A deep neural network improves endoscopic detection of early gastric cancer without blind spots

Authors
Lianlian Wu1,2,3,* , Wei Zhou1,2,3,* , Xinyue Wan1,2,3, Jun Zhang1,2,3, Lei Shen1,2,3, Shan Hu4, Qianshan Ding1,2,3, Ganggang Mu1,2,3, Anning Yin1,2,3, Xu Huang1,2,3, Jun Liu1,3, Xiaoda Jiang1,2,3, Zhenguang Wang1,2,3, Yunchao Deng1,2,3, Mei Liu5, Rong Lin6, Tingsheng Ling7, Peng Li8, Qi Wu9, Peng Jin10, Jie Chen11, Honggang Yu1,2,3

Institutions
1 Department of Gastroenterology, Renmin Hospital of Wuhan University, Wuhan, China
2 Key Laboratory of Hubei Province for Digestive System Disease, Renmin Hospital of Wuhan University, Wuhan, China
3 Hubei Provincial Clinical Research Center for Digestive Disease Minimally Invasive Incision, Renmin Hospital of Wuhan University, Wuhan, China
4 School of Resources and Environmental Sciences of Wuhan University, Wuhan, China
5 Department of Gastroenterology, Tongji Hospital of Huazhong University of Science and Technology, Wuhan, China
6 Department of Gastroenterology, Wuhan Union Hospital of Huazhong University of Science and Technology, Wuhan, China
7 Department of Gastroenterology, Nanjing Drum Tower Hospital of Nanjing University, Nanjing, China
8 Department of Gastroenterology, Beijing Friendship Hospital of the Capital University of Medical Sciences, Beijing, China
9 Endoscopy Center, Beijing Cancer Hospital of Peking University, Beijing, China
10 Department of Gastroenterology, Beijing Military Hospital, Beijing, China
11 Department of Gastroenterology, Shanghai Hospital of the Second Military Medical University, Shanghai, China

submitted 10.4.2018
accepted after revision 14.9.2018

ABSTRACT
Background Gastric cancer is the third most lethal malignancy worldwide. A novel deep convolution neural network (DCNN) to perform visual tasks has been recently developed. The aim of this study was to build a system using the DCNN to detect early gastric cancer (EGC) without blind spots during esophagogastroduodenoscopy (EGD).

Methods 3170 gastric cancer and 5981 benign images were collected to train the DCNN to detect EGC. A total of 24549 images from different parts of stomach were collected to train the DCNN to monitor blind spots. Class activation maps were developed to automatically cover suspicious cancerous regions. A grid model for the stomach was used to indicate the existence of blind spots in unprocessed EGD videos.

Results The DCNN identified EGC from non-malignancy with an accuracy of 92.5%, a sensitivity of 94.0%, a specificity of 91.0%, a positive predictive value of 91.3%, and a negative predictive value of 93.8%, outperforming all levels of endoscopists. In the task of classifying gastric locations into 10 or 26 parts, the DCNN achieved an accuracy of 90% or 65.9%, on a par with the performance of experts. In real-time unprocessed EGD videos, the DCNN achieved automated performance for detecting EGC and monitoring blind spots.

* Contributed equally to this work
Conclusions  We developed a system based on a DCNN to accurately detect EGC and recognize gastric locations better than endoscopists, and proactively track suspicious cancerous lesions and monitor blind spots during EGD.

Introduction

Gastric cancer is the third most lethal and the fifth most common malignancy from a global perspective [1]. It is estimated that about 1 million new gastric cancer cases were diagnosed and about 700,000 people died of gastric cancer in 2012, which represents up to 10% of the cancer-related deaths worldwide [1]. The 5-year survival rate for gastric cancer is 5% – 25% in its advanced stages, but reaches 90% in the early stages [2, 3]. Early detection is therefore a key strategy to improve patient survival.

In recent decades, endoscopic technology has seen remarkable advances and endoscopy has been widely used as a screening test for early gastric cancer (EGC) [4]. In one series, 7.2% of patients with gastric cancer had however been misdiagnosed at an endoscopy performed within the previous year, and 73% of these cases arose from endoscopist errors [5].

The performance quality of EGD varies significantly because of cognitive and technical factors [6]. In the cognitive domain, EGC lesions are difficult to recognize because the mucosa often shows only subtle changes, which require endoscopists to be well trained and armed with a thorough knowledge [4, 7]. In addition, endoscopists could be affected by their subjective state during endoscopy, which restricts the detection of EGC to a large extent [8]. In the technical domain, guidelines to map the entire stomach exist, but are often not well followed, especially in developing countries [9, 10]. Therefore, it is important to develop a feasible and reliable method to alert endoscopists to possible EGC lesions and blind spots.

A potential solution to mitigate the skill variations is to apply artificial intelligence (AI) to EGD examinations. The past decades have seen an explosion of interest in the application of AI in medicine [11]. More recently, a method of AI known as a deep convolutional neural network (DCNN), a method transforming the representation at one level into a more abstract level to make predictions, has opened the door to elaborate image analysis [12]. Recent studies have successfully used DCNNs in the field of endoscopy. Chen et al. achieved accurate classification of diminutive colorectal polyps based on colonoscopy images [13], and Byrne et al. achieved real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps using colonoscopy videos [14]. However, in real-time EGC detection alone with blind spot monitoring, the application of DCNN has not yet been researched.

In this work, we first developed a novel system using DCNN to analyze EGD images of EGC and gastric locations. Furthermore, we exploited an activation map to proactively track suspicious cancerous regions and built a grid model for the stomach to indicate the existence of blind spots on unprocessed EGD videos.

Methods

Datasets, data preparation, and sample distribution

The flowchart of the data preparation and training/testing procedure of the DCNN is shown in Fig. 1. Networks playing different roles were independently trained, and their functions, inclusion criteria, exclusion criteria, image views, data sources, and data preparation are described in Table 1 (Fig. e2: (15, 16)), available online in Supplementary materials. The sample distribution is presented in Fig. e3. Images of the same lesion from multiple viewpoints, or similar lesions from...
A grid model for the stomach

In order to automatically remind endoscopists of blind spots during EGD, a grid model for the stomach was developed to present the covered parts. The onion-skin display method was used to build a grid model for the stomach, as previously described [28]. The characteristics of each part of the stomach were extracted and put together to generate a virtual stomach model. The model was set to be transparent before EGD. As soon as the scope was inserted into the stomach, the DCNN model began to capture images and filled them into the corresponding part in the model, coloring the various parts.

Running the DCNN on videos

Frame-wise prediction was used on unprocessed videos using client-server interaction (Fig. e8a). Images were captured at 2 frames per second (fps). Noises were smoothed by the Random Forest Classifier model [29] and the rule of “output results only when three of the five consecutive images show a same result” (Fig. e8b). The time used for outputting a prediction per frame in the videos in the clinical setting includes time consumed in the client (image capture, image resizing, and rendering images based on predicted results), network communication, and the server (reading and loading images, running the three networks, and saving images).

The speed of the DCNN in the clinical setting was evaluated by 926 independent tests, calculating the total time used to output a prediction per frame in the endoscopy center of Renmin Hospital of Wuhan University.

Human subjects

Endoscopists that participated in our tests were under informed consent. This study was approved by the Ethics Committee of Renmin Hospital of Wuhan University, and was registered as trial number ChiCTR1800014809 of the Primary Registries of the WHO Registry Network.
Statistical analysis

A two-tailed unpaired Student’s $t$ test with a significance level of 0.05 was used to compare differences in the accuracy, sensitivity, specificity, and positive and negative predictive values (PPV and NPV, respectively) of the DCNN and endoscopists. Interobserver and intraobserver agreement of the endoscopists and intraobserver agreement of the DCNN were evaluated using Cohen’s kappa coefficient. All calculations were performed using SPSS 20 (IBM, Chicago, Illinois, USA).

Results

The performance of DCNN on identification of EGC

Comparison between the performance of DCNN and endoscopists

Table 4 shows the predictions of the DCNN and endoscopists for identifying EGC. Among 200 gastroscope images, with or without malignant lesions, the DCNN correctly diagnosed malignancy with an accuracy of 92.5%, a sensitivity of 94%, a specificity of 91%, a PPV of 91.3%, and an NPV of 93.8%. Six experts, eight seniors, and seven novices attained an accuracy of 89.7% (standard deviation [SD] 2.2%), 86.7% (SD 5.6%), and 81.2% (SD 5.7%) for each picture, respectively. The accuracy of DCNN was significantly higher than that of all endoscopists.

Table 4 Performance of the deep convolution neural network (DCNN) and endoscopists in the detection of early gastric cancer. All results are given as a percentage (standard deviation).

<table>
<thead>
<tr>
<th></th>
<th>DCNN</th>
<th>Experts (n=6)</th>
<th>Seniors (n=8)</th>
<th>Novices (n=7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>92.50</td>
<td>89.73 (2.15)&lt;sup&gt;1&lt;/sup&gt;</td>
<td>86.68 (5.58)&lt;sup&gt;2&lt;/sup&gt;</td>
<td>81.16 (5.72)&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>94.00</td>
<td>93.86 (7.65)</td>
<td>90.00 (6.05)</td>
<td>75.33 (6.31)&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>Specificity</td>
<td>91.00</td>
<td>87.33 (7.43)</td>
<td>85.05 (16.18)</td>
<td>88.83 (6.03)</td>
</tr>
<tr>
<td>PPV</td>
<td>91.26</td>
<td>91.75 (4.15)</td>
<td>90.91 (5.69)</td>
<td>80.47 (8.75)&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
<tr>
<td>NPV</td>
<td>93.81</td>
<td>92.52 (5.76)</td>
<td>88.01 (6.55)&lt;sup&gt;2&lt;/sup&gt;</td>
<td>82.32 (11.46)</td>
</tr>
</tbody>
</table>

PPV, positive predictive value; NPV, negative predictive value.  
<sup>1</sup> $P<0.01$  
<sup>2</sup> $P<0.05$
Fig. 9 Representative images predicted by the deep convolution neural network (DCNN) in the test set for detection of early gastric cancer (EGC). 

**a** The displayed normal mucosa, superficial gastritis, and mild erosive gastritis were predicted to be non-malignant by the DCNN with confidence levels of 0.98, 0.95, and 0.91, respectively. 

**b** The mucosal images from EGC type 0-IIa, 0-IIb, and 0-IIc were predicted to be malignant by the DCNN with confidence of 0.86, 0.82, and 0.89, respectively. 

**c** Cancerous regions were indicated in the images from **b** after establishing class activation maps (CAMs). The color depth of the CAMs was positively correlated with the prediction confidence.
Fig. 9 shows representative images predicted by the model in the test set. The CAMs highlighted the cancerous regions after images were evaluated by the model.

Comparison between the stability of DCNN and endoscopists

To evaluate the stability of DCNN and endoscopists on identifying EGC, we mixed up all of the test pictures and randomly selected six endoscopists (2 experts, 2 seniors, and 2 novices) to do the same test again. As shown in Table 5, the experts had substantial interobserver agreement (kappa 0.80), and the seniors and novices achieved moderate interobserver agreement (kappa 0.49 and 0.42, respectively). The intraobserver agreement of experts and nonexperts was moderate or better (kappa 0.84 in the expert group and 0.54–0.77 in the nonexpert group). The DCNN achieved perfect intraobserver agreement (kappa 1.0).

The performance of DCNN on classification of gastric locations

Comparison between the performance of DCNN and endoscopists

Table 6 shows the predictions of the DCNN and endoscopists for classifying gastric locations. A group of 10 experts, 16 seniors, and 9 novices classified EGD images into 10 stomach parts with an accuracy of 90.2% (SD 5.1%), 86.8% (5.2%), and 83.3% (10.3%), respectively, and into 26 sites with an accuracy of 63.8% (6.9%), 59.3% (6.4%), and 46.5% (7.2%), respectively. The DCNN correctly identified EGD images into 10 parts with an accuracy of 90% and into 26 parts with an accuracy of 65.9%, showing no significant difference with any of the levels of endoscopists. Fig. 7 and Fig. 10 show representative images in the test set that were predicted by the DCNN in the task of classifying gastric locations into 26 parts and 10 parts, respectively.

Comparison between the stability of DCNN and endoscopists

In the task of classifying gastric locations into 10 parts, all endoscopists achieved substantial interobserver or intraobserver agreement (kappa 0.75–0.96). In the 26-part classification, all endoscopists achieved moderate interobserver or intraobserver agreement (kappa 0.50–0.68) (Table 7). The DCNN achieved perfect intraobserver agreement (kappa 1.0).

Testing of the DCNN in unprocessed gastroscope videos

To explore the ability of the DCNN in detecting EGC and monitoring blind spots in a real-time clinical setting, we checked the model in two unprocessed gastroscope videos. In Video 1, which had no cancerous lesion, the DCNN accurately presented the covered parts synchronized with the process of EGD to verify that the entire stomach was mapped.

In Video 2, which had cancerous lesions, the DCNN alerted about blind spots synchronized with the process of EGD, and automatically indicated the suspicious EGC regions with CAMs. All lesions were successfully detected; however, a false-positive error occurred when the mucosa was covered by unwashed foam.

To test the speed of the DCNN, 926 independent tests were conducted in a clinical setting. The total time to output a prediction using all three networks for each frame was 230 milliseconds (SD 60; range 180–350). In the test of identifying EGC, six experts, eight seniors, and seven novices took 3.29 seconds per picture (SD 0.42), 3.96 (0.80), and 6.19 (1.92), respectively. In the test of classifying gastric locations into 10 parts, 10 experts, 16 seniors, and 9 novices required 4.51 seconds per picture (SD 2.07), 4.52 (0.65), and 4.76 (0.67), respectively; for classification into 26 parts, they took 14.23 seconds per picture (SD 2.41), 19.33 (9.34), and 24.15 (6.93), respectively. The prediction time of the DCNN in the clinical setting was con-
Fig. 10 Representative images predicted by the deep convolution neural network (DCNN) in the test set for the classification of gastric locations into 10 parts, showing the gastric locations determined by the DCNN and their prediction confidence. Class 0, esophagus; 1, squamocolumnar junction; 2, antrum; 3, duodenal bulb; 4, descending duodenum; 5, lower body in forward view; 6, middle-upper body in forward view; 7, fundus; 8, middle-upper body in retroflexed view; 9, angulus.

Table 7 Intra- and interobserver agreement (kappa value) of endoscopists in classifying gastric location into 10 or 26 parts.

<table>
<thead>
<tr>
<th></th>
<th>Expert 1</th>
<th>Expert 2</th>
<th>Senior 1</th>
<th>Senior 2</th>
<th>Novice 1</th>
<th>Novice 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 parts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 1</td>
<td>0.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 2</td>
<td>0.91</td>
<td>0.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior 1</td>
<td></td>
<td></td>
<td>0.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior 2</td>
<td></td>
<td></td>
<td>0.75</td>
<td>0.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novice 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>Novice 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.85</td>
<td>0.93</td>
</tr>
<tr>
<td>26 parts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 1</td>
<td>0.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert 2</td>
<td>0.61</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior 1</td>
<td></td>
<td></td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior 2</td>
<td></td>
<td></td>
<td>0.52</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Novice 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>Novice 2</td>
<td></td>
<td></td>
<td>0.55</td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
siderably shorter compared with the time taken by the endoscopists.

Discussion

Endoscopy plays a pivotal role in the diagnosis of gastric cancer, the third leading cause of cancer death worldwide [1, 4]. Unfortunately, endoscopic diagnosis of gastric cancer at an early stage is difficult, requiring endoscopists first to obtain thorough knowledge and good technique [7]. The training process of a qualified endoscopist is time- and cost-consuming. In many countries, especially in Western Europe and China, the demand for endoscopists familiar with the diagnosis of EGC means they are in short supply, which greatly limits the effectiveness of endoscopy in the diagnosis and prevention of gastric cancer [7–10].

DCNN is one of the most important deep learning methods for computer vision and image classification [12–14]. Chen et al. [13] achieved accurate classification of diminutive colorectal polyps based on images captured during colonoscopy using DCNN, and Byrne et al. [14] achieved real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps on colonoscopy videos. The most recent study [30] used DCNN to detect EGC with an overall sensitivity of 92.2% and a PPV of 30.6% in their dataset. Here, we developed a DCNN system to detect EGC, with a sensitivity of 94% and a PPV of 91.3%, and distinguish gastric locations on a par with the level of expert endoscopists. Furthermore, we compared the competence of the DCNN with endoscopists and used the DCNN in unprocessed EGD videos to proactively track suspicious cancerous lesions without blind spots.

Observing the whole stomach is a basic prerequisite for the diagnosis of gastric cancer at an early stage [7, 16]. In order to avoid blind spots, standardized procedures have been made to map the entire stomach during gastroscopy. The European Society of Gastrointestinal Endoscopy (ESGE) published a protocol including 10 images of the stomach in 2016 [15]. Japanese researchers published a minimum required “systematic screening protocol for the stomach” (SSS) standard including 22 images of the stomach so as not to miss suspicious cancerous lesions [16]. However, these protocols are often not well followed, and endoscopists may ignore some parts in the stomach because of subjective factors or limited operative levels, which can lead to the misdiagnosis of EGC [7, 8, 10].

In the present study, using 24,549 images from EGDs, we developed a DCNN model that accurately and stably recognizes each part of the stomach on a par with the expert level, automatically captures images during endoscopy, and maps these onto a grid model of the stomach to prompt the operator about blind spots. This real-time assistance system will improve the quality of EGD and ensure that the whole stomach is observed during endoscopy, thereby providing an important prerequisite for the detection of EGC.

White-light imaging is the standard endoscopic examination for the identification of gastric cancer lesions, although it is difficult to make an accurate detection of EGC [31, 32]. It has been reported that the sensitivity of white-light imaging in the diag-
nosis of superficial EGC ranges from 33% to 75% [33]. Many new technologies have been developed to improve diagnostic abilities for EGC, including image-enhanced endoscopy and magnifying endoscopy [34]. Plenty of time and money have been put into training endoscopists to become familiar with the characteristics of early cancer lesions under different views [7]. However, while current technologies are still not strong for some secluded lesions, such as the Ilb-type EGC [34], manual diagnosis, on the other hand, greatly depends on the experience and subjective state of the operator performing the endoscopy [8].

This subjective dependence on the operator decreases the accuracy and stability of EGC diagnosis [8]. To overcome this limitation, DCNN, with its strong learning ability and good reproducibility, has gained attention as a clinical tool for endoscopists. In the present study, 3170 gastric cancer and 6541 normal control images from EGD examinations were collected to train a DCNN model with reliable and stable diagnosis ability for EGC. An independent test was conducted to evaluate the diagnostic ability of DCNN and endoscopists. In our study, the DCNN achieved an accuracy of 92.5%, a sensitivity of 94%, a specificity of 91%, a PPV of 91.3%, and an NPV of 93.8%, outperforming all levels of endoscopists.

In terms of stability evaluation, the nonexperts achieved moderate interobserver agreement (kappa 0.42–0.49), and the experts achieved substantial interobserver agreement (kappa 0.80). Because of the subjective interpretation of the EGC characteristics, human learning curves in the diagnosis of EGC exist, and therefore an objective diagnosis is necessary [5, 8]. In our study, we used up-to-date neural network models to develop a DCNN system. It achieved perfect intraobserver agreement (kappa 1.0), while endoscopists had variable intraobserver agreement. Our results indicate that this gastric cancer screening system based on a DCNN has adequate and consistent diagnostic performance, removes some of the diagnostic subjectivity, and could be a powerful tool to assist endoscopists, especially nonexperts, in detecting EGC.

The diagnostic ability and stability of the DCNN seems to outperform that of experienced endoscopists. In addition, the diagnostic time of the DCNN was considerably shorter than that of the endoscopists. It should be noted that the shorter screening time and the absence of fatigue with the DCNN may make it possible to provide quick predictions of EGC following an endoscopic examination. Importantly, the diagnosis of EGC by the DCNN can be achieved completely automatically and online, which may contribute to the development of telemedicine, thereby alleviating the problem of inadequate numbers and experience of doctors in remote regions.

Another strength of this study is that the DCNN is armed with CAMs and a grid model for the stomach to cover suspicious cancerous regions and indicate the existence of potential blind spots. The CAMs are a weighted linear sum of the presence of visual patterns with different characteristics, by which the discriminative regions of target classification are highlighted [27]. As soon as we insert the scope into the stomach, we can determine whether and where EGC is present in the background mucosa using the DCNN with CAMs. In addition, the observed areas are instantaneously recorded and colored in the grid model of the stomach, indicating whether blind spots exist during the EGD. Through these two auxiliary tools, it is possible for the DCNN to proactively track suspicious cancerous lesions without blind spots, so reducing the pressure and workload on endoscopists during real-time EGD.

There are some limitations in the present study. First, the detection of EGC was based on images only in white light, narrow-band imaging (NBI), and blue-laser imaging (BLI) views. With images under more views, such as chromoendoscopy using indigo carmine [7], i-scan optical enhancement [34], and even optical coherence tomography [35], it is possible to design a more universal EGC detection system.

Second, in the control group of the gastric cancer dataset, only normal, superficial gastritis, and mild erosive gastritis mucosa were enrolled. Other benign diseases, such as atrophic gastritis, gastritis verrucose, and typical benign ulcer, could be enrolled in the control group later. In this way, the practicability of the DCNN in detection of EGC will be further improved.

Third, when the DCNN was applied to unprocessed EGD videos, false-positive errors occurred when the mucosa was not washed clean. We plan to train the DCNN to recognize mucosa that is poorly prepared to avoid these mistakes and to transfer the false-positive errors into a suggestion to the operator with regard to cleaning the mucosa.

Fourth, although the DCNN presented satisfactory results in the detection of EGC and monitoring of EGD quality in real-time unprocessed videos, its competence was only quantitatively evaluated in still images, not in videos. We will keep collecting data to assess its ability in unprocessed videos and to provide accurate effectiveness data for the DCNN in a real clinical setting in the near future.

In summary, a computer-aided system based on a DCNN provides automated, accurate, and consistent diagnostic performance for the detection of EGC and the recognition of gastric locations. The CAMs and grid model for the stomach enable the DCNN to proactively track suspicious cancerous lesions without blind spots during EGD. The DCNN is a promising technique in computer-based recognition and is not inferior to experts. It may be a powerful tool to assist endoscopists in detecting EGC without blind spots. More research should be conducted and clinical applications should be tested to further verify and improve this system’s effectiveness.

Acknowledgments
This work was partly supported by a grant from the Research Funds for Key Laboratory of Hubei Province (No. 2016CFA066), the National Natural Science Foundation of China (grant nos. 81672387 [to Yu Honggang]), and the China Youth Development Foundation (grant no. 81401959 [to Zhou Wei] and grant no. 81703030 [to Ding Qianshan]).
Competing interests

None

References