





# Computer-aided demarcation of early gastric cancer: a pilot comparative study with endoscopists

Satoko Takemoto<sup>1</sup>  · Keisuke Hori<sup>2,4</sup> · Sakai Yoshimasa<sup>1</sup> · Masaomi Nishimura<sup>1</sup> · Keiichiro Nakajo<sup>2,3</sup> · Atsushi Inaba<sup>2</sup> · Maasa Sasabe<sup>2</sup> · Naoki Aoyama<sup>2</sup> · Takashi Watanabe<sup>2</sup> · Nobuhisa Minakata<sup>2</sup> · Hiroaki Ikematsu<sup>2,3</sup> · Hideo Yokota<sup>1</sup>  · Tomonori Yano<sup>2,3</sup>

Received: 8 February 2023 / Accepted: 4 May 2023 / Published online: 31 May 2023  
© Japanese Society of Gastroenterology 2023

## Abstract

**Background** Precise area diagnosis of early gastric cancer (EGC) is critical for reliable endoscopic resection. Computer-aided diagnosis (CAD) shows strong potential for detecting EGC and reducing cancer-care disparities caused by differences in endoscopists' skills. To be used in clinical practice, CAD should enable both the detection and the demarcation of lesions. This study proposes a scheme for the detection and delineation of EGC under white-light endoscopy and validates its performance using 1-year consecutive cases.

**Methods** Only 300 endoscopic images randomly selected from 68 consecutive cases were used for training a convolutional neural network. All cases were treated with endoscopic submucosal dissection, enabling the accumulation of a training dataset in which the extent of lesions was precisely determined. For validation, 462 cancer images and 396 normal images from 137 consecutive cases were used. From the validation results, 38 randomly selected images were compared with those delineated by six endoscopists.

**Results** Successful detections of EGC in 387 cancer images (83.8%) and the absence of lesions in 307 normal images (77.5%) were achieved. Positive and negative predictive values were 81.3% and 80.4%, respectively. Successful detection was achieved in 130 cases (94.9%). We achieved precise demarcation of EGC with a mean intersection over union of 66.5%, showing the extent of lesions with a smooth boundary; the results were comparable to those achieved by specialists.

**Conclusions** Our scheme, validated using 1-year consecutive cases, shows potential for demarcating EGC. Its performance matched that of specialists; it might therefore be suitable for clinical use in the future.

**Keywords** Computer-aided diagnosis · Early gastric cancer · White light endoscopy · Precise area demarcation · Delineation of horizontal extent

## Introduction

Early detection of gastric cancer is one of the most effective ways to reduce cancer death. Patients with advanced gastric cancer have a poor prognosis, whereas the 5-year survival rate is greater than 90% when the lesion area is localized as a result of being detected at an early stage [1–4]. Gastrointestinal endoscopy plays a critical role in improving the detection rate of EGC and enables early intervention with endoscopic submucosal dissection (ESD) through precise area diagnosis. However, endoscopic diagnosis including detection and demarcation is difficult even for well-trained gastroenterologists because a lesion's morphology is usually only slightly depressed or elevated and, because of chronic inflammation, lesions exhibit only a minor color change relative to the background [5–7].

✉ Hideo Yokota  
hyokota@riken.jp

<sup>1</sup> Image Processing Research Team, Center for Advanced Photonics, RIKEN, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan  
<sup>2</sup> Department of Gastroenterology and Endoscopy, National Cancer Center Hospital East, Kashiwa, Japan  
<sup>3</sup> Division of Science and Technology for Endoscopy, Exploratory Oncology Research and Clinical Trial Center, National Cancer Center, Kashiwa, Japan  
<sup>4</sup> Department of Internal Medicine, Tsuyama Chuo Hospital, Tsuyama, Japan

Research in computer-aided lesion detection using endoscopic images or videos to assist endoscopists and equalize the diagnostic ability has been increasing in recent years [8–13]. Such research has strong potential to reduce cancer-care disparities caused by differences in endoscopists' proficiency or by differences among equipped endoscopes. In particular, substantial improvements in the detection of gastrointestinal lesions have been reported since the introduction of convolutional neural networks (CNNs) [14]. In a previous study, we developed an automatic detection scheme for EGC; this scheme was based on an efficiently trained CNN using a small training dataset, revealing the approximate extents of lesions [15]. Although this previous study indicated the feasibility of computer-aided lesion detection, the performance of the proposed method was insufficient for practical use because not only lesion detection but also demarcation of lesions is critical for curative resection for both ESD and partial gastrectomy [16–19].

In this study, we enhanced our detection scheme so that more detailed regions can be demarcated while keeping the efficient CNN trained using a small dataset and then carried out delineation of the horizontal extent of lesions. The proposed scheme was validated using 1-year consecutive cases. In addition, we conducted a pilot study with endoscopists to compare their ability to demarcate EGC with that of our enhanced scheme.

## Methods

### Acquisition of training and test datasets with accurately annotated images

To train the CNN, we collected 150 cancer images and 150 normal images randomly selected from 68 consecutive patients (926 images) with EGC who were treated with ESD [20–22] at National Cancer Center Hospital East (NCCHE) from November 2016 to April 2017. In the present study, “normal” means that the image does not contain cancerous regions. We next collected a test dataset completely independent from the training dataset; the test dataset comprised 462 images (137 patients) from consecutive EGC cases that underwent ESD from April 2017 to March 2018 at NCCHE. The dataset also included 396 normal images containing no cancer regions. All the images in both datasets were 24-bit, 1000 × 870 pixels, full-color images originally obtained from endoscopic still images captured under white light (GIF-H290Z or GIF-H260Z; Olympus, Tokyo, Japan) using a standard video endoscopy system (EVIS LUCERA ELITE; Olympus). On the basis of the ESD pathological results, an experienced and certified gastroenterologist (K.H.) annotated the lesions

on all cancer images to train and validate the CNN. The datasets included different lesion characteristics such as size, macroscopic type, location, invasion depth, and histological type (see Table 1 for details).

This study used a retrospective design, and patients' personal information was removed. The protocol was approved by the Institutional Review Board of the National Cancer Center (2017-090). In this study, the patients provided informed consent to participate in the study using an opt-out method through disclosure of a public document on the website of the facility because of the retrospective nature of the study using completely anonymized data.

### Efficient increase in training dataset

A CNN is one of the machine learning tools used in various tasks related to computer vision, such as image classification and segmentation. CNNs generally require huge amounts of training data. By contrast, an extremely small number of endoscopic images—specifically, only 150 cancer and 150 normal images—were collected for training the CNN in this study. The collected images were divided into approximately 1.13 million small images using the

**Table 1** Patient and lesion characteristics of the training and test datasets

Training dataset	
Patient characteristics ( <i>n</i> = 68)	
Sex: male/female	57/11
Median age: year (range)	75 (36–82)
Lesion characteristics ( <i>n</i> = 68)	
Median tumor size: mm (range)	16 (1–52)
Size: 0–10 mm/11–20 mm/ ≥ 21 mm	14/36/18
Macroscopic types: 0-I/0-IIa/0-IIc	3/11/54
Location: Upper/Middle/Lower	18/22/28
Depth: pM/pSM1/pSM2	57/7/4
Histological types: differentiated / undifferentiated	64/4
Test dataset	
Patient characteristics ( <i>n</i> = 137)	
Sex: male/female	117/20
Median age: year (range)	73 (36–85)
Lesion characteristics ( <i>n</i> = 137)	
Median tumor size: mm (range)	15 (1–61)
Size: 0–10 mm/11–20 mm/ ≥ 21 mm	40/61/36
Macroscopic types: types 0-I/0-IIa/0-IIb/0-IIc	2/9/1/125
Location: Upper/Middle/Lower	24/57/56
Depth: pM/pSM1/pSM2	101/18/18
Histological types: differentiated/undifferentiated	126/11

following procedure. First, each image was cropped automatically and randomly into approximately 100 images with a size of  $224 \times 224$  pixels. At least 80% of the pixels of each cropped image labeled as a cancer included lesions to enable the characteristics of lesions (e.g., the pit pattern, margin, and the central depression) to be identified. Conversely, the cropped images labeled as normal contained no cancer regions. Next, data augmentation, which increases the number of images via geometric or appearance image transformation, was performed to increase the cropped images artificially. We used Keras [23], an open-source neural network library, for data augmentation. Finally, 534,926 cancer images and 593,874 normal images with sizes of  $224 \times 224$  pixels were obtained (see Online Resource Fig. S1a).

### Training network by fine-tuning

Fine-tuning [24], a type of transfer learning, is a powerful tool for efficiently training a neural network model. GoogleNet [25], which is one of the simplest models and is trained using the datasets of the ImageNet large-scale visual recognition challenge [26], was adopted in this study as a pre-trained network for fine-tuning. The aforementioned training dataset, totaling approximately 1.13 million images, was then used to tune the pre-trained network to fit the network with the datasets of this study (Online Resource Fig. S1b). The training time was approximately 10 h with an Intel Xeon (2.4 GHz, 128 GB) and NVIDIA GeForce GTX1080 (8 GB  $\times$  2 memory). After this training, the trained network could classify whether an input was cancer or normal (noncancerous). We evaluated the generalization ability of the trained network by cross-validation—a technique used to evaluate machine learning algorithms—and confirmed that the trained network could perform well on previously unseen input (Online Resource Fig. S2).

### Detection and delineation by per-pixel prediction

We implemented a sliding-window procedure that performed an existence probability prediction of lesions for an unknown input image using the trained CNN, leading to semantic segmentation—a technique used for annotating images at the pixel level. The precise segmentation eventually enabled the delineation of lesions (i.e., demarcation of EGC). A short summary is presented as follows. An input endoscopic image with an original size of  $1000 \times 870$  pixels was divided into small blocks of  $W \times H$  (horizontal  $\times$  vertical) images with a size of  $224 \times 224$  pixels. Each block, which was not resized and therefore preserved fine image features, passed through the trained network and was then classified with a predicted

probability between 0.0 and 1.0 (i.e., block-based prediction). A block with high probability was considered to be cancer, whereas a block with low probability was considered to be normal (Online Resource Fig. S3). After all the blocks had been applied to the trained network, the predicted probability of each block was assigned to each pixel inside the original input image as the existence probability of EGC, according to a weighted function (Online Resource Fig. S4). Each block in which the probability prediction was performed was defined by being slid around the input image from the top left to the bottom right, and the variable sliding pitch determined the number of blocks. The neighboring blocks partially overlapped according to the sliding pitch, leading to multiple predictions for each pixel and increasing the prediction reliability. In this study, we divided each input image into  $43 \times 39$  (horizontal  $\times$  vertical) blocks by sliding around the input image every 28 pixels, resulting in 1677 predictions (6.7 s) being performed per input image.

The assigned and summed probabilities based on the image-based prediction were used and then translated into a pseudo-color that was subsequently superimposed onto the input image. The result provided the exact region of EGC as a heat map by selection of pixels whose probability was greater than 0.5 because our trained network is based on the concept of two-class classification, where the predicted pixel with a probability greater than 0.5 is regarded as cancer. If no misdetection is acceptable, the probability should be set lower than 0.5 to reduce the number of misdetections. However, our detection goal is the accurate delineation of EGC, and we therefore set the probability to 0.5. The detected cancer region was depicted as a rainbow-colored region, whereas the noncancerous (normal) region was depicted as a blue-colored region (Online Resource Fig. S1c).

### Evaluation methods and outcome measures

To validate our proposed scheme, we performed two types of evaluations for the test dataset: one evaluation for detection ability and another for demarcation accuracy. Detection ability refers to whether our scheme can correctly detect the presence of cancer in an input, and demarcation accuracy refers to the overlap rate of the detected region with the exact cancer region. The detection ability was calculated on the basis of the true-positive rate (TP), true-negative rate (TN), false-positive rate (FP), and false-negative rate (FN). The TP and FP reflect the numbers of images correctly and incorrectly recognized as EGC, respectively, whereas the TN and FN represent the numbers of images correctly and incorrectly recognized as normal, respectively. We calculated five metrics—sensitivity ( $TP/(TP + FN)$ ), specificity ( $TN/(FP + TN)$ ),

accuracy  $((TP + TN)/(TP + FN + FP + TN))$ , positive predictive value  $(PPV = TP/(TP + FP))$ , and negative predictive value  $(NPV = TN/(FN + TN))$ —for evaluating the detection ability. These evaluations were performed for block-based prediction, image-based prediction, and case-based prediction of the test dataset (Online Resource Table S1). The evaluation of block-based prediction reflects the detection ability before the predicted probabilities are summed across all the blocks in the sliding-window procedure. We examined the detection ability of 1,438,866 blocks obtained from 462 cancer and 396 normal images by counting the number of blocks when the prediction result matched the ground truth. The ground truth, which is co-located with the target block, is regarded as a cancer block when at least 80% of pixels of a block include cancer regions. For the evaluation for image-based prediction, if at least one cancer pixel is detected in the input target image, then the target image is counted as correctly detected. Regarding the evaluation for case-based prediction, for each image of the 137 cases, if at least one cancer image is detected correctly, then the target case is counted as a correctly detected one.

The demarcation accuracy was measured for the cancer images of the test dataset in terms of the mean intersection over union (mIoU) [27], which is a measure commonly used to evaluate the accuracy of semantic segmentation, which can quantify the degree of overlap between a segmented region and a ground truth. The images that show the lesion extent annotated by experienced gastroenterologists on the basis of the pathological examination of ESD were used as the ground truth for calculating the mIoU.

We also compared the automatically delineated regions of our scheme with the regions delineated via manual annotation by six experienced endoscopists. The average experience number of upper gastrointestinal endoscopy procedures performed by the endoscopists was 6000 (min 3000 to max 12,000). The randomly selected 38 cancer images among the successfully detected cases in the test dataset, and their ground truths, which were precisely detected lesions annotated on the basis of pathological examination, were used for the comparison. When annotating each image, the endoscopists were permitted to reference a white-light-only endoscopic image without magnification. The ratio of lesion characteristics was in approximate accordance with that of the test dataset. In terms of each selected image, the mIoU, sensitivity, specificity, accuracy, PPV, and NPV were calculated between the ground truth and the manual annotation.

## Results

### Evaluation of the detection ability using consecutive cases throughout 1 year

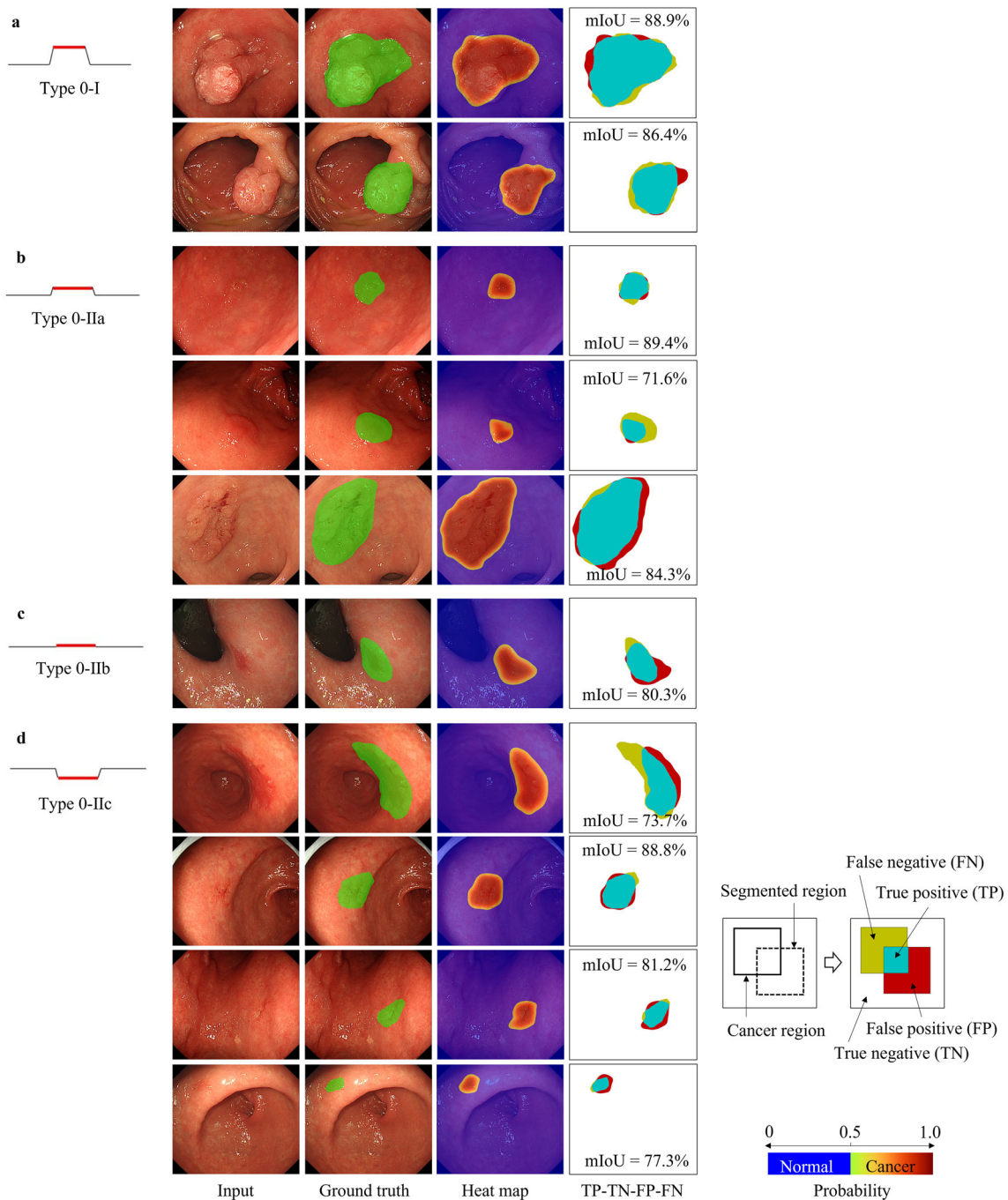
Table 2 presents the detection ability based on the numbers of images and cases. For the image-based result, EGC detection was accomplished in a total of 387 (83.8%) out of 462 cancer images. For the normal cases, 307 (77.5%) out of 396 normal images were predicted correctly. A PPV of 81.3%, NPV of 80.4%, and accuracy of 80.9% were obtained, all of which achieved stable detection accuracy. The PPV of the block-based results was low (Online Resource Table S1), whereas that of the image-based results was improved. This result means that the detection ability was substantially improved in terms of FN occurrence because of the sliding-window procedure. We speculate that multiple predictions performed at each pixel by the sliding-windows procedure led to stable detection accuracy. Notably, from the perspective of the case-based results, our proposed method achieved a detection accuracy of 94.9%.

### Heat map of cancer probability delineating the horizontal extent of lesion

Figure 1a–d shows the results of demarcation for three macroscopic types of gastric cancer: 0-I, 0-IIa, 0-IIb, and 0-IIc. A heat map shows not only the extents of lesions but also the probability of lesions. In addition, the margins of cancer regions can be observed at the pixel level. Although the lesion size ranged from approximately 70 square pixels to 650 square pixels in each image, the demarcation results were accurate irrespective of the lesion size. Especially in

**Table 2** Detection ability based on the numbers of images and cases

Image-based ( <i>n</i> = 858)		Prediction		
		Cancer	Normal	
Ground truth	Cancer	TP: 387	FN: 75	Sensitivity: 83.8%
	Normal	FP: 89	TN: 307	Specificity: 77.5%
		PPV: 81.3%	NPV: 80.4%	Accuracy: 80.9%
Case-based ( <i>n</i> = 137)		Prediction		
		Cancer	Normal	
Ground truth	Cancer	TP: 130	FN: 7	Accuracy: 94.9%



**Fig. 1** Demarcation results of four early gastric cancer (EGC) types: **a** protruding type (type 0-I); **b** superficial elevated type (type 0-IIa); **c** superficial flat type (type 0-IIb); **d** superficial depressed type (type 0-IIc). Each row shows the following images: an input image, a ground truth, a heat map showing the predicted probability of EGC, and a pseudo-colored TP–TN–FP–FN image. In the ground truth images, the manually annotated regions, which were based on the

pathological examination by a gastroenterologist (K.H.), are marked in green. The heat map images, which were produced fully automatically by our proposed scheme, indicate the potential cancer and normal regions detected on the basis of the predicted probability. In the TP–TN–FP–FN images, light blue, white, red, and yellow indicate true positive, true negative, false positive, and false negative, respectively

the cases of types 0-IIa, 0-IIb, and 0-IIc, a few salient features were observed in the shapes. Nevertheless, our proposed scheme provided accurate pixel-level demarcation of the lesion across various macroscopic types of EGC.

In terms of incorrect detection, irregular textured patterns on the gastric surfaces resulted in over-segmentation and mis-segmentations were observed in regions with shadows or in deeper areas (Online Resource Fig. S5).

The mIoU calculated from the successfully detected cancer images (387/462) was 66.5% on average. In general, with respect to the degree of overlap of two regions, we considered an overlap greater than 65% for the target region to be sufficient because moving the two regions diagonally by only 10% gives an mIoU of 65% (Online Resource Fig. S6).

### Analyzing the results in terms of lesion characteristics

Figure 2 shows the demarcation accuracy expressed using the mIoU in terms of lesion size, depth, location, and histological type. The mIoU was significantly better in images of lesions of pSM2 (vs. pM,  $p < 0.0001$ ; vs. pSM1,  $p < 0.05$ ), lesions at the upper stomach (vs. middle,  $p < 0.05$ ), and lesions of the histologically differentiated type (vs. undifferentiated type,  $p < 0.001$ ). There were no significant differences in mIoU with regard to the lesion size.

We also analyzed the correct-detection rate in terms of lesion characteristics such as the size, depth, location, and histological type of the lesion (Table 3). Similar to the results in Fig. 2, lesions with deeper invasion depths and that were located in the upper part of the stomach and were of the differentiated type were likely to be detected more precisely using our scheme. This tendency of detection ability was the same as that of the endoscopists. Notably, according to the case-based results, our proposed method resulted in a correct-detection rate of approximately 95%.

### Comparison with the manual demarcation by experienced endoscopists

Table 4 shows the results of a comparison between the computer detection and the averaging results of six endoscopists in terms of six evaluation metrics. In the three evaluation metrics of mIoU, sensitivity, and NPV, our scheme demonstrated superior diagnostic ability compared with that of the endoscopists. However, in the other three metrics of specificity, accuracy, and PPV, the diagnostic ability of the endoscopists was superior to that of our scheme. The small differences regarding specificity and accuracy verified that the diagnostic ability of our scheme was comparable to that of the endoscopists.

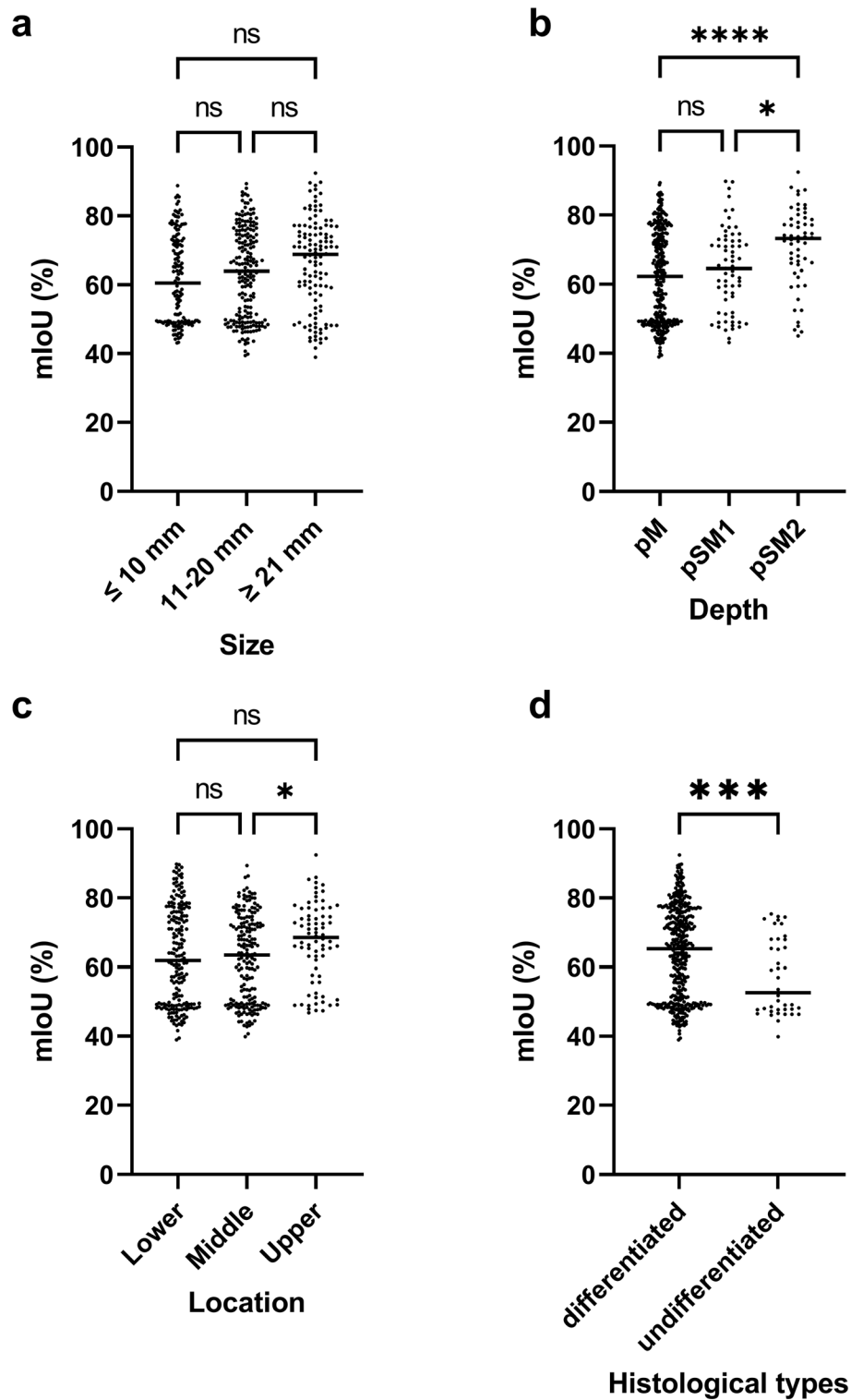
Figure 3 shows three representative examples of lesion demarcation. Similar to the trends shown in Table 3, our scheme tends to lead to over-segmentation compared with the lesion demarcations of the endoscopists. However, the lesion demarcations of the endoscopists tend to be mis-segmented. The results show that the annotated regions of lesions varied substantially from each other.

## Discussion

We developed a scheme for the computer-aided demarcation of EGC that demonstrates area diagnosis ability similar to that demonstrated by endoscopists. The demarcation accuracy of the correctly detected EGC, expressed using mIoU, was 66.5%, which is considered a sufficient degree of overlap with exact cancer regions because mIoU is an extremely strict evaluation metric. For lesions with ambiguous boundaries, such as those in EGC, our pixel-level lesion delineation will assist endoscopists in understanding the spread of lesions and aid in the development of an appropriate treatment plan for ESD.

In a previously reported method that enabled the real-time detection of upper gastrointestinal cancer, greater than 90% sensitivity and specificity were achieved using approximately 1 million images (approximately 84,000 cases) [28]. However, the target lesions for automatic detection were primarily advanced cancer (81% of the total), which deviated from the concept of automatic detection at the early stage of gastric cancer. Ping et al. proposed an automatic delineation of EGC [29]. However, their aim was to delineate the lesion area only in the cancer image; they did not attempt to detect the presence of lesions. We attribute the difficulties of automatic EGC detection and delineation to two primary reasons. First, there is a lack of large amounts of high-quality data for use in training datasets [30]. By contrast, our scheme does not focus on collecting large amounts of training data but instead uses precise training data based on pathological examination—in this study, originally from only 150 cancer images and 150 normal images. The high-quality training data appear to have positively affected the training accuracy, even for one of the simplest CNNs such as GoogLeNet. Our scheme, which overcomes the problem regarding training data collection, has the potential to be readily applied in a wide range of detection tasks in other hospitals and clinics, even with other gastrointestinal diseases. Normally, the appearance of images acquired by different facilities or devices differs because of differences in the imaging conditions and methods; thus, re-training using the images from those facilities or devices is necessary for accurate detection. In such cases, the ability to train with a small training dataset is considered a major advantage. Second, many previous studies utilized a one-stage detection framework typified by the single-shot multibit detector (SSD) [31]. We speculate that the SSD-based scheme might be linked to the low PPV [10, 28]. The SSD uses low-resolution images derived from compressed images (e.g., the SSD-300 model compresses the original image to  $300 \times 300$  pixels); thus, the detailed texture pattern of the images is missing and the lesions with flat

**Fig. 2** Differences in mean intersections over unions (mIoUs) of each lesion characteristic. Scatter plots of mIoUs in terms of **a** lesion size, **b** lesion depth, **c** lesion location, and **d** histological type of lesion. The differentiated lesions include tub, tub1, tub2, pap, and carcinoma, and the undifferentiated lesions include sig, por, and por2. In each plot, the *p*-value was calculated using the Kruskal–Wallis test for the size, depth, and location and using the Mann–Whitney *U* test for the histological types. \**p*<0.05, \*\*\**p*<0.001, \*\*\*\**p*<0.0001. *ns* no significance, *p*≥0.05



and indecisive boundaries, typified by EGC, may not be correctly detected. Our scheme, by contrast, detected the minor differences in the local texture patterns around the lesion boundary as a result of multiple predictions derived from the sliding-window procedure performed using a

CNN trained using uncompressed images. Our particularly striking result is that the occurrence of both FPs and FNs was strongly suppressed for both cancer and normal images and that high-performance results in both PPV and NPV were eventually achieved.

**Table 3** Correct-detection rate based on the lesion characteristics

Lesion characteristics	Correct-detection (Image-based)	Correct-detection (Case-based)
Size (mm)		
≤ 10	104/147 (70.7%)	38/40 (95.0%)
11–20	162/197 (82.2%)	58/61 (95.1%)
≥ 21	103/118 (87.3%)	34/36 (94.4%)
Depth		
pM	264/341 (77.4%)	95/101 (94.1%)
pSM1	54/65 (83.1%)	17/18 (94.4%)
pSM2	51/56 (91.1%)	18/18 (100.0%)
Location		
Lower	151/195 (77.4%)	51/56 (91.1%)
Middle	151/191 (79.1%)	55/57 (96.5%)
Upper	67/76 (88.2%)	24/24 (100.0%)
Histological types		
Differentiated	344/422 (81.5%)	121/126 (96.0%)
Undifferentiated	25/40 (62.5%)	9/11 (81.8%)

**Table 4** Comparison of the delineation accuracy with six endoscopists

	mIoU	Sensitivity	Specificity	Accuracy	PPV	NPV
Our scheme	68.7	69.9	94.0	91.7	62.9	96.3
Average of six endoscopists	68.3 (4.0)	49.8 (12.5)	98.9 (1.8)	93.3 (0.9)	89.0 (10.2)	93.8 (1.3)

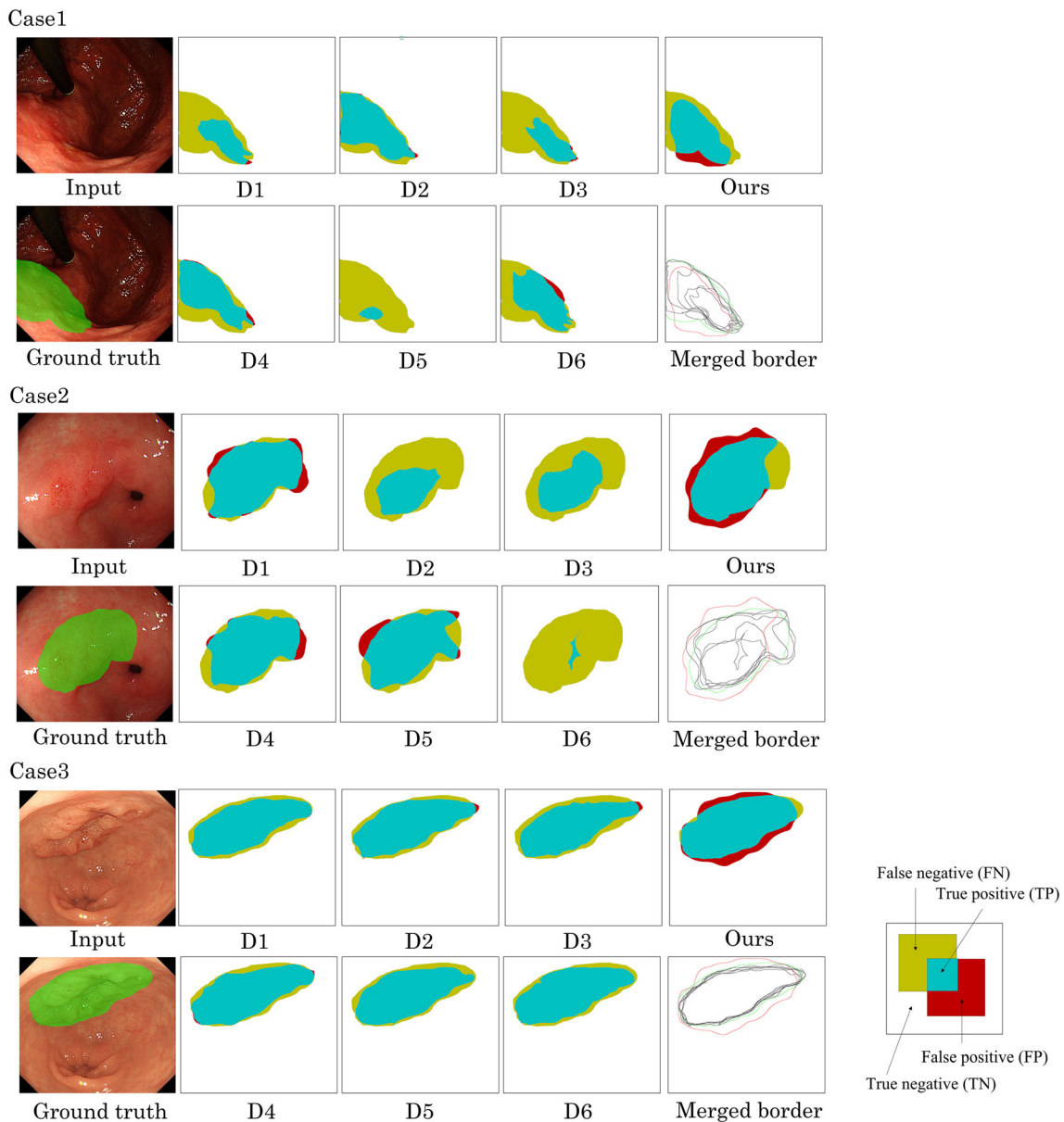
All results are given as percentage (standard deviation)

Although we did not explore methods to shorten the detection time, numerous options are available for this purpose. For example, our scheme can offer coarse or fine segmentation results based on the sliding pitch selected for the sliding windows (Online Resource Fig. S7). As for the probability prediction of each block in the sliding-window procedure, the coarse sliding pitch is first adopted for rough prediction and then only the blocks with high probabilities are re-predicted for the fine prediction. We speculate that this approach will shorten the detection time to enable real-time diagnosis in the future. Alternatively, this pitch selection could be applied to meet the needs of endoscopists to balance segmentation quality and execution time. For example, the advantage of making the sliding pitch small is that more accurate segmentation results with a smooth outline can be obtained, which would be useful for area diagnosis before ESD in our method. Conversely, a large sliding pitch has the advantage of decreased execution time as the number of predictions decreases. The large sliding pitch might contribute to the application of mass screening programs or real-time detection using high-quality video endoscopy in the future. Because a tradeoff exists between the segmentation level and execution time, parameters that suit the needs of the endoscopist must be selected. In addition, the sliding-window procedure can be

performed in parallel via GPU implementation, thereby accelerating the detection speed.

This study had several limitations. First, our dataset was biased in the number of training and test images according to the lesion characteristics. For example, the number of undifferentiated-type images was less than one-tenth that of differentiated-type images. This difference might be responsible for the lower detection accuracy of the undifferentiated-type images compared with that of the differentiated-type images. The effect of this bias on the detection accuracy should be investigated in future work. Second, the proposed scheme was evaluated using retrospective datasets at a single institution with a relatively small number of patients. Additional studies with larger sample sizes from multiple institutions would increase the generalization capability of our proposed scheme. Third, the scheme was trained and tested on images from limited image environments captured using an Olympus video endoscopy system. Additional studies using datasets from multiple image environments may lead to hardware-agnostic prediction. Fortunately, our scheme does not require a large amount of training data and can be readily applied to images from other institutions and imaging environments. In addition, we recently developed a system to accumulate annotated endoscopic images from a daily





**Fig. 3** Comparison of the demarcation accuracy of the proposed method with that achieved by experienced endoscopists. Three representative examples (Cases 1–3) of the computer detection and those of manual detection by six endoscopists. In each result, the upper row shows the input endoscopic image, the results of manual detection by three endoscopists (D1–D3), and the results of detection

by our method (Ours). The lower row shows the ground truth indicating the region of EGC, the results of manual detection by three endoscopists (D4–D6), and the merged border of regions detected by six endoscopists and by our method. In the merged border image, the green line shows the real border of EGC and the red line shows the border detected by our scheme

reporting system, which could make training data collection more convenient in the future [32].

In conclusion, we demonstrated that our scheme achieved a high performance of computer-aided area diagnosis of EGC via a CNN trained using an extremely small number of endoscopic images. A particularly striking result is that our scheme offered pixel-level demarcation of EGC with well-balanced accuracy for both cancer and normal images and demonstrated area diagnosis ability similar to that of experienced endoscopists. In the

future, we believe that our scheme will offer substantial assistance in the decision-making of gastroendoscopists for curative resection of EGC by ESD.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00535-023-02001-x>.

**Acknowledgements** This work was supported by JSPS KAKENHI Grant Number JP21K12742 and by the National Cancer Center research and development fund (29-A-10, 2020-A-10).

**Funding** Japan Society for the Promotion of Science, JP21K12742, Satoko Takemoto, The National Cancer Center research and development fund, 29-A-10, Tomonori Yano, 2020-A-10, Tomonori Yano.

## References

- Sano T, Coit DG, Kim HH, et al. Proposal of a new stage grouping of gastric cancer for TNM classification: international gastric cancer association staging project. *Gastric Cancer*. 2017;20:217–25.
- Katai H, Ishikawa T, Akazawa K, Registration Committee of the Japanese Gastric Cancer Association, et al. Five-year survival analysis of surgically resected gastric cancer cases in Japan: a retrospective analysis of more than 100,000 patients from the nationwide registry of the Japanese gastric cancer association (2001–2007). *Gastric Cancer*. 2018;21:144–54.
- Ono H, Kondo H, Gotoda T, et al. Endoscopic mucosal resection for treatment of early gastric cancer. *Gut*. 2001;48:225–9.
- Necula L, Matei L, Dragu D, et al. Recent advances in gastric cancer early diagnosis. *World J Gastroenterol*. 2019;25:2029–44.
- Menon S, Trudgill N. How commonly is upper gastrointestinal cancer missed at endoscopy? A meta-analysis. *Endosc Int Open*. 2014. <https://doi.org/10.1055/s-0034-1365524>.
- Hosokawa O, Hattori M, Douden K, et al. Difference in accuracy between gastroscopy and colonoscopy for detection of cancer. *Hepatogastroenterology*. 2007;54:442–4.
- Choi KS, Jun JK, Park EC, et al. Performance of different gastric cancer screening methods in Korea: a population-based study. *PLoS ONE*. 2012;7: e50041.
- Mori Y, Kudo SE, Misawa M, et al. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. *Ann Intern Med*. 2018;169:357–66.
- Urban G, Tripathi P, Alkayali T, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. *Gastroenterology*. 2018;155:1069–1078.e8.
- Hirasawa T, Aoyama K, Tanimoto T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. *Gastric Cancer*. 2018;21:653–60.
- Min JK, Kwak MS, Cha JM. Overview of deep learning in gastrointestinal endoscopy. *Gut Liver*. 2019;13:388–93.
- Ebigbo A, Palm C, Probst A, et al. A technical review of artificial intelligence as applied to gastrointestinal endoscopy: clarifying the terminology. *Endosc Int Open*. 2019;7:E1616–23.
- Choi J, Shin K, Jung J, et al. Convolutional neural network technology in endoscopic imaging: artificial intelligence for endoscopy. *Clinical Endosc*. 2020;53:117–26.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–44.
- Sakai Y, Takemoto S, Hori K, et al. Automatic detection of early gastric cancer in endoscopic images using a transferring convolutional neural network. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2018; 4138–4141.
- Pimentel-Nunes P, Dinis-Ribeiro M, Ponchon T, et al. Endoscopic submucosal dissection: European society of gastrointestinal endoscopy (ESGE) guideline. *Endoscopy*. 2015;47:829–54.
- Nagahama T, Yao K, Uedo N, et al. Delineation of the extent of early gastric cancer by magnifying narrow-band imaging and chromoendoscopy: a multicenter randomized controlled trial. *Endoscopy*. 2018;50:566–76.
- Nagahama T, Yao K, Maki S, et al. Usefulness of magnifying endoscopy with narrow-band imaging for determining the horizontal extent of early gastric cancer when there is an unclear margin by chromoendoscopy (with video). *Gastrointest Endosc*. 2011;74:1259–67.
- Nonaka K, Namoto M, Kitada H, et al. Usefulness of the DL in ME with NBI for determining the expanded area of early-stage differentiated gastric carcinoma. *World J Gastrointest Endosc*. 2012;4:362–7.
- Gotoda T, Yamamoto H, Soetikno RM. Endoscopic submucosal dissection of early gastric cancer. *J Gastroenterol*. 2006;41:929–42.
- Isomoto H, Shikuwa S, Yamaguchi N, et al. Endoscopic submucosal dissection for early gastric cancer: a large-scale feasibility study. *Gut*. 2009;58:331–6.
- Gotoda T, Jung HY. Endoscopic resection (endoscopic mucosal resection/ endoscopic submucosal dissection) for early gastric cancer. *Dig Endosc*. 2013;25:55–63.
- Chollet F, Keras. GitHub repository. 2015. <https://github.com/fchollet/keras>
- Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. *J Big Data*. 2016. <https://doi.org/10.1186/s40537-016-0043-6>.
- Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. In: 2015 IEEE Conference on computer vision and pattern recognition (CVPR). 2015; <https://doi.org/10.1109/CVPR.2015.7298594>.
- Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis*. 2015;115:211–52.
- Everingham M, Gool L, Williams CK, et al. The PASCAL visual object classes (VOC) challenge. *Int J Comput Vis*. 2010;88:303–38.
- Luo H, Xu G, Li C, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. *Lancet Oncol*. 2019;20:1645–54.
- An P, Yang D, Wang J, et al. A deep learning method for delineating early gastric cancer resection margin under chromoendoscopy and white light endoscopy. *Gastric Cancer*. 2020;23:884–92.
- Jin P, Ji X, Kang W, et al. Artificial intelligence in gastric cancer: a systematic review. *J Cancer Res Clin Oncol*. 2020;146:2339–50.
- Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector. *Lecture Notes in Computer Science*. Springer, Cham. 2016; 9905:21–37.
- Hori K, Ikematsu H, Yamamoto Y, et al. Detecting colon polyps in endoscopic images using artificial intelligence constructed with automated collection of annotated images from an endoscopy reporting system. *Dig Endosc*. 2022;34:1021–9.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.