Automatic detection of early gastric cancer in endoscopic images using a transferring convolutional neural network

Y. Sakai, S. Takemoto, K. Hori, M. Nishimura, H. Ikematsu, T. Yano and H. Yokota

Abstract— **Endoscopic image diagnosis assisted by machine learning is useful for reducing misdetection and interobserver variability. Although many results have been reported, few effective methods are available to automatically detect early gastric cancer. Early gastric cancer have poor morphological features, which implies that automatic detection methods can be extremely difficult to construct.**

In this study, we proposed a convolutional neural network-based automatic detection scheme to assist the diagnosis of early gastric cancer in endoscopic images. We performed transfer learning using two classes (cancer and normal) of image datasets that have detailed texture information on lesions derived from a small number of annotated images. The accuracy of our trained network was 87.6%, and the sensitivity and specificity were well balanced, which is important for future practical use. We also succeeded in presenting a candidate region of early gastric cancer as a heat map of unknown images. The detection accuracy was 82.8%. This means that our proposed scheme may offer substantial assistance to endoscopists in decision making.

I. Introduction

Gastric cancer accounts for one of the highest morbidity rates among all kinds of cancers. Nevertheless, people with early gastric cancer rarely feel any symptoms. Even when the salient symptoms appear as the cancer progresses, they are similar to those of gastritis and gastric ulcers; thus, it is difficult for patients to realize that they are already in the advanced stages of gastric cancer. Therefore, the early detection of gastric cancer using endoscopic images is required. However, image diagnosis of early gastric cancer is difficult even for gastroenterologists, and the diagnosis accuracy depends on the experience of the gastroenterologist.

Recently, for the purpose of image diagnosis assistance based on machine learning techniques, many studies on automatic polyp detection by endoscopy have been published [1]. However, to the best of our knowledge, the technique for early gastric cancer detection has not been established for practical use, which can be attributed to two reasons. First, the data related to early gastric cancer that can be applied to machine learning have not been maintained sufficiently. Second, many early gastric cancers have poorer morphological features than those of progressive gastric cancers, which renders automatic detection very difficult. The morphological features of early gastric cancer are primarily classified into the following three types: superficial elevated type (type 0-IIa), superficial depressed type (type 0-IIc), and protruding type (type 0-I). Fig. 1 shows the sample endoscopic images of each type. The red rectangles indicate each gastric cancer. Even for gastroenterologists, detecting lesions of types 0-IIa and 0-IIc is occasionally difficult because of their poor morphological changes. Therefore, the diagnosis of early gastric cancer relies on identifying the slight changes in mucosal color and the irregular texture patterns of the underlying submucosa vessels.

A remarkable method for observing this texture pattern is the image enhanced endoscopy (IEE) such as narrow-band imaging (NBI) and blue-laser imaging (BLI), which are highly effective under the use of magnification in producing a definite diagnosis of gastrointestinal cancer compared with the conventional white-light imaging (WLI) endoscopy [2]. Several studies have reported the usefulness of IEE for detecting gastrointestinal neoplasm. NBI with magnifying endoscopy is useful for detecting esophageal and pharyngeal cancers [3]; recently, BLI is reported to be useful for detecting colon neoplasm [4]. However, the efficacy of IEE for detecting gastric cancer has not been reported because of its low brightness. Moreover, the morphological changes in gastric cancer are difficult to detect by the background mucosal change in gastritis; hence, an automatic detection technique for early gastric cancer using WLI endoscopy is strongly required.

In this study, we focus on the computer-aided diagnosis of early gastric cancer that can assist gastroenterologists in decision making. Prior to this work, we acquired approximately 1,000 images of early gastric cancer, especially those of types 0-I, 0-IIa, and 0-IIc captured by WLI, and more than 200 of these images had been provided with the ground truth indicating regions of lesions. The remaining images were certified as "normal" by a gastroenterologist. All the lesions were certified to be early gastric cancer based on the

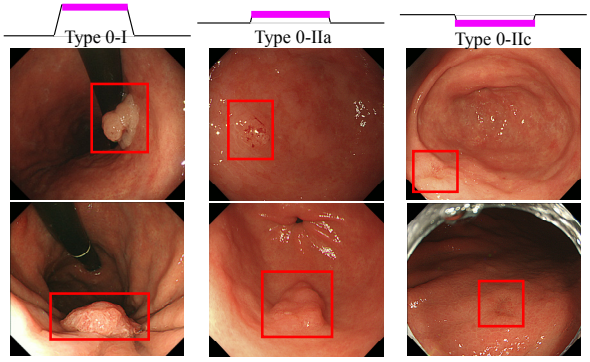

Figure 1 Primary types of early gastric cancer

Y. S., S. T., M. N. and H. Y*. are with Image Processing Research Team, RIKEN Center for Advanced Photonics. 2-1 Hirosawa, Wako, Saitama, 351-0198 Japan (*corresponding author to provide phone: +81-48-462-1293 e-mail: hyokota@riken.jp).

K. H., H. I. and T. Y. are with Department of Gastroenterology and Endoscopy, National Cancer Center Hospital East. 6-5-1 Kashiwanoha, Kashiwa, Chiba, 277-8577 Japan.

Y. S. is also with Graduate School of Information Systems, the University of Electro-Communications. 1-5-1 Chofugaoka, Chofu, Tokyo 182-8585 Japan.

978-1-5386-3646-6/18/$31.00 ©2018 IEEE

retroactive observation after an endoscopic submucosal dissection (ESD) [5].

We herein present our preliminary work on the automatic detection of early gastric cancer based on a convolutional neural network (CNN) [6]. The major contributions of this work are as follows: 1) successful automatic detection of early gastric cancer with poor morphological features, which are occasionally difficult to detect even by endoscopists; 2) construction of our CNN-based detection scheme on a small number of image datasets; and 3) provide readers with some examples of misdetected early gastric cancer. The characteristics of misdetected images could be useful for researchers to construct more powerful methodologies of early gastric cancer detection for practical use.

The remainder of the paper is organized as follows: In section 2, some related works are described. Our proposed scheme and methods are described in section 3. The experimental results and discussion are presented in section 4.

II. Related Work

The research on automatic cancer detection for endoscopic diagnosis has primarily focused on colonoscopic polyp detection (e.g., [7]). Many have utilized the morphological features of polyps, such as the edge shape. A study has reported the construction of an appearance model of polyps, for which region segmentation was performed based on the degree of matching with the model [8]. Shortly thereafter, most methods focused on the strategy to combine feature extraction and its classifiers. Dramatic changes in detection accuracy, however, have been occurring since the upsurge in the utility of the CNN (e.g. [9]). The CNN automatically extracts, learns, and classifies features directly from the input datasets. However, it is well known that CNN-based methods require the preparation of a large number of annotated datasets. Therefore, this is a significant challenge to overcome especially for medical images. Hence, some studies utilized transfer learning, which is a technique that utilizes a pre-trained network using any other large number of learning datasets. Transfer learning has a higher accuracy than a network fully trained from scratch, especially for medical images [10]. Recently, the Single Shot MultiBox Detector [11], a state-of-the-art method for object detection, was applied to detect various kinds of gastric cancers with 92.2 % sensitivity and a positive predictive value (PPV) of 30.6 % [12]. The reasons for the low PPV compared to the high sensitivity might be that early gastric cancer has few morphological features and is similar to gastritis. To mitigate these problems, we propose a transferring CNN model that is fine-tuned via the detailed texture information of two kinds of images: cancer and normal. The proposed model achieved well-balanced accuracy in terms of sensitivity and specificity, and was able to reveal the approximate locations of early gastric cancers.

III. Our Approach

Our proposed scheme is divided into four segments: preparation of training datasets, transfer learning for the CNN, prediction of lesion presence, and visualization of cancer likelihood as a heat map (Fig. 2). All the segments and their related methods are explained in the following subsections.

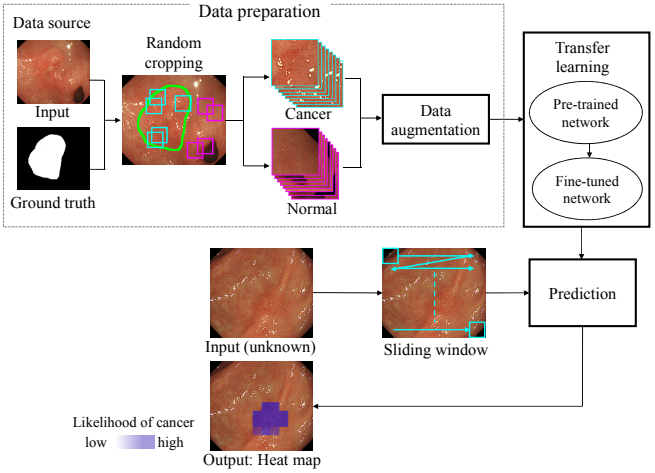

Figure 2 Our proposed prediction scheme for early gastric cancer detection

A. Preparation of training datasets

Our image datasets were originally obtained from an endoscopic video under white light (GIF-H290Z or GIF TYPE H260Z, Olympus Optical, Tokyo, Japan) and a standard video endoscopy system (EVIS LUCERA ELITE, Olympus Optical). In this work, we used the 24-bit full-color images of size 1000 × 870 pixels extracted from the video frames. The images, including the lesions, were confirmed as cancer images only for early gastric cancer, which were obtained by tracing these back to 58 patients who had been treated with ESD. A total of 926 images were collected, of which 228 images included more than one lesion in early gastric cancer. The remaining were noncancerous images (that is, normal images). All the lesions in the cancer images were identified manually by a gastroenterologist. The identified lesions were translated into binary images and used as the ground truth, as shown in Fig. 2.

As early gastric cancers have few morphological features, the detailed texture delivers valuable information for the training. To obtain their information, we cropped approximately 100 images of size 224 × 224 pixels randomly from each of the 100 cancer images that were selected from the 228 cancer images. Each cropped image included over 80% cancerous regions, according to the corresponding ground truth. Subsequently, we were able to obtain 9,587 cancer images for the training datasets. Conversely, 9,800 normal images of size 224 × 224 pixels, which did not include cancerous regions, were randomly cropped from the entire normal and cancer images as the training datasets. For the test datasets, 4,653 cancer images and 4,997 normal images were obtained from the unused cancer and normal images similarly.

For training a CNN, a large number of training datasets are required for reliable learning. Acquiring the annotated training datasets, however, is highly expensive, especially from medical images. As an alternative, data augmentation is generally used to increase the number of training datasets via a geometric or appearance transformation. Keras [13], an open-source neural network library written in Python, was used for data augmentation. Using nine kinds of augmentations, including rotation, shift, shear, zoom, and flip, which were applied twice, we obtained 172,555 cancer images and 176,388 normal images of size 224 × 224 pixels.

B. *Transfer learning and validation*

Transfer learning is a powerful tool for tuning the CNN parameters, and is commonly used to avoid full tuning from scratch [14]. Transfer learning requires two domains of annotated datasets: source domain and target domain. The source domain typically has a large number of annotated datasets and is used for training a CNN from scratch, which is called a pre-trained network. To satisfy the present task, a target domain was used for fine-tuning the parameters of the pre-trained network. GoogLeNet [15], consisting of 22 convolutional layers, was adopted in this study as the pre-trained network, which was trained on the datasets of ImageNet large scale visual recognition challenge [16]. As the target domain, the training datasets mentioned in the previous subsection were used for constructing a fine-tuned network to predict the occurrence of cancer from an image.

For the fine-tuning, we trained the initial network for 50 epochs with learning rates of 0.0001 and 0.00001 before and after 34 epochs, respectively. The network parameters were optimized using stochastic gradient descent with a mini-batch size of 32. As mentioned in the previous subsection, we used 172,555 and 176,388 cancer and normal images, respectively, for the fine-tuning. We then applied the new network to test the datasets mentioned earlier, i.e., 4,653 and 4,997 images, and evaluated them using three metrics: sensitivity, specificity, and accuracy (shown in Table 1). Furthermore, the result shows the correct detection with a high PPV of 93.4%. Both the training and the test were performed on the NVIDIA deep learning GPU training system (DIGITS, Version 5). The training time was approximately 10 h on the Intel Xeon (2.4 GHz, 128GB memory), and the NVIDIA GeForce GTX1080 (8GB × 2 memory). Additionally, we examined the ROC curve to measure the performance of both networks with and without data augmentation. As shown in Fig. 3, the network with data augmentation had a higher performance ability in this prediction.

TABLE I. Evaluation results of the trained network

		Prediction		
		Positive	Negative	
Ground truth	Positive (cancer)	True positive (TP) 3,723	False negative (FN) 930	Sensitivity 80.0%
	Negative (normal)	False positive (FP) 262	Ture negative (TN) 4,735	Specificity 94.8%
				Accuracy 87.6%

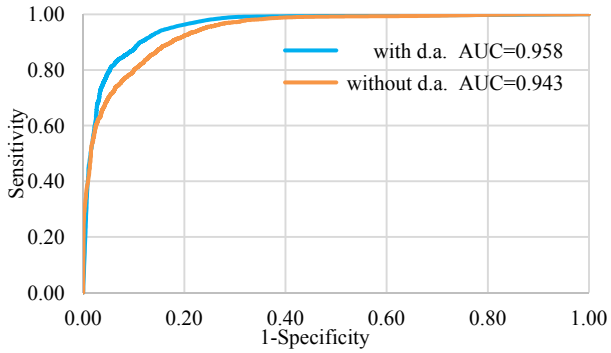

Figure 3 ROC curve of the fine-tuned network

C. *Detecting lesions on sliding window*

Using the fine-tuned network, an automatic detection of early gastric cancer was performed on a total of 926 endoscopic images of size 1000 × 870 pixels. The detection procedure is as follows: Based on a sliding window, from the top left to the bottom right, each image was divided into block images of 10 horizontal images by 9 vertical images, where the size of each is 224 × 224 pixels. The neighboring blocks partly overlapped. Each block image was applied to the network, and then the likelihood of early gastric cancer, which means the existence probability, was assigned. The output was translated into a pseudo color whose density indicates the difference in the likelihood, and was superimposed on the input image. After these procedures, we were able to obtain the heat map, which assists endoscopists with the useful information for image diagnosis.

IV. Results and Discussion

Some examples of our detection results of early gastric cancer are shown in Fig. 4. Given the three types of cancers in our datasets, the results are shown according to each type. The manually identified regions are marked by green, and blue indicates the potentially cancerous regions by our prediction. The target size can range from approximately 50 squared pixels to approximately 900 squared pixels. Especially in the case of types 0-IIa and 0-IIc, few salient features occurred in their shapes. Nevertheless, our proposed scheme accomplished the accurate detection in the various images across the three types of early gastric cancer.

We examined the detection accuracy by counting the number of images when at least one block matches the ground truth of the images with the existing cancer. Consequently, the detection success was accomplished in a total of 205 images (89.9%) out of 228 cancer images. The accuracy, when excluding the images used for learning, was 82.8%. Conversely, a total of 491 images (70.3%) out of 698 normal images were correctly predicted as "normal." When considering the correspondence of each block, 15,102 blocks out of 20,520 blocks (73.6%) were correctly predicted in the cancer images, and 62,096 blocks out of 62,820 blocks (98.8%) were correctly predicted in the normal images. The processing time was 4 ms per image, except for the time required to input/output the image.

Fig. 5 shows the representative examples of the misprediction. Figs. 5a and 5b show the over-detection (false positive) and the misdetection (false negative) in the cancer images, respectively. Conversely, Figs. 5c and 5d show the over-detection in the normal images. The over-detected regions were not confirmed as truly normal because target biopsies had not been conducted from the regions; however, the prescreening for other gastric cancers before the treatment of ESD for the primary region has been conducted, and additionally a retrospective check by experienced gastroenterologist for the obtained datasets confirmed them to be noncancerous regions. As shown in the enlarged images of Figs. 5b and 5d, the over-detected regions have obvious irregular texture patterns on their surface, which might have negatively affected the prediction. Moreover, misdetection may occur when the target regions are out of focus or are located in deeper areas. To improve the detection accuracy, it

might be useful to deliberately add such poor-quality data in the ground truth.

In summary, we demonstrated that our preliminary CNN-based prediction scheme achieved high accuracy in early gastric cancer detection using a small number of learning datasets. A particularly striking result is that our prediction scheme had a well-balanced accuracy for both cancer and normal images. In the future, we believe that it is possible to improve our prediction results by adding images with surface features similar to that of gastritis to the training datasets.

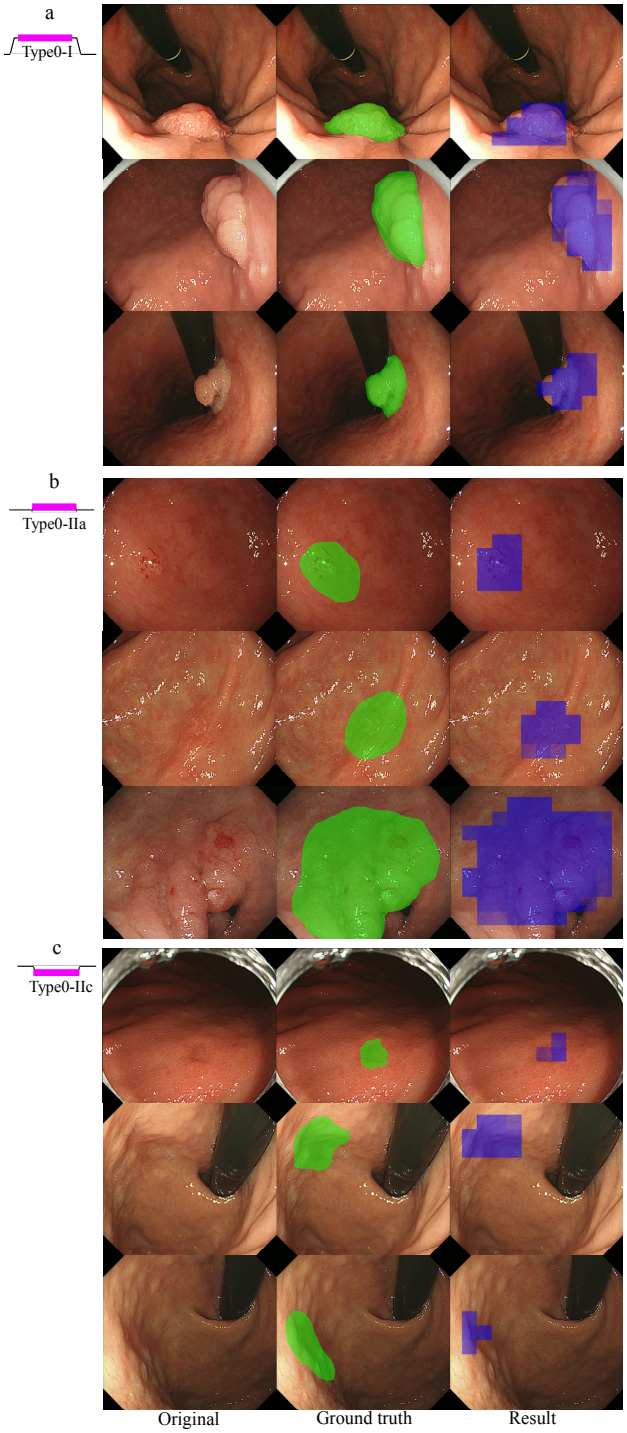

Figure 4 Predicted regions of early gastric cancer

ETHICS

This study was approved by the Institutional Review Board of RIKEN (Wako3 27-10) and National Cancer Center Hospital East (2016-350, 2017-090).

REFERENCES

[1] B. Taha et al., "Automatic polyp detection in endoscopy videos: A survey," In *IASTED Int. Conf. on BioMed*, Feb. 2017.

[2] Y. Ezoe et al., "Magnifying narrowband imaging is more accurate than conventional white-light imaging in diagnosis of gastric mucosal cancer ," *Gastroenterology*, 141, 6, 2017-2025, Dec. 2011.

[3] M. Muto et al., "Early detection of superficial squamous cell carcinoma in the head and neck region and esophagus by narrow band imaging: a multicenter randomized controlled trial," *J Clin Oncol*, 28, 9, 1566-72, Mar. 2010.

[4] H. Ikematsu et al., "Detectability of colorectal neoplastic lesions using a novel endoscopic system with blue laser imaging: a multicenter randomized controlled trial," *Gastrointestinal Endoscopy*, 86, 2, pp.386-394, Aug. 2017.

[5] T. Gotoda et al., "Endoscopic submucosal dissection of early gastric cancer," *J Gastroenterol*, pp. 929-942, Oct. 2006.

[6] Y. Lecun et al., "Gradient-based learning applied to document recognition," *Proc. of the IEEE*, pp. 2278-2324, 1998.

[7] N. Tajbakhsh et al., "Automated Polyp Detection in Colonoscopy Videos Using Shape and Context Information," *IEEE Trans. Medical Imaging*, 35, 2. pp. 630-644, Feb. 2016.

[8] J. Bernal et al., "Towards automatic polyp detection with a polyp appearance model," *Patten Recognition*, 45, 9, pp. 3166-3182, Sep. 2012.

[9] R. Zhu et al., "Lesion detection of endoscopy images based on convolutional neural network features," In *CISP*, pp. 372-376, Oct. 2015.

[10] N. Tajbakhsh et al., "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," *IEEE Trans. Medical Imaging*, 35, 5, pp. 1299-1312, Mar. 2016.

[11] W. Liu et al., "SSD: Single Shot MultiBox Detector," in *CVPR*, Dec. 2016.

[12] T. Hirasawa et al., "Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images," *Gastric Cancer*, Jan. 2018.

[13] F. Chollet et al., "Keras," https://github.com/fchollet/keras 2015.

[14] S.J. Pan et al., "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, 22, 10, pp. 1345-1359, Oct. 2010.

[15] C. Szegedy et al., "Going deeper with convolutions," In *CVPR*, Sep. 2014.

[16] O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," *Int. J. of Computer Vision*, 115, 3, pp. 211-252, Dec. 2015.

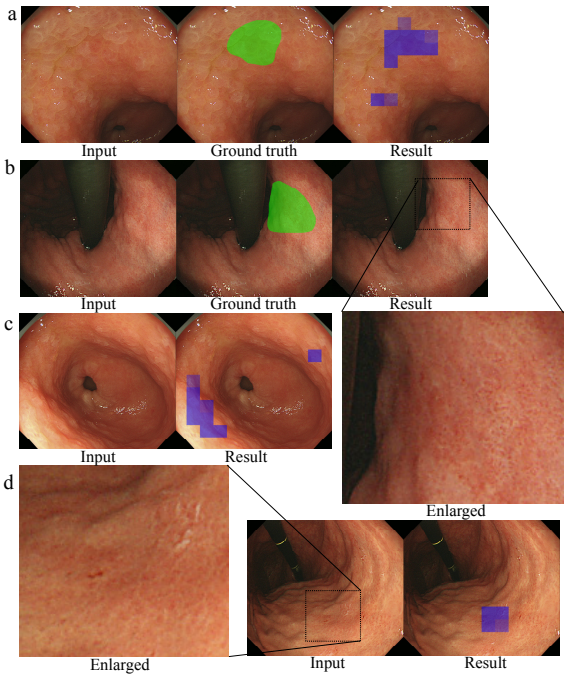

Figure 5 Examples of mispredicted images