrasound images are relatively similar in appearance and structure, especially two classes (Mucinous cystadenoma, High grade serous).

Therefore, for faster and more accurate prediction, our approach is to employ a CNN to detect and classify ovarian tumors on ultrasound images. From there, it is possible to build an application to detect and classify ovarian tumors that can be processed on a regular computer and has low computational space (CPU-only desktops). Among many existing CNN-based detectors, YOLO (You Only Look Once) is a reasonable solution. YOLO is a deep learning network that allows detecting accurately objects. Especially YOLO is capable of operating on low-resource computers and running on CPUs [12]. However, when applying deep learning for computer vision tasks, the availability and data

distribution have a huge impact on the performance. In the medical domain, we may face a few training data and imbalance issues. Therefore, before implementation of the fine-tuning and evaluation process, detect, segment, and classify ovarian tumors in ultrasound images. We perform data augmentation with some simple methods (e.g. scaling, rotation, and flipping of images) on the training set of the original OTU_2D set (*OTU_2D-OS*) of the MMOTU dataset. The augmented dataset is called "*OTU_2D-DAS*".

In this paper, we conduct a comprehensive study to (1) evaluate the effectiveness of detecting with or without ovarian tumors in images; (2) detect and classify 8 classes of ovarian tumors by fine-tuning the ovarian tumor detection model and classifying ovarian tumors by YOLOv5 [13], and YOLOv7 [14]; Two evaluation studies (1) and (2) were performed on the *OTU_2D-OS* and *OTU_2D-DAS*. Below are the findings regarding the quantitative assessment, analysis, and hurdles encountered in the classification of ovarian tumors.

II. Related works

In recent years, there have been several works on automatic diagnosis based on different kinds of medical images. In medicine, several reports have shown AI to have high accuracy in diagnostics, such as in head CT scans, skin cancer, and retinopathy in diabetic patients. In gynecology, the application of AI has been previously studied, such as the automatic diagnosis of Pap-smear and digital colposcopy. In the last decade, the study of CAD has progressed given the remarkable development of computer science. Therefore, the study aimed to use AI to predict pathological diagnosis, using blood biomarkers, in patients with ovarian tumors.

Srivastava et al. [15] have used a fine-tuned VGG-16 network to detect whether an ovarian cyst is present or not in the image. In fine-tuning, an already existing pre-trained model is modified with a similar type of task that it earlier performed. This helps in taking benefit of the feature extraction that happens in the front layers of the network. The authors have used the VGG-16 model and modified its last four layers by training it on the ovarian cyst dataset using a CNN. This is done to achieve a high accuracy of 92.11%.

In the research of Akazawa et al. [16], the real-time dataset generated at the Institute of Tokyo Women's Medical University Medical Center East was used under the approval of the Institutional Review Board (IRB). Sample Size: 53 ovarian cancers, 23 borderline malignant tumors 126 benign ovarian tumors. Accuracy of five machine learning algorithms. The highest accuracy was 0.80 in the XGBoost algorithm, followed by 0.78 in a random forest, 0.67 in logistic regression, and 0.62 in support vector machine.

Qi et al. [17] proposed an MMOTU dataset with a set of *OTU_2D-OS* and *OTU_CEUS* consist 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.

Data augmentation aims to generate additional data which is used to train the model and has been shown to improve performance when validated on a separate unseen dataset [18]. This approach has become commonplace so to help understand the types of data augmentation techniques used in state-of-the-art deep learning models, so conducted a systematic review of the literature where data augmentation was utilized on medical images to train a deep learning model. In this article, we employ data augmentation methods and various versions of YOLO to detect ovarian tumors. Experimental results have shown that the combination of data augmentation and YOLO is a promising approach for accurately detecting and delineating ovarian tumors in medical images. This technology has the potential to greatly aid doctors in accurately diagnosing and pinpointing the location of tumors, providing valuable advantages for the treatment and care of ovarian diseases.

III. Detecting and classifying ovarian tumors in ultrasound images

Detecting and classifying ovarian tumors is a very important task in the process of building a system to help doctors diagnose ovarian cancer. In this paper, we propose to conduct a comparative study on detecting and classifying ovarian tumors on ultrasound images, as illustrated in Fig. 1. In Fig. 1, the input is *OTU_2D-OS*, and through data augmentation, we get the *OTU_2D-DAS*. Then use YOLO to fine-tune the ovarian tumors detection and classification model on the *OTU_2D-OS* and the *OTU_2D-DAS*. The first output is the result of detecting the with or without ovarian tumors on ultrasound image, the second output is the result of classifying 8 classes of ovarian tumors. The results of the quantitative evaluation, analysis, and challenges of ovarian tumor classification are also presented below.

1. Background

Currently, many deep learning networks have been studied and applied to the problem of object detection and classification. In which YOLO version (v) 1 [19] is a typical network with average accuracy, has very fast computation speed, and can perform computations on CPU. From 2020

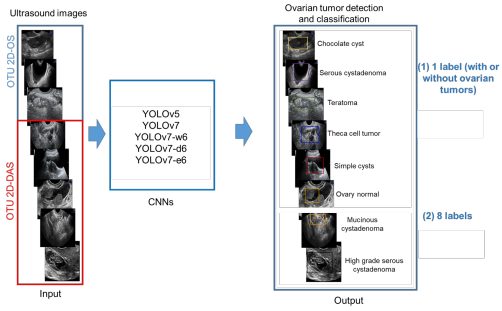


Figure 1. Our proposed framework for comparative study on ovarian tumors detection with and without data augmentation.

and early 2023, YOLOv5 [13] and YOLOv7 [14] networks were developed and announced.

YOLOv5 [20]: YOLOv5 focuses on speed and ease of use. YOLOv5 model architecture [20] has three important parts: Backbone, Neck, and Head, as illustrated in Fig. 2.

- The backbone aims to extract rich features from the input image. A Cross Stage Partial Network (CSPNet) is used as a model backbone. CSPNets are faster than deeper networks. YOLOv5 improves YOLOv4's CSPResBlock [21] into a new module, little more than a Convolution layer called a C3 module. In YOLOv5 still use CSPDarkNet53 for predicting the original object candidates according to the boxes. For the activation function, YOLOv4 [21] employs Mish or LeakyReLU in its lightweight version, whereas YOLOv5 utilizes SiLU.

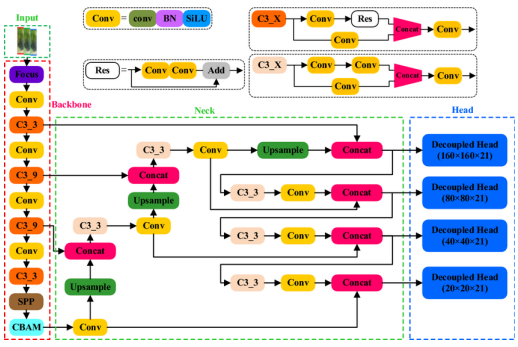


Figure 2. YOLOv5 architecture [20].

- The Neck constructs feature pyramids. This component helps the model to detect and identify the same objects at different sizes and scales. This leads to better performance with unlearned data. Different types of feature pyramids are used by many models, like FPN, BiFPN, and PANet which is used in YOLOv5. YOLOv5 adopts a module similar to SPP, but twice as fast and calls it SPP-Fast (SPPF). Instead of using parallel MaxPooling as in SPP, YOLOv5 SPPF uses sequential MaxPooling. Furthermore, the kernel size in the MaxPooling of SPPF is all 5 instead of (5,9,13) like the SPP of YOLOv4.
- The Head performs the final detection. It applies anchor boxes on generated features and outputs final vectors containing class probabilities, objectiveness scores, and bounding boxes. YOLOv5 has the same model head as YOLOv3 and YOLOv4.

Other changes in YOLOv5 compared to YOLOv4: about data augmentation: mosaic augmentation, copy-paste augmentation, mixup augmentation; about loss function: add scale factor for Objectness Loss; about anchor boxes: auto anchor using genetic algorithm; about removing grid sensitivity but the formula is different; about exponential moving average weight.

YOLOv7 [22]: YOLOv7 surpasses all known object detectors in both speed and accuracy in the range from 5 FPS to 160 FPS and has the highest accuracy at 0.568 AP among all known real-time object detectors with 30 FPS or higher on GPU V100.

The architecture of YOLOv7 is illustrated in Fig 3.

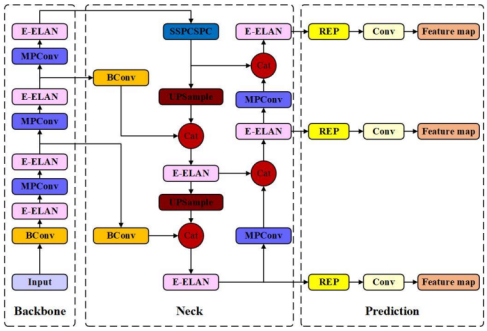


Figure 3. YOLOv7 architecture [14].

This architecture also includes three main parts.

- The backbone used the ELAN (YOLOv7-p5, YOLOv7-p6) and E-ELAN (YOLOv7-E6E).
- The neck used the SPPCSPC + (CSP-OSA)PANet (YOLOv7-p5, YOLOv7-p6) + RepConv.
- The head used the YOLOR + Auxiliary Head (YOLOv7-p6).

YOLOv7 has some improvements over than previous versions. One of the main improvements of YOLOv7 is the use of anchor boxes. Anchor boxes, which consist of predefined boxes with various aspect ratios, are utilized to detect objects of diverse shapes. YOLOv7 uses nine anchor boxes, allowing YOLO to detect a wider range of object shapes and sizes than previous versions, thus helping to decrease the number of wrong identifications. An important improvement in YOLO v7 is the use of a

new loss function called "focal loss". Previous versions of YOLO used the standard cross-entropy loss function, which is known to be less effective at detecting small objects. Focal loss solves this problem by reducing loss weights for well-classified examples and focusing on hard examples such as hard-to-detect objects. YOLOv7 also has different versions like YOLOv7-p5, YOLOv7-p6. Which YOLOv7-p5 group including YOLOv7-tiny, YOLOv7, and YOLOv7x is to use the images with size (640×640) for training. Which YOLOv7-p6 group including YOLOv7-w6, YOLOv7-e6, and YOLOv7-d6 is to use the images with size (1280×1280) for training.

2. Data augmentation

Augmenting medical image data is increasingly prevalent in deep learning applications [18]. Basic techniques like scaling, rotation, and flipping of images are commonly employed in model design. As more advanced methods such as deformable augmentation with randomized displacement fields and deep learning-based approaches like GANs are developed, they are increasingly integrated into model design practices. Although realism can be a desired outcome for augmented data, it's not always essential. Unrealistic augmentations might still yield a more generalizable model. Data augmentation proves particularly advantageous in scenarios where there's inadequate training data, helping to mitigate model over-fitting.

Data augmentation is a crucial technique in the realm of deep learning, especially when faced with a limited amount of training data. The primary concept behind this technique is to produce new variations of the existing data by implementing minor alterations. The objective is to generate a diverse dataset that enables the model to learn significant features without requiring additional data. By leveraging the diversity of the original dataset, data augmentation can enhance the model's ability to learn generalized features. Common transformations, such as horizontal and vertical flipping, rotation, and mirroring, can be particularly effective for tasks involving images.

When implementing data augmentation in situations where there is a dependency between samples in the data, it is crucial to carefully consider the potential loss of intrinsic characteristics or the creation of misleading samples that could negatively affect the performance of the model. Preprocessing techniques can be utilized to mitigate this issue. Prior to applying data augmentation, the data should be divided into independent sets to ensure the integrity of the data.

In this paper, we use five methods (x5) of data augmentation: (1) flip the vertical image, (2) flip the image horizontally, (3) rotate the image 90 degrees, (4) HUE, and (5) blur the image. They are illustrated in Fig. 4. Raising the hue of an image refers to the process of shifting the colors in the image along the hue spectrum. Hue is one of the components of the HSL (Hue, Saturation, Lightness) color model or the HSV (Hue, Saturation, Value) color model, which are commonly used in image processing. Hue represents the dominant wavelength of light that is perceived as color. By adjusting the hue value of an image, you can change the overall color tone while preserving the brightness and saturation of the image.

Average blurring, also known as box blurring or mean blurring, is a type of blurring operation commonly used in image augmentation. It involves replacing each pixel in an image with the average value of its neighboring pixels within a defined neighborhood. Here's how the average blurring operation works:

- Step 1: Define a kernel or filter size: The kernel size determines the size of the neighborhood used for blurring. It is typically represented as a square or rectangular matrix. The size of the kernel determines the extent of blurring. A larger kernel size will result in a more pronounced blurring effect.
- Step 2: Traverse through the image: For each pixel in the image, calculate the average value of the pixel intensities within the defined neighborhood. This is done by summing up the pixel intensities in the neighborhood and dividing by the total number of pixels in the neighborhood.
- Step 3: Replace the pixel value: Replace the original pixel value with the calculated average value. This step is repeated for every pixel in the image. The average blurring operation effectively smooths out the image by reducing high-frequency details and noise. It is a simple blurring technique that produces a uniform blurring effect across the image. Average blurring can be beneficial during image augmentation for various reasons. It helps to reduce noise and unwanted artifacts, which can improve the robustness and generalization of machine learning models. Additionally, it can be used to create a more visually pleasing appearance or simulate certain real-world effects.

This system classifies the test image into one of eight labels by leveraging features extracted from the test image. These features are processed through classifiers previously trained using the training parameters evaluated by the offline learning system. The offline classification system assesses the training parameters of the classifiers by employing a combination of features extracted from the training dataset and the corresponding ground truth training class labels.

Vikas et al [23], explained that the number of parameters

required corresponds to the data complexity of the operation in the machine learning model that must be carried out by using this kind of augmentation process. Sometimes it might still go wrong even if we have the correct size of the training set. To regularize convolution neural network models, it can be upgraded dataset by including more images dataset, using the data augmentation methods as shown in Fig 4.

By using data augmentation methods in the dataset, the image ratio has been improved at a 1: 9 ratio.

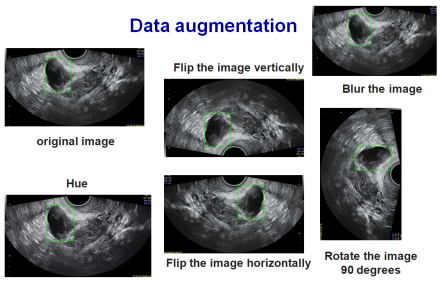


Figure 4. Five augmentation techniques are employed to enrich data in the training set.

IV. Experimental Results

1. Dataset

In this work, we show a follow-up study based on the *OTU_2D-OS*) of the MMOTU dataset [17]. The *OTU_2D-OS* includes 1469 2D ultrasound images and eight labels, as illustrated in Fig. 5.

For example, for an ovarian tumor, in OTU_2D set of MMOTU [17] dataset, some classes have a small amount of data as shown in Fig. 6. (0: Chocolate cyst, 1: Serous cystadenoma, 2: Teratoma, 3: Theca cell tumor, 4: Simple cysts, 5: Ovary normal, 6: Mucinous cystadenoma, 7: High grade serous cystadenoma). The number of samples in the classes is imbalanced, with 226 images in class 0, while class 7 has only 38 images. Imbalanced data often leads to inaccurate predictions for the minority group. This is because the majority class dominates the prediction results, while the minority class is poorly represented. However, accurately predicting a sample belonging to the minority group is much more important than predicting samples from the majority group. To improve prediction results, appropriate adjustments are needed to ensure that the model achieves high accuracy on the minority group.

Before performing data enhancement, we perform data normalization of the *OTU_2D-OS*, with the image size being 320 × 240 pixels. This process includes ground truth standardization. At the same time, we perform data preparation according to two datasets, which are single-label data with or without ovarian tumors (1-label) and 8-label data on ovarian tumor classes as shown in of MMOTU dataset [17]. In particular, the bounding box of the tumors was normalized against the bounding box of the YOLO's bounding box format with the COCO 2017 dataset [24], as presented in Eq. 1.

$$a = \frac{\frac{x\_max+x\_min}{2}}{w\_b}; \quad b = \frac{\frac{y\_max+y\_min}{2}}{h\_b}$$
$$c = \frac{w\_b}{w\_im}; \quad d = \frac{h\_b}{h\_im} \tag{1}$$

where (x_max, y_max) is the coordinates of the top right corner of the ovarian tumor bounding box on the image, (x_max, y_max) is the coordinates of the bottom left corner of the ovarian tumor bounding box on the image, w_b is the width size of the bounding box of the ovarian tumor on the image, h_b is the height size of the ovarian tumor 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 COCO 2017 used for training and evaluation has the form (l, a, b, c, d), where l is the label of the object.

After data augmentation, the *OTU_2D-DAS* obtained the following number of images in Fig. 7. The statistical data before and after augmentation are presented in Tab. I.

We performed data augmentation exclusively on some minor classes (i.e. theca cell tumor, simple cysts, mucinous cystadenoma, and high-grade serous cystadenoma) in the training set of the entire dataset, increasing the sample number of each of such classes by a factor of five. After the augmentation phase, the number of data samples in the training set increased from 1000 to 1852 images. We have precisely summarized the number of original data samples along with the augmented samples in Tab. I.

Data augmentation is a common technique used to enrich training data when the training set is limited. Numerous studies have demonstrated that augmentation contributes to enhancing the performance of image-based tasks. Conventional techniques for augmenting data involve applying geometric and photo-metric transformations to the original images. In our work, we employ five transformations across four classes, as mentioned previously, including: (1) vertical flipping, (2) horizontal flipping, (3) rotating by 90 degrees, (4) adjusting hue, and (5) blurring. To ensure the variation and diversity of the augmented data, we did not apply any additional pre-processing steps to remove dependent data. In our work, we apply five data augmentation techniques to a subset of our training set. The augmented data is utilized to train the network. However, we do not currently evaluate

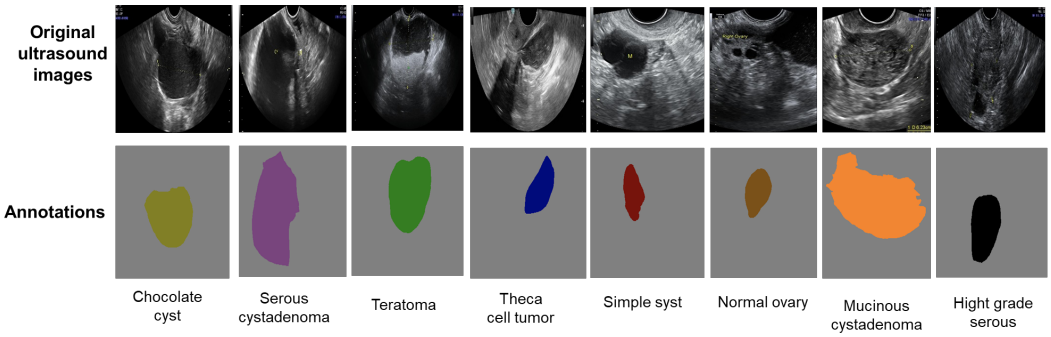


Figure 5. Illustration of eight ovarian tumor categories in the *OTU_2D-OS* dataset.

TABLE I
THE STATISTICS OF ORIGINAL DATA (OTU_2D-OS) AND DATA AFTER ENHANCEMENT (OTU_2D-DAS).

Classes	Index classes	Number of original training samples	Augmentation factor	Number of augmented training samples	Number of testing samples
Chocolate cyst	0	226		226	110
Serous cystadenoma	1	153		153	66
Teratoma	2	228		228	108
Theca cell tumor	3	57	x5	285	31
Simple cysts	4	47	x5	235	19
Ovary normal	5	180		180	87
Mucinous cystadenoma	6	71	x5	355	33
High grade serous cystadenoma	7	38	x5	190	15
Total		1000		1852	469

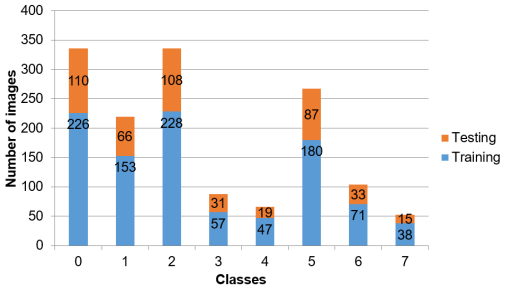


Figure 6. The number of training and testing images in the dataset OTU_2D.

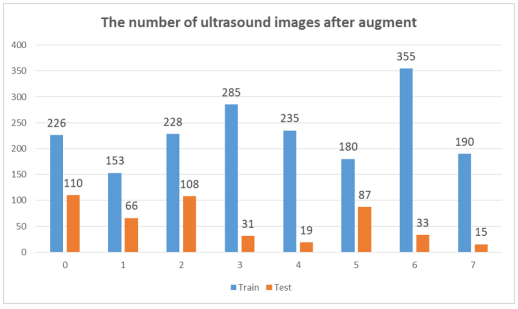


Figure 7. The number of ovarian tumors ultrasound images after augmentation.

the specific role of each transformation. Investigating the role of each augmentation technique will be a focus of our future work. This aspect is also highlighted in the conclusion section of the revised manuscript.

2. Evaluation metrics and measurement

We use *IOU* (Intersection Over Union) - metric of evaluation to evaluate the performance of detection of ovarian tumors as Fig. 8, Precision (P) as Eq. 2, Recall (R) as Eq. 2, and mAP (mean Average Precision) as Eq. 3.

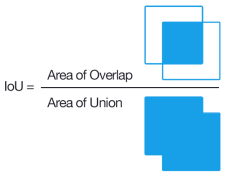


Figure 8. The formula to calculate the Intersection Over Union (IOU).

$$P = \frac{TP}{TP + FP}; \qquad R = \frac{TP}{TP + FN}; \tag{2}$$

$$mAP = \frac{1}{|\ \text{Classes}\ |} \sum_{c \in \text{Classes}} \frac{TP(c)}{TP(c) + FP(c)} \tag{3}$$

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 no predicted

ground truth data area. The threshold of IOU we use is 0.5.

In this work, we implemented the examined models on a server featuring an NVIDIA GeForce RTX 2080 Ti, 12GB GPU, for the purposes of fine-tuning, training, and testing. The programs were written in the Python language (≥3.7 version) with the support of the CUDA 11.2/cuDNN 8.1.0 libraries. In addition, there are some libraries such as Matplotlib, OpenCV, Mmcls≥0.20.1, Packaging, Numpy, Prettytable, PyTorch 1.5+, etc. Especially the source code is inherited and developed on the YOLOv5 ([1]) and YOLOv7 ([2]). The models of YOLOv5 and YOLOv7 are fine-tuned with 100 epochs. The source code to fine-tune the ovarian tumor detection and classification model of YOLOv5 and YOLOv7 is shown in links ([3]),([4]), respectively.

3. Results and Discussions

Ovarian cancer stands as the foremost cause of death among women, underscoring the critical significance of early diagnosis in the medical realm. In this paper, we conducted a comparative study utilizing YOLO for the segmentation of ovarian tumors. In this study, two main problems were solved: the detection and classification of 1-label (1 class) and 8-label (8 classes) of ovarian tumors in the OTU_2D subset of the MMOTU dataset.

The results of ovarian tumor detection and one class classification (1-label) with YOLOv5, YOLOv7, YOLOv7-w6, YOLOv7-d6, and YOLOv7-e6 on the *OTU_2D-OS* and *OTU_2D-DAS* are shown in Tab. II. The best detection and classification results are from YOLOv5 with ($P = 0.91, R = 0.89$) of the *OTU_2D-OS* and ($P = 0.90, R = 0.90$) of the *OTU_2D-DAS*. It can be seen that the results in Tab. II, are not improved on the *OTU_2D-OS* and *OTU_2D-DAS* sets. This result is due to the augmented data process being just some simple processing methods of image processing on the original data image, so the change of the distinguishing features with or without ovarian tumor is not great. And this evaluation is only concerned with or without ovarian tumors on the image so the detection results are not improved. Specifically, in Tab. II with the P metric, the average results of the models with the *OTU-OS* dataset are 0.87, while with the *OTU-DAS* dataset, it's 0.86. With the R metric, the average results of the models with the *OTU-OS* dataset are also 0.87, and with the *OTU-DAS* dataset, it's 0.86. As for the mAP50 metric, the average results of the models with the *OTU-OS* dataset are 0.90, while with the *OTU-DAS* dataset, it's 0.91, which is 1% higher compared to the non-augmented data.

Figure 9(a) presents the distribution of ovarian tumor classification results on the training data of the *OTU_2D-OS* and *OTU_2D-DAS* when using YOLOv7-e6 to classify with or without ovarian tumors in ultrasound images. We show the results of YOLOv7-e6 because based on the results of Tab. II, YOLOv7-e6 has the best results among the variants of YOLOv7. In Fig. 9(b), the training results of YOLOv7-e6 are not much improved on the original data ($R = 0.97, P = 0.901$) and the augmented data ($R = 0.98, P = 0.942$).

Figure 10(a) and Figure 10(b) presents the distribution of ovarian tumor classification results on the training data of the *OTU_2D-OS* and *OTU_2D-DAS* when using YOLOv5 to classify with or without ovarian tumors in ultrasound images.

The results in Tab. II also show that R is much lower than P, which shows that the with or without ovarian tumors classifier of YOLO is highly wrong classified. Because R is calculated as the number of correct classifications divided by the number of correct classifications plus the number of false negatives. If R is higher then the lower the number of missed ovarian tumors.

The results of ovarian tumor detection and eight classes classification (8-labels) with YOLOv5, YOLOv7, YOLOv7-w6, YOLOv7-d6, and YOLOv7-e6 on the *OTU_2D-OS* and *OTU_2D-DAS* are shown in Tab. III.

In our work, we applied five data augmentation techniques to a subset of our training set. The augmented data is utilized to train the network. In the original dataset *OTU_2D-OS*, there are 1239 gray-scales images and 216 color images (doppler images). Among five data augmentation techniques, only "adjusting hue" affects the color of 216 color images. To evaluate the impact of "adjusting hue" technique on the final result, we conducted an additional experiment where only 4 augmentation techniques without "hue adjusting" (vertial flipping, horizontal flipping, rotating by 90 degrees, and blurring) to create a dataset *OTU_2D-DAS** (without hue adjusting). We select Yolov5 for testing and experimental result is shown in Table IV. In average, without "hue adjusting" augmentation, the recall is reduced significantly from 0.72 to 0.675. However, mAP50 remains similar. In conclusion, hue adjusting helps to improve the recall of detection. We have included this experiment in the revised manuscript.

Table V shows the results of ovarian tumor detection and eight classes classification (8-labels) on the *OTU_2D-OS* by YOLOv5 (epoch = 100) with threshold $IOU = 0.25$ and $IOU = 0.75$. The average results of detecting and

[1] https://github.com/ultralytics/yolov5

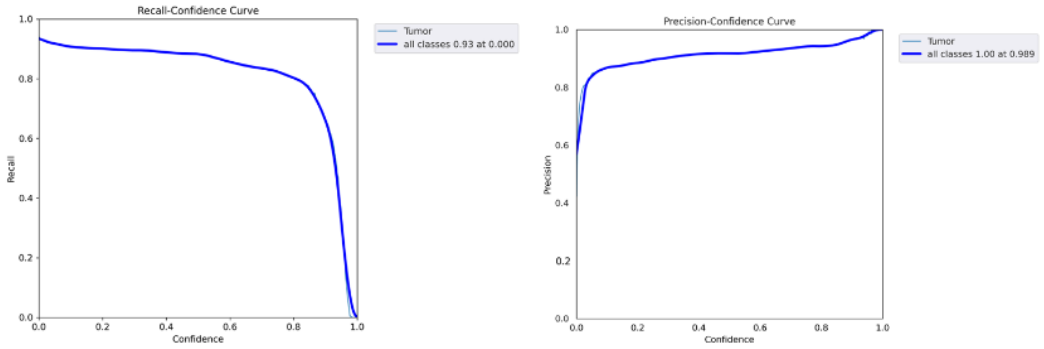
[2] https://github.com/WongKinYiu/yolov7

[3] https://drive.google.com/file/d/1aJS8hEisvQu-ugBDYxq5e9_Zw5EceisB/view?usp=sharing

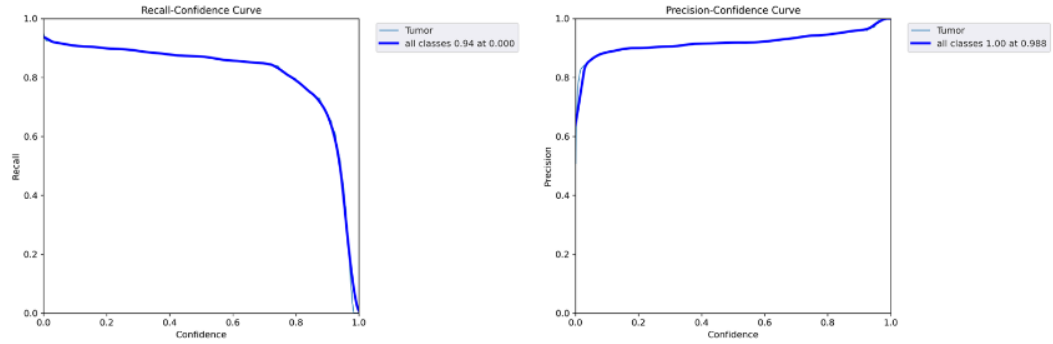
[4] https://drive.google.com/file/d/1Qg1umKPCToi2WyOsKJ_Kq4Gc1Mz-lE7O/view?usp=sharing

TABLE II
THE RESULTS OF OVARIAN TUMOR DETECTION AND BINARY CLASSIFICATION (1-LABEL) ON THE *OTU_2D-OS* AND *OTU_2D-DAS* BY YOLOV5, YOLOV7, AND OTHER YOLOV7 VARIATIONS (ALL ARE TRAINED WITH THE NUMBER OF EPOCHS = 100).

Class/ Measurements	Dataset	Precision (*P*)					Recall (*R*)					*mAP50*				
		YOLO v5	YOLO v7	YOLO v7-w6	YOLO v7-d6	YOLO v7-e6	YOLO v5	YOLO v7	YOLO v7-w6	YOLO v7-d6	YOLO v7-e6	YOLO v5	YOLO v7	YOLO v7-w6	YOLO v7-d6	YOLO v7-e6
Tumor	OTU 2D-OS	**0.91**	0.86	0.83	0.84	0.90	**0.89**	0.87	0.82	0.88	0.87	**0.92**	0.91	0.86	0.92	0.92
	OTU 2D-DAS	**0.90**	0.84	0.84	0.90	0.84	**0.90**	0.83	0.87	0.86	0.89	**0.93**	0.89	0.91	0.92	0.91



(a) Distribution of classification results on the training data of the original data when using YOLOv7-e6 to classify with or without ovarian tumors in ultrasound images



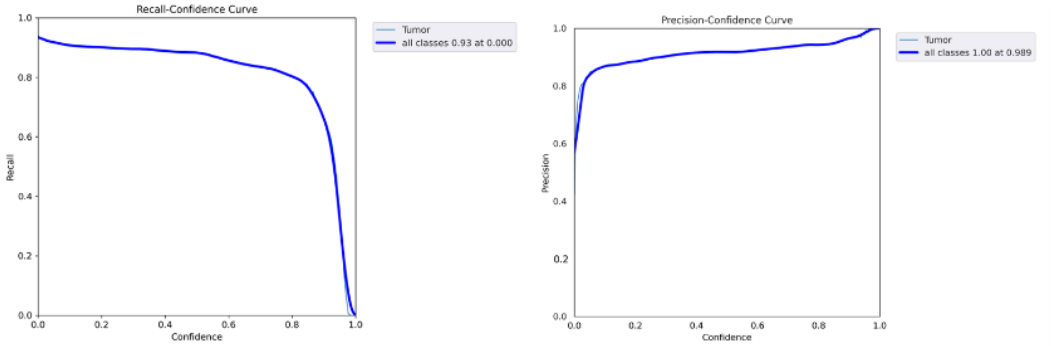
(b) Distribution of classification results on the training data of the data augmentation when using YOLOv7-e6 to classify with or without ovarian tumors in ultrasound images

Figure 9. Illustration of the distribution of ovarian tumor classification results on the training data of the *OTU_2D-OS* and *OTU_2D-DAS* when using YOLOv7-e6 to classify with or without ovarian tumors in ultrasound images.

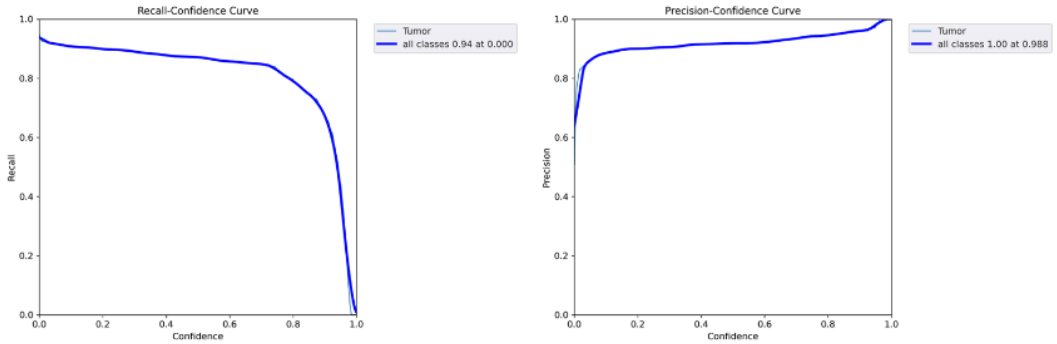
classifying 8 ovarian tumor classes of YOLOv5 are ($P = 0.853, P = 0.83, P = 0.748$), ($R = 0.683, R = 0.66, R = 0.6$), ($mAP25 = 0.781, mAP50 = 0.75, mAP75 = 0.628$) with threshold $IOU = 0.25, IOU = 0.5, IOU = 0.75$, respectively. The detection and classification results are gradually decreasing with increasing IOU.

Table VI shows the results of ovarian tumor detection and eight classes classification (8-labels) on the *OTU_2D-DAS* by YOLOv5 (epoch = 100) with threshold $IOU = 0.25$ and $IOU = 0.75$, respectively. The results of detecting and classifying ovarian tumors also decrease as IOU increases as shown in Tab. V.

Figure 11 presents the distribution of ovarian tumor classification results on the training data of the *OTU_2D-OS* and *OTU_2D-DAS* when using YOLOv7-e6 to classify 8 classes of ovarian tumors in ultrasound images. In Fig. 11, the training results of YOLOv7-e6 are not much improved on the original data ($R = 0.95, P = 0.96$) and the augmented data ($R = 0.914, P = 0.972$). Figure 11 also shows the distribution of prediction results R and P of labels (3: Theca cell tumor, 4: Simple cysts, 6: Mucinous cystadenoma, 7: High grade serous cystadenoma) which have ten times more confidence than labels (0: Chocolate cyst, 1: Serous cystadenoma, 2: Teratoma, 5: Ovary normal). Therefore, the results of classifying 8 ovarian tumor classes of labels (0: Chocolate cyst, 1: Serous cystadenoma, 2: Teratoma, 5: Ovary normal) are better than labels (3: Theca cell tumor, 4: Simple cysts, 6: Mucinous cystadenoma, 7: High grade serous cystadenoma). The results in Tab. III are also presented on YOLOv5, and YOLOv7 and the results show



(a) Distribution of ovarian tumor classification results on the training data of the original data when using YOLOv5 to classify with or without ovarian tumors in ultrasound images



(b) Distribution of ovarian tumor classification results on the training data of the data augment when using YOLOv5 to classify with or without ovarian tumors in ultrasound images

Figure 10. Illustration of the distribution of ovarian tumor classification results on the training data of the *OTU_2D-OS* and *OTU_2D-DAS* when using YOLOv5 to classify with or without ovarian tumors in ultrasound images.

that the detection and classification results of YOLOv5 have the highest results at 83%.

Figure 12 shows the confusion matrix in classifying 8 ovarian tumor classes on ultrasound images of the *OTU_2D-OS* and *OTU_2D-DAS* using YOLOv5. On the *OTU_2D-OS* (left), the prediction result of label 7 (High grade serous cystadenoma) is mistaken as background with a rate of 0.2, and label 0 (Chocolate cyst) is also mistakenly classified as background with a rate of 0.33. On the *OTU_2D-DAS* (right), the prediction result of label 7 (High grade serous cystadenoma) is mistaken as background with a rate of 0.33, and label 1 (Serous cystadenoma) is also mistakenly classified as background with a rate of 0.24.

Figure 13 shows the results of R in comparison to the distribution of YOLOv5 to classifier 8 ovarian tumor classes when training on the *OTU_2D-OS* and *OTU_2D-DAS* with 100 epochs and 469 epochs. In this paper, we performed YOLOv5 training to build a model to detect and classify 8 classes on the *OTU_2D-OS* and *OTU_2D-DAS* with a number of epochs of 500. However, when training to epoch 469^{th}, the training process stopped due to the model can not be improved further, because, after each parameter adjustment at each epoch, the model will be evaluated on the validation set. This shows that the ovarian tumor detection and classification model that we trained at 100 epochs has reached the highest possible performance, it can not get better with more epochs of training. Therefore, it is reasonable to stop training the model at 100 epochs to compare the results.

From the performance results, overfitting gradually decreased, leading to higher accuracy, which is better than CNN models using *OTU_2D-OS*. For instance, when using the YOLOv7 model with *OTU_2D-DAS*, the Precision metric increased by 0.15. Similarly, when using the YOLOv5 model, the average R metric increased by 0.06 compared to using *OTU_2D-OS*. It can be said that the issue of detecting and classifying ovarian tumors is very important in the process of examining and diagnosing ovarian tumors for patients. The results in Tab. II, Tab. III show that R is low. This proves that many ovarian tumors have not been detected and classified properly. This is very dangerous in detecting and diagnosing ovarian cancer. For example, a woman has an ovarian tumor but it is not detected through medical examination, and this person will not receive early treatment for the tumor, which will lead to the tumor becoming larger and metastasizing to organs, and others in

TABLE III
THE RESULTS OF OVARIAN TUMOR DETECTION AND EIGHT CLASSES CLASSIFICATION (8-LABELS) ON THE ***OTU_2D-OS*** AND ***OTU_2D-DAS*** BY YOLOV5, YOLOV7, AND OTHER YOLOV7 VARIATIONS (NUMBER OF EPOCHS FOR TRAINING THE MODELS = 100).

Classes/ Measurements	**Dataset**	**Precision (*P*)**					**Recall (*R*)**					***mAP*50**				
		YOLO v5	**YOLO v7**	**YOLO v7-w6**	**YOLO v7-d6**	**YOLO v7-e6**	**YOLO v5**	**YOLO v7**	**YOLO v7-w6**	**YOLO v7-d6**	**YOLO v7-e6**	**YOLO v5**	**YOLO v7**	**YOLO v7-w6**	**YOLO v7-d6**	**YOLO v7-e6**
Chocolate cyst	OTU 2D-OS	0.82	0.57	0.58	0.55	0.53	0.65	0.76	0.75	0.75	0.80	0.74	0.69	0.70	0.70	0.71
	OTU 2D-DAS	0.71	0.72	0.66	0.71	0.73	0.77	0.72	0.74	0.70	0.77	0.76	0.74	0.75	0.74	0.76
Serous cystadenoma	OTU 2D-OS	0.91	0.44	0.42	0.37	0.39	0.64	0.86	0.85	0.94	0.86	0.81	0.62	0.56	0.57	0.62
	OTU 2D-DAS	0.62	0.61	0.53	0.61	0.54	0.82	0.70	0.70	0.61	0.74	0.78	0.71	0.61	0.69	0.62
Teratoma	OTU 2D-OS	0.93	0.61	0.62	0.54	0.41	0.76	0.82	0.77	0.81	0.90	0.88	0.78	0.82	0.81	0.81
	OTU 2D-DAS	0.80	0.80	0.82	0.81	0.65	0.86	0.69	0.76	0.74	0.82	0.89	0.79	0.83	0.81	0.80
Theca cell tumor	OTU 2D-OS	0.82	0.49	0.38	0.38	0.27	0.65	0.34	0.19	0.52	0.36	0.68	0.41	0.28	0.31	0.22
	OTU 2D-DAS	0.79	0.80	0.76	0.64	0.54	0.60	0.68	0.61	0.57	0.58	0.68	0.74	0.61	0.64	0.65
Simple cysts	OTU 2D-OS	0.83	0.99	0.00	0.15	0.12	0.74	0.11	0.00	0.37	0.16	0.72	0.27	0.13	0.15	0.12
	OTU 2D-DAS	0.87	0.80	0.63	0.27	0.26	0.74	0.63	0.58	0.42	0.58	0.72	0.69	0.60	0.45	0.38
Ovary normal	OTU 2D-OS	0.85	0.39	0.52	0.54	0.47	0.61	0.63	0.75	0.81	0.82	0.76	0.52	0.69	0.68	0.65
	OTU 2D-DAS	0.74	0.70	0.56	0.59	0.66	0.70	0.48	0.81	0.75	0.71	0.76	0.61	0.72	0.66	0.68
Mucinous cystadenoma	OTU 2D-OS	0.81	0.28	0.24	0.21	0.18	0.67	0.49	0.55	0.82	0.67	0.74	0.40	0.30	0.29	0.27
	OTU 2D-DAS	0.84	0.67	0.53	0.61	0.53	0.65	0.48	0.39	0.57	0.48	0.77	0.60	0.53	0.50	0.55
High grade serous cystadenoma	OTU 2D-OS	0.65	0.39	0.33	0.27	0.23	0.61	0.20	0.27	0.13	0.30	0.64	0.22	0.17	0.16	0.20
	OTU 2D-DAS	0.77	0.65	0.69	0.65	0.38	0.66	0.40	0.44	0.38	0.33	0.67	0.53	0.44	0.47	0.34
AVERAGE	OTU 2D-OS	**0.83**	0.52	0.39	0.38	0.32	**0.66**	0.53	0.51	0.64	0.61	**0.75**	0.49	0.46	0.46	0.45
	OTU 2D-DAS	**0.77**	0.72	0.65	0.61	0.54	**0.72**	0.60	0.63	0.59	0.63	**0.75**	0.67	0.63	0.62	0.60

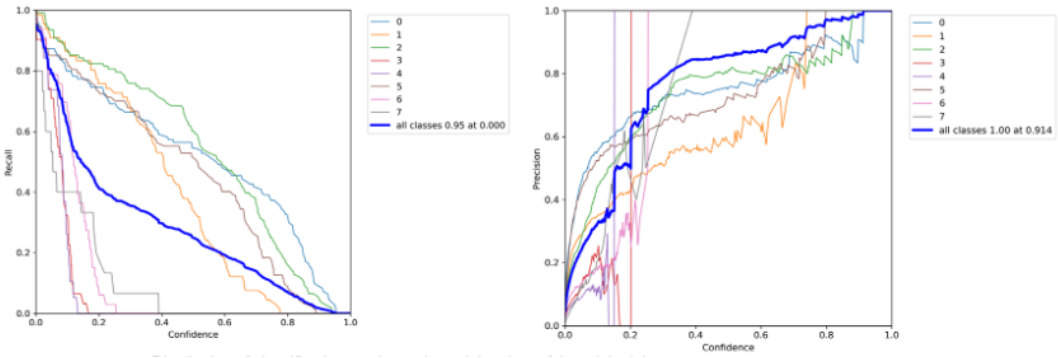
TABLE IV
COMPARISON OF PERFORMANCE OF YOLO-V5 WITH AND WITHOUT "HUE ADJUSTING" AUGMENTATION TECHNIQUE

Class ID	**Precision**		**Recall**		**mAP50**	
	OTU_2D -DAS	OTU_2D -DAS* (without hue adjusting)	OTU_2D -DAS	OTU_2D -DAS* (without hue adjusting)	OTU_2D -DAS	OTU_2D -DAS* (without hue adjusting)
0	0.71	0.811	0.77	0.703	0.76	0.789
1	0.62	0.727	0.82	0.769	0.78	0.806
2	0.8	0.885	0.86	0.785	0.89	0.887
3	0.79	0.794	0.6	0.613	0.68	0.669
4	0.87	0.769	0.74	0.684	0.72	0.741
5	0.74	0.789	0.7	0.643	0.76	0.752
6	0.84	0.815	0.65	0.666	0.77	0.719
7	0.83	0.782	0.66	0.533	0.67	0.661
Average	0.77	0.797	**0.72**	**0.675**	0.75	0.757

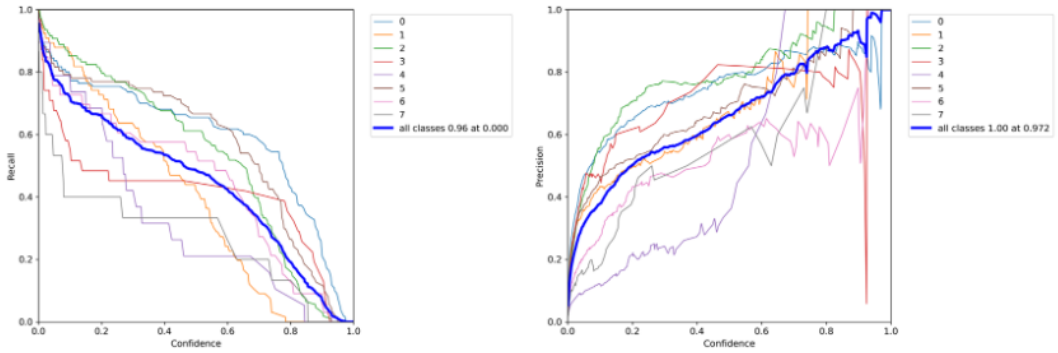
the body When detected, the disease is in a late stage and very difficult to treat. This reduces the patient's chance of survival. In medical examination and treatment, especially when detecting ovarian tumors, it is often better to accept "a mistake than to miss" it. Thereby increasing the ability to detect early and promptly treat ovarian tumors. The study also highlights significant obstacles in constructing a support system for diagnosing ovarian tumors. Moving forward, we will persist in enhancing CNNs for detecting ovarian tumors and assessing their performance on additional datasets related to ovarian tumors.

The processing time for detecting and classifying 8 ovarian tumor classes on the *OTU_2D-OS* when performing calculations on GPU, is shown in Tab. VII.

Based on the results in Tab. II, Tab. III, Fig. 11, Fig. 12, Fig. 13, it can be seen that the results of detecting ovarian tumors on ultrasound images with YOLOv5 are better than those of YOLOv7. This problem occurs because the



(a) Distribution of ovarian tumor classification results on the training data of the original data when using YOLOv7-e6 to classify to 8 class ovarian tumors in ultrasound image



(b) Distribution of ovarian tumor classification results on the training data of the original data when using YOLOv7-e6 to classify to 8 class ovarian tumors in ultrasound image

Figure 11. Illustration of the distribution of ovarian tumor classification results on the *OTU_2D-OS* and *OTU_2D-DAS* training data when using YOLOv7-e6 to classify to 8 class ovarian tumors in ultrasound images.

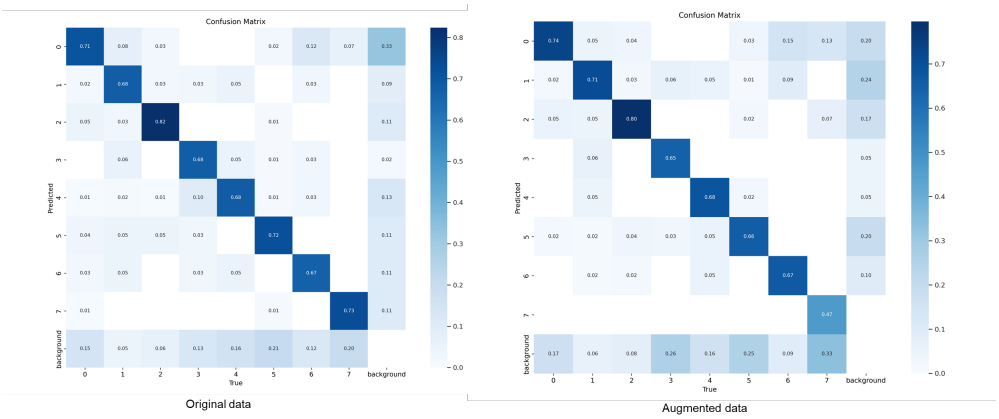
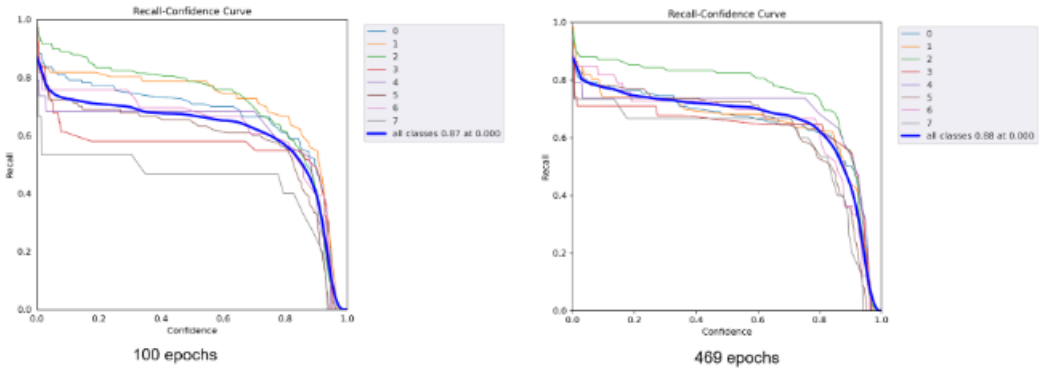


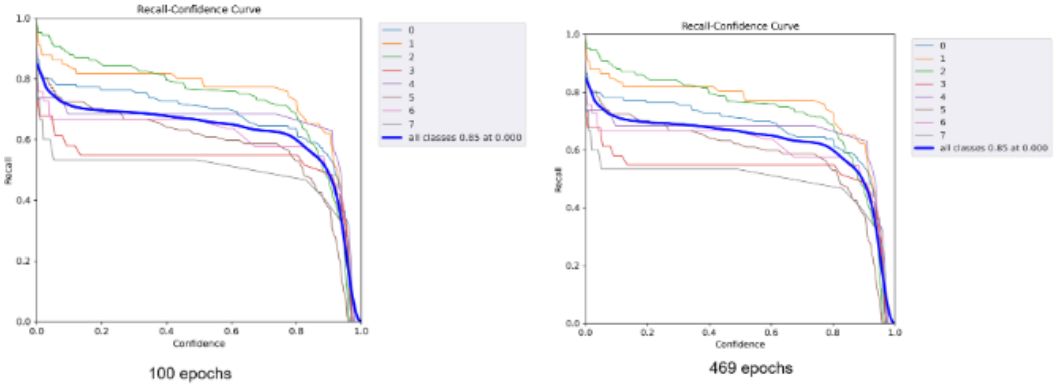
Figure 12. The confusion matrix in classifying 8 ovarian tumor classes on ultrasound images of the *OTU_2D-OS* and *OTU_2D-DAS* using YOLOv5.

difference between YOLOv5 and YOLOv7 is mainly about computational speed in object detection, higher resolution allows YOLO v7 to detect smaller objects. In this paper, the size of the normalized input image is 320×240 pixels. When used in YOLOv5, the size of the image is changed to 640×640 pixels, and when using YOLOv7, the input image size is mainly changed to 1280×1280 pixels. Therefore, the object prediction on a much larger image makes the learning set less efficient. Hence, the object detection result of YOLOv7 is lower than that of YOLOv5. The results in Tab. III have increased 0.116 because the data has been augmented for the training set, which only YOLOv7 has increased results and YOLOv5 has not improved results.

Figure 15 shows the detection and classification of the 8 classes of ovarian tumors. The label of the class and the confidence coefficient of the prediction are shown on the



(a) Distribution of ovarian tumor classification results on the training data of the original data when when training on the OTU 2D-OS and OTU 2D-DAS with 100 epochs and 469 epochs



(b) Distribution of ovarian tumor classification results on the training data of the original data when training on the OTU 2D-OS and OTU 2D-DAS with 100 epochs and 469 epochs

Figure 13. The recall distribution of YOLOv5 when classifying 8 ovarian tumor classes when training on the *OTU_2D-OS* and *OTU_2D-DAS* with 100 epochs and 469 epochs.

bounding box. Each label will be represented by a distinct color of the bounding box.

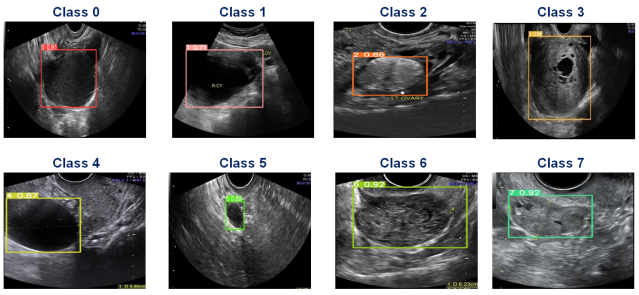


Figure 14. The result of detection and classification by YOLOv5 with 8 classes of ovarian tumors.

Figure 16 shows a wrong classification result using YOLOv5, on an area of ovarian tumor with 2 delineated bounding boxes. Where the bounding box on the outside is the wrong prediction.

During the experimentation process, there were some failure cases in which the proposed method could not detect tumors as shown in Fig. 17, or incorrect delineation of the

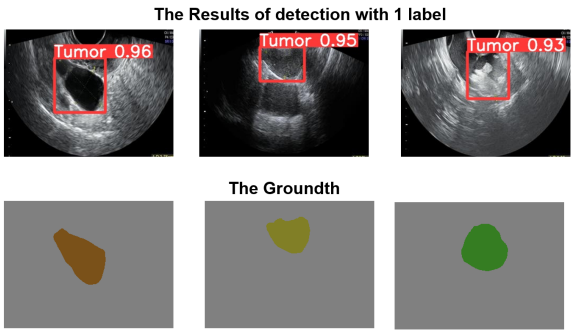


Figure 15. The result of detection and classification by YOLOv5 with 1-label.

regions. It can be seen that the case in Fig. 17 is very challenging, the tumor area data is weakly differentiated from the normal data area of the ultrasound image.

V. Conclusion

In this work, we augmented only the training dataset without making any changes to the testing set. This is a common practice when applying augmentation techniques

TABLE V
THE RESULTS OF OVARIAN TUMOR DETECTION AND EIGHT CLASSES CLASSIFICATION (8-LABELS) ON THE *OTU_2D-OS* BY YOLOV5 (EPOCH = 100) WITH THRESHOLD IOU = 0.25 AND IOU = 0.75.

IOU	Measurement/ Index classes	P	R	mAP25
IOU=0.25	0	0.849	0.665	0.81
	1	0.913	0.635	0.827
	2	0.943	0.766	0.896
	3	0.818	0.645	0.681
	4	0.838	0.737	0.721
	5	0.87	0.616	0.799
	6	0.807	0.667	0.736
	7	0.786	0.736	0.779
	Average	**0.853**	**0.683**	**0.781**
	Measurement/ Index classes	**P**	**R**	**mAP75**
IOU=0.75	0	0.778	0.609	0.644
	1	0.869	0.605	0.742
	2	0.749	0.609	0.65
	3	0.777	0.613	0.628
	4	0.84	0.737	0.715
	5	0.555	0.391	0.347
	6	0.771	0.636	0.696
	7	0.643	0.6	0.602
	Average	**0.748**	**0.6**	**0.628**

TABLE VI
THE RESULTS OF OVARIAN TUMOR DETECTION AND EIGHT CLASSES CLASSIFICATION (8-LABELS) ON THE *OTU_2D-DAS* BY YOLOV5 (EPOCH = 100) WITH THRESHOLD IOU = 0.25 AND IOU = 0.75.

IOU	Measurement/ Index classes	P	R	mAP25
IOU=0.25	0	0.765	0.809	0.839
	1	0.701	0.788	0.799
	2	0.827	0.806	0.864
	3	0.822	0.548	0.644
	4	0.913	0.632	0.755
	5	0.722	0.782	0.731
	6	0.961	0.755	0.839
	7	0.713	0.499	0.615
	Average	**0.803**	**0.702**	**0.761**
	Measurement/ Index classes	**P**	**R**	**mAP75**
IOU=0.75	0	0.599	0.625	0.603
	1	0.707	0.788	0.786
	2	0.658	0.639	0.617
	3	0.824	0.548	0.64
	4	0.922	0.626	0.713
	5	0.483	0.517	0.349
	6	0.961	0.748	0.813
	7	0.616	0.43	0.453
	Average	**0.721**	**0.615**	**0.622**

to enrich the training set for deep learning models. The testing set should be kept in its original form. In the future, we will conduct more experiments with larger datasets and compare the results with the current findings.

We perform an initial comparative study in tumor detection (classifying one class with or without a tumor) and

TABLE VII
THE PROCESSING TIME (FPS) FOR DETECTING AND CLASSIFYING 8 OVARIAN TUMOR CLASSES ON THE ***OTU_2D-OS*** WHEN PERFORMING CALCULATIONS ON GPU AND CPU.

Methods	Type	YOLO v5	YOLO v7
Processing time	GPU	166	**204**
Processing time	CPU	0.7	**2.04**

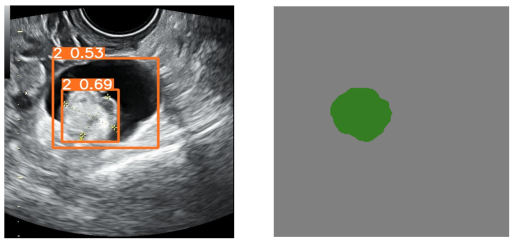


Figure 16. The result shows a misclassification result using YOLOv5, on an area of ovarian tumor with 2 delineated bounding boxes.

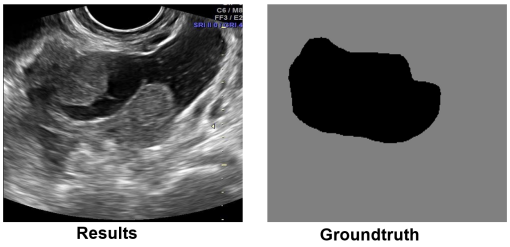


Figure 17. Illustrating the case of undetectable ovarian tumor on ultrasound image.

classifying eight classes of ovarian tumors on ultrasound using YOLOv5 and YOLOv7-p6. With mAP of ovarian tumors detection and one class classification is 0.86 to 0.92 of YOLOv5 and YOLOv7-p6. From 0.45 to 0.75 of mAP when ovarian tumors were detected and eight classes of classification.

At the same time, we also fine-tuned the tumor detection and classification model on an augmented training dataset. The results show that data augmentation increases the results by 0.116. In particular, we also show YOLO's superior computation time compared to other methods for ovarian tumor detection and classification (YOLOv7's processing speed is 204fps on GPU and 2.04fps on CPU). In the future, other deep learning networks could be investigated and improved for tumor detection and and classification.

ACKNOWLEDGMENT

This research is funded by Ministry of Science and Technology (MOST) under grant number KC-4.0-45/19-25 "Research and development of a computer-aided support system for ovarian cancer diagnosis using ultrasound images".

REFERENCES

[1] N. C. Institute, "Cancer Stat Facts: Ovarian Cancer," https://seer.cancer.gov/statfacts/html/ovary.html, 2023, [Online; accessed 14-July-2023].

[2] Vietnam, "Overcrowding strains HCM City hospitals, doctors," https://en.vietnamplus.vn/vietnam-to-publicly-rank-quality-of-hospitals/87505.vnp, 2023, [Online; accessed 14-July-2023].

[3] J. R. Astley, J. M. Wild, and B. A. Tahir, "FUNCTIONAL IMAGING OF THE LUNG SPECIAL FEATURE : Artificial intelligence in functional imaging of the lung," *Br J Radiol. 2022 Apr 1; 95(1132): 20201107*, no. April 2021, 2022.

[4] Z. Jin, T. Gan, P. Wang, Z. Fu, C. Zhang, Q. Yan, X. Zheng, X. Liang, and X. Ye, "Deep learning for gastroscopic images: computer-aided techniques for clinicians," *BioMedical Engineering Online*, vol. 21, no. 1, pp. 1–41, 2022. [Online]. Available: https://doi.org/10.1186/s12938-022-00979-8

[5] P.-T. Nguyen, M.-Q. Le, Q.-T. Dao, V. A. Tran, V.-H. Dao, and T.-H. Tran, "Automatic classification of upper gastrointestinal tract diseases from endoscopic images," in *2022 11th International Conference on Control, Automation and Information Sciences (ICCAIS)*. IEEE, 2022, pp. 442–447.

[6] T.-H. Tran, D. H. Vu, D. H. Tran, K. L. Do, P. T. Nguyen, V. T. Nguyen, L. T. Nguyen, N. K. Ho, H. Vu, and V. H. Dao, "Dcs-unet: Dual-path framework for segmentation of reflux esophagitis lesions from endoscopic images with u-net-based segmentation and color/texture analysis," *Vietnam Journal of Computer Science*, vol. 10, no. 02, pp. 217–242, 2023.

[7] P. Xue, J. Wang, D. Qin, H. Yan, Y. Qu, S. Seery, Y. Jiang, and Y. Qiao, "Deep learning in image-based breast and cervical cancer detection: a systematic review and meta-analysis," *npj Digital Medicine*, vol. 5, no. 1, 2022.

[8] E. Mahoro and M. A. Akhloufi, "Applying Deep Learning for Breast Cancer Detection in Radiology," *Current Oncology*, vol. 29, no. 11, pp. 8767–8793, 2022.

[9] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift fur Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019.

[10] B. Usha and S. Sandya, "Abnormality Detection in Ovarian Ultrasound Images using Active Contours," *Journal of Biomedical Engineering and Medical Imaging*, vol. 1, no. 4, pp. 14–23, 2014.

[11] T.-l. Pham, V.-h. Le, T.-h. Tran, and D. H. Vu, "Comprehensive study on Semantic Segmentation of Ovarian Tumors from Ultrasound Images," in *Conference on Information Technology and its Applications*, 2023.

[12] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *https://arxiv.org/pdf/1804.02767.pdf*, 2018. [Online]. Available: http://arxiv.org/abs/1804.02767

[13] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2021-October, pp. 2778–2788, 2021.

[14] G. Weng, D. Liu, L. Liu, Q. Wan, and S. Wu, "Research on an improved yolov7 algorithm for road surface crack detection," 2023.

[15] 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.

[16] M. Akazawa and K. Hashimoto, "Artificial intelligence in ovarian cancer diagnosis," *Anticancer research*, vol. 40, no. 8, pp. 4795–4800, 2020.

[17] 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," *CoRR*, vol. abs/2207.06799, 2022.

[18] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, "A review of medical image data augmentation techniques for deep learning applications," *Journal of Medical Imaging and Radiation Oncology*, vol. 65, no. 5, pp. 545–563, 2021.

[19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[20] Y. Zhu, S. Li, W. Du, and Y. Du, "Identification of table grapes in the natural environment based on an improved yolov5 and localization of picking points," *Precision Agriculture*, vol. 24, no. (4), pp. 1–22, 2023.

[21] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv*, 2020.

[22] 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," *https://arxiv.org/pdf/2207.02696.pdf*, pp. 7464–7475, 2022. [Online]. Available: http://arxiv.org/abs/2207.02696

[23] N. Sathiadhas Puvaneswari, "Detection of polycystic ovarian syndrome using convolutional neural network in conjunction with transfer learning models," Ph.D. dissertation, Dublin, National College of Ireland, 2022.

[24] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.

Thi-Loan Pham received Bachelor degree at Faculty Information Technology Hanoi Pedagogical University 2 (2007). She received M.Sc. degree at University of Engineering and Technology (2012). Currently, she is a lecture of HaiDuong University and is a PhD student of School of Electrical and Electronic Engineering (SEEE), Hanoi University of Science and Technology, Vietnam. Her research interests include Ovarian cancer on Ultrasound images.

Email: phamthiloan2011@gmail.com

Van-Hung Le received M.Sc. degree at Faculty Information Technology- Hanoi National University of Education (2013). He received PhD degree at International Research Institute MICA HUSTC-NRS/UMI - 2954 - INP Grenoble (2018). Currently, he is a lecture of Tan Trao University. His research interests include Computer vision, RANSAC and RANSAC variation.

Email: van-hung.le@mica.edu.vn

Thi-Lan Le graduated in Information Technology from Hanoi University of Science and Technology (HUST), Vietnam. She obtained an MS. degree in Signal Processing and Communication from HUST, Vietnam. In 2009, she received her Ph.D. degree at INRIA Sophia Antipolis, France in video retrieval. She is currently associate professor at School of Electrical and Electronic Engineering (SEEE), HUST, Vietnam. Her research interests include images processing, computer vision, content-based indexing and retrieval, video understanding and human-robot interaction.
Email: Thi-Lan.Le@mica.edu.vn

Thanh-Hai Tran graduated an engineer's degree in Information Technology from Hanoi University of Science and Technology (HUST) in 2001. She holds a M.S. degree and a Ph.D degree in Imagery Vision Robotic from Grenoble INP, France, in 2002 and 2006 respectively. Currently, she is a lecturer/researcher at School of Electronics and Telecommunications and Computer Vision Department, International Research Institute in Multimedia, Information, Communication and Application, HUST. Her main research interests are visual object recognition, video understanding, human-robot interaction and text detection for applications in Computer Vision.
Email: hai.tranthithanh1@hust.edu.vn

Duy-Hai Vu is an associate professor of biomedical engineering, director of Biomedical Electronics Center at Hanoi University of Science and Technology. He received the PhD degree from the Hanoi University of Science and Technology in 2013 and became the associate professor of biomedical engineering in 2016. His research ranges are bio-signal processing and applications, medical devices and technology, integrated electronic medical record, medical IoT and AI for health. Manufacturing equipment used in the covid-19 pandemic: BKVM-HF1 (HFNC), BK-Vent (Ventilator), BK-NICO (ICG Cardiac Output Measurement Equipment.
Email: hai.vuduy@hust.edu.vn

Hai Vu received BE in Electronics and Telecommunications in 1999 and M.E. in Information Processing and Communication in 2002, both from HUST, Hanoi, Vietnam. He received Ph.D. in Computer Science from Osaka University, Japan, in 2009. He has been a lecturer/researcher in HUST, Hanoi, Vietnam since 2012. His current research interests are in Computer Vision, Pattern Recognition, and Human-Computer Interactions.
Email: hai.vu@hust.edu.vn