HOKKAIDO UNIVERSITY

Title	Self-supervised learning for gastritis detection with gastric X-ray images
Author(s)	Li, Guang; Togo, Ren; Ogawa, Takahiro et al.
Citation	International Journal of Computer Assisted Radiology and Surgery, 18, 1841-1848 https://doi.org/10.1007/s11548-023-02891-5
Issue Date	2023-04-11
Doc URL	https://hdl.handle.net/2115/91736
Rights	This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/s11548-023-02891-5
Type	journal article
File Information	Guang_IJCARS_2023.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Noname manuscript No.
(will be inserted by the editor)

Self-Supervised Learning for Gastritis Detection with Gastric X-ray Images

Guang Li [1] · **Ren Togo** [2] · **Takahiro Ogawa** [2] · **Miki Haseyama** [2]

Received: date / Accepted: date

Abstract *Purpose:* Manual annotation of gastric X-ray images by doctors for gastritis detection is time-consuming and expensive. To solve this, a self-supervised learning method is developed in this study. The effectiveness of the proposed self-supervised learning method in gastritis detection is verified using a few annotated gastric X-ray images.

Methods: In this study, we develop a novel method that can perform explicit self-supervised learning and learn discriminative representations from gastric X-ray images. Models trained based on the proposed method were fine-tuned on datasets comprising a few annotated gastric X-ray images. Five self-supervised learning methods, *i.e.*, SimSiam, BYOL, PIRL-jigsaw, PIRL-rotation, and SimCLR, were compared with the proposed method. Furthermore, three previous methods, one pretrained on ImageNet, one trained from scratch, and one semi-supervised learning method, were compared with the proposed method.

Results: The proposed method′s harmonic mean score of sensitivity and specificity after fine-tuning with the annotated data of 10, 20, 30, and 40 patients were 0.875, 0.911, 0.915, and 0.931, respectively. The proposed method outperformed all comparative methods, including the five self-supervised learning and three previous methods. Experimental results showed the effectiveness of the proposed method in gastritis detection using a few annotated gastric X-ray images.

Conclusions: This paper proposes a novel self-supervised learning method based on a teacher-student architecture for gastritis detection using gastric X-ray images. The proposed method can perform explicit self-supervised learning and learn discriminative representations from gastric X-ray images. The proposed method exhibits potential clinical use in gastritis detection using a few annotated gastric X-ray images.

Keywords Deep learning · Medical image analysis · Gastric X-ray examination · Self-supervised learning

Guang Li
guang@lmd.ist.hokudai.ac.jp

Ren Togo
togo@lmd.ist.hokudai.ac.jp

Takahiro Ogawa
ogawa@lmd.ist.hokudai.ac.jp

Miki Haseyama
mhaseyama@lmd.ist.hokudai.ac.jp ✉

[1] Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan

[2] Faculty of Information Science and Technology, Hokkaido University, Sapporo, Japan

1 Introduction

Gastric cancer is one of the most severe malignant tumors worldwide, with the fifth highest incidence and third highest mortality rates [1]. The prevalence of gastric cancer is declining globally, nonetheless varying across regions [20]. The percentage of people who develop gastric cancer is higher in Eastern Asia than in other regions. Among the predisposing factors of gastric cancer, chronic gastritis is considered the leading risk factor [27]. Therefore, we focus on chronic gastritis in this study.

Chronic gastritis is considered the first step of gastric mucosal changes, leading to gastric cancer. Therefore, gastritis screening has been used for identifying patients with the risk of gastric cancer [11]. Different

methods to evaluate gastritis exist, for example, double-contrast upper gastrointestinal barium X-ray radiography, upper gastrointestinal endoscopy, and serum anti-*H. pylori* IgG titer [26]. Among such methods, the screening of gastric X-ray images is still the most simple and widely used method, and thus, it is suitable for mass screening in East Asia. However, diagnosis requires technical knowledge and time for screening. Computer-aided diagnosis (CAD) systems that automatically analyze gastric X-ray images and provide supporting information for physicians can overcome these problems.

With the development of supervised learning based on deep convolutional neural networks (DCNNs) [10], significant progress has been achieved in medical image analysis [23]. Gastritis detection from gastric X-ray images with high reliability using DCNN-based CAD systems has been proven in our previous studies [26]. Similar to most medical image analysis studies, our previous methods are based on supervised learning, and therefore, their performance depends on a large number of manually annotated gastric X-ray images [9]. However, generating the annotations of complex gastric X-ray images typically requires expert knowledge. Thus, it is expensive and time-consuming [28]. Consequently, the scarcity of expert-level annotations is one of the main limitations impacting real-world applications in gastritis detection.

In recent years, self-supervised learning has been extensively researched [19]. Unlike supervised learning that requires manually annotated labels, self-supervised learning benefits from image characteristics (*e.g.*, color and texture) without the manual annotation of labels [8]. Subsequently, a model trained with self-supervised learning can be used to fine-tune the gastric X-ray image dataset. As fine-tuning after self-supervised learning requires only a small portion of annotated data, a feasible solution is provided for insufficient data annotations in gastritis detection. For example, a previous study [5] predicted the rotation degrees of images, and the study [22] learned discriminative representations by playing a jigsaw game on images. Because these methods are designed for ordinary images, they cannot learn sufficient discriminative representations from gastric X-ray images.

Herein, we propose a novel self-supervised learning method based on a teacher-student architecture for gastritis detection using gastric X-ray images. We present cross-view and cross-model losses that enable explicit self-supervised learning and learn discriminative representations from gastric X-ray images. Experimental results show that the proposed method achieves a high detection performance for gastritis detection using only a few annotations. To the best of our knowledge, this is the first study to prove the effectiveness of self-supervised learning in the field of gastritis detection using gastric X-ray images.

Our contributions are summarized as follows:

- We propose a novel self-supervised learning method for learning discriminative representations from gastric X-ray images.
- We realize high detection performance on a complex gastric X-ray image dataset with only a few annotations.

2 Methods

2.1 Gastric X-ray image preprocessing

The resolution of the patient-level gastric X-ray images in our dataset is 2,048 × 2,048 pixels. Further, the dataset contains only a few images. To fully use the semantic information of the gastric X-ray images, we divide each patient-level image into patches and manually annotate them with the following three labels:

- $\mathcal{O}$: patches outside the stomach (outside patches),
- $\mathcal{N}$: patches extracted from negative (non-gastritis) X-ray images inside the stomach (negative patches),
- $\mathcal{P}$: patches extracted from positive (gastritis) X-ray images inside the stomach (positive patches).

Figure 2 shows examples of patient-level gastric X-ray images. As shown in the figure, the stomach without gastritis (Figure 2-(a)) includes straight folds and uniform mucosal surface patterns. Conversely, the stomach with gastritis (Figure 2-(b)) has non-straight folds and coarse mucosal surface patterns. Figure 3 illustrates the examples of divided gastric X-ray image patches. Figure 3-(a) illustrates outside patches in $\mathcal{O}$. Figure 3-(b) displays negative patches in $\mathcal{N}$. Figure 3-(c) shows positive patches in $\mathcal{P}$.

2.2 Self-supervised learning

An overview of the proposed method is shown in Fig. 1. The teacher-student architecture comprises two networks with the same structure, where the weights of the teacher network are an exponential moving average of the weights of the student network [25]. Encoder f_θ, projector p_θ, and predictor g_θ belong to the student network. Encoder f_ψ and projector g_ψ belong to the teacher network. The teacher-student architecture is designed for learning discriminative representations (*e.g.*, gastric folds or mucosal surfaces) from gastric patches.

Given an input gastric X-ray patch x, two transformations t_1 and t_2 are randomly sampled from the

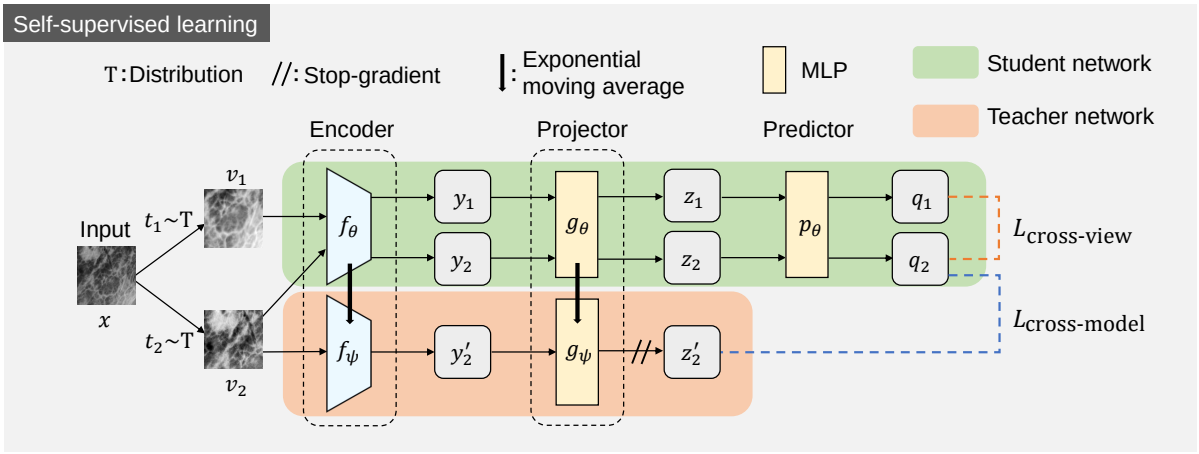
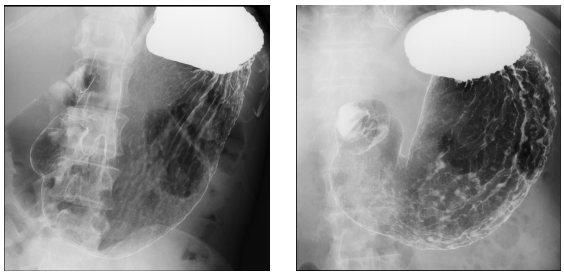
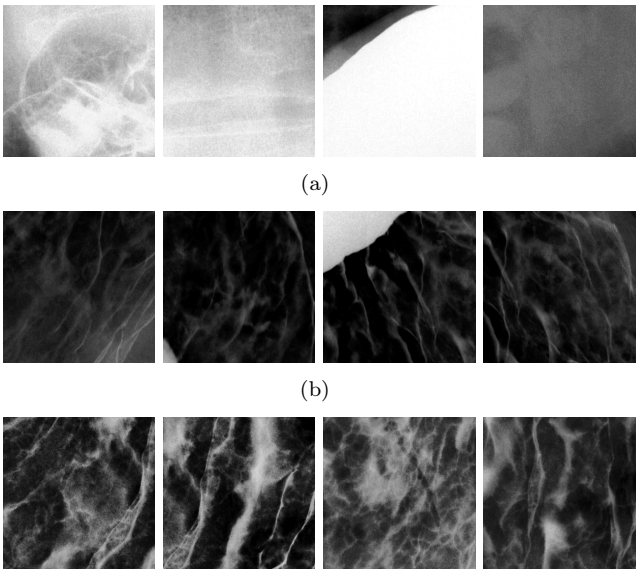


Fig. 1 Overview of the proposed method. Our method minimizes (1) a cross-view loss between the gastric features of two views from the student network and (2) a cross-model loss between the gastric features of the same view from the teacher-student networks. MLP represents multilayer perceptron.



(a) (b)

Fig. 2 Examples of patient-level gastric X-ray images: (a) negative image and (b) positive image.



(a) (b) (c)

Fig. 3 Examples of patches: (a) outside patches in $\mathcal{O}$, (b) negative patches in $\mathcal{N}$, and (c) positive patches in $\mathcal{P}$.

distribution T to generate two views $v_1 = t_1(x)$ and $v_2 = t_2(x)$. These transformations are combined with standard data augmentation methods such as cropping, resizing, flipping, and Gaussian blurring, which can increase the diversity of gastric features. The view v_1 is transmitted into the encoder f_θ and transformed by the projector g_θ in the student network. Accordingly, the view v_2 is transmitted into the encoder f_ψ and transformed by the projector g_ψ in the teacher network, where z'_2 is the obtained gastric feature. One copy of v_2 is transmitted into the student network for calculating the final loss. Further, predictor p_θ is used to transform two views to gastric features q_1 and q_2 in the student network. The predictor p_θ and projectors g_θ, g_ψ are MLP with the same structure [2]. Since the computed gastric features are high-dimensional, MLP aims to reduce the dimensionality of the gastric features for fast learning.

Finally, self-supervised learning is performed by reducing the distance between the gastric features from the student network and reducing the distance between the gastric features from the teacher-student networks.

Cross-view loss. The cross-view loss defined by the following equation compares the representations of gastric features from the student network, which penalizes gastric features for different views from the same network:

$$\begin{aligned} L_{\text{cross-view}} &= ||\hat{q}_1 - \hat{q}_2||_2^2 \\ &= 2 - 2 \cdot \frac{\langle q_1, q_2 \rangle}{||q_1||_2 \cdot ||q_2||_2}, \end{aligned} \tag{1}$$

where $\hat{q}_1 = q_1/||q_1||_2$ and $\hat{q}_2 = q_2/||q_2||_2$ represent the normalized gastric features of v_1 and v_2 from the stu-

dent network, respectively.

Cross-model loss. The cross-model loss defined by the following equation compares the representations of the same view from the teacher-student networks, which penalizes gastric features for the same view from different networks:

$$L_{\text{cross-model}} = ||\hat{q}_2 - \hat{z}'_2||_2^2 = 2 - 2 \cdot \frac{\langle q_2, z'_2 \rangle}{||q_2||_2 \cdot ||z'_2||_2}, \quad (2)$$

where $\hat{z}'_2 = z'_2/||z'_2||_2$ represents the normalized gastric features of v_2 from the teacher network. The predictor is only used in the student network to render the teacher-student architecture asymmetric, which can prevent learning from collapsing [6]. The consistency among views from the teacher-student networks enables learning discriminative representations (*e.g.*, gastric folds or mucosal surfaces) from the gastric patches. The weights of the student network (θ) are updated by minimizing the total loss. The total loss $L_{\theta,\psi}$ and optimizing process are defined as follows:

$$L_{\theta,\psi} = L_{\text{cross-view}} + L_{\text{cross-model}}, \quad (3)$$

$$\theta \leftarrow \text{Opt}(\theta, \nabla_\theta L_{\theta,\psi}, \alpha), \quad (4)$$

Opt and α represent an optimizer and the learning rate, respectively. The weight-updating process of the teacher network (ψ) is defined as follows:

$$\psi \leftarrow \tau\psi + (1-\tau)\theta, \quad (5)$$

where τ represents the degree of moving average. The teacher network is not updated using backpropagation, because the stop-gradient operation prevents the collapse of self-supervised learning [3]. Implementing self-supervised learning, the encoder of the student network (f_θ) can learn sufficient discriminative representations from the gastric patches and can be used for fine-tuning and gastritis detection.

2.3 Fine-tuning and gastritis detection

After network training using self-supervised learning, we fine-tune the encoder of the student network (f_θ) with only a few annotated images, which can reduce the labor of doctors labeling. In the test phase, we also divide the patient-level gastric X-ray images into patches, as described in Section 2.1. Thereafter, we load the divided patches into the fine-tuned DCNN model and predict the labels of these patches. Subsequently, we calculate the number of patches estimated as $\tilde{\mathcal{N}}$ and $\tilde{\mathcal{P}}$. We do not count the number of patches estimated as $\tilde{\mathcal{O}}$ because these patches outside the stomach are not related to the final gastritis detection. Finally, we estimate the label of a patient-level gastric X-ray image as follows:

$$y^{\text{test}} = \begin{cases} 1 \text{ if } \frac{\tilde{\mathcal{P}}}{\tilde{\mathcal{N}}+\tilde{\mathcal{P}}} \geq \sigma \\ 0 \text{ otherwise} \end{cases}, \quad (6)$$

where $\tilde{\mathcal{N}}$ and $\tilde{\mathcal{P}}$ represent the number of estimated negative and positive patches, respectively. σ represents a threshold that can be adjusted according to different experimental conditions. If $y^{\text{test}} = 1$, the estimated label of a patient-level gastric X-ray image is positive, and the estimated label is negative if $y^{\text{test}} = 0$.

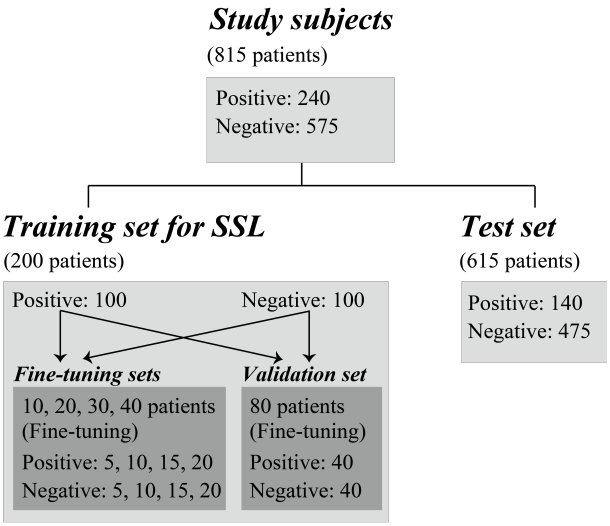


Fig. 4 Details of the partitioned datasets used in the present study. SSL represents the self-supervised learning process.

3 Experiments

3.1 Dataset

In this study, the medical data were provided by The University of Tokyo Hospital in Japan [26]. The medical data include gastric X-ray images of 815 patients subjects (240 positive and 575 negative). The resolution of these gastric X-ray images is 2,048 × 2,048 pixels. Each image has a ground truth label (positive/negative), determined based on the diagnostic results of an X-ray inspection and endoscopic examination. We used 200 patients′ images (100 positive and 100 negative) as the training set and the rest as the test set. As shown in Section 2.1, we first divided the gastric X-ray images into patches. The patch size and sliding interval were set to 299 and 50 pixels, respectively, as described in our previous study [9]. The patches in the training data

Table 1 Comparison with different self-supervised learning methods.

	10 patients			20 patients			30 patients			40 patients		
Method	Sen	Spe	HM	Sen	Spe	HM	Sen	Spe	HM	Sen	Spe	HM
PM	0.957	0.806	0.875	0.964	0.863	0.911	0.936	0.895	0.915	0.964	0.901	0.931
CM1	0.950	0.739	0.831	0.964	0.771	0.857	0.921	0.863	0.891	0.907	0.928	0.918
CM2	0.964	0.758	0.849	0.907	0.920	0.913	0.807	0.956	0.875	0.964	0.861	0.910
CM3	0.721	0.806	0.761	0.943	0.507	0.659	0.879	0.743	0.805	0.864	0.737	0.795
CM4	0.607	0.735	0.665	0.929	0.316	0.472	0.879	0.440	0.586	0.886	0.632	0.738
CM5	0.707	0.274	0.395	0.521	0.617	0.565	0.407	0.762	0.531	0.479	0.861	0.615

Table 2 Comparison with our previous methods. "-" represents there are no results reported in [18].

Method	10 patients	20 patients	30 patients	40 patients	200 patients
PM	0.875	0.911	0.915	0.931	0.954
CM6	0.860	0.870	-	-	0.922
CM7	0.382	0.759	0.435	0.870	0.954
CM8	0.348	0.477	0.563	0.644	0.876

Table 3 Hyperparameters of the proposed method.

Parameter	Value
Epoch	80
Batch size	256
Learning rate (α)	0.03
momentum	0.9
weight decay	0.0004
moving average (τ)	0.996
mlp hidden size	4096
projection size	256
View size	128
Threshold (σ)	0.5

were annotated as $\mathcal{O}$, $\mathcal{N}$, and $\mathcal{P}$ by a radiological technologist. Note that if the area inside the stomach in the patch was less than 1%, the patch was annotated as $\mathcal{O}$. If the area inside the stomach in the patch was greater than 85%, the patch was annotated as $\mathcal{N}$ or $\mathcal{P}$. We discarded the rest of the patches in the training data. The number of obtained $\mathcal{O}$, $\mathcal{N}$, and $\mathcal{P}$ patches were 48,385, 42,785, and 45,127, respectively.

The obtained 200 patients′ patches were used as the training set for the self-supervised learning process, and their label information was not used. We split the 200 patients′ patches into 120 and 80 patients′ patches (half positive and half negative) as the training and validation sets, respectively, for the fine-tuning process. Thereafter, we randomly selected 10, 20, 30, and 40 patients′ patches (half positive and half negative) from the training set as the fine-tuning sets of the fine-tuning process. Details of the partitioned data used in this study are shown in Figure 4.

3.2 Implementation

For distribution T, we used sequential data augmentation methods for random view generation, including cropping, resizing, flipping, and Gaussian blurring [15, 16]. ResNet50 [7] was used as the encoder. The optimizer used in the present method was the SGD optimizer, whose learning rate α, momentum, and weight decay were 0.03, 0.9, and 0.0004, respectively. We performed 80 epochs in the self-supervised learning process. Hyperparameters of the proposed method are shown in Table 3.

In the fine-tuning process, we used the weights of the trained encoder in the student network (f_θ) as initial weights. We assume that the effect of self-supervised learning is positively correlated with the final gastritis detection performance. To verify the effectiveness of the proposed method (PM), we compared it with the following methods. CMs 1–5 are several self-supervised learning methods. CMs 6–8 are our previous methods.

CM1. We used the weights of the trained encoder of SimSiam [3] as initial weights for the fine-tuning process.

CM2. We used the weights of the trained encoder of BYOL [6] as initial weights for the fine-tuning process.

CM3. We used the weights of the trained encoder of PIRL [21] based on the jigsaw pretext task as initial weights for the fine-tuning process.

CM4. We used the weights of the trained encoder of PIRL [21] based on the rotation prediction pretext task as initial weights for the fine-tuning process.

CM5. We used the weights of the trained encoder of SimCLR [2] as initial weights for the fine-tuning process.

CM6. A semi-supervised learning method based on tri-training [18].

CM7. A supervised learning method based on transfer learning [9].
CM8. We used random weights (trained from scratch) as initial weights for the fine-tuning process.

We performed self-supervised learning and fine-tuned the trained encoder on all four fine-tuning sets. In all the experiments, we selected the model that exhibited the highest accuracy on the validation set and tested gastritis detection performance on the test set of 615 patients′ data. Note that the settings of SimSiam and BYOL used in our experiments were strictly the same as those in the PM. The settings of PIRL and SimCLR used in our experiments were based on the suggestions in the original papers [2, 21]. We used the same settings in the fine-tuning process across all the experiments.

In the test phase, we experimentally set the threshold σ to 0.5 for realizing a high gastric detection performance. We used sensitivity (Sen), specificity (Spe), and harmonic mean (HM) as the evaluation metrics.

$$\mathrm{Sen} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{7}$$

$$\mathrm{Spe} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}, \tag{8}$$

$$\mathrm{HM} = \frac{2 \times \mathrm{Sen} \times \mathrm{Spe}}{\mathrm{Sen} + \mathrm{Spe}}, \tag{9}$$

where TP, TN, FP, and FN represent the number of true positive, true negative, false positive, and false negative, respectively. As Sen and Spe exhibit a tradeoff relationship, we consider HM as the final evaluation metric.

3.3 Results

The experimental results are shown in Table 1 and 2. Table 1 shows the patient-level gastritis detection results after fine-tuning with the annotated data of 10, 20, 30, and 40 patients in the experiments. According to Table 1, the PM outperforms other comparative methods in terms of gastritis detection. The PM′s HM score of sensitivity and specificity after fine-tuning with the annotated data of 10, 20, 30, and 40 patients are 0.875, 0.911, 0.915, and 0.931, respectively. The average HM scores of the PM are greater than those of the other self-supervised learning methods (CMs1–5) with 0.034, 0.021, 0.153, 0.293, and 0.351 on four randomly selected training sets, respectively. Experimental results showed that our method achieves a high gastritis detection performance with only a few annotations, which can substantially reduce the number of manually annotated labels required.

Table 2 shows the patient-level gastritis detection results after fine-tuning with different numbers of the annotated data patients for comparison with our previous methods. For semi-supervised learning based on tri-training, we directly use the results reported in [18] for a reasonable comparison. As per Table 2, our method not only drastically outperforms previous methods with a small amount of annotated data but also achieves excellent detection performance as the number of annotated data increases.

4 Discussion

4.1 Contributions to clinical fields

Different methods are available for the evaluation of gastritis: double-contrast upper gastrointestinal barium X-ray radiography, upper gastrointestinal endoscopy, and serum anti-*H. pylori* IgG titer [26]. Among such methods, the screening of gastric X-ray images is still the most simple and widely used method, and thus, it is suitable for mass screening in East Asia. However, the number of clinicians who can perform radiological diagnosis is decreasing owing to diverse inspection approaches [24]. Therefore, the introduction of DCNN-based CAD systems is expected in gastritis detection.

As shown in our previous studies, gastritis can be detected from gastric X-ray images with high reliability by DCNN-based CAD systems [26]. Similar to most medical image analysis studies, our previous methods are based on supervised learning [9]. The performance of our previous methods depends on a large number of manually annotated gastric X-ray images. Generating annotations of complex gastric X-ray images typically requires expert knowledge. Thus, it is expensive and time-consuming [28]. The proposed self-supervised learning method addressed this challenging problem and achieved a high detection performance with only a few annotated gastric X-ray images. The method can reduce the labor required for annotating gastric X-ray images. Although our research focuses on gastric X-ray images, we expect that the findings and methodology of this study can be applied to other types of medical images such as laparoscopic, endoscopy, and computed tomography images. Furthermore, because the self-supervised learning process is task-agnostic, we can also apply the trained encoder to other tasks, such as few-shot learning, semantic segmentation, and anomaly detection.

Our research has several limitations. The experimental data used in this study were obtained from a single medical facility, which could affect the robustness of

our method. Several studies have shown that additional data enable increasing the representation learning performance in self-supervised learning [28]. In the future, we will explore the effect of data from multiple medical facilities on self-supervised learning. Our previous study [12, 14, 17] about dataset distillation can improve the effectiveness and security of medical data sharing among different medical facilities, which fits well with self-supervised learning.

4.2 Contributions to technical fields

Early self-supervised learning methods usually learn discriminative representations based on pretext tasks. For example, the study [5] predicts the rotation degrees of images, and discriminative representations were learned by playing a jigsaw game on images [22]. Because these pretext tasks are monotonous and not complex, sufficient discriminative representations cannot be learned from images [19]. PIRL [21] proposes combining the jigsaw or rotation pretext task with learned invariant representations, outperforming supervised learning in certain tasks.

Recently, several self-supervised learning methods have achieved excellent performance on the large-scale natural image dataset ImageNet [4]. Specifically, SimSiam [3] and BYOL [6] perform self-supervised learning by directly reducing the distance between the representations of two views from the Siamese networks. These methods are efficient for processing gastric X-ray images with high-resolution and complex semantic information. Therefore, we propose a novel self-supervised learning method based on a teacher-student architecture (a variant of the Siamese networks) for gastritis detection using gastric X-ray images. We introduce cross-view and cross-model losses, which can perform explicit self-supervised learning and learn discriminative representations from gastric X-ray images [13]. To the best of our knowledge, this study is the first to prove the effectiveness of self-supervised learning in the field of gastritis detection using gastric X-ray images. Experimental results show that the PM can achieve a high patient-level detection performance in gastritis detection with only a few annotations.

5 Conclusion

We propose a novel self-supervised learning method based on a teacher-student architecture for gastritis detection using gastric X-ray images. We present cross-view and cross-model losses that enable explicit self-supervised learning and learn discriminative representations from gastric X-ray images. The experimental results showed that the proposed method achieved a high patient-level gastritis detection performance with only a few annotations.

Ethical approval

No ethics approval is required.

Declaration of competing interest

None declared.

Acknowledgments

We express our thanks to Nobutake Yamamichi of the Graduate School of Medicine, The University of Tokyo, and Katsuhiro Mabe of the Junpukai Health Maintenance Center.

Fundings

This study was supported in part by AMED Grant Number JP21zf0127004, the Hokkaido University-Hitachi Collaborative Education and Research Support Program, and the MEXT Doctoral program for Data-Related InnoVation Expert Hokkaido University (D-DRIVE-HU) program. This study was conducted on the Data Science Computing System of Education and Research Center for Mathematical and Data Science, Hokkaido University.

References

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A.: Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians **68**(6), 394–424 (2018)
2. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: Proceedings of the International Conference on Machine Learning (ICML) (2020)
3. Chen, X., He, K.: Exploring simple siamese representation learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2021)
4. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248–255 (2009)
5. Gidaris, S., Singh, P., Komodakis, N.: Unsupervised representation learning by predicting image rotations. In: Proceedings of the International Conference on Learning Representations (ICLR) (2018)

6. Grill, J.B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., Doersch, C., Avila Pires, B., Guo, Z., Gheshlaghi Azar, M., Piot, B., Kavukcuoglu, K., Munos, R., Valko, M.: Bootstrap your own latent-a new approach to self-supervised learning. In: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS) (2020)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016)
8. Jing, L., Tian, Y.: Self-supervised visual feature learning with deep neural networks: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (2020)
9. Kanai, M., Togo, R., Ogawa, T., Haseyama, M.: Gastritis detection from gastric x-ray images via fine-tuning of patch-based deep convolutional neural network. In: Proceedings of the IEEE International Conference on Image Processing (ICIP), pp. 1371–1375 (2019)
10. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015)
11. Lee, Y.C., Chiang, T.H., Chou, C.K., Tu, Y.K., Liao, W.C., Wu, M.S., Graham, D.Y.: Association between helicobacter pylori eradication and gastric cancer incidence: a systematic review and meta-analysis. Gastroenterology **150**(5), 1113–1124 (2016)
12. Li, G., Togo, R., Ogawa, T., Haseyama, M.: Soft-label anonymous gastric x-ray image distillation. In: Proceedings of the IEEE International Conference on Image Processing (ICIP), pp. 305–309 (2020)
13. Li, G., Togo, R., Ogawa, T., Haseyama, M.: Cross-view self-supervised learning via momentum statistics in batch normalization. In: Proceedings of the IEEE International Conference on Consumer Electronics – Taiwan (ICCE-TW) (2021)
14. Li, G., Togo, R., Ogawa, T., Haseyama, M.: Compressed gastric image generation based on soft-label dataset distillation for medical data sharing. Computer Methods and Programs in Biomedicine **227**, 107189 (2022)
15. Li, G., Togo, R., Ogawa, T., Haseyama, M.: Self-knowledge distillation based self-supervised learning for covid-19 detection from chest x-ray images. In: Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process., pp. 1371–1375. IEEE (2022)
16. Li, G., Togo, R., Ogawa, T., Haseyama, M.: Tribyol: Triplet byol for self-supervised representation learning. In: Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process., pp. 3458–3462. IEEE (2022)
17. Li, G., Togo, R., Ogawa, T., Haseyama, M.: Dataset distillation for medical dataset sharing. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), Workshop, pp. 1–6 (2023)
18. Li, Z., Togo, R., Ogawa, T., Haseyama, M.: Chronic gastritis classification using gastric x-ray images with a semi-supervised learning method based on tri-training. Medical & Biological engineering & Computing **58**(6), 1239–1250 (2020)
19. Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., Tang, J.: Self-supervised learning: Generative or contrastive. IEEE Transactions on Knowledge and Data Engineering (2021)
20. Maktabi, M., Köhler, H., Ivanova, M., Jansen-Winkeln, B., Takoh, J., Niebisch, S., Rabe, S.M., Neumuth, T., Gockel, I., Chalopin, C.: Tissue classification of oncologic esophageal resectates based on hyperspectral data. International Journal of Computer Assisted Radiology and Surgery **14**(10), 1651–1661 (2019)
21. Misra, I., Maaten, L.v.d.: Self-supervised learning of pretext-invariant representations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6707–6717 (2020)
22. Noroozi, M., Favaro, P.: Unsupervised learning of visual representations by solving jigsaw puzzles. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 69–84 (2016)
23. Shen, D., Wu, G., Suk, H.I.: Deep learning in medical image analysis. Annual Review of Biomedical Engineering **19**, 221–248 (2017)
24. Sugano, K.: Screening of gastric cancer in asia. Best Practice & Research Clinical Gastroenterology **29**(6), 895–905 (2015)
25. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), pp. 1195–1204 (2017)
26. Togo, R., Yamamichi, N., Mabe, K., Takahashi, Y., Takeuchi, C., Kato, M., Sakamoto, N., Ishihara, K., Ogawa, T., Haseyama, M.: Detection of gastritis by a deep convolutional neural network from double-contrast upper gastrointestinal barium x-ray radiography. Journal of Gastroenterol. **54**(4), 321–329 (2019)
27. Wroblewski, L.E., Peek, R.M., Wilson, K.T.: Helicobacter pylori and gastric cancer: factors that modulate disease risk. Clinical Microbiology Reviews **23**(4), 713–739 (2010)
28. Zhou, Z., Sodha, V., Pang, J., Gotway, M.B., Liang, J.: Models genesis. Medical Image Analysis **67**, 101840 (2021)