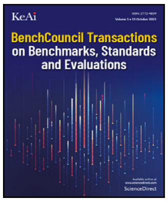
Contents lists available at ScienceDirect

BenchCouncil Transactions on Benchmarks, Standards and Evaluations

journal homepage: https://www.keaipublishing.com/en/journals/benchcouncil-transactions-on-benchmarks-standards-and-evaluations/


KeAi
BenchCouncil Transactions on Benchmarks, Standards and Evaluations

MVDI25K: A large-scale dataset of microscopic vaginal discharge images

Lin Li [a,b], Jingyi Liu [c], Fei Yu [a], Xunkun Wang [a,*], Tian-Zhu Xiang [d,*]

[a] *QINGDAO HUA JING BIOTECHNOLOGY CO., LTD., No. 77 Keyun Road, Qingdao, China*
[b] *Ocean University of China, Qingdao, China*
[c] *Qingdao University of Science and Technology, Qingdao, China*
[d] *Inception Institute of Artificial Intelligence, United Arab Emirates*

ARTICLE INFO

Keywords:
Vaginal discharge detection
Object segmentation
Object detection
Benchmark dataset
Medical images

ABSTRACT

With the widespread application of artificial intelligence technology in the field of biomedical images, the deep learning-based detection of vaginal discharge, an important but challenging topic in medical image processing, has drawn an increasing amount of research interest. Although the past few decades have witnessed major advances in object detection of natural scenes, such successes have been slow to medical images, not only because of the complex background and diverse cell morphology in the microscope images, but also due to the scarcity of well-annotated datasets of objects in medical images. Until now, in most hospitals in China, the vaginal diseases are often checked by observation of cell morphology using the microscope manually, or observation of the color reaction experiment by inspectors, which are time-consuming, inefficient and easily interfered by subjective factors. To this end, we elaborately construct the first large-scale dataset of **m**icroscopic **v**aginal **d**ischarge **i**mages, named ***MVDI25K***, which consists of 25,708 images covering 10 cell categories related to vaginal discharge detection. All the images in *MVDI25K* dataset are carefully annotated by experts with bounding-box and object-level labels. In addition, we conduct a systematical benchmark experiments on *MVDI25K* dataset with 10 representative state-of-the-art (SOTA) deep models focusing on two key tasks, *i.e.*, object detection and object segmentation. Our research offers the community an opportunity to explore more in this new field.

1. Introduction

Obstetrics and gynecology infectious diseases, such as vaginitis, cervicitis, and endometritis, often trouble women's health. It is reported in [1] that, the incidence of obstetrics and gynecology infectious diseases accounts for about 40% of the female population in China. The vaginal discharge examination is the most direct and effective way to detect the above diseases. For instance, the presence of trichomonas in the secretions can be used to determine whether a patient has trichomonas vaginitis, the presence of clue cells indicates that the patient has bacterial vaginosis, and the presence of candida albicans determine whether the patient has vulvovaginal candidiasis. As mentioned in [2], an increase in the number of leukocytes in vaginal secretions is a strong predictor of bacterial vaginosis or cervical infection. Besides, whether there are epithelial cells in the secretions is also a sign to judge whether the secretion sampling is qualified.

For a long time, manual inspection methods, observing the smear through a high-power microscope to conduct the diagnose, have dominated. As it is known, however, they suffer from some defects, such as time-consuming, labor-intensive, inefficient, and easy to be interfered by subjective factors. Recently, deep learning has prospered in object detection of natural scenes, indicating its great potential in the detection of vaginal discharge. However, there is a long way to go, which can be attributed to two key aspects. Firstly, there are considerable differences in the morphology, number, and distribution of cells in vaginal secretions, due to the differences between not only individuals but also different life stages of the same person. Obviously, it poses numerous difficulties for automatic and robust vaginal discharge detection, such as complex background, scale variations, extremely nonuniform object densities, large aspect ratios, and nonrigid changes of cell shape, as shown in Fig. 1. Most importantly, deep learning greatly relies on the large-scale well-annotated datasets, which has long been lacking in the medical community and thus hinder further research in this field.

To facilitate the study of vaginal discharge detection, we provide two contributions. First, we elaborately constructed a novel large-scale dataset of microscopic vaginal discharge images, called *MVDI25K*, which contains 25,708 microscope images covering 10 object classes of cells related to vaginal discharge detection. To the best of our knowledge, *MVDI25K* is the first large-scale dataset for vaginal discharge detection. It has several distinctive features:

[*] Corresponding authors.
E-mail addresses: maksimljc@163.com (X. Wang), tianzhu.xiang19@gmail.com (T.-Z. Xiang).

https://doi.org/10.1016/j.tbench.2021.100008
Received 6 August 2021; Received in revised form 11 October 2021; Accepted 20 October 2021
Available online 11 November 2021

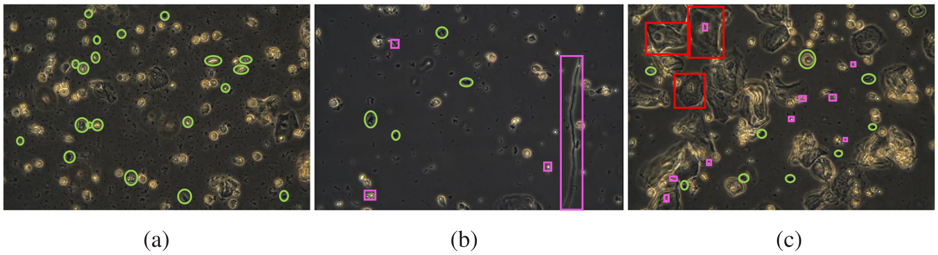


Fig. 1. Various examples of challenging images from *MVDI25K* dataset. The objects labeled with green bounding boxes are all impurities, including cell debris (such as epithelial cell debris), naked nucleus (no cytoplasm and cell membrane), drugs, crystals, starch granules, oil drop, *etc*. In (a), excessive impurities make the complex background, and bring various interferences to object detection. (b) shows the objects with different aspect ratios. Pink bounding boxes label the four types of candida, *i.e.*, candida1, candida2, candida3 and hyphae, among which the longest is hyphae. In (c), the epithelial cells labeled with red bounding boxes show the changeable cell morphology, and the large scale variations compared to the candida cells with pink bounding boxes.

- *Hierarchical categories*. All objects in the microscopic images are labeled into ten cell classes related to vaginal discharge detection, *e.g.*, epithelial cell, clue cell, leukocyte, and lactobacillus, *etc*. Specially, considering the diverse morphology of candida, we annotated them into four sub-classes according to morphological differences. The hierarchical categories could benefit the accurate and fine-grained object detection.
- *Diverse annotations*. The objects in *MVDI25K* dataset are hierarchically annotated with category labels, bounding boxes labels, and object-level masks, which can greatly facilitate different medical image processing tasks, such as object localization, object detection and object/cell segmentation.
- *High quality*. All the images in the dataset are collected by Leica and Olympus phase-contrast microscopes and megapixel dedicated medical Basler cameras with the size from 1536 × 1536 to 2064 × 3088. The phase-contrast microscope facilitates the acquisition of clearer and more realistic cell images. Moreover, cross checking by multiple experts and volunteers is conducted to maintain accuracy, reliability and consistency during the whole annotation process. These high-quality data and annotations could help providing deeper insight into the performance of algorithms.

Second, based on the established *MVDI25K*, we present a comprehensive study on 10 state-of-the-art baselines for vaginal discharge detection. We provide detailed experimental analyses in two scenarios, *i.e.*, object detection and object segmentation. Based on the evaluation results, we find that vaginal discharge detection is very challenging and still far from being solved, leaving much room for improvement. We hope that our research will give a strong boost to growth in this new field.

The remainder of the paper is organized as follows. We review the current medical datasets, medical object detection, and medical object segmentation in Section 2. In Section 3, we present details on the proposed *MVDI25K* dataset, including collection manner, annotation pipeline, and data statistics. Then, we describe benchmark experiments from the aspects of object detection and object segmentation, and provide both quantitative and qualitative experimental analysis in Section 4. Finally, we draw conclusions in Section 5.

2. Related work

In this section, we briefly review some closely related works, including current medical datasets, medical object detection, and medical object segmentation.

2.1. Medical image dataset

In general, X-rays, Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Positron Emission Computed Tomography (PET) are the four most widely used image-assisted means to help clinicians diagnose diseases, assess prognosis, and plan operations. Thus, a series of corresponding medical image datasets are constructed. Table 1 summarizes their details.

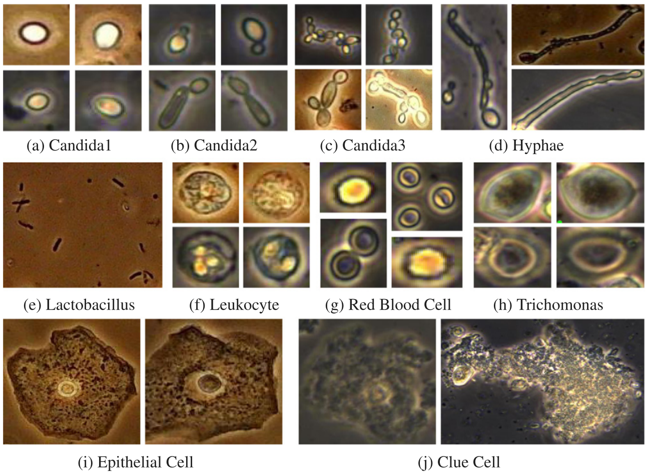


Fig. 2. Examples of 10 classes of cell images in *MVDI25K* dataset.

As is known to all, the well-annotated dataset plays an important role in data-driven medical image processing research. However, to the best of our knowledge, there is few datasets collected from microscope imaging for vaginal discharge research, which may hinder further research in this field. To this end, in this paper, we constructed the first large-scale dataset of vaginal discharge images providing professional annotations. Compared with Peng's dataset [11], the proposed *MVDI25K* provides more images with diverse and rich annotations. It is worth noting that collecting the microscopic image dataset is more difficult than datasets of other medical imaging equipment, because the image quality of microscope imaging is greatly affected by various factors, e.g., focus adjustment.

2.2. Deep models for medical object detection

Object detection, to identify and locate objects in an image or video, is a longstanding problem in computer vision. Recently, with the development of deep learning, many researchers in the medical image processing community have adapted deep object detectors developed for natural images to medical images.

As suggested in [12,13], object detection can be roughly divided into two categories: two-stage algorithms, such as R-CNN [14] and its variants, and one-stage algorithms, such as YOLO [15] and SSD [16]. Due to its high efficiency and good performance, YOLO and its variants have attracted extensive attention in the medical imaging community.

Table 1
Medical image datasets.

Dataset	Object	Year	Number	X-ray	CT	MRI	PET	MImg	BBox.	Obj.	Cate.
ABIDE [3]	Brain	2013	1,112*			✓				✓	✓
OASIS-3 [4]	Brain	2019	3,776			✓	✓			✓	✓
DDSM [5]	Breast	2000	10,239	✓						✓	✓
MURA [6]	Upper Limb	2018	40,561	✓							✓
LIDC-IDRI [7]	Lung	2006	244,527	✓	✓					✓	✓
LUNA16[a]	Lung	2016	888		✓					✓	✓
NSCLC [8]	Lung	2018	1,355		✓		✓			✓	✓
DeepLesion [9]	Lung *etc.*	2018	928,020		✓	✓	✓		✓		✓
ChestX-ray14 [10]	Chest	2017	112,120	✓					✓		✓
Peng' Dataset [11]	Vaginal Dis.	2021	3,645					✓			✓
MVDI25K (Ours)	Vaginal Dis.	2021	25,708					✓	✓	✓	✓

BBox.: Bounding-box annotation. Obj.: Object-level annotation. Cate.: Categories. MImg: Microscopic image. Vaginal Dis.: Vaginal discharge. *: number of sub-datasets.
[a] https://luna16.grand-challenge.org/.

Based on the chest CT scans from Lung Image Database Consortium, YOLO-based model was applied to efficiently and accurately identify lung nodules in [17]. The high-quality gallstone CT image dataset was established in [18], and YOLO-v3 achieved good performance on the detection of granular gallstones and muddy gallstones. To effectively fighting against COVID-19, an improved model based on YOLOv2 and ResNet-50 was designed to detect medical masks with high accuracy [19].

2.3. *Deep models for medical object segmentation*

In the past few years, convolutional neural networks have been the most commonly-used architecture in state-of-the-art models for medical image segmentation, such as FCN [20], U-Net [21], and Deeplab series [22].

Among them, the U-net plays an important role and has been applied to numerous fields, such as using NAS-Unet to segment Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and ultrasound with high quality [23], and medical object segmentation including liver, brain and lung tissue and tumor segmentation [24–26], cell segmentation [27], optic disc segmentation [28], *etc*.

Another related topic that deserves attention is camouflage object detection [29,30], which is to segment objects which have a similar pattern (e.g. texture, color and direction) to their natural or man-made environment. In our vaginal discharge detection, trichomonas has a strong camouflage compared to other cell morphologies. Thus camouflage object detection could shed new light on trichomonas detection.

3. Proposed dataset

Our *MVDI25K* dataset contains 25,708 microscopic images belonging to ten object/cell classes related to vaginal discharge detection. The images are carefully selected to cover diverse challenging cases, *e.g.*, complex background, large-scale variation, and nonuniform object density. Examples can be found in Figs. 1 and 2. We will describe the details of *MVDI25K* in terms of three aspects, *i.e.*, data collection, annotation pipeline, and data statistics, as follows.

3.1. *Data collection*

We build a high-quality dataset, *MVDI25K*, images of which are collected from the *HJ500* Discharge Analysis Workstation with two sources. One is directly photographed from the fresh samples collected by many hospitals across the country, and the other is captured by ourselves using the specimens we collected from other partner hospitals. The images are acquired by Leica and Olympus phase-contrast microscope and the megapixel dedicated medical Basler camera. Our dataset covers *315* hospitals distributed in more than *20* provinces in China, and 26 of them are tertiary hospitals including Beijing Tiantan Hospital and Hubei Maternity and Child Health Hospital. The entire collection work last nearly 9 weeks. We just use the devices to collect the image data independently, and do not collect any patient information. The images are free from copyright and loyalties and will be available at: https://zenodo.org/record/5523661.

Besides, the images are acquired by two types of microscopes, Leica and Olympus microscope. Both microscopes adopt a phase contrast field of view, which is more suitable for microscopy examination than ordinary optical microscopes, and even unstained cells can be observed more clearly and brighter.

3.2. *Professional annotation*

To facilitate the study of vaginal discharge detection, we provide bounding-box and object-level annotations for each image in our *MVDI25K* dataset. We hire seven professional annotators, and six of them are divided into three groups. Each group is responsible for the annotation and meantime they need to cross-check the label results from other groups. After finishing the annotation process, the team leader (the seventh annotation expert) will carefully conduct the final validation to ensure high-quality annotation.

3.2.1. *Categories*

We establish a hierarchical taxonomic system for the proposed dataset. We first choose seven major cell categories such as *epithelial cell*, *clue cell*, *leukocyte*, *candida*, *red blood cell*, *lactobacillus*, and *trichomonas*. Considering the large morphology differences of candida cells, shown in the first row of Fig. 2, we divided the candida into four sub-classes, namely *candida1*, *candida2*, *candida3*, and *hyphae*. Finally, we integrate these classes into 10 cell/ object classes. The taxonomic structure of our *MVDI25K* is given in Fig. 3(a). The example of the word cloud is shown in Fig. 3(b). We believe that the fine-grained classification of candidas would play a positive role in accurate vaginal discharge detection.

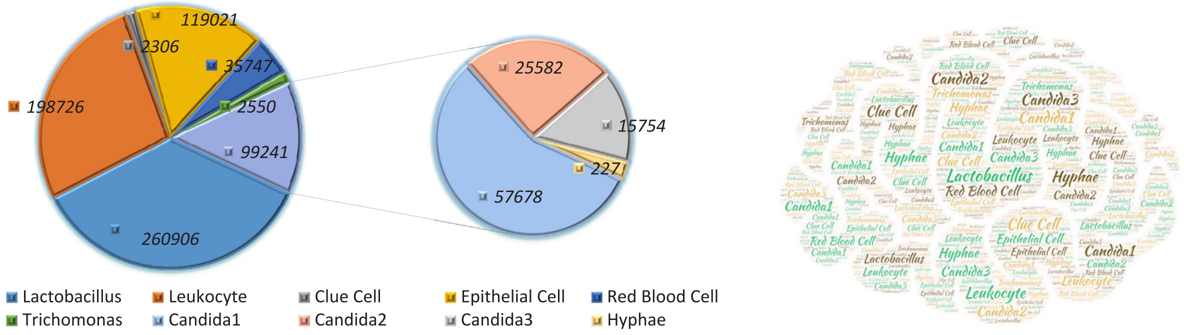
3.2.2. *Bounding-box annotation*

Bounding box is widely used in object detection and localization. To extend *MVDI25K* for the object proposal task, we carefully annotate the bounding boxes around the objects in each image. Finally, we obtained total 718,497 object instances from 25,708 microscopic images. Some examples of annotated patches are shown in Fig. 4(a).

3.2.3. *Object-level annotation*

High-quality pixel-level annotations are necessary for *MVDI25K* dataset. Here we focus on the *Trichomonas* category, the disease who causes is one of the most common obstetrics and gynecology infectious diseases.[1] Besides, the Trichomonas cell is usually active and thus difficult to be observed clearly from vaginal secretion microscope sample,

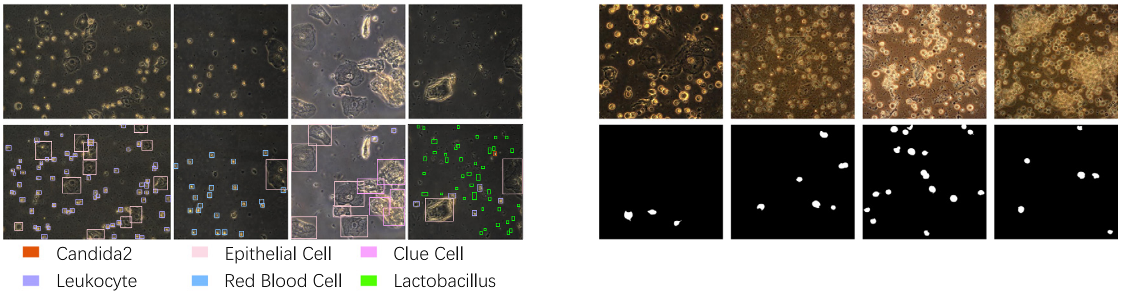
[1] https://news.un.org/en/story/2019/06/1039891.



(a) The 10 cell/object classes in *MVDI25K* and the number of labels in each class.

(b) Word cloud of *MVDI25K*.

Fig. 3. Categories of *MVDI25K*.



(a) Examples of bounding-box annotations in the proposed *MVDI25K* dataset. As can be seen, the labeled objects are with diverse scales, nonuniform distributions, and nonrigid shape changes, which bring challenges for object detection.

(b) Examples of object-level annotations for Trichomonas in the proposed *MVDI25K* dataset. As can be seen, the labeled Trichomonas are often small, and with some occlusion or overlap. In addition, they are generally easily concealed in their background, that is, very similar to their background, which may largely confuse the algorithms.

Fig. 4. Annotations of *MVDI25K*.

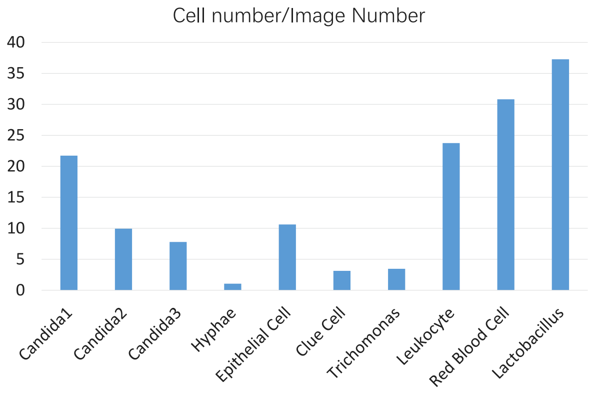


Fig. 5. Multiple Objects. The number of each object in each image.

which poses a challenge for data collection. Through observation, we found that most of the Trichomonas cells seem to have a similar pattern, *e.g.*, texture, color and shape, to other cells in the background, that is, they have a certain degree of concealment, which is easy to cause confusion to the detector. What is more, by data annotation, it can be seen that the Trichomonas cell generally has the following challenging attributes: (1) dense objects: more than 10 objects in a single image; (2) small object: too small size compared with its large background; (3) occlusion/overlap: incomplete object contour due to the occlusion of other cells or impurities and multiple cells overlap; (4) irregular shape: cell contains tiny parts (*e.g.* small tails). As a result, it may become a hot potato when using deep models to detect this type of cells.

To this end, Trichomonas detection deserves more effort and thus we provide meticulous object-level annotations for Trichomonas. We adopt *Photoshop* as the annotation tool to label the object-level masks. In this way, we obtain a total of 2,550 object-level annotations from 912 Trichomonas images. Some examples can be seen in Fig. 4(b).

3.3. Dataset features and statistics

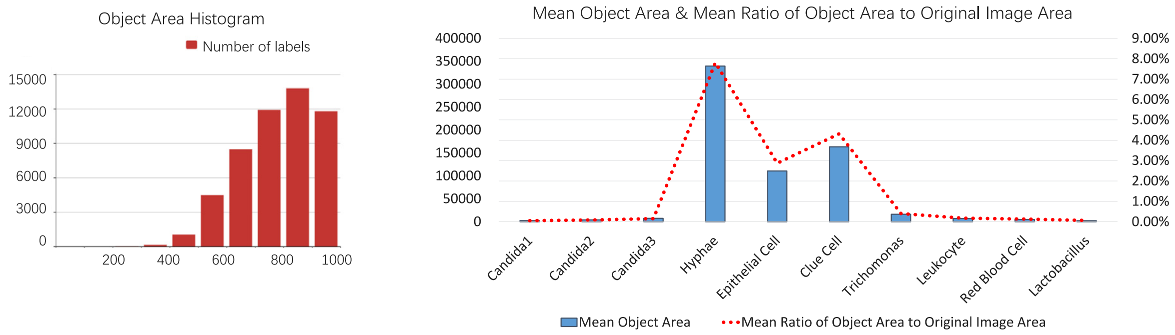
To provide deeper insights into our *MVDI25K*, we present its several important characteristics in below.

3.3.1. Multiple objects

In this paper, we define multiple objects as cells of the same type in one image with a number equal to or greater than two. Note that the multi-object value is the total number of a certain type of object divided by the number of images containing this object. As shown in Fig. 5, hypha is with the low multi-object attribute value 1.07, while the other cell classes are larger than 3. The top-3 is the Lactobacillus, red blood cell, and Leukocyte, which are 37.27, 30.82, and 23.75, respectively.

3.3.2. Small object

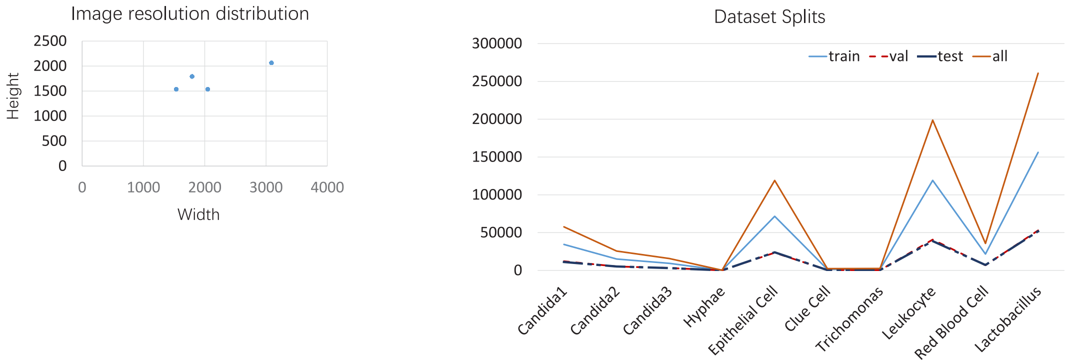
Small objects, As we know, is defined as (a) the objects whose absolute size is less than 32 × 32, or (b) the objects whose width and height are less than 1/10 of the width and height of the whole image. Generally speaking, the small object is easily overwhelmed by the noisy background. In addition, for deep models, its feature information will disappear when the network gradually goes deeper, which lets many deep models be cast into the shade. Thus, detection of small objects is a challenging issue.



(a) Number of pixels per object/cell.

(b) Object pixel ratio.

Fig. 6. The attribution analysis of small object in the proposed *MVDI25K*.



(a) Image resolution distribution of *MVDI25K* dataset (unit for the axis: pixel).

(b) Dataset Splits

Fig. 7. Annotations of *MVDI25K*.

We counted these two small target attributes in the *MVDI-25K* dataset, shown in Fig. 6(a). It shows the number of cells whose absolute pixel is less than or equal to 1000. A total of 51,549 small objects are labeled in the proposed dataset, accounting for 7.17%. Fig. 6(b) shows the mean area of each type of cells and the mean ratio of their area in one images. Obviously, except for hyphae, clue cell and epithelial cell, the area ratio of other classes of cells are far less than 1%, the threshold of small objects.

3.3.3. *Resolution distribution*

High-resolution images generally provide more object details for deep model training and thus facilitate to yield outstanding detection performance when testing [31]. When collecting data, we carefully adjusted the microscope settings to obtain high-resolution images. Fig. 7(a) shows the resolution distribution of our dataset. Specifically, the four resolutions of the images are: 1536 ×1536, $1536 \times 2048, 1792 \times 1792, 2064\times$ 3088, and their proportions are: 30.03%, 0.47%, 42.94%, 26.56%.

3.3.4. *Dataset splits*

To provide a large quantity of training data for learning-based approaches, we divided 25,708 images into training set, validation set and test set, with a ratio of 6:2:2. It should be noted that it is impossible to split the dataset based on cell class and then select randomly from each class, because each image contains at least 2 or 3 classes of cells. In order to ensure the same distribution of the training set, the validation set, and the test set, we first select the images with the least number of cells (hyphae), and then select them randomly. Then the images containing the second-fewest cell classes (Trichomonas) are selected at random. Follow this rule until all types of cells have been split. Fig. 7(b) presents the final split results of different cell categories. Consequently, the dataset split satisfies the same distribution of training set, validation set and test set.

4. Benchmark experiments

Based on the established *MVDI25K*, we systematically benchmark 10 representative models on two key tasks, object detection and object segmentation. From the evaluation results, we conduct some in-depth analysis and present several insightful conclusions which may inspire further research.

4.1. *Object detection experiments*

4.1.1. *Dataset settings*

Based on the data split rules described in Section 3.3.4, we split the whole images into 15,428 training set, 5,143 validation set, and 5,137 test set, respectively, with corresponding bounding box ground-truth.

4.1.2. *Training protocols*

In this benchmark experiment, we collect the released codes of three representative object detection models, that is YOLOv5-s, YOLOv5-m, and YOLOv5-x [32], and re-train these models with the training set of *MVDI25K*. The images are set to 640 × 640 as model input. The initial learning rate is 1e-2, and GIoU loss gain is 0.05, and class loss gain is 0.5. For optimizer, the momentum is set to 0.937, and weight decay is set to 0.0005. The batch size of YOLOv5-s, YOLOv5-m and YOLOv5-x are set to 36, 24 and 8 respectively. The total training is 300 epoch on a NVIDIA GeForce RTX 2080Ti with 11 GB memory.

4.1.3. *Evaluation metrics*

We apply three widely-used metrics to evaluate object detection performance. These metrics include precision (P), recall (R), and mean average precision (mAP).

P is the accuracy rate, that is the percentage of the correct positive classes account for all positive classes detected, *i.e.*, P=$TP/(TP+FP)$.

Table 2
Quantitative results of object detection on our *MVDI25K* dataset. Cls1-10: Candida1, Candida2, Candida3, Hyphae, Epithelial Cell, Clue Cell, Trichomonas, Leukocyte, Red Blood Cell, and Lactobacillus. The first three categories with the worst performance are shown in red, blue, and green fonts.

Models	Metric	All	$Cls1$	$Cls2$	$Cls3$	$Cls4$	$Cls5$	$Cls6$	$Cls7$	$Cls8$	$Cls9$	$Cls10$
YOLOv5-s [32]	P	0.464	0.480	0.407	0.409	0.198	0.690	0.267	0.473	0.596	0.643	0.480
	R	0.894	0.966	0.942	0.950	0.476	0.982	0.811	0.915	0.985	0.961	0.956
	mAP@.5	0.801	0.882	0.853	0.860	0.307	0.964	0.572	0.825	0.942	0.936	0.872
	mAP@.5:.95	0.560	0.520	0.540	0.557	0.185	0.809	0.417	0.599	0.715	0.754	0.502
YOLOv5-m [32]	P	**0.502**	0.507	0.449	0.458	0.202	0.747	0.287	0.536	0.648	0.678	0.513
	R	0.905	0.971	0.959	0.967	0.548	0.976	0.831	0.908	0.980	0.957	0.957
	mAP@.5	0.820	0.891	0.897	0.899	0.370	0.965	0.573	0.840	0.945	0.939	0.879
	mAP@.5:.95	0.588	0.541	0.590	0.613	0.243	0.816	0.430	0.618	0.732	0.773	0.521
YOLOv5-x [32]	P	0.493	0.512	0.476	0.495	0.183	0.717	0.263	0.505	0.629	.657	0.487
	R	**0.919**	0.977	0.971	0.972	0.619	0.979	0.843	0.915	0.983	0.962	0.967
	mAP@.5	**0.837**	0.901	0.920	0.919	0.434	0.966	0.605	0.851	0.947	0.941	0.883
	mAP@.5:.95	**0.607**	0.561	0.617	0.640	0.266	0.819	0.465	0.650	0.741	0.782	0.532

Noted *TP* means that both the predicted value and the true value are both 1, and *FP* denotes that the true value is 0 and the predicted value is 1.

R is the recall rate, that is the percentage of the correct positive class accounts for all the true positive classes, *i.e.*, *R=TP/(TP+FN)*. Noted *FN* means that the true value is 1 and the predicted value is 0.

AP is the area under the *P-R* curve, and *mAP* is the average of *AP* of each category. We adopt *mAP@.5* where ".5" indicates the threshold for judging *IoU* as a positive or negative sample, and *mAP@0.5:0.95* which means the *AP* average under a series of thresholds that start at 0.5 and increase to 0.95 in steps of 0.05.

4.1.4. Quantitative evaluation

As can be seen in Table 2, from the "all" column, the overall missed detection rate of the outstanding YOLOv5-x model is only 8.1%(*i.e.*, 1-91.9%), but its cost exceeds that of the YOLOv5-m model. False positive samples accounted for 50.7% (*i.e.*, 1-49.3%), which means that the number of false positive cells detected is almost the same as the number of true cells. In fact, the requirements for accurate cell identification in medical images are relatively high. Even if R and mAP have reached a high level, the low P metrics means too many false positive samples, which is a very important but still challenging problem in medical image detection.

As shown in Table 2, in all evaluation items, the fourth (Hyphae) and sixth classes (Clue Cell) are the two worst performing categories. Objectively speaking, in our *MVDI25K* dataset, the numbers of images containing hyphae and clue cells are very small, especially only 213 images contain hyphae. For the seventh classes cell (Trichomonas), due to the small number of images, and sometimes the similar appearance to leukocyte, it is also a difficult class in object detection. As we all know, different types of objects are inherently unevenly distributed in nature. It can also be seen from this experiment that the object detection of unbalanced categories is a very important but challenging issue, which is worthy of further study.

4.1.5. Qualitative evaluation

Five representative detection results are shown in Fig. 8. In the first line, the epithelial cell in the middle left is disturbed by the complex background, causing the failed detection for three models. In addition, three models misjudged its left cell as Trichomonas. The second row contains two types of typical red blood cells, dark side and bright side, as well as intact and broken white blood cells. At the same time, leukocytes in complete and fragmented form. YOLOv5-m model is the best for detecting multiple leukocytes in the upper left corner of the image. The image in the third row is a typical environment where a large number of lactobacillus exist. From the perspective of confidence, YOLOv5-m performed slightly better than YOLOv5-x. The difficulty in the identification of the fourth line of pictures is that the morphology of Trichomonas and some leukocytes are very similar, especially the three adjacent cells in the upper right corner. None of the three models can completely accurately identify leukocytes and Trichomonas. The last image contains intact epithelium and deformed epithelium. It is difficult to identify deformed epithelium, and all three models are missed.

4.2. Object segmentation experiments

4.2.1. Dataset settings

In this benchmark experiment, our dataset provides a total of 2550 object-level annotations from 912 Trichomonas images. We split these images into 730 images for training and 182 for testing, with corresponding object-level ground-truth.

4.2.2. Training protocols

In this study, we evaluate eight representative, recently published and state-of-the-art deep models for object segmentation or concealed object detection, including BASNet [33], CPD [34], SCRN [35], U^{2}Net [36], F3Net [37], GateNet [38], PraNet [39], and SINet [29]. We collect the released codes of these models and re-train them on our dataset with 50 epochs on a NVIDIA GeForce RTX 2080Ti GPU. During the training stage, the batch size is set to 20, and the maximum learning rate is 0.05. For the Adam optimizer, the momentum is 0.9 and the weight decay is 5e-4. When the memory is insufficient, the batch size and epoch are changed to 10 and 100, respectively.

4.2.3. Evaluation metrics

To provide a comprehensive evaluation, six widely-used metrics are employed to quantitatively compare the eight deep models for object segmentation on the proposed *MVDI25K*, with the evaluation toolbox provided by [39], including structural similarity measure (S_α, with α =0.5) [40], enhanced-alignment measure ($meanE_\phi$ and $maxE_\phi$), and F_β measure (wF_β, $meanF_\beta$ and $maxF_\beta$).

- S-Measure calculates the structure similarities from the object-aware and region-aware aspects, between objects in ground-truth (GT) maps and predicted maps:

$$S_\alpha = \alpha * S_o + (1-\alpha) * S_r, \tag{1}$$

where

$$S_o = \frac{2 * E(pre)}{E(pre)^2 + 1 + \sigma + e}, \tag{2}$$

S_r indicates that the four regions are cut into four regions according to the position of the center of gravity, and the area of the entire image occupied by the pixels of the four regions is used as the weight, and the weighted average of the structure similarities of the four regions is calculated.

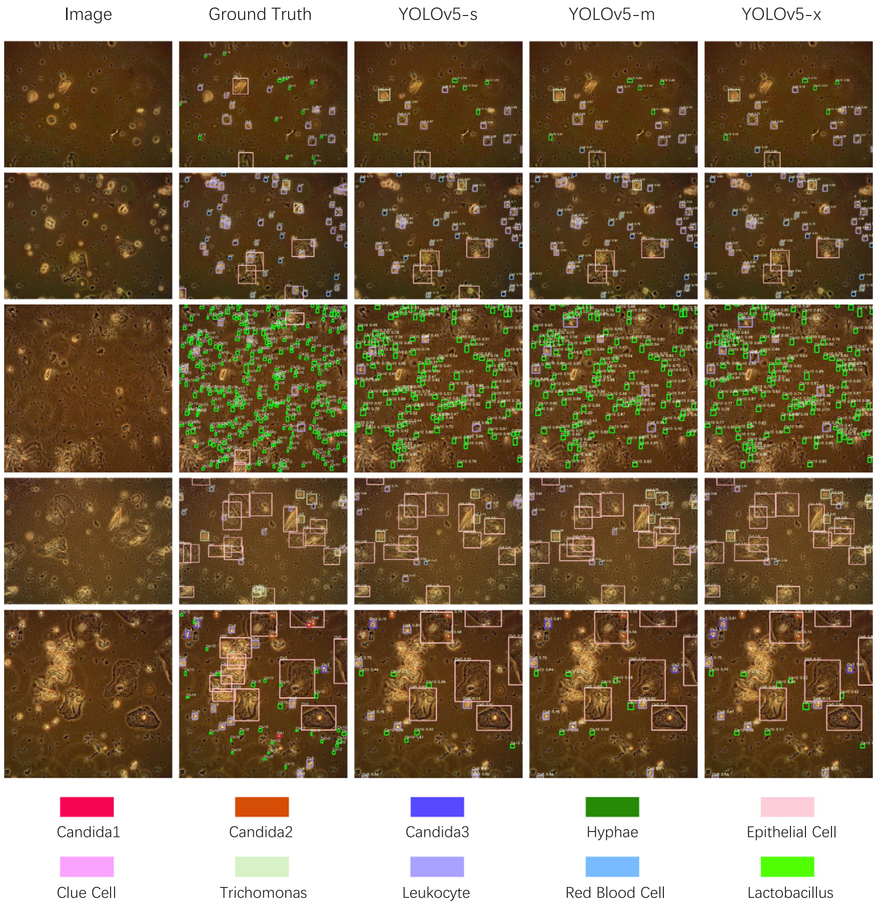


Fig. 8. Qualitative examples of object detection with three representative YOLO models evaluated on our *MDVI25K* dataset.

- E-Measure is a cognitive vision-inspired metric, which measures both the local and global similarities between two binary maps [41]. It combines local pixel values and image-level average values to capture both image-level statistics and local pixel matching information. Specifically, it is defined as:

$$E_{\phi} = \frac{1}{w * h} \sum_{x=1}^{w} \sum_{y=1}^{h} \phi FM(x, y) \tag{3}$$

  where w and h are the width and the height of the map, respectively. $\phi FM = f(\xi FM)$, the value of ξFM depends on the similarity between feature map and ground truth, $f(x) = \frac{1}{4}(1+x)^2$. The rest of the specific derivation procedure is given by [41]. Here, we introduce mean/maximal E-measure, *i.e.*, $meanE_{\phi}$ and $maxE_{\phi}$, to provide a more comprehensive evaluation.
- F-measure is essentially a region-based similarity metric, which is based on weighted precision and recall values.

$$F_{\beta} = \frac{(1+\beta^2)R * P}{R + \beta^2 P}, \tag{4}$$

  where β^2 is a parameter to trade-off recall and precision, and it is usually set to 0.3. Here, Here, we introduce 3 variants of this metric, namely wF_{β}, $meanF_{\beta}$ and $maxF_{\beta}$, for a comprehensive evaluation.

4.2.4. Quantitative evaluation

Table 3 shows the evaluation results of all models on our dataset. Overall, PraNet is the best performing models compared to the others. PraNet obtained the best results on four metrics, S_{α}, wF_{β}, $meanF_{\beta}$ and $maxF_{\beta}^{w}$, especially on S_{α}. Its value for the S_{α} metric is 0.159 above the mean (0.648) and 0.014 above the second place (0.793). BASNet and U^{2}Net performed the best on metrics $meanE_{\phi}$ and $maxE_{\phi}$, respectively. The $meanE_{\phi}$ of BASNet was 0.158 higher than the average (0.711) and the $maxE_{\phi}$ of U^{2}Net was 0.043 higher compared to the average (0.852).

Although some deep models have achieved seemingly-good results on some metrics, they are still far from obtaining satisfactory performance. In addition, trichomonas is extremely active in microscope samples, so when observed with a microscope, trichomonas usually has two states of blur and clarity. Our current dataset is mainly focused on labeling clear trichomonas. We will continue to label the active and non-clearly trichomonas to construct an increasingly challenging dataset.

4.2.5. Qualitative evaluation

Fig. 9 shows seven representative results, including original images, ground truth, and object segmentation results of deep models. Specifically, our images contain multiple dense objects (first row), transgressions (second row), overlaps (third row), partial occlusion (4th and 5th rows), complex shapes (5th row), small objects and cases with extremely high similarity between objects and backgrounds, which bring many difficulties for object segmentation. From the segmentation map, it can be seen that the these models are in a less favorable situation for trichomonas edge recognition.

Obviously, the segmentation results in the first row are angular and ambiguous. For trichomonas that are partially occluded by other objects (more than 30% of their own area), there is a tendency to miss detection. For example, the trichomonas near the middle-left position in the fourth row, all deep models failed to detect it. Meanwhile, most of the above models cannot accurately capture the complex shape of the object. Trichomonas in the upper left corner of the fourth row and the middle right of the fifth row cannot be clearly and completely identified and all models miss their small tails or flagella. In addition, for objects

Table 3
Quantitative results of object segmentation on our *MVDI25K* dataset. "↑" indicates the higher the score the better. The best results are in boldface.

Models	S_α ↑	$meanE_\phi$ ↑	$maxE_\phi$ ↑	wF_β ↑	$meanF_\beta$ ↑	$maxF_\beta$ ↑
2019 BASNet [33]	0.793	**0.868**	0.890	0.585	0.634	0.653
2019 CPD [34]	0.616	0.608	0.716	0.253	0.328	0.400
2019 SCRN [35]	0.641	0.771	0.893	0.217	0.517	0.587
2020 U^{2}Net [36]	0.691	0.739	**0.895**	0.344	0.489	0.612
2020 F3Net [37]	0.791	0.856	0.881	0.557	0.623	0.661
2020 PraNet [39]	**0.807**	0.856	0.880	**0.613**	**0.658**	**0.677**
2020 SINet [29]	0.520	0.726	0.868	0.044	0.472	0.578

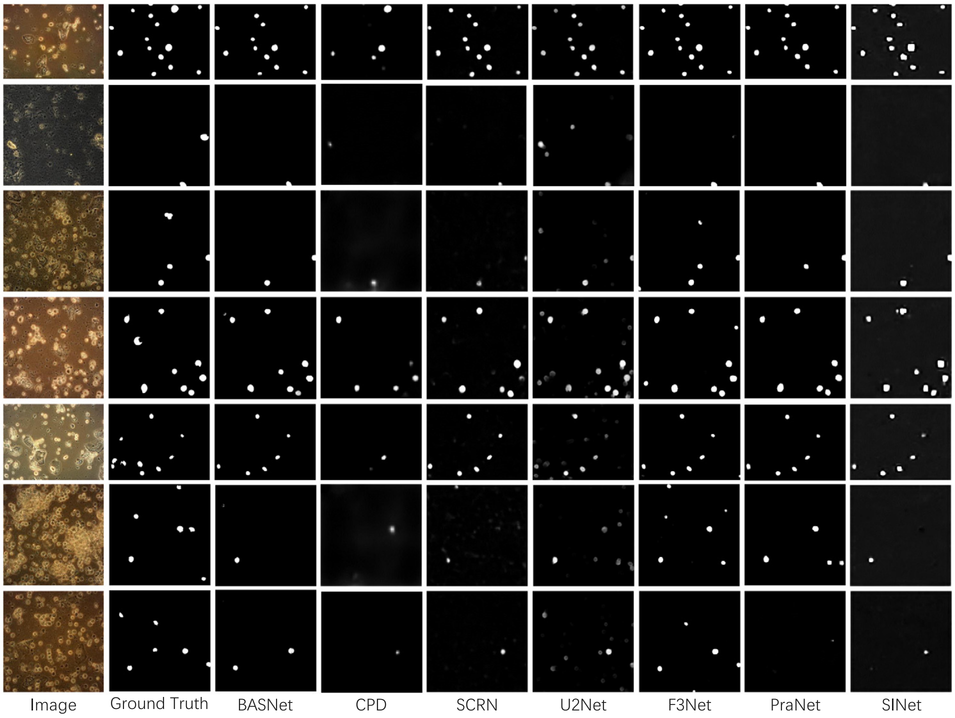


Fig. 9. Qualitative examples of object segmentation with the existing representative models evaluated on our *MDVI25K* dataset.

with similar colors and shapes (such as leukocytes and trichomonas), most of models tend to confuse them.

5. Conclusion

In this paper, we have presented the first comprehensive benchmark study on vaginal discharge detection. Specifically, we have constructed the first large-scale and challenging dataset of microscopic vaginal discharge, *MVDI25K*, containing 25,708 images with diverse and high-quality annotations. Then, we conducted a systematical benchmark experiments on 10 representative SOTA deep models on two key tasks, *i.e.*, object detection and object segmentation, and provided some insightful discussions. The benchmark indicates that vaginal discharge detection is far from being solved. We hope the studies presented in this work would facilitate the development of this field.

References

[1] R. Chen, Q. Liao, Attach importance to female reproductive tract infection and vaginal microecological diagnosis and treatment in China, Chin. J. Lab. Med. 41 (4) (2018) 251–253.

[2] W. Geisler, S. Yu, M. Venglarik, J. Schwebke, Vaginal leucocyte counts in women with bacterial vaginosis: relation to vaginal and cervical infections, Sex. Transm. Infect. 80 (5) (2004) 401–405.

[3] C. Craddock, Y. Benhajali, C. Chu, F. Chouinard, A. Evans, A. Jakab, B.S. Khundrakpam, J.D. Lewis, Q. Li, M. Milham, et al., The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives, Front. Neuroinformatics 7 (2013).

[4] P.J. LaMontagne, T.L. Benzinger, J.C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant, J. Hassenstab, K. Moulder, A. Vlassenko, et al., OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease, MedRxiv (2019).

[5] K. Bowyer, D. Kopans, W. Kegelmeyer, R. Moore, M. Sallam, K. Chang, K. Woods, The digital database for screening mammography, in: Third International Workshop on Digital Mammography, Vol. 58, pp. 27.

[6] P. Rajpurkar, J. Irvin, A. Bagul, D. Ding, T. Duan, H. Mehta, B. Yang, K. Zhu, D. Laird, R.L. Ball, et al., Mura: Large dataset for abnormality detection in musculoskeletal radiographs, 2017, arXiv preprint arXiv:1712.06957.

[7] S.G. Armato III, G. McLennan, L. Bidaut, M.F. McNitt-Gray, C.R. Meyer, A.P. Reeves, B. Zhao, D.R. Aberle, C.I. Henschke, E.A. Hoffman, et al., The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans, Med. Phys. 38 (2) (2011) 915–931.

[8] S. Bakr, O. Gevaert, S. Echegaray, K. Ayers, M. Zhou, M. Shafiq, H. Zheng, J.A. Benson, W. Zhang, A.N. Leung, et al., A radiogenomic dataset of non-small cell lung cancer, Sci. Data 5 (1) (2018) 1–9.

[9] K. Yan, X. Wang, L. Lu, R.M. Summers, DeepLesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning, J. Med. Imaging 5 (3) (2018) 036501.

[10] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, R.M. Summers, Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2097–2106.

[11] S. Peng, H. Huang, M. Cheng, Y. Yang, F. Li, Efficiently recognition of vaginal micro-ecological environment based on convolutional neural network, in: 2020 IEEE International Conference on E-Health Networking, Application & Services, HEALTHCOM, IEEE, 2021, pp. 1–6.

[12] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen, Deep learning for generic object detection: A survey, Int. J. Comput. Vis. 128 (2) (2020) 261–318.

[13] Z. Zou, Z. Shi, Y. Guo, J. Ye, Object detection in 20 years: A survey, 2019, arXiv preprint arXiv:1905.05055.

[14] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587.

[15] J. Redmon, S. Divvala, R. Girshick, 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.

[16] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, Ssd: Single shot multibox detector, in: European Conference on Computer Vision, Springer, 2016, pp. 21–37.

[17] J. George, S. Skaria, V. Varun, et al., Using YOLO based deep learning network for real time detection and localization of lung nodules from low dose CT scans, in: Medical Imaging 2018: Computer-Aided Diagnosis, Vol. 10575, International Society for Optics and Photonics, 2018, p. 105751I.

[18] S. Pang, T. Ding, S. Qiao, F. Meng, S. Wang, P. Li, X. Wang, A novel YOLOv3-arch model for identifying cholelithiasis and classifying gallstones on CT images, PLoS One 14 (6) (2019) e0217647.

[19] M. Loey, G. Manogaran, M.H.N. Taha, N.E.M. Khalifa, Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection, Sustainable Cities Soc. 65 (2021) 102600.

[20] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[21] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[22] L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, 2017, arXiv preprint arXiv:1706.05587.

[23] Y. Weng, T. Zhou, Y. Li, X. Qiu, Nas-unet: Neural architecture search for medical image segmentation, IEEE Access 7 (2019) 44247–44257.

[24] V. Cherukuri, P. Ssenyonga, B.C. Warf, A.V. Kulkarni, V. Monga, S.J. Schiff, Learning based segmentation of CT brain images: application to postoperative hydrocephalic scans, IEEE Trans. Biomed. Eng. 65 (8) (2017) 1871–1884.

[25] W. Li, et al., Automatic segmentation of liver tumor in CT images with deep convolutional neural networks, J. Comput. Commun. 3 (11) (2015) 146.

[26] Y. Onishi, A. Teramoto, M. Tsujimoto, T. Tsukamoto, K. Saito, H. Toyama, K. Imaizumi, H. Fujita, Multiplanar analysis for pulmonary nodule classification in CT images using deep convolutional neural network and generative adversarial networks, Int. J. Comput. Assist. Radiol. Surg. 15 (1) (2020) 173–178.

[27] T.-H. Song, V. Sanchez, H. EIDaly, N.M. Rajpoot, Dual-channel active contour model for megakaryocytic cell segmentation in bone marrow trephine histology images, IEEE Trans. Biomed. Eng. 64 (12) (2017) 2913–2923.

[28] H. Fu, J. Cheng, Y. Xu, D.W.K. Wong, J. Liu, X. Cao, Joint optic disc and cup segmentation based on multi-label deep network and polar transformation, IEEE Trans. Med. Imaging 37 (7) (2018) 1597–1605.

[29] D.-P. Fan, G.-P. Ji, G. Sun, M.-M. Cheng, J. Shen, L. Shao, Camouflaged object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2777–2787.

[30] L. Li, B. Dong, E. Rigall, T. Zhou, J. Donga, G. Chen, Marine animal segmentation, IEEE Trans. Circuits Syst. Video Technol. (2021).

[31] Y. Zeng, P. Zhang, J. Zhang, Z. Lin, H. Lu, Towards high-resolution salient object detection, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 7234–7243.

[32] G. Jocher, K. Nishimura, T. Mineeva, R. Vilariño, Yolov5, Code Repos. (2020) https://github.com/ultralytics/yolov5.

[33] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, M. Jagersand, Basnet: Boundary-aware salient object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 7479–7489.

[34] Z. Wu, L. Su, Q. Huang, Cascaded partial decoder for fast and accurate salient object detection, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3907–3916.

[35] Z. Wu, L. Su, Q. Huang, Stacked cross refinement network for edge-aware salient object detection, in: The IEEE International Conference on Computer Vision (ICCV), 2019.

[36] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O.R. Zaiane, M. Jagersand, U2-net: Going deeper with nested U-structure for salient object detection, Pattern Recognit. 106 (2020) 107404.

[37] J. Wei, S. Wang, Q. Huang, F^{3}Net: Fusion, feedback and focus for salient object detection, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 07, 2020, pp. 12321–12328.

[38] X. Zhao, Y. Pang, L. Zhang, H. Lu, L. Zhang, Suppress and balance: A simple gated network for salient object detection, in: European Conference on Computer Vision, Springer, 2020, pp. 35–51.

[39] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, L. Shao, PraNet: Parallel reverse attention network for polyp segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 263–273.

[40] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, A. Borji, Structure-measure: A new way to evaluate foreground maps, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4548–4557.

[41] D.-P. Fan, C. Gong, Y. Cao, B. Ren, M.-M. Cheng, A. Borji, Enhanced-alignment measure for binary foreground map evaluation, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18), 2018, pp. 698–704.