EBioMedicine 65 (2021) 103238

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Research paper

A deep learning-based system for bile duct annotation and station recognition in linear endoscopic ultrasound

Liwen Yao[a,b,c,1], Jun Zhang[a,b,c,1], Jun Liu[e,1], Liangru Zhu[e], Xiangwu Ding[f], Di Chen[a,b,c], Huiling Wu[a,b,c], Zihua Lu[a,b,c], Wei Zhou[a,b,c], Lihui Zhang[a,b,c], Bo Xu[f], Shan Hu[d], Biqing Zheng[d], Yanning Yang[g,1,*], Honggang Yu[a,b,c,**,1]

[a] Department of Gastroenterology, Renmin Hospital of Wuhan University, Wuhan, China
[b] Key Laboratory of Hubei Province for Digestive System Disease, Renmin Hospital of Wuhan University, Wuhan, China
[c] Hubei Provincial Clinical Research Center for Digestive Disease Minimally Invasive Incision, Renmin Hospital of Wuhan University, Wuhan, China
[d] Wuhan EndoAngel Medical Technology Company, Wuhan, China
[e] Department of Gastroenterology, Wuhan Union Hospital, Huazhong University of Science and Technology, Wuhan, China
[f] Wuhan Puai Hospital, Wuhan, China
[g] Department of Ophthalmology, Renmin Hospital of Wuhan University, 99 Zhangzhidong Road, Wuhan 430060, Hubei Province, China

A R T I C L E I N F O

Article History:
Received 12 October 2020
Revised 22 January 2021
Accepted 25 January 2021
Available online xxx

Keywords:
Endoscopic ultrasound
Biliary tract
Training
Deep learning

S U M M A R Y

Background: Detailed evaluation of bile duct (BD) is main focus during endoscopic ultrasound (EUS). The aim of this study was to develop a system for EUS BD scanning augmentation.
Methods: The scanning was divided into 4 stations. We developed a station classification model and a BD segmentation model with 10681 images and 2529 images, respectively. 1704 images and 667 images were applied to classification and segmentation internal validation. For classification and segmentation video validation, 264 and 517 videos clips were used. For man-machine contest, an independent data set contained 120 images was applied. 799 images from other two hospitals were used for external validation. A crossover study was conducted to evaluate the system effect on reducing difficulty in ultrasound images interpretation.
Findings: For classification, the model achieved an accuracy of 93.3% in image set and 90.1% in video set. For segmentation, the model had a dice of 0.77 in image set, sensitivity of 89.48% and specificity of 82.3% in video set. For external validation, the model achieved 82.6% accuracy in classification. In man-machine contest, the models achieved 88.3% accuracy in classification and 0.72 dice in BD segmentation, which is comparable to that of expert. In the crossover study, trainees' accuracy improved from 60.8% to 76.3% ($P < 0.01$, 95% C.I. 20.9–27.2).
Interpretation: We developed a deep learning-based augmentation system for EUS BD scanning augmentation.
Funding: Hubei Provincial Clinical Research Center for Digestive Disease Minimally Invasive Incision, Hubei Province Major Science and Technology Innovation Project, National Natural Science Foundation of China.

1. Introduction

Endoscopic ultrasound (EUS) has excellent performance for the diagnosis of biliary disease, such as choledocholithiasis, bile duct (BD) obstruction, ampullary carcinoma and common BD carcinoma. In BD evaluation, EUS is closest to endoscopic retrograde cholangio pancreaticography (ERCP), which is the gold standard [1,2]. For choledocholithiasis diagnosis, the sensitivity was 0.97 for EUS and 0.0.87 fro magnetic resonance cholangiopancreatography [3].

Multi-station imaging techniques is the standard scanning procedure in EUS-BD evaluation. The stations contain anatomical landmarks which could be used to locate the transducer and to identify areas that have not been scanned. EUS of the BD can be done from the stations as follows: Station 1: the fundus of stomach (liver); Station 2: body of stomach (and antrum); Station 3: duodenal bulb; Station 4: descending duodenum [4-6]. Comprehensive evaluation of BD is frequently the main focus of imaging during EUS and in such situations, multi-station imaging is necessary to scan the whole BD [7].

Abbreviations: BD, bile duct; EUS, endoscopic ultrasound; DCNN, deep convolutional neural network; ERCP, Endoscopic Retrograde Cholangiopancreatography

** Corresponding author at: Department of Gastroenterology, Renmin Hospital of Wuhan University. 99 Zhangzhidong Road, Wuhan 430060, Hubei Province, China.
* Corresponding author.
E-mail addresses: ophyyn@163.com (Y. Yang), yuhonggang@whu.edu.cn (H. Yu).
[1] These authors contribute equally to this work.

https://doi.org/10.1016/j.ebiom.2021.103238
2352-3964/

Research in context

Evidence before this study

We searched PubMed for papers published between Jan 1, 2001, and March 1, 2020, with the keywords "machine learning", "artificial intelligence" OR"deep learning" AND "endoscopic ultrasound". No restrictions on study type or language were implemented. Our search retrieved studies on the use of deep learning for computer aided diagnosis of pancreas lesions but no studies to improve the ultrasonographics interpretation in biliary system.

Added value of this study

We constructed a deep learning-based system, BP MASTER, for real-time endoscopic ultrasound biliary scanning augmentation. This system was followed by internal- and external validation in images or videos, and subsequently compared with the performance of endoscopists. The effect of the system on eliminating the difficulty of ultrasonographic interpretation was evaluated among trainees in prospectively collected endoscopic ultrasound videos. Our study confirmed the feasibility of using deep learning for endoscopic ultrasound biliary augmentation.

Implications of all the available evidence

Endoscopic ultrasound provides improved imaging functions and has provided multiple treatment method in biliary disease, but the endoscopic ultrasound systems have been hesitantly adopted by some gastroenterologists due to its steep learning curve and relying too much on the operator. Our study shows that the BP MASTER system can recognize the standard station for bile duct scanning and prompt the physicians for the corresponding operation instruction. Moreover, the system can also segment the bile duct with high precision and automatically measure the bile duct diameter. With the system's augmentation, the trainees improved their accuracy of station recognition. The BP MASTER system has potential to play an important role in endoscopic ultrasound biliary scanning augmentation.

However, EUS is one of the most challenging endoscopic procedures to learn and requires integration of both cognitive and endoscopic skills [8,9]. The cognitive portion of the procedure is exceedingly difficult to learn. Most experienced endosonographers believe that the key to acquiring competence in both components of the EUS procedure is pattern recognition obtained through repetitive examinations. Such experience can be acquired only at a training center performing a large volume of cases. Because few centers provide this experience, other training options are needed [10]. Therefore, an augmentation system is very needed when performing EUS BD examination and training. Ideally, a station recognition model could provide the information of transducer location as well as the operation instruction. A BD annotation model could help endosonographers to visualize the BD.

EUS-guided biliary puncture is an emerging technique that combines the advantages of the endoscopic and percutaneous approaches, without the inconveniences and discomfort of an indwelling external catheter [11]. Puncture route selection is critical in successful BD puncture cases [12]. The choice among different routes is mainly based on the degree and location of the duct dilation [13,14]. Three routes were proposed for BD puncture. The first route is transmural puncture of the intrahepatic BD by transesophageal and transgastric puncture. The second and third methods are transduodenal puncture of the extrahepatic BD via the proximal duodenum and the second portion of the duodenum, respectively. The prerequisite for choosing a suitable puncture route is to determine the obstruction position and degree of the duct dilation. Station recognition and BD annotation augmentation has potential to improve the comprehensive evaluation of BD.

In recent years, deep learning has made tremendous progress in the field of digestive endoscopy [15]. Previous work from our group showed that Deep Convolutional Neural Networks (DCNNs), one of the representative algorithms of deep learning, could accurately recognize the stations of EUS pancreas in real-time manner [16]. However, the role of deep learning in EUS biliary scanning remains unknown.

In our current study, we constructed a deep learning-based system, BP MASTER, for real-time stations recognition and BD annotation in linear EUS. There were two reasons why the radial images were not utilized: First, linear EUS was superior in the delineation of the area from the hepatic portal region to the superior BD [17]. Second, EUS-guided biliary puncture is conducted by linear EUS while the radial EUS can only applied for diagnosis. For station recognition, deep learning-based image classification model was constructed. For BD annotation, a BD segmentation model was constructed to detect the BD within the digital image from the endoscopy processor and output the BD boundary as a green line. This system was followed by internal- and external validation in images or videos, and subsequently compared with the performance of EUS endoscopists. The effect of the system on eliminating the difficulty of ultrasonographic interpretation was evaluated among trainees in prospectively collected EUS videos. The purpose of this study is to explore the role of deep learning in linear EUS BD scanning.

2. Method

2.1. System framework

Four DCNN models were incorporated into BP MASTER system to achieve two main functions: First, to position the station where the transducer is located and provide the corresponding operation instructions. Second, to annotate the CBD and provide diameter measurement when endoscopists froze the frames. DCNN1 was applied to filter out white light images and input the ultrasound images to DCNN2. DCNN2 was applied to classify ultrasound images into standard and non-standard categories, and activate DCNN3 with standard images. DCNN3 was used to recognize BD stations. DCNN4 was used to segment and annotate BD (Fig. 1). The STARD 2015 reporting guidelines was followed when writing this work.

2.2. Datasets and preprocessing

For DCNN1 training and validation, 2000 white light images of gastroscopy and 2000 EUS images were applied at a 9 to 1 ratio. For unqualified images filtering, 10001 standard station and 17335 unqualified EUS images were used to train, 1735 standard station and 1412 unqualified EUS images were used to test DCNN2. The criteria for unqualified images were jointly negotiated by two EUS experts, including: obscure, large lesions, kidney, spleen, abdominal aorta, elastography, and extremely dilated bile/pancreatic duct. Representative images of the unqualified images were shown in Fig. S1.

Five data sets were used for training, internal validation and external validation of BP MASTER system:

(1) 10681 images from 443 EUS procedures were used to train the model for BD station recognize (DCNN1). 2529 images contained complete and clear BD from the same procedures were applied to train the model for BD annotation. The average age of the patients is 55 years old (standard deviation is 12.6). The proportion of men in this dataset is 49.7% (220/443). All the images

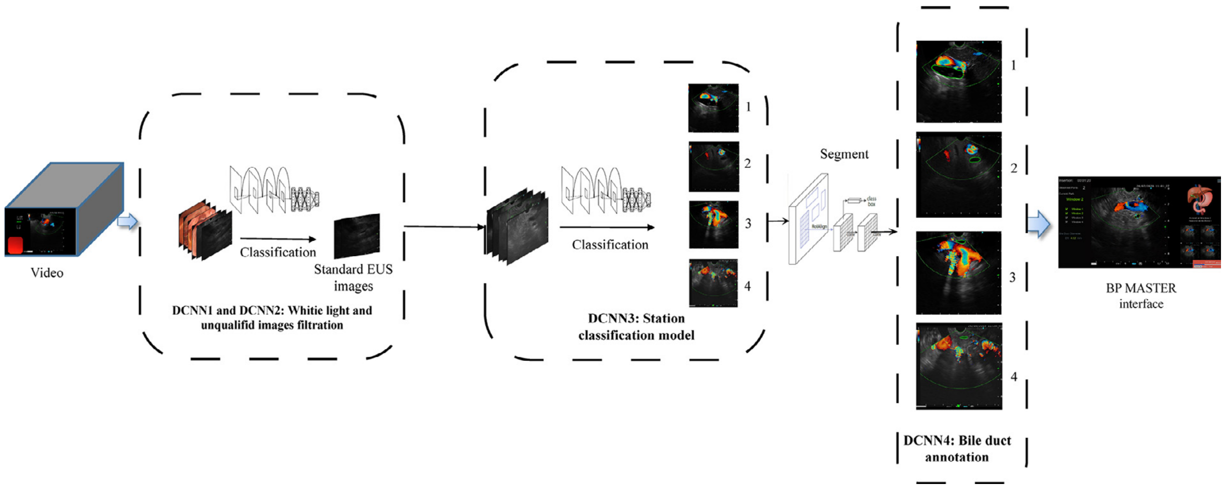

Fig. 1. BP MASTER system framework: DCNN1 was applied to filter out white light images and input the ultrasound images to DCNN2. DCNN2 was applied to classify ultrasound images into standard and non-standard categories, and activate DCNN3 with standard images. DCNN3 was used to recognize stations. DCNN4 was used to segment and annotate bile duct.

were from Wuhan Renmin Hospital during December 2016–July 2020.

(2) 1704 images from 44 EUS procedures from Renmin Hospital of Wuhan University during October 2019–December 2019 were used for internal validation. 264 video clips contained 33280 frames from the same procedures were applied for station recognition video validation. 251 positive video clips contained 13751 frames with each frame contained BD and 300 negative video clips contained 12771 frames without BD were used to test the performance of DCNN4. The average age of the patients is 50.6 years old (standard deviation is 13.8). The proportion of men in this dataset is 47.7% (21/44). All the images were from Wuhan Renmin Hospital during October 2019–December 2019.

(3) 120 images from 44 EUS procedures from Renmin Hospital of Wuhan University during September 2019 - May 2020 were used to compare the performance of DCNN3 and DCNN4 with that of EUS experts (man-machine contest). The average age of the patients is 56 years old (standard deviation is 12.1). The proportion of men in this dataset is 61.4% (27/44).

(4) For the external validation, an external testing data set contained 799 images from 20 examinations (Wuhan Union Hospital) and 89 examinations (Wuhan Puai hospital) were collected. The average age of the patients is 59 years old (standard deviation is 10.3). The proportion of men in this dataset is 31.2% (34/109).

The sample distribution for each data set was shown in Table 1. The 4 stations and its representative images predicted by the DCNN models were shown in Fig. 2. Images from the same person were not split among the data sets. The procedures were performed by Olympus EU-ME1 and EU-ME2 (Olympus Medical Systems Co., Tokyo, Japan) processors and adapted endoscopes.

2.3. Annotation

Two EUS experts A and B from Wuhan Renmin Hospital labeled each images and video clips with negotiation. Their labels were used as gold standard for all the training and validation.

For man-machine contest, expert C, senior endoscopists D, E and F were required to classified each image in the comparison data set and then, annotate the BD based on the classification results. Both endoscopists and model results were compared with ground truth annotated by expert A and B.

For annotators level of expertise, expert endoscopists were defined as who had at least 10 years while senior endoscopists were defined as 5 years of experience in performing EUS examination and treatment.

2.4. Training of DCNN models

We used ResNet for image classification and Unet++ for image segmentation. Both networks were trained on an NIVIDIA GeForce GTX 2080. The technical details and neural network architecture were illustrated in supplementary. For DCNN1, 2 and 3, ResNet-50, a mature DCNN architectures pretrained by data from ImageNet (1.28 million images from 1000 object classes), were used to train DCNN1, 2 and 3. We replaced the final classification layer with another fully connected layer using transfer learning, retrained them

Table 1
Baseline information.

	Patient (n)	Station 1	Station 2	Station 3	Station 4	Total
DCNN3 training set (frames)	443	1518	5768	1071	2324	10681
DCNN4 training set (frames)	443	360	692	799	678	2529
DCNN3 internal validation set (frames)	44	312	619	333	440	1704
DCNN3 video validation set (clips/frames)	44	42/4295	76/10498	96/10134	50/8281	264 /33208
DCNN4 internal validation set (frames)	44	72	160	283	206	721
DCNN4 video validation set (clips/frames)	44	69/4762	76/4484	43/1931	63/2576	251 /13753
Man-machine contest set (frames)	44	30	30	30	30	120
External validation set (frames)	109	335	148	204	112	799
Crossover study set (videos)	29	22	29	29	23	29

DCNN, deep convolutional neural network.

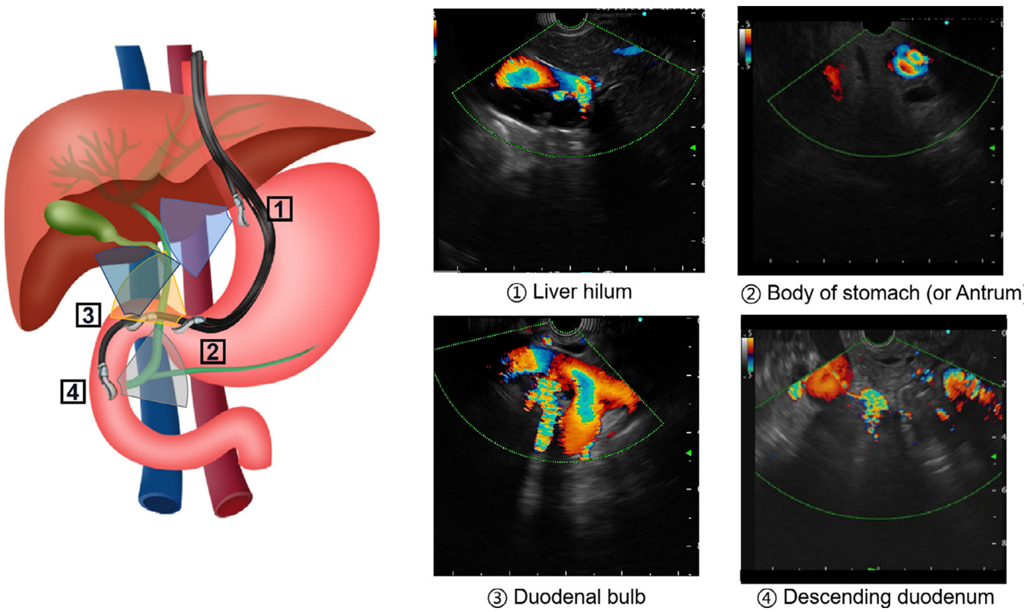

Fig. 2. A schematic illustration of the stations about the visualization of bile duct in linear EUS and its representative images predicted by the DCNN3.

using our datasets, and fine-tuned the parameters to fit our needs. The dataset was randomly divided into 5 subsets and one subset was validated individually with the remaining for training in Google's TensorFlow [18]. Three method, dropout [19], data augmentation [20] and early stopping [21], were used to minimize the overfitting risk.

For BD annotation, UNet++, a novel and powerful architecture for medical image segmentation was implemented to develop DCNN4 [22,23]. With the original EUS image as the input, the resolution is 512 $\times$ 512, and the expert-marked map as the output, UNet ++ is used to train and test DCNN4 in image-to-image manner in Keras. According to the result of internal validation, we get the best segmentation threshold by increasing 2 each time and the threshold was set as 0.55 (Fig. S5) Besides BD annotation, DCNN4 also provide diameter measurement result when endoscopists froze the videos. The details of how to identify whether the videos were frozen and algorithm of diameter measurement were provided in supplementary.

2.5. Construction of BP MASTER System

For station recognition prediction, we used the Random Forest Classifier model [24] and the rule of 'output results only when three of the five consecutive images show a same result' to smooth noises. The FPS (frames per second) for running the system in videos was 4.78 on a GPU. The speed of the DCNN in the clinical setting to output a prediction per frame in the endoscopy center of Renmin Hospital of Wuhan University was 200-300ms, including time consumed in the client (image capture, image re-sizing, and rendering images based on predicted results), network communication, and the server (reading and loading images, running the three networks, and saving images). For BD annotation, the system was set to segment and output the result at 15 FPS. All the models were trained and ran on a server with a GPU NVIDIA RTX2080Ti (with 8 GB GPU memory).

2.6. Ultrasonographics interpretation study

2.6.1. Prospective data set collection

To evaluate the effect of BP MASTER, we prospectively consecutively enrolled patients undergoing EUS examinations and their corresponding videos were collected between July 2020 to August 2020. The study was approved by the Ethics Committee of Renmin Hospital of Wuhan University (WDRY2019-K091) and under trail registration number ChiCTR1900028648 of the Primary Registries of the WHO Registry Network. Informed consents were obtained from each participant. Patient with lower gastrointestinal EUS, radial EUS or no standard station scanned were excluded.

2.6.2. Study procedure

With the prospectively collected videos, a crossover study was performed to evaluate BP MASTER effect in improving trainees station recognition and BD segmentation. 8 primary trainees and 4 advanced trainees were included in this study. The primary trainees who participated had already more than one-year gastroenterology fellowship experience and none had any prior experience or training in EUS while the advanced trainees had handled at least 100 training EUS. All the trainees were required to read the reference of CBD multi-station scanning and were provided 20 typical images of each station for learning a week in advance [4].

Clinicians were provided videos and were requested to record the time point of first recognizing each station. Using a crossover design, the trainees were randomly and equally divided into 2 groups. The randomization was generated by a random grouping software. Group A first read the videos and images without BP MASTER augmentation, and group B first read with BP MASTER augmentation. After a washout period of 2 weeks, the arrangement was reversed such that group A performed read with augmentation and group B read the videos without (Fig. 4). With the model augmentation, readers had the option to take it into consideration or disregard it based on judgment. The time point and accuracy at which the BP MASTER first recognized each station as well as the segmentation result were also recorded.

2.7. Statistical analysis

For station classification evaluation, we used accuracy as metric, which defined as the number of correctly classified images divided by the total number of images. Similarly, per frame accuracy was defined as the correctly classified frames divided by the total number of frames.

The standard deviation was calculated as: $SD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - u)^2}$

For segmentation evaluation, intersection over union (IOU) was defined as the relative area of overlap between the predicted bounding box (A) and the ground-truth (B) bounding box. The ground truth was labeled by the Expert A and B:

$$\mathrm{IOU} = \frac{|A \cap B|}{|A| \cup |B|}$$

$$\mathrm{Dice}\ (\mathrm{F1\ score}) = \mathrm{IOU}$$

When the IOU>threshold, the prediction is true positive; When the IOU<threshold, the prediction is false positive; When the model segmentation area = 0, it is false negative.

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

Inter-observer agreement of the endoscopists and the DCNN were evaluated using Cohen's kappa coefficient.

For the crossover study, we compared the time point accuracy for each trainee with or without augmentation.

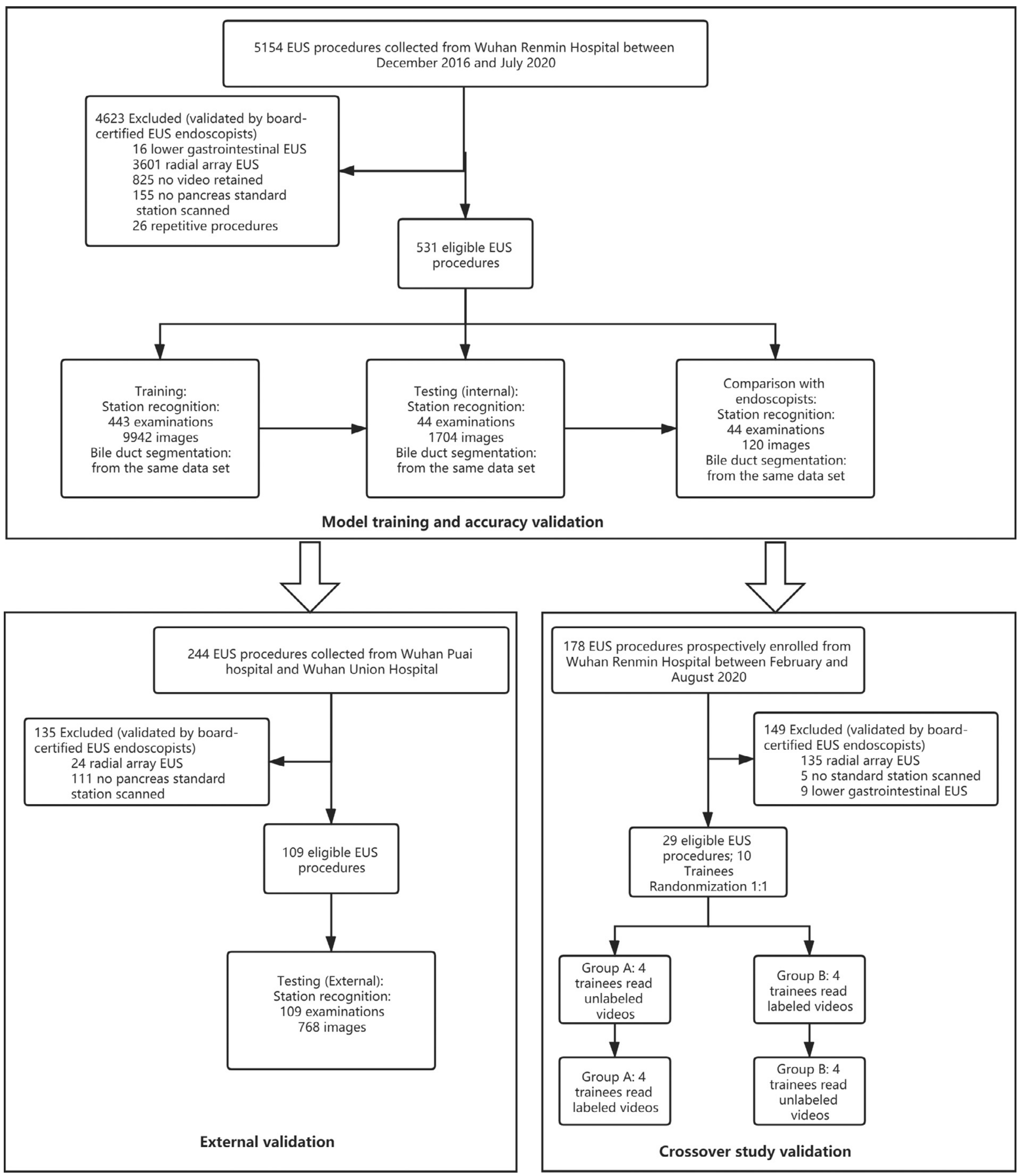

Fig. 3. Flowchart of the study development and validation.

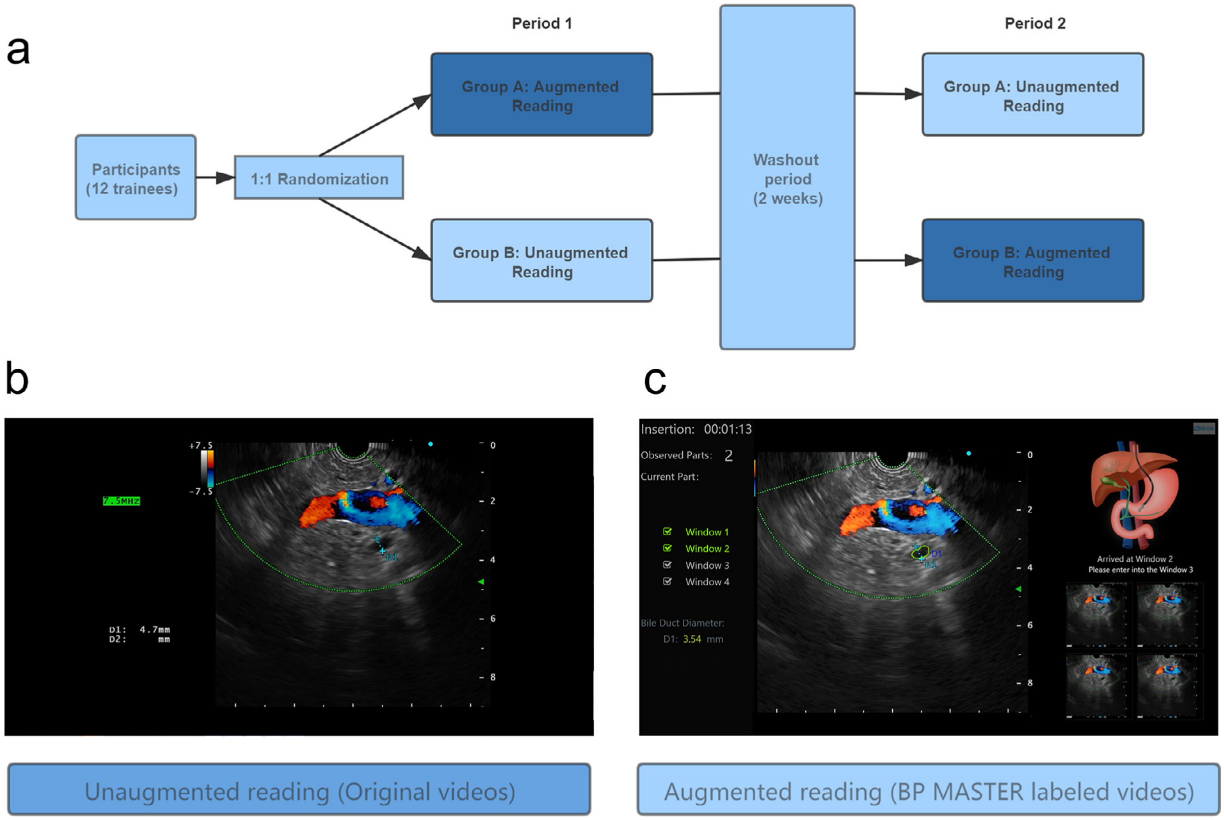

Fig. 4. The crossover study design: a. Study design. 12 trainees were divided into 2 groups to perform reads with and without model augmentation in random order, with a 2-week washout period between. b. Unaugmented read, with original EUS videos. c. Augmented read, videos with model labeled. EUS, endoscopic ultrasound.

To assess whether the trainees achieved significant increases in performance with model augmentation, McNemar's test was performed on the differences in aforementioned metric across all 12 trainees. $P < 0.05$ was considered statistically significant. All calculations were performed using SPSS 23 (IBM, Chicago, Illinois, USA).

2.8. Role of the funding source

The funder had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

3. Results

3.1. Internal and external validation

For white light and ultrasound images classification, DCNN1 achieved an accuracy of 100%. In the standard and non-standard images classification, the DCNN2 achieved an accuracy of 87.4%. The confusion matrixes of DCNN1 and 2 were shown in Fig. S2 and S3, respectively.

For DCNN3, the model had an accuracy of 93.3% in image validation set (Fig. S4). In video validation set, the model had a per-frame accuracy of 90.1% (Table 2). As for the BD segmentation performance, DCNN4 had a Dice of 0.77. The recall and precision at 50% IOU were 85% and 98.2%. In video validation set, the sensitivity among positive video clips was 89.5% and the specificity among negative video clips was 82.3% (Table 3).

Among the retrospectively collected images from Wuhan Puai Hospital and Wuhan Union Hospital, the DCNN3 achieved an accuracy of 82.6%.

Table 2
DCNN3 station recognition accuracy in internal, external and video validation.

	Internal validation	External validation	Video validation
Station 1 (%)	86.3	83.6	82.5
Station 2 (%)	99.5	82.4	95.9
Station 3 (%)	93.7	76.5	91.4
Station 4 (%)	89.8	91.1	87.1
Total	93.3	83.9	90.1

3.2. Man-machine contest

In the testing dataset for man-machine contest, DCNN3 correctly classified the BD stations with an accuracy of 88.3%. The accuracy for expert C, endoscopists D, E and F was 90%, 85.8%, 74.2% and 84.2%, respectively. For the BD annotation, the model had a dice of 0.59. Among the images that contained BD, the Dice was 0.72 for models and 0.74, 0.65, 0.67, 0.65 for endoscopists, respectively (Fig. 5). Among all images, the dice for expert C, endoscopists D, E and F was 0.61, 0.54, 0.55 and 0.54 (Table S1). The inter-observer agreement between DCNN3 and experts was shown in Table S2.

3.3. Crossover study

In the crossover study, trainees achieved a time point accuracy of 60.8% without augmentation as a group. With augmentation, the accuracy significantly improved from 60.8% to 76.3% ($P < 0.01$, 95% C.I. 20.9–27.2). The underlying model had an accuracy of 86.2%. The performances of individual trainees were reported in the Table 4. All the 12 trainees have made significant improvement with the augmentation.

Table 3
DCNN4 segmentation performance.

	Internal validation					Video validation	
	DICE	Precision>0.5	Recall>0.5	Precision>0.3	Recall>0.3	Sensitivity (%)	Specificity (%)
Station 1	0.83	0.94	0.99	0.98	0.99	89.5	–
Station 2	0.72	0.75	0.98	0.84	0.98	89	–
Station 3	0.76	0.81	0.98	0.9	0.98	90.2	–
Station 4	0.69	0.87	0.69	0.93	0.85	82.6	–
Total	0.77	0.85	0.95	0.92	0.97	88.1	82.30%

Table 4
Trainees' station recognition accuracy with and without augmentation.

		Without augmentation (%)	With augmentation (%)	Increase (%)(95% C.I.)	P value
	Model	-	86.2	–	–
Group A	Trainee A	72.4	83.6	11.2 (5.2–21.6)	<0.01
	Trainee B	43.1	73.3	30.2 (17.6–41.4)	<0.01
	Trainee C	45.7	62.9	17.2 (4.42–29.3)	<0.01
	Trainee D	56	77.6	21.6 (9.46–32.8)	<0.01
	Trainee E*	70	78.3	8.3 (−2.7 to 19.6)	<0.01
	Trainee F*	65.8	77.5	11.7 (0.5–23.3)	<0.01
Group B	Trainee G	57.8	73.3	15.5 (3.3–27.1)	<0.01
	Trainee H	69.8	83.6	13.8 (2.9–24.3)	<0.01
	Trainee I	63.8	74.1	10.3 (−1.6 to 21.9)	<0.01
	Trainee J	42.2	62.1	19.8 (7.0–31.8)	<0.01
	Trainee K*	69.2	82.5	13.3 (2.8–24.4)	<0.01
	Trainee L*	73.3	86.7	13.4 (2.6–23.0)	<0.01
	Total	60.8	76.3	15.5 (20.9–27.2)	<0.01

* : These trainees were advanced trainees.
Group A: Augmented reading first; Group B: Unaugmented reading first.

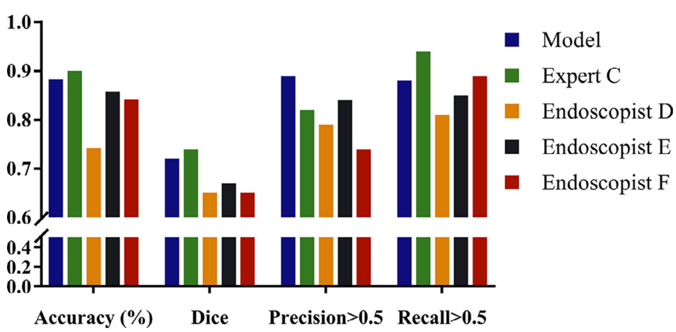

Fig. 5. The accuracy, Dice, recall and precision in the man-machine contest. In the man-machine contest, the accuracy for DCNN 3, expert C, endoscopists D, E and F was 88.3%, 90%, 85.8%, 74.2% and 84.2%, respectively. Among the images with bile duct, the dice for DCNN 4, expert C, endoscopists D, E and F was 0.72, 0.74, 0.65, 0.67, 0.65, respectively.

4. Discussion

In this study, we constructed an artificial intelligence-assisted linear EUS system, which can recognize the standard station for BD scanning and prompt the physicians for the corresponding operation instruction. Moreover, the system can also segment the BD with high precision and automatically measure the BD diameter, which could simplify the physician's operation. In the comparison with endoscopists, the DCNN1 accuracy was better than that of senior EUS endoscopists and was comparable with EUS expert.

EUS provides improved imaging functions and has been available on the market since the 1980s, but the EUS systems have been hesitantly adopted by some gastroenterologists due to its steep learning curve and relying too much on the operator [25]. Although efforts to shorten the EUS learning curve have been ongoing, and some specific centers use computer-based simulators or live animal models to improve the learning curve, EUS is still not fully applied globally [26–28]. In particular, although EUS-BD has a significant clinical impact on the treatment of patients, the performance of EUS-BD is still limited to tertiary referral centers [29]. Since EUS strongly relies on the training, skills and experience of endoscopists, the development of real-time ultrasonographics interpretation system is essential for the widespread adoption of EUS.

The standard stations contain specific anatomical landmarks and represent the precise location where the transducer was scanning. Therefore, the stations could serve as navigation marks under ultrasonographics. Among the stations, there are specific operating techniques. The physician can complete ultrasound endoscopic scanning by identifying the standard station. On the other hand, different parts of BD can be observed from each station and the station recognition can remind the part that the endoscopists have missed. Therefore, in recent years, the EUS training has gradually focused on standard station scanning education. Wani et al developed a scoring tool to evaluate the learning curve of advanced ultrasound endoscopy trainees [30,31]. The tool utilizes a 4-point scoring system: 1 (superior) = achieves independently, 2 (advanced) = achieves with minimal verbal instruction, 3 (intermediate) = achieves with multiple verbal instructions or hands-on assistance, and 4 (novice) = unable to complete requiring trainer to take over. The tool is scored based on the scanning performance of the advanced students at each station. If the endoscopists can obtain the positioning information and the corresponding operation method from a real-time augmentation system, the endoscopists can reach the competence in EUS in a shorter time. In our crossover study, the augmentation from our system has significantly improved the accuracy of the station recognition and BD segmentation by the endoscopists. The results from the crossover study indicated that the system has the potential to shorten the learning curved in the future.

Since the initial report on the use of BD puncture after failed ERCP in 2004, several studies have reported BD puncture as an effective salvage technique for achieving biliary cannulation after failed ERCP [32–35]. The BD puncture techniques comprise three methods that are based on the approach route: TG, from the second portion of the duodenum in a short endoscopic position, and from the bulb of the duodenum in a long endoscopic position. Though there is no formal consensus for how to decide between intrahepatic or extrahepatic approach, studies have suggested that endoscopists should choose the approach according to the bile duct anatomy. Therefore, comprehensive BD scanning is critical for the routes selection. The system in our current study can contribute to a comprehensive BD scanning and thus, can contribute to the dilation and stricture evaluation. Moreover, the function of automatically measure the diameter of the BD can further improve the diagnostic sensitivity of BD dilation.

For the popularity of the system, it can ensure stable and smooth operation on a computer with an RTX3070 graphics card. The price of such a computer configuration is about $2000 which is totally affordable for a practicing gastroenterologist in a private practice. The system could run totally automatically and giving real-time instruction for endoscopists. Therefore, this system will be easy to spread among practicing gastroenterologists in private practice.

There are several limitations of our study. First, though the accuracy of this system has been fully validated, the effect of this system was only tested in an augmentation reading study. In the future, a randomized study on evaluating the effect of the system was needed. Secondly, lesion identification function has not been added to this system. That is because though EUS can evaluate the nature of the stenosis and dilatation of BD, its accuracy is not as good as that of spyglass and the role of EUS in the biliary system is mainly focus on screening and treatment. However, the lesions identification system is under development in our unpublished study.

In conclusion, we constructed an EUS BD scanning augmentation system based on deep learning. The accuracy of this system has been validated both internal and external. In the future, this system has potential to play an important role in EUS training and quality control.

Contributors

YNY and HGY conceived and designed the study; LHZ, ZHL, HLW, XWD, LRZ, BX, WZ, DC collected the images; SH and BQZ trained and tested the models; LWY, JL and JL contributed to images annotation, data analysis and manuscript writing; all authors have reviewed and approved the final manuscript for submission. All authors read and approved the final version of the manuscript. All authors contribute to critical revision of the manuscript.

Data sharing

Individual de-identified participant data that underlie the results reported in this article and study protocol will be shared for investigators whose proposed use of the data has been approved by an independent review committee. Data can only be used to achieve aims in the approved proposal. Data disclosure begins 9 months after and ends 36 months after article publication. To gain access, data requesters will need to sign a data access agreement. Proposals should be directed to the corresponding author.

Declaration of Competing Interest

None.

Acknowledgments

Project of Hubei Provincial Clinical Research Center for Digestive Disease Minimally Invasive Incision (grant no. 2018BCC337); Hubei Province Major Science and Technology Innovation Project (grant no. 2018-916-000-008); National Natural Science Foundation of China (grant no. 81770899).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ebiom.2021.103238.

References

[1] Giljaca V, Gurusamy KS, Takwoingi Y, Higgie D, Poropat G, Štimac D, et al. Endoscopic ultrasound versus magnetic resonance cholangiopancreatography for common bile duct stones. Cochrane Database Syst Rev 2015(2):CD011549.
[2] Polkowski M, Regula J, Tilszer A, Butruk E. Endoscopic ultrasound versus endoscopic retrograde cholangiography for patients with intermediate probability of bile duct stones: a randomized trial comparing two management strategies. Endoscopy 2007;39(4):296–303.
[3] Meeralam Y, Al-Shammari K, Yaghoobi M. Diagnostic accuracy of EUS compared with MRCP in detecting choledocholithiasis: a meta-analysis of diagnostic test accuracy in head-to-head studies. Gastrointest Endosc 2017;86(6):986–93.
[4] Sharma M, Pathak A, Shoukat A, Rameshbabu CS, Ajmera A, Wani ZA, et al. Imaging of common bile duct by linear endoscopic ultrasound. World J Gastrointest Endosc 2015;7(15):1170–80.
[5] Irisawa A, Yamao K. Curved linear array EUS technique in the pancreas and biliary tree: focusing on the stations. Gastrointest Endosc 2009;69(2 Suppl):S84–S9.
[6] Committee EFS, Yamao K, Irisawa A, Inoue H, Matsuda K, Kida M, et al. Standard imaging techniques of endoscopic ultrasound-guided fine-needle aspiration using a curved linear array echoendoscope. Digest Endosc 2007;19:S180–205.
[7] Domagk D, Oppong KW, Aabakken L, Czakó L, Gyökeres T, Manes G, et al. Performance measures for ERCP and endoscopic ultrasound: a European Society of Gastrointestinal Endoscopy (ESGE) quality improvement initiative. Endoscopy. 2018;50(11):1116–27.
[8] Wani S, Keswani R, Hall M, Han S, Ali MA, Brauer B, et al. A prospective multicenter study evaluating learning curves and competence in endoscopic ultrasound and endoscopic retrograde cholangiopancreatography among advanced endoscopy trainees: the rapid assessment of trainee endoscopy skills study. Clin Gastroenterol Hepatol 2017;15(11).
[9] Shahidi N, Ou G, Lam E, Enns R, Telford J. When trainees reach competency in performing endoscopic ultrasound: a systematic review. Endosc Int Open 2017;5(4): E239–E43.
[10] Kim GH, Bang SJ, Hwang JH. Learning models for endoscopic ultrasonography in gastrointestinal endoscopy. World J Gastroenterol 2015;21(17):5176–82.
[11] Tsuchiya T, Itoi T, Sofuni A, Tonozuka R, Mukai S. Endoscopic ultrasonography-guided rendezvous technique. Digest Endosc 2016;28:96–101.
[12] Okuno N, Hara K, Mizuno N, Hijioka S, Tajika M, Tanaka T, et al. Endoscopic ultrasound-guided rendezvous technique after failed endoscopic retrograde cholangiopancreatography: which approach route is the best? Intern Med 2017;56 (23):3135–43.
[13] Sarkaria S, Sundararajan S, Kahaleh M. Endoscopic ultrasonographic access and drainage of the common bile duct. Gastrointest Endosc Clin 2013;23(2):435–52.
[14] Gupta K, Perez-Miranda M, Kahaleh M, Artifon EL, Itoi T, Freeman ML, et al. Endoscopic ultrasound-assisted bile duct access and drainage: multicenter, long-term analysis of approach, outcomes, and complications of a technique in evolution. J Clin Gastroenterol 2014;48(1):80–7.
[15] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25(1):44–56.
[16] Zhang J, Zhu L, Yao L, Ding X, Chen D, Wu H, et al. Deep-learning–based pancreas segmentation and station recognition system in EUS: development and validation of a useful training tool (with video). Gastrointest Endosc 2020.
[17] Kaneko M, Katanuma A, Maguchi H, Takahashi K, Osanai M, Yane K, et al. Prospective, randomized, comparative study of delineation capability of radial scanning and curved linear array endoscopic ultrasound for the pancreaticobiliary region. Endosc Int Open 2014;2(3):E160.
[18] Tensorflow: a system for large-scale machine learning. In: Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, editors. 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16); 2016.
[19] Baldi P, Sadowski P. The dropout learning algorithm. Artif Intell 2014;210:78–122.
[20] Data augmentation for improving deep learning in image classification problem. In: Mikołajczyk A, Grochowski M, editors. 2018 international interdisciplinary PhD workshop (IIPhDW). IEEE; 2018.
[21] Duvenaud D, Maclaurin D, Adams R, editors. Early stopping as nonparametric variational inference. Artificial Intelligence and Statistics; 2016.
[22] Deng J, Dong W, Socher R, Li L-J, Kai L, Li F-F. ImageNet: a large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition; 2009. p. 248–55.
[23] Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer; 2018. p. 3–11.
[24] Deep residual learning for image recognition. In: He K, Zhang X, Ren S, Sun J, editors. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
[25] Wani S, Coté GA, Keswani R, Mullady D, Azar R, Murad F, et al. Learning curves for EUS by using cumulative sum analysis: implications for American Society for Gastrointestinal Endoscopy recommendations for training. Gastrointest Endosc 2013;77(4):558–65.

[26] Bhutani MS, Wong RF, Hoffman BJ. Training facilities in gastrointestinal endoscopy: an animal model as an aid to learning endoscopic ultrasound. Endoscopy 2006;38(9):932–4.
[27] Fritscher-Ravens A, Cuming T, Dhar S, Parupudi SVJ, Patel K, Ghanbari A, et al. Endoscopic ultrasound-guided fine needle aspiration training: evaluation of a new porcine lymphadenopathy model for in vivo hands-on teaching and training, and review of the literature. Endoscopy 2013;45(2):114–20.
[28] Konge L, Clementsen PF, Ringsted C, Minddal V, Larsen KR, Annema JT. Simulator training for endobronchial ultrasound: a randomised controlled trial. Eur Respir J 2015;46(4):1140–9.
[29] Eisen GM, Dominitz JA, Faigel DO, Goldstein JA, Petersen BT, Raddawi HM, et al. Guidelines for credentialing and granting privileges for endoscopic ultrasound. Gastrointest Endosc 2001;54(6):811–4.
[30] Wani S, Han S, Simon V, Hall M, Early D, Aagaard E, et al. Setting minimum standards for training in EUS and ERCP: results from a prospective multicenter study evaluating learning curves and competence among advanced endoscopy trainees. Gastrointest Endosc 2019;89(6).
[31] Wani S, Hall M, Keswani RN, Aslanian HR, Casey B, Burbridge R, et al. Variation in aptitude of trainees in endoscopic ultrasonography, based on cumulative sum analysis. Clin Gastroenterol Hepatol 2015;13(7) 1318-25. e2.
[32] Brauer BC, Chen YK, Fukami N, Shah RJ. Single-operator EUS-guided cholangiopancreatography for difficult pancreaticobiliary access (with video). Gastrointest Endosc 2009;70(3):471–9.
[33] Maranki J, Hernandez A, Arslan B, Jaffan A, Angle J, Shami V, et al. Interventional endoscopic ultrasound-guided cholangiography: long-term experience of an emerging alternative to percutaneous transhepatic cholangiography. Endoscopy 2009;41(06):532–8.
[34] Kim Y, Gupta K, Mallery S, Li R, Kinney T, Freeman ML. Endoscopic ultrasound rendezvous for bile duct access using a transduodenal approach: cumulative experience at a single center. A case series. Endoscopy 2010;42(06):496–502.
[35] Shah JN, Marson F, Weilert F, Bhat YM, Nguyen-Tang T, Shaw RE, et al. Single-operator, single-session EUS-guided anterograde cholangiopancreatography in failed ERCP or inaccessible papilla. Gastrointest Endosc 2012;75(1):56–64.