Stochastic Gastric Image Augmentation for Cancer Detection from X-ray Images

Hideaki Okamoto*, Quan Huu Cap*, Takakiyo Nomura†, Hitoshi Iyatomi*, Jun Hashimoto†
Email: {hideaki.okamoto.2u, huu.quan.cap.78}@stu.hosei.ac.jp
6bmud010@mail.u-tokai.ac.jp, iyatomi@hosei.ac.jp, junhashi@tokai-u.jp
*Applied Informatics, Graduate School of Science and Engineering, Hosei University, Tokyo, Japan
†Department of Radiology, Tokai University School of Medicine, Isehara, Japan

***Abstract*—X-ray examinations are a common choice in mass screenings for gastric cancer. Compared to endoscopy and other common modalities, X-ray examinations have the significant advantage that they can be performed not only by radiologists but also by radiology technicians. However, the diagnosis of gastric X-ray images is very difficult and it has been reported that the diagnostic accuracy of these images is only 85.5%. In this study, we propose a practical diagnosis support system for gastric X-ray images. An important component of our system is the proposed on-line data augmentation strategy named stochastic gastric image augmentation (sGAIA), which stochastically generates various enhanced images of gastric folds in X-ray images. The proposed sGAIA improves the detection performance of the malignant region by 6.9% in F1-score and our system demonstrates promising screening performance for gastric cancer (recall of 92.3% with a precision of 32.4%) from X-ray images in a clinical setting based on Faster R-CNN with ResNet101 networks.**

***Index Terms*—gastric cancer, X-ray images, data augmentation, convolutional neural networks, computer-aided diagnosis**

I. Introduction

Gastric cancer is the third most common cancer, with 1,000,000 new diagnoses and 780,000 deaths annually [1]. Advanced gastric cancer has a poor prognosis but early prognosis before metastasis is good, so early detection and appropriate treatment are important. In general, an X-ray examination and endoscopy are used for gastric cancer diagnosis. After a cancer diagnosis has been confirmed, other medical modalities, such as CT and PET/CT, are used for examination of cancer progression if necessary. Endoscopy is now widely used and has demonstrated superior detection capability to the other techniques, with a sensitivity of 95.4% [2]. Nevertheless, X-ray imaging has a great advantage. Endoscopy must be performed by only radiologists, whereas X-ray imaging can be done with both radiologists and technicians. However, gastric X-ray diagnosis is challenging, especially for inexperienced radiologists and the acquisition of this diagnostic skill takes a very long time compared to other diagnostic imaging methods, such as CT, MRI and PET/CT. Accordingly, the diagnosis ability varies greatly between inexperienced and expert radiologists. According to statistics, the diagnostic accuracy of gastric X-ray examination is 85.5% [2].

978-1-7281-0858-2/19/$31.00 ©2019 IEEE

In medical practice, one of the basics of gastric X-ray diagnosis is finding inflammation of the gastric mucosa surface [3]. Gastric folds of healthy subjects show thin, smooth and parallel characteristics, while their appearance changes in thickness and size, showing irregular bends, pits, breaks in their inner membrane (i.e. mucosa) due to gastric diseases caused by inflammation, infection, ulcers and so on [4]. Based on the above medical findings, several automated diagnosis or diagnosis support systems have been proposed to alleviate the burden of radiologists and incorrect diagnoses. Ishihara et al. developed an image-based classifier of H. pylori infection, one of the major causes of gastric cancer, from a total of 2,100 patient data with eight images for each patient. They focused on mucosal patterns and gastric folds after the infection and calculated a large number of statistical features (i.e. a total of 7,760 per case), including texture, shape and gradient-based ones from the image. Their diagnosis model with a traditional SVM classifier attained a sensitivity of 89.5% and a specificity of 89.6% [5]. In addition, Abe et al. focused on the symptoms of gastric folds on gastric X-ray images and designed their original handcrafted features based on an original image processing method. They reported excellent performance in detecting abnormal lesion locations (e.g. precision of 89.3% with a recall of 100%) with traditional classifiers, such as LDA and SVMs [6]. The design of their handcrafted features looks reasonable, but since the number of images for performance validation is very small (88 cases including 13 abnormal cases) and the images taken are in research conditions (i.e. controlled geometry), the performance in a practical environment needs to be further verified.

In recent years, deep learning techniques have demonstrated their outstanding performance and have dramatically changed the traditional machine learning schema. Convolutional neural networks (CNNs) are today's indispensable elements of deep learning, and have achieved excellent results, not only in computer vision but also in wide areas, including medical applications [7]. The most preferred benefit of CNNs is that they automatically capture efficient features for their target tasks from the training. In a related study, CNNs were used to diagnose gastric cancer from X-ray images and showed a sensitivity of 89.5% and a specificity of 93.5% based on images from 2,100 patients, each consisting of eight images [8]. Togo et al. achieved a precision of 98.3% and a recall of 96.2% according to their CNN classification using 815 patients data (6,520 images) in a study using double-enhanced upper gastrointestinal barium radiography [9]. Although these methods show superior results, they assume that the ROI for diagnosis (i.e. candidate of lesions) has been shown in advance. In practical situations, it is desirable for the system to be able to detect ROIs in order to eliminate medical oversight.

In a computer vision field, several integrated strategies, such as region-based CNN (R-CNN) [10], Faster R-CNN [11] and single-shot multibox detector (SSD) [12] based on CNNs, have been proposed. These techniques have also applied in wide variety of computer vision tasks. With these sophisticated techniques, Li et al. analyzed a total of 1,231 chest X-ray images with R-CNN to detect eight types of lung disease [13]. They defined correct detection using an intersection over union (IoU) greater than 0.5 between the detected bounding box and the ground truth box. Their system attained 7-84% disease localization accuracy for each disease, with an average of 26.9% under the five-fold validation. On gastric endoscopic images, the SSD-based system based on 13,684 images achieved a sensitivity of 92.2% and a specificity of 30.6% [14]. As can be seen from the literature above, the integrated task of target detection and classification is more difficult than that needs only classification. In the classification only task, only controlled images (e.g. on pre-determined geometry) are subject to analysis. This is namely that the ROI must be inside of the image, so the diversity of the image is much less than the task needs to determine both. To the best of our knowledge, few investigations offer an integrated process for ROI detection and diagnosis on gastric X-ray images. In this context, we aim to realize an efficient diagnosis support system that simultaneously detects the involved areas from the gastric X-ray image and diagnoses each area.

The major challenge of our study is to achieve practical accuracy using a limited number of available gastric X-ray images for training. Usually, the abovementioned deep-learning based integrated methods require a significant amount of training data with appropriate annotation, but it is very expensive especially in medical tasks like this. In addition, the gastric X-ray images involved in the study were taken in clinical settings, not research settings. In other words, our image conditions, such as geometry, angle and size are not strictly controlled, making them very diverse and difficult to investigate compared to research-quality images used in the literature. We believe that achieving satisfactory performance in this difficult clinical setting is a condition for realizing a practical diagnostic support system. Therefore, this study addresses these issues by introducing our effective on-line data augmentation based on medical knowledge, called stochastic gastric image augmentation (sGAIA), rather than simply relying on the power of deep learning.

The contributions of this study are summarized as follows:

- To the best of our knowledge, this is the first practical screening system for gastric cancer from X-ray images in a clinical setting.
- Our proposed on-line data augmentation strategy based on medical knowledge, sGAIA, has realized a practical accurate screening system for gastric cancer.

II. Methods

In this study, we propose a practical diagnostic support system for gastric X-ray images in clinical settings with various difficulties. In order to achieve this goal with a limited number of training images, we propose sGAIA, an on-line data augmentation based on medical evidence, and combined it with an excellent object detection algorithm, Faster R-CNN [11].

A. Stochastic gastric image augmentation (sGAIA)

The proposed sGAIA is an effective on-line data augmentation strategy to enhance the gastric folds of X-ray images with reference to the literature [6]. We expect that the accuracy of the diagnosis would improve if the detector is trained on images in which areas relevant to diagnosis are emphasized. In their method, they detected gastric fold candidate regions based on the gradient and high-frequency component of the given gastric X-ray image. Then, based on the determined gastric fold regions, they calculated their hand-crafted image features for the next diagnosis stage. Specifically, let the gastric X-ray input image be $I(x, y)$, their gradient $\nabla I(x, y)$, and the high-frequency components of the original image, $H(x, y)$. They normalized $\nabla I(x, y)$ and $H(x, y)$ into [0, 1] and performed the K-means clustering with K=2. This process divides the image $I(x, y)$ into two regions per pixel. They regard the region belonging to the cluster with larger mean gradient and high-frequency component as the gastric fold edge candidate. However, since this method was designed to detect gastric folds in a controlled condition (i.e. geometry of images is almost identical among dataset), obtaining the necessary performance would not be easy in a clinical setting. In addition, this algorithm is developed for preprocessing, an effective collaboration with recent deep learning techniques was not considered. That is, this method yields one candidate of gastric region image per image. The proposed sGAIA is a fast and effective on-line data augmentation method that can be combined with deep learning techniques and overcomes the drawbacks of the above method. Figure 1 shows the schematic of sGAIA. The key feature of the proposal is to statistically determine the gastric fold candidate regions and avoid time-consuming clustering. The details of sGAIA consist of the following four steps:

- (Step 1) Calculation of edge strength
  Firstly, a contrasting enhanced image of the original gastric X-ray image $I(x, y)$, $I_e(x, y)$ is generated with equalizing the histogram of a grayscale image. Then, we obtain its gradient and high-frequency component, $\nabla I_e(x, y)$, and $H_e(x, y)$, respectively. Each of them is normalized into [0, 1] and we calculate the normalized edge strength $E(x, y)$ as follows. Here, $\overline{x}$ is the normalized value of xto[0,1].

$$E(x,y) = (\overline{\nabla I_e(x,y)} + \overline{H_e(x,y)})/2$$

- (Step 2) Calculation of selection probability of gastric edge candidate region
  We select the gastric edge candidate region with probabilistic manner per pixel. We obtain the probability map of the gastric fold edge, $p(x,y)$, using the pre-determined table, as shown in Table I and $E(x,y)$. Note that the probability in this table is determined according to our preliminary experiments.
- (Step 3) Determination of gastric fold edge region
  According to the probability map $p(x,y)$, we determine a binary gastric fold edge region. Since this region is noisy, we perform a morphological opening operation to eliminate too many small isolated regions. Let $G(x,y)$ be a binary mask representing gastric fold edge regions obtaining by this process.
- (Step 4) Enhancement of gastric edge region
  Finally, an enhanced gastric fold edge image $A(x,y)$ is obtained with the following linear combination.

$$A(x,y) = I_e(x,y) + \alpha\, G(x,y) + \beta$$

  Here, α and β are hyper parameters. They were selected in the range of α from 0.9 to 1.0 and β from -15 to -5 according to the results of preliminary experiments.

On this basis, the proposed sGAIA stochastically determines the location and strength of the gastric fold edge regions each time. Since it provides different location and strength of them at each trial, sGAIA has a high affinity with on-line data augmentation in the training of deep neural networks.

B. Detection of malignant regions from gastric X-ray images

Figure 2 illustrates our overall system architecture. We used Faster R-CNN [11] to localize the malignant regions from gastric X-ray images taken in a clinical setting. Faster R-CNN provides an integrated process of detection of target objects and their classification. It is generally slower than the methods proposed later, such as R-FCN [15] and SSD [12] but are known to be more accurate than these techniques [16]. In addition, this weakness is greatly reduced when the number of areas to be detected is limited like our target task. Therefore, we choose this model in this study. The input of our system receives 2,048 $\times$ 2,048 pixels. In the Faster R-CNN model, we fine-tune ResNet101 [17] pre-trained with the ImageNet dataset for generating feature maps and the last layer of conv4_x block (conv4_23) were used as the input for region proposal networks. Note here, the image size was reduced into 600 $\times$ 600 pixels before feeding it to the network. The configuration of fully connected layers is the same as original study, and the number of output node is one that represents the malignancy score.

III. Experiments

The image set used in this study consists of a total of 3,832 gastric X-ray images in a clinical setting from 105 patients (672 images with 1,126 lesion annotation by radiologists and 3,160 images without lesions) acquired from Tokai University School of Medicine, Japan. Each image used in these experiments is 8-bit grayscale and has a resolution of 2,048 $\times$ 2,048 pixels. We created the smallest rectangle that encloses the lesion areas based on the given annotations. In order to build a robust lesion detector in terms of various lesion sizes, shapes and resolutions, on-line data augmentations (random crop, random flip, random rotation and brightness augmentation) are performed on the training data and we treat this as a performance baseline. In the training of our Faster-RCNN model, the learning rate, batch size and number of training iterations are set to 0.001, 4, and 100, respectively. The momentum SGD [18] is used as an optimizer and its exponential decay rate of the first order moment is set to 0.9. In the training and evaluation, we used patient-based five-

TABLE I
Selection probability as gastric fold edge candidate region

Normalized edge strength e=$E(x,y)$	Probability (%)
$0 \le e < 0.4$	0
$0.4 \le e < 0.5$	30
$0.5 \le e < 0.6$	50
$0.6 \le e < 0.8$	70
$0.8 \le e \le 1.0$	95

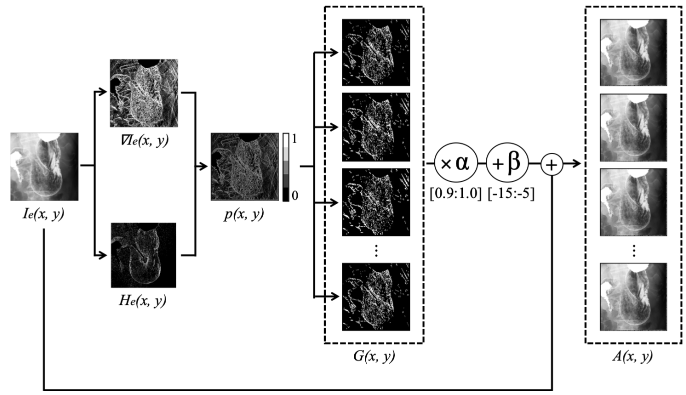


Fig. 1. Overview of stochastic gastric image augmentation

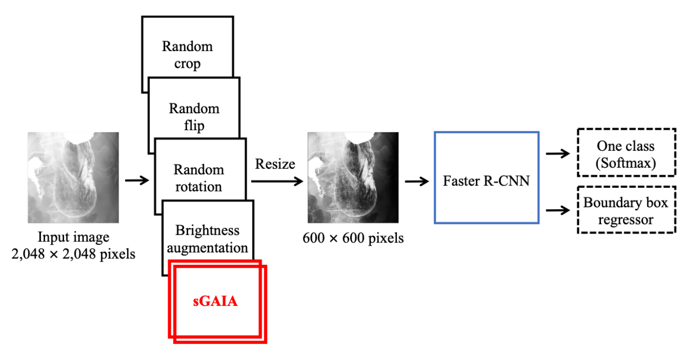


Fig. 2. Overview of proposed detection system

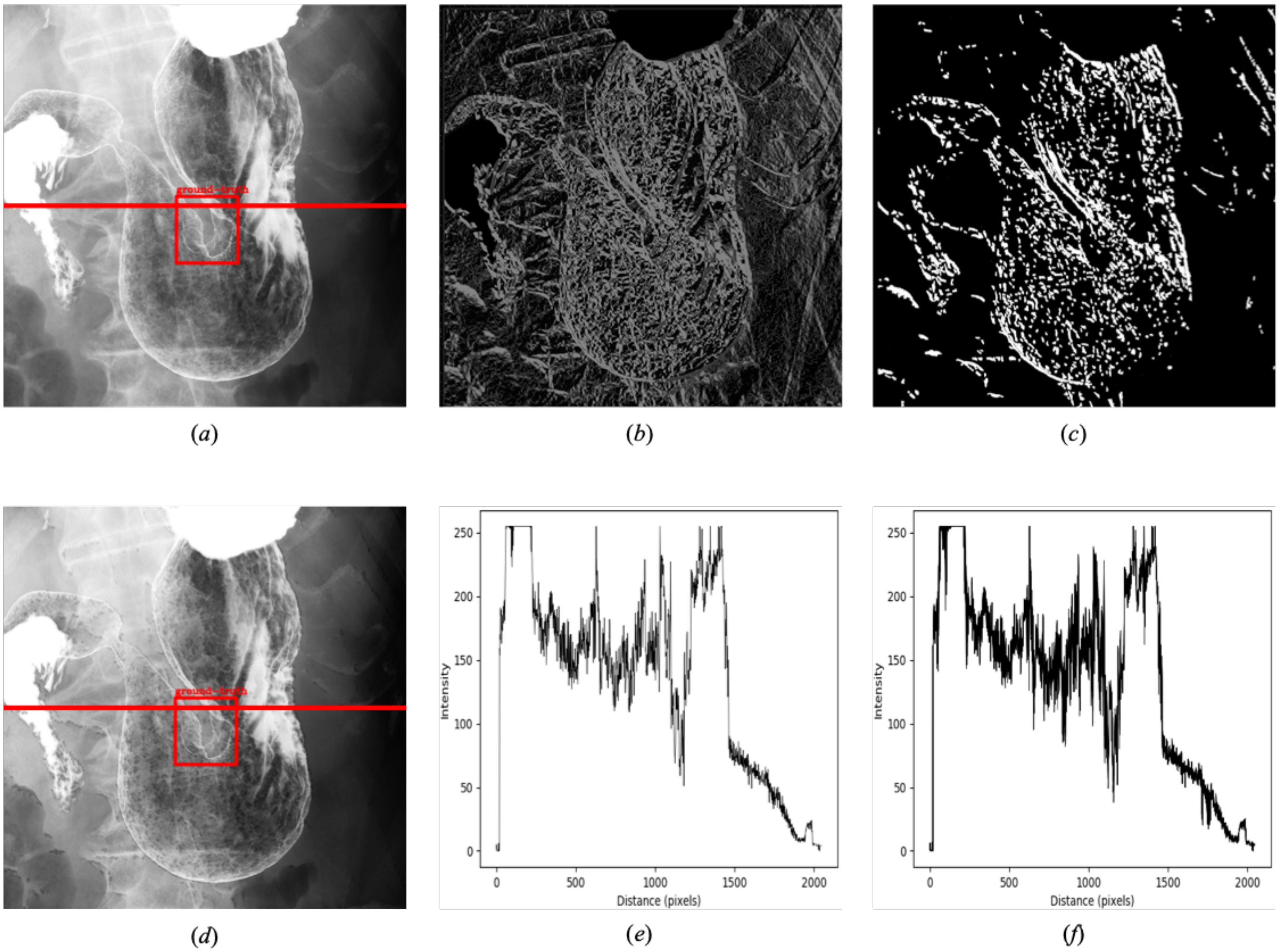


Fig. 3. Example of augmented images generation process by sGAIA and comparison of intensity
(a) $I_e(x, y)$, (b) $p(x, y)$, (c) One example of $G(x, y)$,
(d) One example of $A(x, y)$, (e) Line profile of (a), (f) 10 line profiles of images generated by sGAIA

group cross-validation, i.e., the images from the same patient is not divided into the training and evaluation. Precision, recall and F1-score are used as performance metrics in our system. Our system provides candidate lesion boundary boxes with associated confidence score and we accept those boxes above the predefined detection threshold α. The larger α is selected, the fewer the detection boxes and vice versa. In order to confirm the effectiveness of the proposed sGAIA, we conducted three evaluation experiments:

- **Baseline**: Train our network with only four-types of on-line data augmentations described above (random crop, random flip, random rotation and brightness augmentation).
- **Preprocess based on [6]**: We enhanced the fold regions of the given X-ray image with the process used in [6] and those enhanced images and above data augmentations were used for training. Specifically, the fold candidate regions (i.e. binary map) were determined by K-means algorithm, $G'(x, y)$ and the enhanced images $A'(x, y)$ were calculated in the similar way as the (step 4) of sGAIA. Here, we used α = 0.95, β = -10, the best combination in preliminary experiments.
- **sGAIA**: Train our network with above augmentation with the proposed sGAIA. Here, we randomly choose the hyper parameter from 0.9 $\leq \alpha \leq$ 1.0 and -15 $\leq \beta \leq$ -5.

IV. Results and Discussions

Figure 3 shows an example how sGAIA produces augmented images $A(x, y)$ from the original contrast enhanced images $I_e(x, y)$ and their line profiles. Here, Figs. 3(e) and (f) are the line profile of the $I_e(x, y)$ and ten line profiles of images generated by sGAIA, respectively. The final detection performance are summarized in Table II with typical detection threshold α and the receiver operating characteristics curves are in Fig. 4. Figure 5 shows a comparison of the detection results with baseline and the sGAIA. From the comparison of

TABLE II
DETECTION RESULTS FOR THE THREE EXPERIMENTS

		α =0.3 (%)	α =0.2 (%)	α =0.15 (%)
Baseline	precision	43.0	26.8	12.7
	recall	59.1	88.4	93.8
	F1-score	49.8	41.1	22.4
Preprocess based on [6]	precision	46.6	30.1	13.7
	recall	61.3	89.6	95.9
	F1-score	52.9	45.1	24.0
sGAIA	precision	48.9	32.4	14.2
	recall	63.5	92.3	97.5
	F1-score	55.3	**48.0**	**24.8**

the line profiles, we can find that the proposed sGAIA generates a variety of suitable augmented images. We confirmed from these results that the proposed sGAIA largely improved the lesion detection performance and showed superior performance to the baseline and the method that achieved excellent results in the past study [6]. At the detection threshold α = 0.20, our proposed system showed good balance of detection performance for malicious regions (recall of 92.3% with a precision of 32.4%). At this threshold, the recall rate is almost 7% better than radiologists reported in literature (85.5%) and the ratio of true positive:false positive is suppressed to 1:3. In other words, it is highly likely that 1/3 of the detected regions by our system are malicious. Since the purpose of this study is to act as a practical screening system for gastric X-ray images, this result is promising. In addition, since the proposed sGAIA is an on-line augmentation technique, it therefore does not affect the execution time during testing. The processing time required to process one X-ray image is 0.360 s. From this point of view, the proposed sGAIA is very suitable for application as a screening system. In contrast, our proposal with sGAIA may fail if normal classifier successfully identify the location of the malignancy lesion, such as the bottom case of Fig. 5. We believe this is caused by inappropriate enhancement of the training images introduced by sGAIA. In the future, we will improve these problems by developing methods that focus more on and emphasize the area around the lesion.

V. CONCLUSIONS

In this study, we proposed a stochastic gastric image augmentation method for the automated detection of gastric cancer from X-rays, in which we emphasis gastric folds prior to training. Based on experiments, we have confirmed that our proposed augmentation technique improves considerably the detection performance of gastric cancer tumors from X-ray images in a clinical setting.

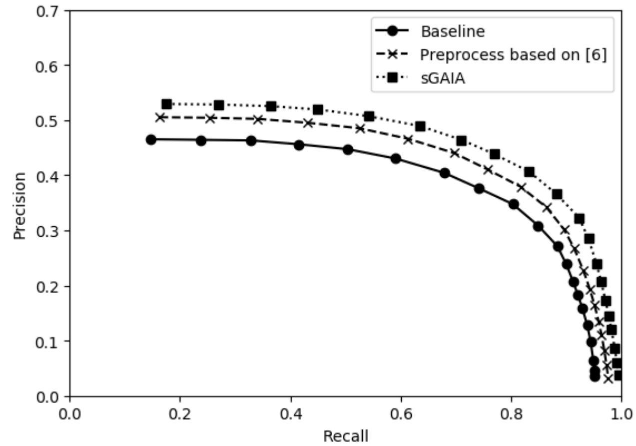
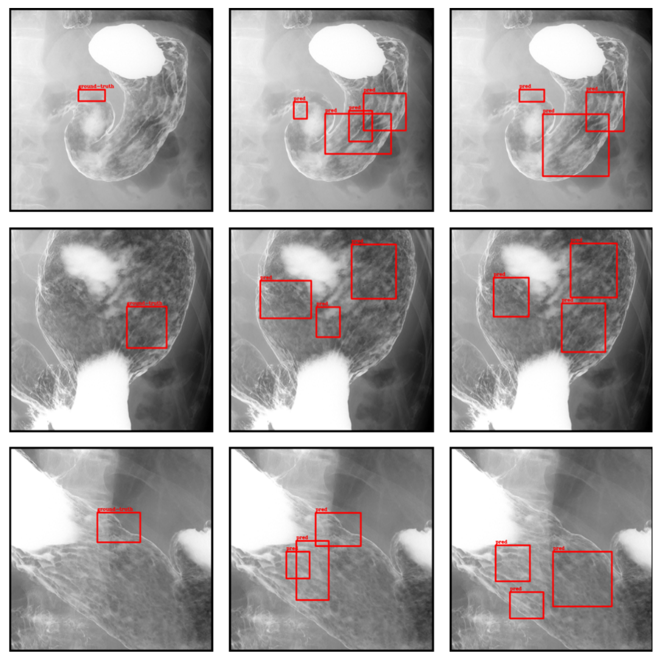


Fig. 4. Precision-recall curves for the three experiments



(a) (b) (c)

Fig. 5. Example results of images that benefited from sGAIA (top, middle), image that failed prediction due to sGAIA (bottom)
(a) Grand-truth, (b) Baseline, (c) sGAIA

REFERENCES

[1] GLOBOCAN2018, "CANCER FACT SHEETS, Stomach(C16)," http://gco.iarc.fr/today/data/factsheets/cancers/7-Stomach-fact-sheet.pdf, 2018.
[2] C. Hamashima, M. Okamoto, M. Shabana, et al, "Sensitivity of endoscopic screening for gastric cancer by the incidence method," Int J Cancer, vol. 133, no. 3, pp. 653-659, 2013.
[3] N. Uemura, S. Okamoto, S. Yamamoto, N. Matsumura, S. Yamaguchi, et al, "Helicobacter pylori infection and the development of gastric cancer," N Engl J Med 345: 784-789, 2001.
[4] P. Correa, M. B. Piazuelo, and M. C. Camargo, "The future of gastric cancer prevention," Gastric cancer, vol. 7, no. 1, pp. 9-16, 2004.
[5] K. Ishihara, T. Ogawa, and M. Haseyama, "Helicobacter pylori infection detection from multiple x-ray images based on combination use of support vector machine and multiple kernel learning," IEEE International Conference on Image Processing (ICIP), pp. 4728-4732, 2015.
[6] K. Abe, H. Nakagawa, M. Minami, and H. Tian, "Features for discriminating normal cases in mass screening for gastric cancer with double contrast x-ray images of stomach," Journal of Biomedical Engineering and Medical Imaging, vol. 1, no. 6, 2014.
[7] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M.

B. Gotway, and J. Liang, “Convolutional neural networks for medical image analysis: Full training or fine tuning?,” IEEE Trans. Medical Imaging, vol. 35, no. 5, pp. 1299-1312, 2016.
[8] K. Ishihara, T. Ogawa, and M. Haseyama, “Detection of gastric cancer risk from x-ray images via patch-based convolutional neural network,” IEEE International Conference on Image Processing (ICIP), pp. 2055-2059, 2017.
[9] R. Togo, N. Yamamichi, K. Mabe, Y Takahashi, C. Takeuchi, M. Kato, N. Sakamoto, K. Ishihara, T. Ogawa, and M. Haseyama, "Detection of gastritis by a deep convolutional neural network from double-contrast upper gastrointestinal barium X-ray radiography," Journal of Gastroenterology, pp. 1-9, 2018.
[10] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
[11] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: towards real-time object detection with region proposal networks,” Neural Information Processing Systems (NIPS), 91-99, 2015.
[12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “SSD: Single shot multibox detector,” European Conference on Computer Vision, pp. 21-37, 2016.
[13] Z. Li, C. Wang, M. Han, Y. Xue, W. Wei, L. Li, and F. Li, “Thoracic disease identification and localization with limited supervision,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
[14] T. Hirasawa, K. Aoyama, T. Tanimoto, et al, “Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images,” Gastric Cancer (2018) 21:653-660, 2018.
[15] J. Dai, Y. Li, K. He, and J. Sun, “R-FCN: Object detection via region-based fully convolutional networks,” Neural Information Processing Systems (NIPS), 2016.
[16] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy, “Speed/accuracy trade-offs for modern convolutional object detectors,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
[17] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[18] N. Qian, “On the momentum term in gradient descent learning algorithms,” Neural networks, vol. 12, no. 1, pp. 145-151, 1999.