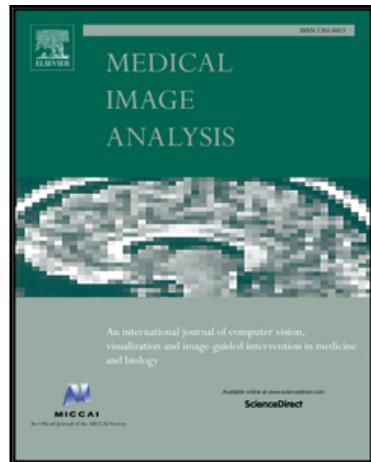


## Journal Pre-proof

DigestPath: a Benchmark Dataset with Challenge Review for the Pathological Detection and Segmentation of Digestive-System

Qian Da , Xiaodi Huang , Zhongyu Li , Yanfei Zuo, Chenbin Zhang, Jingxin Liu, Wen Chen, Jiahui Li, Dou Xu, Zhiqiang Hu, Hongmei Yi, Yan Guo, Zhe Wang, Ling Chen, Li Zhang, Xianying He, Xiaofan Zhang, Ke Mei, Chuang Zhu., Weizeng Lu, Linlin Shen, Jun Shi, Jun Li, Sreehari S, Ganapathy Krishnamurthi, Jiangcheng Yang, Tiancheng Lin, Qingyu Song, Xuechen Liu, Simon Graham, Raja Muhammad Saad Bashir, Canqian Yang, Shaofei Qin, Xinmei Tian, Baocai Yin, Jie Zhao, Dimitris N. Metaxas, Hongsheng Li, Chaofu Wang, Shaoting Zhang



PII: S1361-8415(22)00132-3  
DOI: <https://doi.org/10.1016/j.media.2022.102485>  
Reference: MEDIMA 102485

To appear in: *Medical Image Analysis*

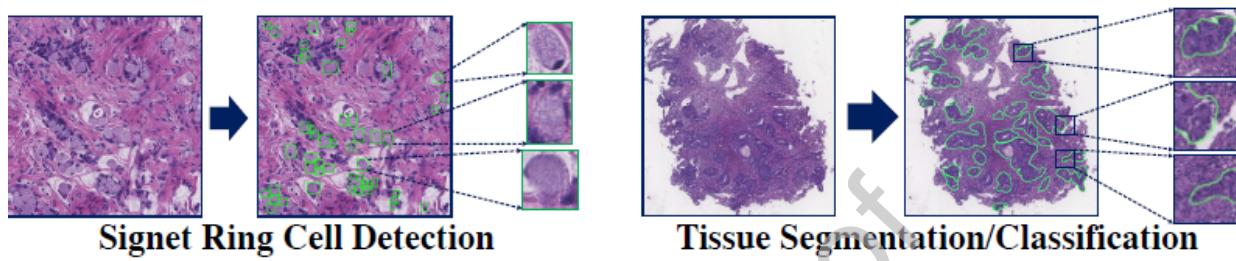
Received date: 2 October 2021  
Revised date: 8 April 2022  
Accepted date: 20 May 2022

Please cite this article as: Qian Da , Xiaodi Huang , Zhongyu Li , Yanfei Zuo, Chenbin Zhang, Jingxin Liu, Wen Chen, Jiahui Li, Dou Xu, Zhiqiang Hu, Hongmei Yi, Yan Guo, Zhe Wang, Ling Chen, Li Zhang, Xianying He, Xiaofan Zhang, Ke Mei, Chuang Zhu., Weizeng Lu, Linlin Shen, Jun Shi, Jun Li, Sreehari S, Ganapathy Krishnamurthi, Jiangcheng Yang, Tiancheng Lin, Qingyu Song, Xuechen Liu, Simon Graham, Raja Muhammad Saad Bashir, Canqian Yang, Shaofei Qin, Xinmei Tian, Baocai Yin, Jie Zhao, Dimitris N. Metaxas, Hongsheng Li, Chaofu Wang, Shaoting Zhang, DigestPath: a Benchmark Dataset with Challenge Review for the Pathological Detection and Segmentation of Digestive-System, *Medical Image Analysis* (2022), doi: <https://doi.org/10.1016/j.media.2022.102485>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**GRAPHICAL ABSTRACT**

	Signet Ring Cell Detection		Tissue Segmentation/Classification	
	Positive	Negative	Malignant	Benign
Training Images	77	383	250	410
Testing Images	27	100	90	122
Total	104	583	340	532
Average Image Size	2000 × 2000		5000 × 5000	
Magnification	X40, 0.258um/pixel		X20, 0.475um/pixel	



## HIGHLIGHTS

- We organized challenges for the digestive system on MICCAI.
- We released two well-annotated datasets: cell detection and tissue segmentation.
- We report the top-performing methods and results of two challenge tasks.



Contents lists available at ScienceDirect

## Medical Image Analysis

journal homepage: [www.elsevier.com/locate/media](http://www.elsevier.com/locate/media)

## DigestPath: a Benchmark Dataset with Challenge Review for the Pathological Detection and Segmentation of Digestive-System

Qian Da<sup>a,\*</sup>, Xiaodi Huang<sup>b,\*</sup>, Zhongyu Li<sup>c,\*</sup>, Yanfei Zuo<sup>d,\*</sup>, Chenbin Zhang<sup>b</sup>, Jingxin Liu<sup>d</sup>, Wen Chen<sup>b</sup>, Jiahui Li<sup>b</sup>, Dou Xu<sup>c</sup>, Zhiqiang Hu<sup>b</sup>, Hongmei Yi<sup>a</sup>, Yan Guo<sup>d</sup>, Zhe Wang<sup>e</sup>, Ling Chen<sup>e</sup>, Li Zhang<sup>f</sup>, Xianying He<sup>v,w</sup>, Xiaofan Zhang<sup>n,u</sup>, Ke Mei<sup>g</sup>, Chuang Zhu<sup>g</sup>, Weizeng Lu<sup>h</sup>, Linlin Shen<sup>h</sup>, Jun Shi<sup>i</sup>, Jun Li<sup>i</sup>, Sreehari S<sup>j</sup>, Ganapathy Krishnamurthi<sup>j</sup>, Jiangcheng Yang<sup>n</sup>, Tiancheng Lin<sup>n,o</sup>, Qingyu Song<sup>m</sup>, Xuechen Liu<sup>m</sup>, Simon Graham<sup>k,l</sup>, Raja Muhammad Saad Bashir<sup>l</sup>, Canqian Yang<sup>n,o</sup>, Shaofei Qin<sup>n,o</sup>, Xinmei Tian<sup>p</sup>, Baocai Yin<sup>q</sup>, Jie Zhao<sup>v,w</sup>, Dimitris N. Metaxas<sup>l</sup>, Hongsheng Li<sup>r,s</sup>, Chaofu Wang<sup>a,\*\*</sup>, Shaoting Zhang<sup>a,u,\*\*</sup>

<sup>a</sup>Department of Pathology, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

<sup>b</sup>SenseTime Research, Shanghai, China

<sup>c</sup>School of Software Engineering, Xi'an Jiaotong University, Xi'an, China

<sup>d</sup>Shanghai Histo Pathology Diagnostic Center, Shanghai, China

<sup>e</sup>Department of Pathology, Xijing Hospital, Fourth Military Medical University, Xi'an, China

<sup>f</sup>Shanghai Songjiang District Central Hospital, Shanghai, China

<sup>g</sup>Center for Data Science, Beijing University of Posts and Telecommunications, Beijing, China

<sup>h</sup>Computer Vision Institute, Shenzhen University, Shenzhen, China

<sup>i</sup>Hefei University of Technology, Hefei, China

<sup>j</sup>Indian Institute of Technology Madras, Chennai, India

<sup>k</sup>Mathematics for Real-World Systems Centre for Doctoral Training, University of Warwick, UK

<sup>l</sup>Department of Computer Science, University of Warwick, UK

<sup>m</sup>Real Doctor AI Research Centre, Zhejiang University, Hangzhou, China

<sup>n</sup>Shanghai Jiao Tong University, Shanghai, China

<sup>o</sup>Shanghai Institute for Advanced Communication and Data Science, Shanghai, China

<sup>p</sup>University of Science and Technology of China, Hefei, China

<sup>q</sup>iFLYTEK, Hefei, China

<sup>r</sup>Multimedia Laboratory, The Chinese University of Hong Kong, Hong Kong, China

<sup>s</sup>Centre for Perceptual and Interactive Intelligence (CPII) Ltd, Hong Kong, China

<sup>t</sup>Department of Computer Science, Rutgers University, Piscataway, NJ, USA

<sup>u</sup>Shanghai Artificial Intelligence Laboratory, Shanghai, China

<sup>v</sup>National Telemedicine Center of China, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

<sup>w</sup>National Engineering Laboratory for Internet Medical Systems and Applications, Zhengzhou, China

## ARTICLE INFO

### Article history:

2000 MSC: 41A05, 41A10, 65D05, 65D17

**Keywords:** Benchmark Dataset, Tissue Segmentation, Cell Detection, Digestive System Cancer, Grand Challenge

## ABSTRACT

Examination of pathological images is the golden standard for diagnosing and screening many kinds of cancers. Multiple datasets, benchmarks, and challenges have been released in recent years, resulting in significant improvements in computer-aided diagnosis (CAD) of related diseases. However, few existing works focus on the digestive system. We released two well-annotated benchmark datasets and organized challenges for the digestive-system pathological cell detection and tissue segmentation, in conjunction with the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). This paper first introduces the two released datasets, i.e., signet ring cell detection and colonoscopy tissue segmentation, with the descriptions of data collection, annotation, and potential uses. We also report the set-up, evaluation metrics, and top-performing methods and results of two challenge tasks for cell detection and tissue segmentation. In particular, the challenge received 234 effective submissions from 32 participating teams, where top-performing teams developed advancing approaches and tools for the CAD of digestive pathology. To the best of our knowledge, these are the first released publicly available datasets with corresponding

challenges for the digestive-system pathological detection and segmentation. The related datasets and results provide new opportunities for the research and application of digestive pathology.

© 2022 Elsevier B. V. All rights reserved.

## 1. Introduction

Gastrointestinal (GI) cancer is a term for the group of cancers that affect the GI tract, including stomach, small intestine, large intestine, and GI-related organs of liver, pancreas, rectum, etc. GI cancer has the highest incidence rates and is four of the top ten cancer types for estimated incidence in both men and women worldwide (Bray et al., 2018).

This paper addresses two GI cancer types of signet-ring cell carcinoma (SRCC) and colorectal cancer (CRC). SRCC is a histopathologic subtype of GI cancer, which defines that more than 50% of the cells containing abundant intracellular mucin (Hamilton and Aaltonen, 2000). It is a highly malignant adenocarcinoma commonly found in the stomach and colon, accounting for 8 – 30% and nearly 1% of gastric and colorectal cancers (Pernot et al., 2015; Messerini et al., 1995). CRC begins in the colon, and the rectum is the third and fourth most common cancer in the United States and the United Kingdom, respectively (Mármol et al., 2017; Grabovac et al., 2020). Colorectal cancer is the fourth most common cancer, accounting for 1.79% of total deaths in 2017 in China (Yin et al., 2019).

In clinical diagnosis, most digestive system cancers require endoscopy, followed by biopsy of suspicious tissue. The pathological examination is considered the golden standard in determining the presence and nature of varieties of cancers. Numerous reports have identified that SRCC has a poor prognosis and survival (Kepil et al., 2019; Belli et al., 2014). Therefore, early detection and diagnosis are essential. While the diagnosis of CRC is made by endoscopic biopsy for symptomatic patients, histopathology is considered the best prognostic marker for CRC (Treanor and Quirke, 2007).

Benefiting from advanced microscopy imaging techniques, digital pathology has become increasingly popular in clinical uses (Al-Janabi et al., 2012), which allows remote examination of high-resolution whole-slide images (WSI) with fine-grained details from tissues to cells. However, manual examination and analysis of WSIs is still a time-consuming task for most pathologists since the WSI can be up to the size of  $100,000 \times 100,000$  pixels. Besides, board-certified pathologists are usually in shortage worldwide, resulting in labor-intensive daily work to examine many WSIs. These challenges prevent accurate diagnosing and screening from being widely adopted in developing countries/regions with not enough pathologists but many patients.

In recent years, deep learning technology dominates the research fields of computer vision and machine learning based on large-scale and well-annotated datasets. Frontier deep neural networks can successfully solve practical problems such as image classification, detection, segmentation, etc., achieving comparable and even better performance with human experts (LeCun et al., 2015; Wen et al., 2016). Computer-aided diagnosis (CAD) has been widely investigated and applied for medical image analysis (Litjens et al., 2017; Shen et al., 2017; Li et al., 2018b) with the rapid expansion of deep learning. For digital pathology, multiple datasets and challenges are released for diverse diagnosing and exploration tasks, including directly screening cancers from WSIs, classifying WSIs with benign/malignant or specific sub-types, the segmentation of cells/nucleus, the classification of detected cells, etc (Bejnordi et al., 2017; Veta et al., 2019; Naylor et al., 2018). Accordingly, a significant number of CAD methods based on deep neural networks are developed, including fully supervised (Bejnordi et al., 2017), weakly-supervised (Campanella

\*Equal Contribution

\*\*Corresponding Authors: Shaoting Zhang (zhangshaoting@pjlab.org.cn), Chaofu Wang (wangchaofu@126.com).

et al., 2019), semi-supervised (Shi et al., 2020), and unsupervised methods (Hou et al., 2019). These methods can assist pathologists in diagnosing and screening cancers in the breast, lymph node, lung, blood, and so on (Peikari et al., 2017; Bandi et al., 2018; Sirinukunwattana et al., 2017).

Although extensive efforts have been made for pathological CAD, few current works focus on examining cancers in the digestive system (Yu et al., 2021). Similar to the challenges mentioned above, examination of digestive-system cancers also lacks pathologists but with many patients, especially in developing countries/regions. On the other side, for examining specific types of cancer, the corresponding large-scale and well-annotated datasets are prerequisites to support the training of robust and accurate deep learning models. Therefore, new datasets and methods are urgent for examining digestive-system cancers.

To this end, we address two typical challenging tasks in the pathological CAD of the digestive-system, i.e., signet ring cell detection and colonoscopy tissue segmentation. In particular, signet-ring cell carcinoma and colon cancer are deadly and common cancers. This paper first introduces two new large-scale datasets for the above two tasks, respectively, well-annotated by broad-certified pathologists. Our challenge, named DigestPath, is composed of 20,000 annotated signet-ring cells with bounding boxes and 1,000 pixel-level segmented colonoscopy tissue slices from in total of 690 patients. Furthermore, based on the two released datasets, we present details for the grand challenge on Digestive-System Pathological Detection and Segmentation in conjunction with MICCAI. The detailed description includes the challenge organization, evaluation, top-performance methodologies for the two tasks. The challenge attracts 32 teams from all over the world with 234 effective submissions, where the newly developed methods have greatly improved the CAD of related cancers. DigestPath is the first released grand challenge for the digestive-system cancers of signet-ring cell detection and colonoscopy tissue segmentation to the best of our knowledge. We hope that the datasets and algorithms provided in this challenge will be helpful in the pathological CAD research community.

The remaining paper is organized as follows: Section 2 briefly review relevant frontiers of datasets, challenges, and methods for the pathological CAD. Section 3 presents the introduction of our released two public available datasets, i.e., signet-ring cell detection and colonoscopy tissue segmentation. The organization and evaluation of the “Digestive-System Pathological Detection and Segmentation” grand challenge is given in Section 4, followed by detail descriptions of the top-performing competing solutions in Section 5. Afterward, Section 6 discusses challenges, opportunities, potential use cases from this challenge, and finally, the conclusion.

## 2. Review of Computer-aided Diagnosis of Pathological Images

Recently, many works focus on the computational analytics of pathological images, where large-scale benchmark datasets, competitive challenges, and varieties of methods have been released to boost the clinical diagnosis. Here, we briefly review recent advances in datasets, challenges, and methods related to pathological CAD.

### 2.1. Datasets and Challenges

Datasets with annotations play critical roles in the analytics of medical images. There are many existing public datasets available in recent years for the pathological images collected from multiple tissues and organs, with the image annotations from the whole-slide level to the cell level. We summarize the currently available pathological image datasets and corresponding challenges with related information in Table. 1, including provided data size, image modalities, related challenge tasks, and diseases. According to Table. 1, existing datasets focus on analyzing large-sized whole-slide images to image patches and cell images, with challenging classification, detection, segmentation, and registration tasks. However, the existing datasets haven't demonstrated the diagnosis of varieties of diseases, which mainly focus on diagnosing breast cancer, lung cancer, and lymph node. Few works have been addressed

Name of datasets	Number of samples	Image modalities	Challenge tasks	Related Diseases
GlaS (Sirinukunwattana et al., 2017)	330	H&E images	Gland segmentation	Colorectal cancer
TUPAC16 (Veta et al., 2019)	1076	Whole-slide images	Tumor proliferation assessment, mitosis detection	Breast cancer
CAMELYON16 (Bejnordi et al., 2017)	270	Whole-slide images	Detection and classification of breast cancer metastases	Breast cancer metastases
CAMELYON17 (Bandi et al., 2018)	1000	Whole-slide images	Detection and classification of breast cancer metastases	Breast cancer metastases
Kaggle 2018 Data Science Bowl (Caicedo et al., 2019)	670	Nuclei images	Nucleus detection	Variety of cell types, modalities
BACH (Aresta et al., 2019)	400 microscopy images 30 whole-slide images	Microscopy images Whole-slide images	Classification Pixel-wise Labeling	Breast cancer
MoNuSeg (Kumar et al., 2019)	29,000	Tissue images with nuclear boundary	Nuclear segmentation	Multiple organs
LYON19 (Swiderska-Chadaj et al., 2019)	441	Region of interests	Lymphocyte detection	Lymphocyte
C-NMC 2019 (Gupta et al., 2017)	15114	Cell images	Cell classification	B-ALL white blood cancer
BreastPathQ (Peikari et al., 2017)	3700	Image patches	Cancer cellularity scoring	Breast cancer
ANHIR (Borovec et al., 2018)	50	Whole-slide images	Non-rigid registration	Lesions, lung-lobes, mammary-glands
ACDC-LUNGHP (Li et al., 2018a)	200	Whole-slide images	Detecting and classifying of lung cancer	Lung cancer
PatchCamelyon (Veeling et al., 2018)	327,680	Color images (96 x 96px)	Classification of metastatic tissue	Lymph node
Gleason 2019 (Nir et al., 2018; Karimi et al., 2019)	331	Tissue micro-array	Gleason grading	Prostate cancer
PAIP2019 (Kim et al., 2021)	100	Whole-slide images	Liver cancer segmentation viable tumor burden estimation	Liver cancer
LYSTO (Ciompi et al., 2019)	14,000	Image patches	Assessment of lymphocytes	Lymphocytes
ECDP (Conde-Sousa et al., 2021)	360	Whole-slide image	Identification of HER2	Breast cancer

Table 1: Summary of recently released pathological image datasets with their corresponding challenges.

for diseases that consider pathology the golden standard, such as the digestive system (including stomach cancer, esophageal cancer, colon cancer, etc.). Accordingly, more related datasets and challenges should be released and investigated for the CAD of pathological digestive-system cancers.

## 2.2. Review of Methods for Pathological CAD

Based on the above-released datasets and challenges, many methods have been proposed to analyze pathological images in recent years. Related practices have been covered research tasks of classification, retrieval, segmentation, detection, with the image annotations of cells, image patches, and whole-slide images. Primarily based on the well-developed deep learning techniques in recent years, methods for pathological image analysis have achieved many successes. Some methods can achieve comparable and even better performance with board-certified pathologists. As our released datasets and challenges mainly focus on the signet ring cell detection and colonoscopy tissue segmentation/classification, we mostly review representative methods and review articles on cell detection, tissue classification/segmentation, and WSI diagnosis and analysis.

**Cell Detection:** Cell/Nucleus detection plays a critical role in describing the molecular, morphological information, providing quantitative analytical results to support clinical diagnosis. Before the prevalence of deep learning, many traditional methods have been investigated to detect cell/nucleus. For cell detection, traditional representative methods mainly include Distance Transform (Maurer et al., 2003), Morphology Operation (Soille, 2013), LoG Filtering (Byun et al., 2006; Zhang et al., 2015a), Maximally Stable Extremal Region (MSER) (Lu et al., 2013), Hough Transform (Lee and Street, 2003), Radial Symmetry Transform (Veta et al., 2011). These methods have been applied to detect many types of cell/nuclei, such as breast cancer nuclei, RNAi fluorescence cells, pap smear images. Xing et al. (Xing and Yang, 2016) provide a comprehensive review for the nucleus/cell detection in digital pathology and microscopy images. In recent years, with the ever-increasing techniques of deep neural networks, the study of cell detection has achieved great successes through the learning of provided datasets with specific tasks. Representation methods mainly include typical deep convolutional neural networks and their variants for pixel-level classification (e.g., AlexNet (Krizhevsky et al.,

2012), VGGNet (Simonyan and Zisserman, 2014), GoogLeNet (Szegedy et al., 2015), ResNets (He et al., 2016)), object detection networks (e.g., R-CNN (Girshick et al., 2014), Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2015), Mask R-CNN (He et al., 2017)), and segmentation networks (e.g., FCN (Long et al., 2015), U-Net (Ronneberger et al., 2015)). The deep learning-based methods have been widely implemented to detect mitosis in breast cancer (Chen et al., 2016; Li et al., 2019a), lung cancer (Qu et al., 2020), neuron cells from fluorescence (Falk et al., 2019), 3D glial cells (Yang et al., 2016), etc.

**Tissue Classification and Segmentation:** With the advent of imaging techniques, digital scanners can obtain high-throughput histopathological images for accurate clinical diagnosis. Due to the large size of WSIs (e.g., more than  $100000 \times 100000$  pixels), many research works start with the analysis of image patches extracted from WSI images. Traditional methods employ hand-crafted features, such as SIFT (Lowe, 2004), Color Histogram (Dalal and Triggs, 2005), tackling the classification and retrieval for the histopathological image pathes. For example, Zhang et al. (Zhang et al., 2014b,a) first introduce the learning to hashing to effectively index and retrieve large-scale histopathological images and cells (Shi et al., 2016; Wang et al., 2017; Zhang et al., 2015b). With the multiple released WSI image datasets and challenges, methods have been developed to tackle the WSI images through deep neural networks. This kind of method can either cut WSI with small image patches, train deep models for image patches and compose segmentation results for integral WSIs (Bejnordi et al., 2017; Bandi et al., 2018), or directly segment lesions from WSIs.

**WSI Diagnosis and Analysis:** In addition to the cell-level and tissue-level analysis, pathologists usually need a diagnostic conclusion at the WSI level. Due to the requirements of gigapixel manual annotation, weakly supervised deep learning methods are commonly used to diagnose WSI-level to reduce the labor-intensive manual annotation. For example, Lu et al. (Lu et al., 2021) developed clustering-constrained-attention multiple-instance learning (CLAM), which extends attention-based multiple-instance aggregation to general multi-class weakly supervised WSI classification without requiring any pixel-level annotation, ROI extraction, or sampling. Campanella et al. (Campanella et al., 2019) presented a multiple instance learning-based deep learning system that uses only the reported diagnoses as labels for training. The proposed method was tested on 44,732 WSIs, with the cancer types of prostate cancer, basal cell carcinoma, and breast cancer metastases. Moreover, Schmauch et al. (Schmauch et al., 2020) proposed HE2RNA, a deep learning algorithm to predict RNA-seq expression based on WSIs, without expert annotation. In addition to weakly supervised deep learning methods (Shao et al., 2021; Li et al., 2021b), self-supervised (Li et al., 2021a), semi-supervised (Shi et al., 2020), and unsupervised learning (Muhammad et al., 2019; Chen et al., 2021) methods are also commonly used for the diagnosis and analysis at WSI-level.

According to the above review of datasets, challenges, and methods, there are mainly three problems that need to be considered for further analysis of pathological images. Firstly, existing methods rely heavily on the corresponding datasets with well annotations under the supervised framework. However, there are no related datasets for many specific tasks that can be used for pathological CAD. Secondly, the style of pathological images can be quite different for the same datasets, since the datasets are generally collected from multiple medical centers or hospitals with diverse imaging conditions. Thirdly, a few of the current datasets, challenges, and methods focus on the pathological CAD of the digestive-system, such as the signet ring cell detection and colonoscopy tissue segmentation, which play critical roles in clinical diagnosis. Therefore, new benchmark datasets, challenges, and methods should be released and developed to boost pathological CAD study further.

### 3. Digestive-system Pathological Detection and Segmentation Datasets

The DigestPath releases two datasets for diagnosing digestive-system pathological images, i.e., the signet ring cell detection dataset and the colonoscopy tissue segmentation and classification dataset. Both were collected from four medical institutions in

	Signet Ring Cell Detection		Tissue Segmentation/Classification	
	Positive	Negative	Malignant	Benign
Training Images	77	383	250	410
Testing Images	27	100	90	122
Total	104	583	340	532
Average Image Size	$2000 \times 2000$		$5000 \times 5000$	
Magnification	X40, 0.258um/pixel		X20, 0.475um/pixel	

Table 2: The composition of DigestPath datasets: signet ring cell detection dataset, and colonoscopy tissue segmentation dataset.

China, i.e., Ruijin Hospital, Xijing Hospital, Shanghai Songjiang District Central Hospital, and Histo Pathology Diagnostic Center. These medical institutions are located in southeast and northwest China, which shown patients from different areas. In these two datasets, patients ranged in age from 20 to 70 years and the male to female ratio was about 1:1. All slides were acquired from 2017 - 2019 with hematoxylin and eosin (HE) stains and scanned using the KFBIO FK-Pro-120 slide scanner. The annotation team for DigestPath challenges has four experienced pathologists from the three hospitals and two expert pathologists from the Histo Pathology Diagnostic Center. All the datasets were annotated by experienced pathologists (with over five years of experience in gastric and intestinal (GI) pathology). Then these datasets were reviewed by expert pathologists (over ten years experience in GI pathology). For the disagreement resolution protocols between pathologists, the difficult/disagreement samples were decided by two expert pathologists together. To ensure an impartial evaluation, only the training set of each task was released to the participants. Here, we introduce the two datasets respectively for more details.

### 3.1. Signet Ring Cell Detection Dataset

Early detection of signet ring cells leads to a massive improvement of SRCC patients' survival rate (Li *et al.*, 2019c; Da *et al.*, 2022). Particularly, signet ring cell is a kind of specific GI cancer, a cell with a large vacuole, which has a different appearance from other GI cell types. In clinical diagnosis, recognizing signet ring cells is an important step to support the final conclusion of cancer screening and grading. However, manual checking by pathologists in WSIs is time-consuming and labor-intensive. In this challenge, we release a refined annotated signet-ring cell carcinoma dataset to support the research for pathological CAD of signet ring cell detection. To the best of our knowledge, this is the first public database for studying the problem of signet ring cell detection.

A total of 155 patients' WSIs were collected from two organs of gastric mucosa and intestine. Tumor cells were poorly differentiated, staging from TIS to T1. All whole slide images were stained by hematoxylin and eosin and scanned at X40 objective magnification (0.248 $\mu\text{m}/\text{pixel}$ ). We selected 682 images of size  $2000 \times 2000$  pixels from WSIs to reduce the computational burden and the annotation complexity. The dataset consists of 104 positive images with signet ring cells and 583 negative images without them. Four experienced pathologists perform the annotation on positive images. Each signet ring cell is labeled by a rectangle bounding box tightly surrounding the cell. There are 14,859 cells annotated, and each labeled cell is indeed a signet ring cell. However, positive images may have signet ring cells miss labeling. The negative images have no signet ring but could contain other types of tumor cells. We randomly split the dataset into training and testing. The composition of the dataset is shown in Table 2. In addition, we also provide samples of annotated full images and parts of an image with cell bounding boxes in Fig. 1.

### 3.2. Colonoscopy Tissue Segmentation Dataset

Colonoscopy pathology examination can find early-stage colon tumors from small tissue slices. Pathologists need to daily examine hundreds of tissue slices, which is a time-consuming and exhausting work. Especially, colonoscopy tissue includes different

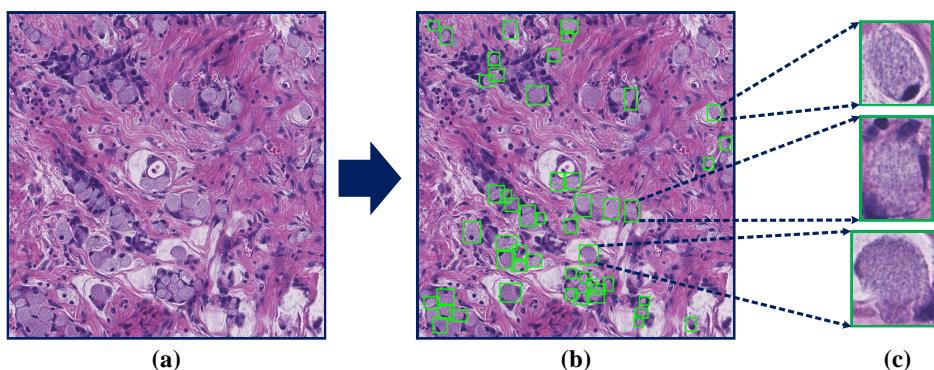


Fig. 1: An example image with annotated bounding boxes of signet ring cells: (a) a histopathological image patch; (b) the image with partially annotated bounding boxes (in green) of signet ring cells; (c) three selected signet ring cells.

stages and subtypes. For the developmental stages, tumors were classified as inflammatory, low-grade intraepithelial neoplasia, high-grade intraepithelial neoplasia, and carcinoma. For the subtypes, tumors were classified as papillary adenocarcinoma, mucinous adenocarcinoma, poorly cohesive carcinoma, and signet ring cell carcinoma. In clinical diagnosis, accurately outlining and quantifying malignant regions is an important step for pathologists, which can quickly localize and identify suspicious regions to improve diagnostic efficiency. Besides, the proportion of malignant regions is also an important metric for the final conclusion. Here, we release a publicly available dataset on colonoscopy tissue segmentation and screening, aiming at automatic lesion segmentation and classification of the whole tissue (benign vs. malignant).

The criteria for distinguishing between malignant (positive) and benign (negative) is challenging. According to the WHO classification of tumours in the digestive system, we regard the following diseases as malignant lesions: high-grade intraepithelial neoplasia and adenocarcinoma, including papillary adenocarcinoma, mucinous adenocarcinoma, poorly cohesive carcinoma, and signet ring cell carcinoma. To increase the challenge's difficulty, we add six low-grade intraepithelial neoplasia samples into the negative set. Notice that in clinical diagnosis, pathologists would face more complicated situations. We provide an example of the colonoscopy tissues with segmented annotations and magnified lesion regions in Fig. 2.

A total of 872 tissue sub-images from 476 patients were collected, with an average size of  $5000 \times 5000$  pixels were extracted from both benign and malignant areas to cover as much variety of tissue appearance as possible. After slide scanning, the WSIs were subsequently re-scaled to  $\times 20$  magnification with a pixel resolution of  $0.475 \mu\text{m}$ . Two expert pathologists review the sub-images to ensure no uncertain tissues between benign and malignant. Then, the malignant images were manually annotated at pixel-level by four experienced pathologists, followed by expert pathologist examination. The dataset was separated into training and testing of 660 and 212 images. We also ensure that the images in the training set would not be from the same WSIs as those in the testing set. The details of dataset composition are provided in Table. 2.

#### 4. Challenge Introduction

The “Digestive-System Pathological Detection and Segmentation Challenge”, abbreviated as “DigestPath”, is in conjunction with MICCAI. The challenge page can be found in <https://digestpath2019.grand-challenge.org/>. After the challenge, we keep maintaining an online platform for new submissions and keep the leaderboard of DigestPath.

Mainly, DigestPath challenge includes two tasks, i.e., signet ring cell detection, and colonoscopy tissue segmentation and classification. The organizing committee consists of researchers from the medical research institutes, industries, hospitals, and

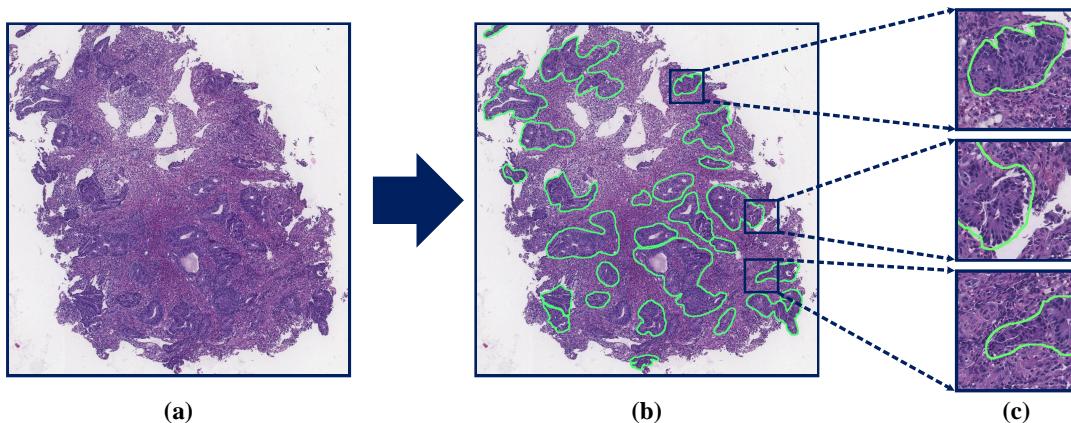


Fig. 2: Illustrated sample of a full image tissue and its segmented annotations of lesion regions: (a) a full histopathological image tissue; (b) segmented annotations of colonoscopy lesion tissues; (c) three magnified lesion regions.

universities, supporting the challenge in both clinical and technical sides. The challenge was also hosted on the Grand Challenges in Biomedical Imaging Platform, one of the most popular platforms for medical and biomedical image analytics challenges. Here, we introduce the organization of the two tasks, mainly including the evaluation metrics and the submission rules.

#### 4.1. Evaluation Metrics

We provided the evaluation metrics for the two tasks in the DigestPath web page for fair comparison.

**Evaluation of Signet Ring Cell Detection:** Each team's submission will be ranked by 3 evaluation metrics separately first: 1) Instance-level Recall; 2) Normal Region False Positives; 3) Free-response Receiver Operating Characteristic (FROC). These metrics can reflect the capability of deep models in the recognition and differentiation of signet ring cells. We consider these metrics as equal in the final ranking. The final rank of each team is the average of each evaluation metrics' rank. Computational details are introduced in the following:

For the Instance-level Recall, there exist overcrowded regions of the signet ring cell, with various appearances. It is impossible to get perfect annotation as shown in Fig. 1. In addition to the labeled cells, other cells may be signet ring cells but are unlabeled in an overcrowded region. Pathologists can only guarantee that the labeled cells are signet ring cells, while the unlabeled cells may also be. Thus, we seriously consider the instance-level recall in this problem when precision is more than 20%. There are two types of images in test data. Positive images contain signet ring cells, where negative images don't. The instance-level recall is the sum of matched ground-truth boxes divided by the number of ground truth boxes, ranging from 0 to 1.

The Normal Region False Positives is the average number of false-positive predictions in the negative images. FPs is written as  $\text{Max}(100 - \text{Normal region false positives}, 0)$ . Besides the instance-level recall, the metric of normal region false positives can further reflect the capability of the deep model in the recognition and differentiation of signet ring cells.

FROC is the average recall of different versions of predictions. We can get various versions of the prediction array by adjusting the confidence threshold. When normal region false positives are 1, 2, 4, 8, 16, 32, FROC can be computed by the average recall of predictions.

We also provide metrics formulations:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \times 100\% \quad (1)$$

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \times 100\% \quad (2)$$

$$\text{FPs} = \max(100 - \text{Normal Region False Positives}, 0) \times 100\% \quad (3)$$

$$\text{FROC} = \text{Average(Recall)} \quad (4)$$

For the FPs, we search confidence threshold from 0.0 to 1.0 and get Recalls when Normal Region False Positives equal [1, 2, 4, 8, 16, 32]. Precision and Recall are calculated on only Positives Images. FPs are calculated on only Normal Regions. We reject submissions if the Precision is lower than 20%. Competitors were asked to write their prediction results to XML files for submission.

**Evaluation of Colonoscopy Tissue Segmentation and Classification:** Each team's submission is first ranked by the two evaluation metrics (Dice Similarity Coefficient (DSC) and Area under the curve (AUC)) separately. To assign each team's final rank, we weigh the two metrics equally and use the average of the two individual ranks. In the following, we introduce the details of the two metrics:

The DSC metric measures area overlap between segmentation results and annotations. DSC is computed by where A is the sets of foreground pixels in the annotation and B is the corresponding sets of foreground pixels in the segmentation result, respectively:

$$\text{DSC} = \frac{2|A \cap B|}{|A| + |B|} \times 100\% \quad (5)$$

AUC equals the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one (assuming “positive” ranks higher than “negative”). The AUC is formulated as follows:

$$\text{AUC} = \int_{x=0}^1 \text{TPR}(\text{FPR}^{-1}(x)) dx = P(X_1 > X_0) \quad (6)$$

where  $X_1$  is the score for a positive instance.  $X_0$  is the score for a negative instance. TPR means true-positive rate. FPR means false-positive rate. The mask files are .jpg formatted images. During the competition, we strongly advise participants to get a binary mask using threshold 128 to train their model. For the evaluation, we also use 128 as the threshold to get the binary mask of their predictions' mask, then compute DSC scores.

#### 4.2. Submission Criterion

For the submission, participants should containerize their algorithms with Docker<sup>1</sup> and submit these to organizers for evaluation. The test data will not be released to the public. The challenge organizers will run submitted methods on all test data. This guarantees that the test data remains secret and cannot be included in the training procedure. This workflow has proven successful in previous MICCAI challenges (e.g. IVDM3Seg Challenge (Zheng, Guoyan and Li, Shuo and Belavy, Daniel, 2020)). To help those

---

<sup>1</sup><https://www.docker.com/>

participants who have no experience with Docker, we also provided an easy-to-follow example<sup>2</sup> to containerize their algorithms and a video tutorial<sup>3</sup> for task2.

Besides the submitted methods, each top-ranking team are invited to submit a 4-page report using LNCS and an oral speech at MICCAI to describe their approach. After the competition, we also make the challenge publicly available for submission with the new platform in <https://digestpath2019.grand-challenge.org/>.

In addition, competitors could search the docker evaluation status for each submission after the evaluation is finished. A search result could be: 1) whether the docker is valid. If it is invalid, it will return a short error log; 2) If it is validated ( $DSC > 5\%$ ), it will only return that the submission is valid. No specific numeric result will be shown to the participants before the challenge deadline. After the deadline, the leaderboard will be released, containing the top-10 teams of each task. Other teams could search for their highest score by the SEARCH form.

## 5. Challenging Solutions

This section provides detailed solutions of top-5 ranked teams from the two challenging tasks. Particularly, we first summarize the leaderboard with the top-ranked performance of the two tasks, respectively. Then, detailed solutions are presented, as well as discussions and technical specifications.

### 5.1. Task-1: Signet Ring Cell Detection

This task focuses on detecting all signet ring cells, returning rectangle bounding boxes surrounding each cell on given pathological images from the WSI of stomach and intestine tissues. Participating teams employed and proposed novel methods in tackling this task through recent advances in deep neural networks.

To achieve the signet ring cell detection, there are mainly three challenges with this dataset: 1) because of the difficulty of manual annotation, pathologists may miss some signet ring cells. This dataset is noisy with its positive images not fully annotated. 2) The dataset (both positive and negative images) can contain other kinds of tumor cells, which may demonstrate similar morphology with the signet ring cells. 3) As the first public dataset for the signet ring cell detection, there is no other related database that can assist the training of the detection model. In considering these challenges, participating teams developed advanced methods, and achieved excellent performance in signet ring cell detection.

#### 5.1.1. Leaderboard of Task-1

We display the top-10 submission results for the Task-1 of signet ring cell detection. As shown in Table. 3, the results are ordered according to the average rank of Recall, FPNormal, and FROC metrics. The presented results demonstrate significant variance in the performance from top-1 to top-10. Under the metrics of Recall and FROC, the top-2 ranked teams achieved much better performance than others, i.e., 0.8774 and 0.7424 compared to 0.4732 and lower scores. Under the metric of FPNormal, the top-4 ranked teams achieved excellent performance than others, i.e., nearly 100 FPNormal compared to 90 and lower FPs. Under the 3 metrics, the top-10 ranked teams showed large gaps in signet ring cell detection results, i.e., Recall scores ranging from 0.8774 to 0.0796, FPNormal scores ranging from 100 to 14.98, and FROC ranging from 0.8774 to 0.0561. The diversities on the top-10 submission results indicate the challenging of the task and the variations of solutions.

Detailed solutions of the top-5 teams on the Task-1 are provided in the supplementary material.

<sup>2</sup><https://drive.google.com/file/d/1Gx3CnqRH1PnwaKWJ1A2uXW-UfL1MyOxj/view>

<sup>3</sup><https://drive.google.com/file/d/1OZPgfUMrlVVUUqbryQ5QTCKTHZxODfr/view>

Team	Recall	Recall Rank	FP\@normal	FP\@normal Rank	FROC	FROC Rank	Final Rank
zju_realdoctor	0.8774	1	100.00	1	0.8774	1	1
SJTU_MedicalCV	0.7424	2	99.80	2	0.7424	2	2
mirl_task1	0.4732	5	97.79	4	0.4599	3	3
szucv517	0.4493	6	99.21	3	0.4493	4	4
HFUTB906	0.2741	8	88.94	6	0.1999	5	5
DeepBlueAI	0.4991	3	14.98	10	0.1718	7	6
ForeverYoung	0.4350	7	17.00	8	0.1787	6	7
MCPRL_218	0.4783	4	15.20	9	0.1524	8	7
chaibrypil	0.1349	9	90.96	5	0.0827	9	8
Murasame	0.0796	10	73.90	7	0.0561	10	9

Table 3: The top-10 submission results of Task-1: signet ring cell detection.

Team	Approach	Ensemble	Loss	Augmentation	Patch Size	Patch Strategy	Unlabelled Cells	Data Unbalance	Other
zju_realdoctor	RetinaNet	Y	FocalLoss	TTA	512 × 512	RC	cascaded self-training	resampling strategy	modified NMS
SJTU_MedicalCV	Dcov RetinaNet Faster-RCNN Cascaded-RCNN	Y	GHM L1 loss FocalLoss	RF, RC, RS	800 × 800	sliding window	-	pos:neg=1:3	NMS
mirl	Mask RCNN ResNet-X FPN	N	BCE ECL	RR, RT, RS RF, CJ	512 × 512	OLP	-	hard-mining	-
szucv517	DSFD pyramidbox	Y	MultiBox	RC	640 × 640	sliding window	crop cells with background	-	SE-Resnet18
HFUTB906	YOLO v3-SPP	N	-	rescale, RF, pad	1024 × 1024	sliding window	-	-	-

Table 4: The top-5 solutions with key components of Task-1: signet ring cell detection. RF: Random Flip, RR: Random Rotation, RS: Random Scaling, RT: Random Translation, RC: Random Crop, CJ: Color Jitter, OLP: Overlapping Patch, BCE: Binary Cross Entropy with Logits, ECL: Cross-Entropy Loss, Y: Yes, N: No, - indicates not mentioned.

### 5.1.2. Results and Discussions

According to the introduction of the above top-5 solutions for the task of signet ring cell detection, we summarize their key components in Table. 4. As presented in Table. 4, the top-5 teams employed varieties of backbone network architectures, including RetinaNet (Lin et al., 2017b), Mask R-CNN (He et al., 2017), DSFD (Li et al., 2019b), YOLOv3 (Redmon and Farhadi, 2018), etc, where the top-2 ranked teams both employed RetinaNet (Lin et al., 2017b) as their backbone network. For the model ensemble, only the *mirl\_task1* (ranked 3) and HFUTB906 (ranked 5) haven't reported their ensemble strategy. Considering the tackling of image patches, three of the teams (SJTU\_MedicalCV, szucv517, HFUTB906) use the sliding window strategy. The other two teams (*zju\_realdoctor* and *mirl\_task1*) use the random crop and overlapping patch, respectively. All the top-5 solutions have employed multiple data augmentation strategies to improve the data size of training images, such as random flip, random rotation, random scaling, color jitter. Among the top-5 solutions, only the szucv517 reported the strategy in addressing the unlabelled cells, i.e., crop all cells with the background. Four teams reported the loss functions they employed, including Focal Loss, GHM, L1 Loss, Binary Cross-Entropy Loss, MultiBoxLoss, etc., consistent with their backbone network architectures. Instead of directly tackling original image patches, the five teams cropped patches from the size of  $512 \times 512$  to  $1024 \times 1024$ . Moreover, three teams reported the strategy for the data unbalance problem (i.e., *zju\_realdoctor*, *SJTU\_MedicalCV*, and *mirl\_task1*), where re-sampling, reset the ratio of positive and negative samples, and hard-mining strategies have been employed.

According to these solutions, we can find that the top-2 ranked teams (*zju\_realdoctor* and *SJTU\_MedicalCV*), especially the top-1 team (*zju\_realdoctor*), achieved much better performance compared with other top-ranked teams. Based on their implementation details, we summarize some strategies that may be successful in the detection task of signet ring cells. For the backbone network architectures, only the top-2 ranked teams employed RetinaNet (Lin et al., 2017b) with the loss function of FocalLoss. This one-stage detector is suitable for signet ring cell detection. The top-2 ranked teams both employed ensemble strategy, a commonly used strategy to improve performance. Both teams employed the Non-Maximum Suppression (NMS) strategy. However, for different purposes, the *zju\_realdoctor* employed NMS to generate new training labels. The *SJTU\_MedicalCV* employed the NMS to process the predicted bounding boxes. For the data augmentation, only the top-1 team reported in using the test time augmentation (TTA) strategy. Only the top-1 team considered unlabelled cells by cascaded self-training. In addition, resampling strategies for unbalanced positive and negative data and random cropping for the patch processing are also explicitly employed in the top-1 team. Given these diverse solutions, signet ring cell detection is still a challenging open problem. There is no single answer that can dominate this problem.

Additionally, we also present the detection results of signet ring cells of top-5 teams in Fig. 3 for the most intuitive comparison, with three randomly selected histopathological image patches. We present three randomly selected histopathological image patches with ground truth annotations of signet ring cell detection in Fig. 3, as well as the detection results of top-5 ranked methods. Compared with ground truth annotations and top-5 ranked results, the performance varies significantly by considering the detected signet ring cells. For example, the results of *zju\_realdoctor* and *SJTU\_MedicalCV* tend to recognize and detect more signet ring cells than ground truth annotation, while the other three methods (i.e., *mirl\_task1*, *szucv517*, and *HFUTB906*) tend to recognize and detect less signet ring cells. By referring to the evaluation metrics of the five methods, these results indicate that the top-2 methods can detect signet ring cells and easily recognize other types of cells as signet ring cells. In contrast, the other three methods cannot identify signet ring cells well. Overall, despite the current results achieved by these top-ranked teams, detecting signet ring cells is still considered a challenging task.

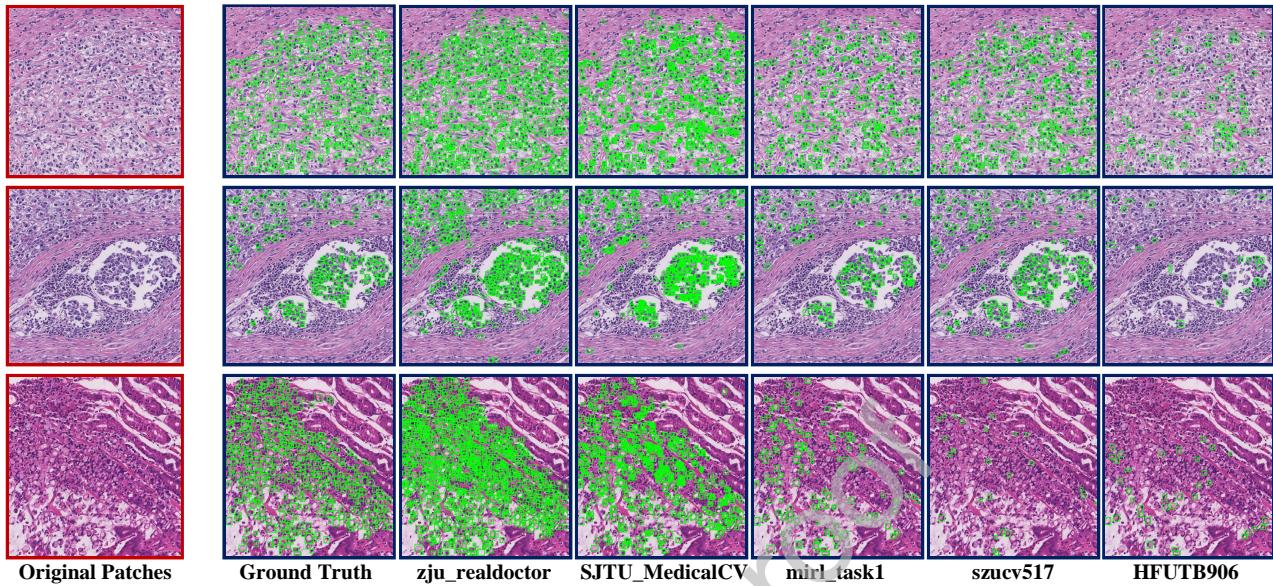


Fig. 3: Three randomly selected original image patches (in red frame) with the results of signet ring cells detection under the top-5 ranked methods, as well as ground truth annotations (in blue frame).

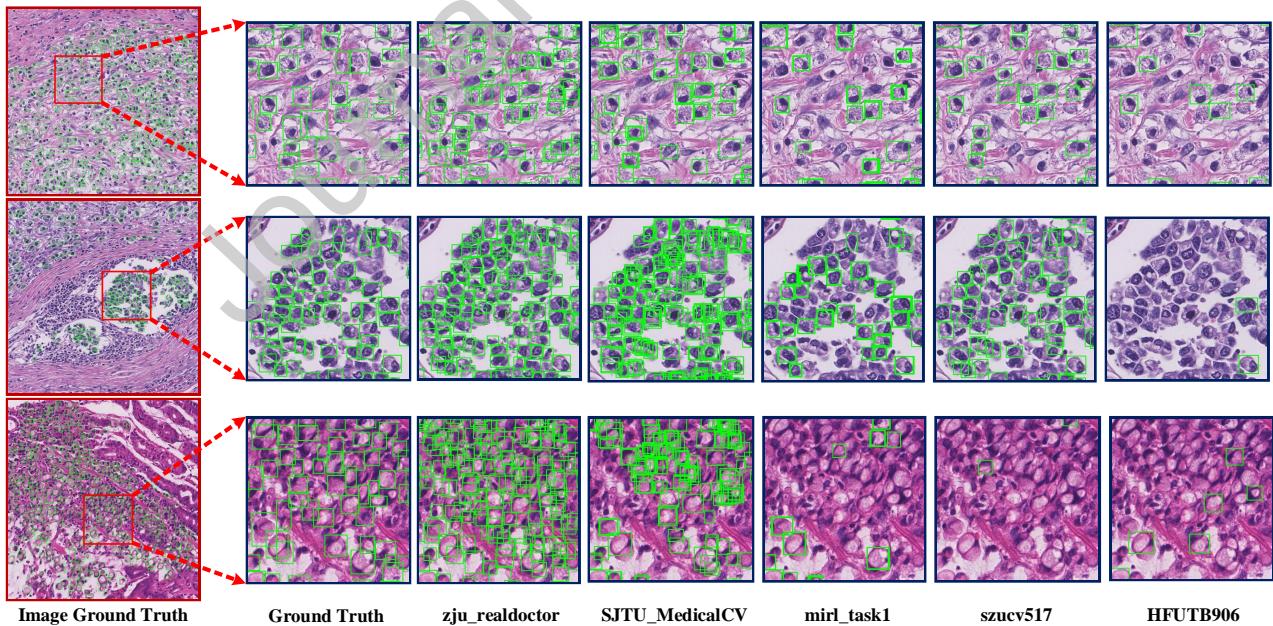


Fig. 4: Three randomly selected images with ground truth (in red frames) and the results of signet ring cells detection of local regions enlarged, under the top-5 ranked methods, as well as ground truth annotations (in blue frames)

Team	DSC	DSC Rank	AUC	AUC Rank	Final Rank
kuangkuang	0.8075	1	1.0000	1	1
zju_realdoctor	0.7789	5	1.0000	1	2
TIA_Lab	0.7878	3	0.9948	4	3
SJTU_MedicalCV	0.7928	2	0.9773	6	4
ustc_czw	0.7862	4	0.9784	5	5
chenpingjun	0.7197	8	0.9974	3	6
MCPRL_218	0.7397	7	0.9745	8	7
path_fitting	0.6794	10	0.9754	7	8
mirl_task2	0.7590	6	0.5164	13	9
Roselia	0.6920	9	0.8886	11	10

Table 5: The top-10 submission results of Task-2: colonoscopy tissue segmentation and classification.

### 5.2. Task-2: Colonoscopy Tissue Segmentation and Classification

This task focuses on automated colonoscopy tissue segmentation and screening, aiming at automatic lesion segmentation and classifying whole tissues (benign vs. malignant). Participating teams employed and proposed novel methods for this task based on frontiers of deep learning techniques.

In this task, there are mainly two challenges: 1) the dataset shows significant variations in terms of appearance because the data are collected from 4 medical centers, especially from several small centers in developing countries/regions. Image style differences can be an obstacle for the screening task; 2) the histopathological images are large, i.e., with the average size of 5,000×5,000 pixels. Some of them are too large to directly train deep neural networks. Accordingly, holding the challenge and releasing many expert-level annotations will attract much attention from the medical imaging community and substantially advance the research on automatic colonoscopy screening.

#### 5.2.1. Leaderboard of Task-2

Table. 5 presents the top-10 submission results of Task-2 of colonoscopy tissue segmentation and classification, which are ordered according to the average rank of the DSC and AUC metrics. In comparison with the leaderboard results of Task-1 in Table. 3, the top-ranked results of colonoscopy tissue segmentation and classification are more close. The top-1 team (*kuangkuang*) achieves the DSC of 0.8075, while the second-highest performance achieves a 0.7928 DSC score (*SJTU\_MedicalCV*) with only a 0.0147 difference. For the metric of AUC, both the *kuangkuang* team and the *zju\_realdoctor* team have achieved a 1.0000 score. The third-highest AUC score is 0.9974, achieved by the *chenpingjun* team. Moreover, the range from rank-1 to rank-10 performance is relatively close under DSC and AUC metrics.

Detailed solutions of the top-5 teams on the Task-2 are provided in the supplementary material.

#### 5.2.2. Results and Discussions

According to the introduction of the above top-5 solutions, we summarize their key components in Table. 6. As presented in Table. 6, the top-5 teams employed varieties of backbone network architectures, including DenseNet (Huang et al., 2017), ResNeXt (Xie et al., 2017), InceptionNet-V3 (Szegedy et al., 2016), U-Net (Ronneberger et al., 2015), EfficientNet (Tan and Le, 2019), etc., where the top-ranked teams employed quite different network architectures. For the model ensemble, only the *SJTU\_MedicalCV* has not employed an ensemble strategy. All the top-5 teams have reported their data augmentation strategy, including random flip, random rotation, random scaling, color jitter, etc. Since this task requires both segment and classify colonoscopy tissues, each team's commonly used loss functions include DiceLoss, FocalLoss, BCE with Logits Loss, Binary Cross-Entropy Loss, etc. All the top-5 solutions have employed the sliding window strategy in tackling image patches, where the

Team	Approach	Ensemble	Loss	Augmentation	Patch Size	Patch Strategy	Other Strategy
kuanguang	DenseNet161	Y	BCEWithLogitsLoss	random folds random brightness contrast grid distortion	1536, 1536	sliding window	3 stages
	ResNext10		CosineEmbeddingLoss				
	InceptionV3		DiceLoss				
	U-Net		FocalLoss				
	ResNet50		Lovasz-Softmax loss				
zju_realdoctor	U-Net	Y	DiceLoss	RC TTA	512, 512	sliding window	increase receptive field
			BceWithLogDiceLoss				
			FocalLoss				
			CrossEntropyLoss				
TIA_Lab	MILD-Net	Y	Categorical_crossentropy DiceLoss	RF RR RS random shear elastic transformation blur, gaussinan noise color augmentation TTA	448, 448	sliding window	focus image and context image as input
SJTU_MedicalCV	EfficientNet DeepLabV3	N	Binary_crossentropy	RF RS RR RC TTA	1024, 1024	sliding window	DSV scSE balance sample strategy
ustc_czw	PSP-Net	Y	-	RC rotation mirror color_jittering random_scale elastic	512, 512	sliding window	strategy for noise data

Table 6: The top-5 solutions with key components of task2: Colonoscopy Tissue Segmentation and Classification. RF: Random Flip, RR: Random Rotation, RS: Random Scaling, RT: Random Translation, RC: Random Crop, Y: yes, N: no- indicates not mentioned.

patch sizes are set from  $448 \times 448$  to  $1536 \times 1536$ . The top-5 solutions also employ other strategies to improve segmentation and classification performance, such as increasing the receptive field, decomposing the problem into three stages, using balance sampling strategies, and addressing noisy data.

Unlike the signet ring cell detection task, the top-ranked teams in colonoscopy tissue segmentation and classification task achieved comparable performance under two evaluation metrics. Four of the top-5 ranked teams employed an ensemble strategy to boost the performance. Unlike the other solution, the ensemble strategy of the top-1 team (kuanguang) is based on three different backbone networks for classification. (i.e., ResNext10 (Xie et al., 2017), InceptionV3 (Szegedy et al., 2016), and DenseNet161 (Huang et al., 2017)), which can achieve higher DSC and AUC scores. The top-1 team developed a three-stage pipeline to achieve the classification and segmentation, i.e., from judging a WSI with benign or malignant to the classification in patch level and the segmentation of positive patches. This coarse-to-fine pipeline is useful for colonoscopy tissue classification and segmentation. Besides, increasing receptive field and data augmentation strategies are commonly used in the top-ranked teams. Despite the comparable performance of top-5 ranked teams, their solutions are diverse in baseline network, loss functions, and segmentation and classification strategies. There is no single answer that can dominate this challenge.

Additionally, we also present the segmentation results of colonoscopy tissue of top-5 ranked teams in Task-2 in Fig. 5, as we do in Task-1 in Fig. 3. Compared with ground truth annotations and top-5 ranked results, we find that the top-ranked methods can well segment colonoscopy tissues at a coarse level, i.e., accurately localizing tissue areas. Generally, we cannot sufficiently identify tissue boundaries at the fine-grained level. Due to the significant variance of tissue appearance, the currently top-ranked methods tend to demonstrate two limitations: 1) wrongly identify parts of patches that are benign tissues; 2) cannot fully identify colonoscopy tissues with missing parts and pixels. Accordingly, segmentation of colonoscopy tissues is still challenging when

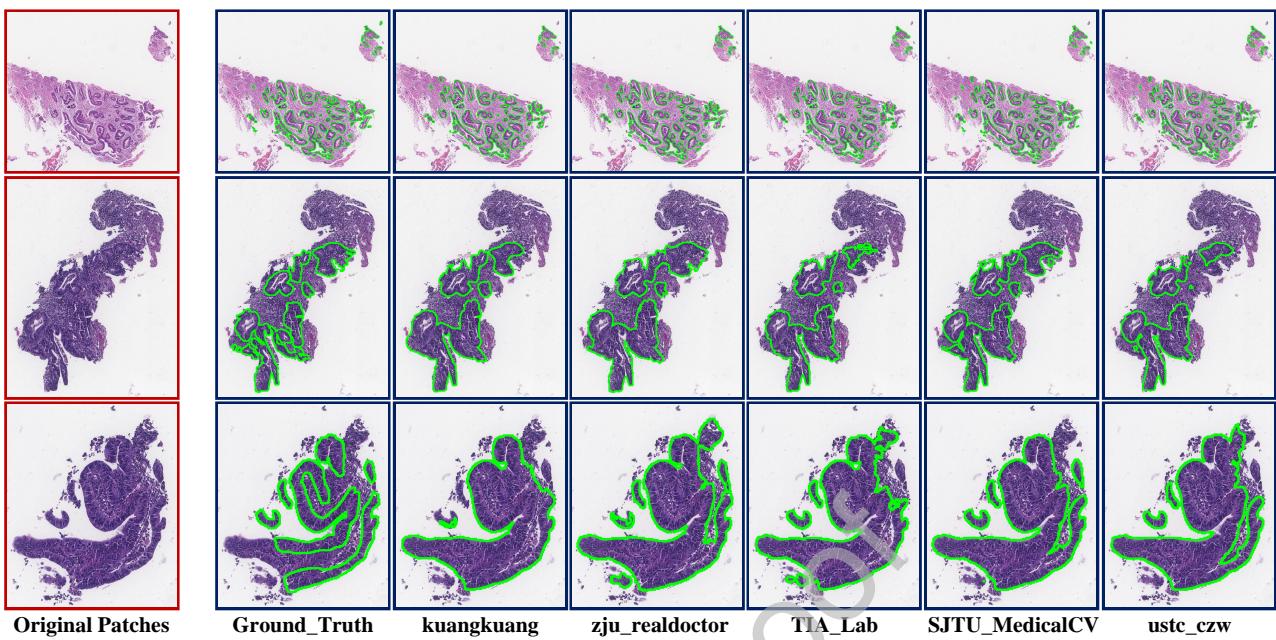


Fig. 5: Three randomly selected original image patches (in red frame) with the results of signet ring cells detection under the top-5 ranked methods, as well as ground truth annotations (in blue frame).

tackling histopathological images with significant variances.

## 6. Conclusion

This paper first presents two benchmark datasets for the computer-aided diagnosis of pathological images towards two representative tasks in digestive-system cancers, i.e., signet ring cell detection, and colonoscopy tissue segmentation and classification. The released two public datasets can indeed improve the research of pathological CAD of digestive-system cancers. We further introduce corresponding grand challenges based on the two datasets, including the challenge organization, evaluation, and detailed solutions. The challenge attracts 32 participating teams with 234 submissions. This challenge facilitates the development of various new methods for cell detection, tissue segmentation, and classification, which can be further explored for the application of other medical image analysis tasks.

## Acknowledgments

This study was supported by the funding of the Science and Technology Commission Shanghai Municipality (19511121400), the National Key Research and Development Project of China (2020YFC2004800), the Science and Technology Project for Innovation Ecosystem Construction of Zhengzhou National Supercomputing Center (201400210400), and the Central Guidance on Local Science and Technology Development Fund of Henan Province. The authors would like to thank the Intel® student Ambassador Program for AI, which provided the necessary computational resources on the Intel® AI DevCloud for running the WSI inference pipeline. The authors are also grateful to the NVIDIA GPU Grant Program for donating Titan-V GPU used in this research. The authors would like to thank Qi Duan, and Shuang Yang from SenseTime Research for their great help in challenging organizations. The authors are thankful to Xiao Mu, Bo Xu, Bin Gu, Zhaoyu Wang, Qiang Li from Shanghai Histo Pathology Diagnostic Center, Weiwei Sun, Minmin Gu, Caijuan Li, Yanqiong Zhang from Shanghai Songjiang District Central Hospital for their work in preparation of the data set.

## References

- Al-Janabi, S., Huisman, A., Van Diest, P.J., 2012. Digital pathology: current status and future perspectives. *Histopathology* 61, 1–9.
- Aresta, G., Araújo, T., Kwok, S., Chennamsetty, S.S., Safwan, M., Alex, V., Marami, B., Prastawa, M., Chan, M., Donovan, M., et al., 2019. Bach: Grand challenge on breast cancer histology images. *Medical Image Analysis*.
- Bandi, P., Geessink, O., Manson, Q., Van Dijk, M., Balkenhol, M., Hermsen, M., Bejnordi, B.E., Lee, B., Paeng, K., Zhong, A., et al., 2018. From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge. *IEEE Transactions on Medical Imaging* 38, 550–560.
- Bejnordi, B.E., Veta, M., Van Diest, P.J., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J.A., Hermsen, M., Manson, Q.F., Balkenhol, M., et al., 2017. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 318, 2199–2210.
- Belli, S., Aytac, H.O., Karagulle, E., Yabanoglu, H., Kayaselcuk, F., Yildirim, S., 2014. Outcomes of surgical treatment of primary signet ring cell carcinoma of the colon and rectum: 22 cases reviewed with literature. *International Surgery* 99, 691–698.
- Borovec, J., Munoz-Barrutia, A., Kybic, J., 2018. Benchmarking of image registration methods for differently stained histological slides, in: 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE. pp. 3368–3372.
- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A., 2018. Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* 68, 394–424.
- Byun, J., Verardo, M.R., Sumengen, B., Lewis, G.P., Manjunath, B., Fisher, S.K., 2006. Automated tool for the detection of cell nuclei in digital microscopic images: application to retinal images. *Mol Vis* 12, 949–60.
- Caicedo, J.C., Goodman, A., Karhohs, K.W., Cimini, B.A., Ackerman, J., Haghghi, M., Heng, C., Becker, T., Doan, M., McQuin, C., et al., 2019. Nucleus segmentation across imaging experiments: the 2018 data science bowl. *Nature Methods* 16, 1247–1253.
- CampANELLA, G., Hanna, M.G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K.J., Brogi, E., Reuter, V.E., Klimstra, D.S., Fuchs, T.J., 2019. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine* 25, 1301–1309.
- Carboneau, M.A., Cheplygina, V., Granger, E., Gagnon, G., 2018. Multiple instance learning: A survey of problem characteristics and applications. *Pattern Recognition* 77, 329–353.
- Chen, C.L., Chen, C.C., Yu, W.H., Chen, S.H., Chang, Y.C., Hsu, T.I., Hsiao, M., Yeh, C.Y., Chen, C.Y., 2021. An annotation-free whole-slide training approach to pathological classification of lung cancer types using deep learning. *Nature Communications* 12, 1–13.
- Chen, H., Dou, Q., Wang, X., Qin, J., Heng, P.A., 2016. Mitosis detection in breast cancer histology images via deep cascaded networks, in: Thirtieth AAAI Conference on Artificial Intelligence.
- Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision, pp. 801–818.
- Ciompi, F., Veta, M., Albarqouni, S., Jiao, Y., Tan, T., Zhang, L., Jeroen van der, L., Nasir, R., 2019. Lymphocyte assessment hackathon. <https://lysto.grand-challenge.org/>.
- Conde-Sousa, E., Vale, J., Feng, M., Xu, K., Wang, Y., Della Mea, V., La Barbera, D., Montahaei, E., Baghshah, M.S., Turzynski, A., et al., 2021. Herohe challenge: assessing her2 status in breast cancer without immunohistochemistry or in situ hybridization. *arXiv preprint arXiv:2111.04738*.
- Da, Q., Deng, S., Li, J., Yi, H., Huang, X., Yang, X., Yu, T., Wang, X., Liu, J., Duan, Q., et al., 2022. Quantifying the cell morphology and predicting biological behavior of signet ring cell carcinoma using deep learning. *Scientific Reports* 12, 1–8.
- Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Ieee. pp. 886–893.
- Falk, T., Mai, D., Bensch, R., Cicek, O., Abdulkadir, A., Marrakchi, Y., Bohm, A., Deubner, J., Jackel, Z., Seiwald, K., et al., 2019. U-net: deep learning for cell counting, detection, and morphometry. *Nature Methods* 16, 67.
- Girshick, R., 2015. Fast r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448.
- Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587.
- Grabovac, I., Smith, L., Jackson, S., Yang, L., 2020. Gastrointestinal cancer. *Encyclopedia of Biomedical Gerontology*, 128–135.
- Graham, S., Chen, H., Gamper, J., Dou, Q., Heng, P.A., Sneed, D., Tsang, Y.W., Rajpoot, N., 2019. Mild-net: Minimal information loss dilated network for gland instance segmentation in colon histology images. *Medical Image Analysis* 52, 199–211.
- Gupta, R., Mallick, P., Duggal, R., Gupta, A., Sharma, O., 2017. Stain color normalization and segmentation of plasma cells in microscopic images as a prelude to development of computer assisted automated disease diagnostic tool in multiple myeloma. *Clinical Lymphoma, Myeloma and Leukemia* 17, e99.
- Hamilton, S.R., Aaltonen, L.A., 2000. World Health Organization classification of tumours: pathology and genetics of tumours of the digestive system. IARC Publications.
- He, K., Gkioxari, G., Dollar, P., Girshick, R., 2017. Mask r-cnn, in: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
- Hou, L., Nguyen, V., Kanevsky, A.B., Samaras, D., Kurc, T.M., Zhao, T., Gupta, R.R., Gao, Y., Chen, W., Foran, D., et al., 2019. Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images. *Pattern Recognition* 86, 188–200.
- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.
- Karimi, D., Nir, G., Fazli, L., Black, P.C., Goldenberg, L., Salcudean, S.E., 2019. Deep learning-based gleason grading of prostate cancer from histopathology imagesrole of multiscale decision aggregation and data augmentation. *IEEE Journal of Biomedical and Health Informatics* 24, 1413–1426.
- Kepil, N., Batur, S., Goksel, S., 2019. Immunohistochemical and genetic features of mucinous and signet-ring cell carcinomas of the stomach, colon and rectum: a comparative study. *International Journal of Clinical and Experimental Pathology* 12, 3483.
- Kim, Y.J., Jang, H., Lee, K., Park, S., Min, S.G., Hong, C., Park, J.H., Lee, K., Kim, J., Hong, W., et al., 2021. Paip 2019: Liver cancer segmentation challenge. *Medical Image Analysis* 67, 101854.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, pp. 1097–1105.
- Kumar, N., Verma, R., Anand, D., Zhou, Y., Onder, O.F., Tsougenis, E., Chen, H., Heng, P.A., Li, J., Hu, Z., et al., 2019. A multi-organ nucleus segmentation challenge. *IEEE Transactions on Medical Imaging*.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444.
- Lee, C.Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z., 2015. Deeply-supervised nets, in: Artificial Intelligence and Statistics, pp. 562–570.
- Lee, K.M., Street, W.N., 2003. An adaptive resource-allocating network for automated detection, segmentation, and classification of breast cancer nuclei topic area: image processing and recognition. *IEEE Transactions on Neural Networks* 14, 680–687.
- Li, B., Li, Y., Eliceiri, K.W., 2021a. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14318–14328.

- Li, J., Chen, W., Huang, X., Yang, S., Hu, Z., Duan, Q., Metaxas, D.N., Li, H., Zhang, S., 2021b. Hybrid supervision learning for pathology whole slide image classification, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 309–318.
- Li, J., Hu, Z., Yang, S., 2019a. Accurate nuclear segmentation with center vector encoding, in: International Conference on Information Processing in Medical Imaging, Springer. pp. 394–404.
- Li, J., Wang, Y., Wang, C., Tai, Y., Qian, J., Yang, J., Wang, C., Li, J., Huang, F., 2019b. Dsfd: dual shot face detector, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5060–5069.
- Li, J., Yang, S., Huang, X., Da, Q., Yang, X., Hu, Z., Duan, Q., Wang, C., Li, H., 2019c. Signet ring cell detection with a semi-supervised learning framework, in: International Conference on Information Processing in Medical Imaging, Springer. pp. 842–854.
- Li, Z., Hu, Z., Xu, J., Tan, T., Chen, H., Duan, Z., Liu, P., Tang, J., Cai, G., Ouyang, Q., et al., 2018a. Computer-aided diagnosis of lung carcinoma using deep learning-a pilot study. arXiv preprint arXiv:1803.05471 .
- Li, Z., Zhang, X., Müller, H., Zhang, S., 2018b. Large-scale retrieval for medical image analytics: A comprehensive review. *Medical Image Analysis* 43, 66–84.
- Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017a. Feature pyramid networks for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125.
- Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017b. Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988.
- Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P., 2018. Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* .
- Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafourian, M., Van Der Laak, J.A., Van Ginneken, B., Sanchez, C.I., 2017. A survey on deep learning in medical image analysis. *Medical Image Analysis* 42, 60–88.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440.
- Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, 91–110.
- Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F., 2021. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature Biomedical Engineering* 5, 555–570.
- Lu, Z., Carneiro, G., Bradley, A.P., 2013. Automated nucleus and cytoplasm segmentation of overlapping cervical cells, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 452–460.
- Mármol, I., Sánchez-de Diego, C., Pradilla Dieste, A., Cerrada, E., Rodriguez Yoldi, M.J., 2017. Colorectal carcinoma: a general overview and future perspectives in colorectal cancer. *International Journal of Molecular Sciences* 18, 197.
- Maron, O., Lozano-Pérez, T., 1998. A framework for multiple-instance learning. *Advances in Neural Information Processing Systems* , 570–576.
- Maurer, C.R., Qi, R., Raghavan, V., 2003. A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, 265–270.
- Messerini, L., Palomba, A., Zampi, G., 1995. Primary signet-ring cell carcinoma of the colon and rectum. *Diseases of the Colon & Rectum* 38, 1189–1192.
- Muhammad, H., Sigel, C.S., Campanella, G., Boerner, T., Pak, L.M., Büttner, S., IJzermans, J.N., Koerkamp, B.G., Doukas, M., Jarnagin, W.R., et al., 2019. Unsupervised subtyping of cholangiocarcinoma using a deep clustering convolutional autoencoder, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 604–612.
- Naylor, P., Lae, M., Reyal, F., Walter, T., 2018. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Transactions on Medical Imaging* 38, 448–459.
- Nir, G., Hor, S., Karimi, D., Fazli, L., Skinnider, B.F., Tavassoli, P., Turbin, D., Villamil, C.F., Wang, G., Wilson, R.S., et al., 2018. Automatic grading of prostate cancer in digitized histopathology images: Learning from multiple experts. *Medical Image Analysis* 50, 167–180.
- Peikari, M., Salama, S., Nofech-Mozes, S., Martel, A.L., 2017. Automatic cellularity assessment from post-treated breast surgical specimens. *Cytometry Part A* 91, 1078–1087.
- Pernot, S., Voron, T., Perkins, G., Lagorce-Pages, C., Berger, A., Taieb, J., 2015. Signet-ring cell carcinoma of the stomach: Impact on prognosis and specific therapeutic challenge. *World Journal of Gastroenterology: WJG* 21, 11428.
- Qu, H., Wu, P., Huang, Q., Yi, J., Yan, Z., Li, K., Riedlinger, G.M., De, S., Zhang, S., Metaxas, D.N., 2020. Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images. *IEEE Transactions on Medical Imaging* 39, 3655–3666.
- Redmon, J., Farhadi, A., 2018. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 .
- Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks, in: Advances in Neural Information Processing Systems, pp. 91–99.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer. pp. 234–241.
- Roy, A.G., Navab, N., Wachinger, C., 2018. Concurrent spatial and channel squeeze & excitationin fully convolutional networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 421–429.
- Schmauch, B., Romagnoni, A., Pronier, E., Saillard, C., Maillé, P., Calderaro, J., Kamoun, A., Sefta, M., Toldo, S., Zaslavskiy, M., et al., 2020. A deep learning model to predict rna-seq expression of tumours from whole slide images. *Nature Communications* 11, 1–15.
- Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al., 2021. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in Neural Information Processing Systems* 34.
- Shen, D., Wu, G., Suk, H.I., 2017. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering* 19, 221–248.
- Shi, J., Wu, J., Li, Y., Zhang, Q., Ying, S., 2016. Histopathological image classification with color pattern random binary hashing-based pcanet and matrix-form classifier. *IEEE Journal of Biomedical and Health Informatics* 21, 1327–1337.
- Shi, X., Su, H., Xing, F., Liang, Y., Qu, G., Yang, L., 2020. Graph temporal ensembling based semi-supervised convolutional neural network with noisy labels for histopathology image analysis. *Medical Image Analysis* 60, 101624.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .
- Sirinukunwattana, K., Pluim, J.P., Chen, H., Qi, X., Heng, P.A., Guo, Y.B., Wang, L.Y., Matuszewski, B.J., Bruni, E., Sanchez, U., et al., 2017. Gland segmentation in colon histology images: The glas challenge contest. *Medical Image Analysis* 35, 489–502.
- Soille, P., 2013. Morphological image analysis: principles and applications. Springer Science & Business Media.
- Swiderska-Chadaj, Z., Pinckaers, H., van Rijthoven, M., Balkenhol, M., Melnikova, M., Geessink, O., Manson, Q., Sherman, M., Polonia, A., Parry, J., et al., 2019. Learning to detect lymphocytes in immunohistochemistry with deep learning. *Medical Image Analysis* 58, 101547.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826.
- Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks, in: International Conference on Machine Learning, PMLR. pp. 6105–6114.

- Treanor, D., Quirke, P., 2007. Pathology of colorectal cancer. *Clinical Oncology* 19, 769–776.
- Veeling, B.S., Limmans, J., Winkens, J., Cohen, T., Welling, M., 2018. Rotation equivariant cnns for digital pathology, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 210–218.
- Veta, M., Heng, Y.J., Stathonikos, N., Bejnordi, B.E., Beca, F., Wollmann, T., Rohr, K., Shah, M.A., Wang, D., Rousson, M., et al., 2019. Predicting breast tumor proliferation from whole-slide images: the tupac16 challenge. *Medical Image Analysis* 54, 111–121.
- Veta, M., Huisman, A., Viergever, M.A., van Diest, P.J., Pluim, J.P., 2011. Marker-controlled watershed segmentation of nuclei in h&e stained breast cancer biopsy images, in: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE. pp. 618–621.
- Wang, C., Shi, J., Zhang, Q., Ying, S., 2017. Histopathological image classification with bilinear convolutional neural networks, in: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE. pp. 4050–4053.
- Wen, Y., Zhang, K., Li, Z., Qiao, Y., 2016. A discriminative feature learning approach for deep face recognition, in: European Conference on Computer Vision, Springer. pp. 499–515.
- Xie, S., Girshick, R., Dollar, P., Tu, Z., He, K., 2017. Aggregated residual transformations for deep neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1492–1500.
- Xing, F., Yang, L., 2016. Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. *IEEE Reviews in Biomedical Engineering* 9, 234–263.
- Yang, L., Zhang, Y., Guldner, I.H., Zhang, S., Chen, D.Z., 2016. 3d segmentation of glial cells using fully convolutional networks and k-terminal cut, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 658–666.
- Yin, J., Bai, Z., Zhang, J., Zheng, Z., Yao, H., Ye, P., Li, J., Gao, X., Zhang, Z., 2019. Burden of colorectal cancer in china, 1990–2017: Findings from the global burden of disease study 2017. *Chinese Journal of Cancer Research* 31, 489.
- Yu, H., Zhang, X., Song, L., Jiang, L., Huang, X., Chen, W., Zhang, C., Li, J., Yang, J., Hu, Z., et al., 2021. Large-scale gastric cancer screening and localization using multi-task deep neural network. *Neurocomputing* 448, 290–300.
- Zhang, X., Liu, W., Dundar, M., Badve, S., Zhang, S., 2014a. Towards large-scale histopathological image analysis: Hashing-based image retrieval. *IEEE Transactions on Medical Imaging* 34, 496–506.
- Zhang, X., Su, H., Yang, L., Zhang, S., 2015a. Fine-grained histopathological image analysis via robust segmentation and large-scale retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5361–5368.
- Zhang, X., Xing, F., Su, H., Yang, L., Zhang, S., 2015b. High-throughput histopathological image analysis via robust cell segmentation and hashing. *Medical Image Analysis* 26, 306–315.
- Zhang, X., Yang, L., Liu, W., Su, H., Zhang, S., 2014b. Mining histopathological images via composite hashing and online learning, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 479–486.
- Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2881–2890.
- Zheng, Guoyan and Li, Shuo and Belavy, Daniel, 2020. <https://ivdm3seg.weebly.com/>.

**CONFLICT OF INTEREST**

o All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version. o This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue. o The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript o The following authors have affiliations with organizations with direct or indirect financial interest in the subject matter discussed in the manuscript: