SOFT-LABEL ANONYMOUS GASTRIC X-RAY IMAGE DISTILLATION

Guang Li, Ren Togo, Takahiro Ogawa, and Miki Haseyama

Graduate School of Information Science and Technology,
Hokkaido University

27th IEEE International Conference on Image Processing

Session: Biomedical and Biological Image Processing I

October 28, 2020

Target of Our Study and Challenges

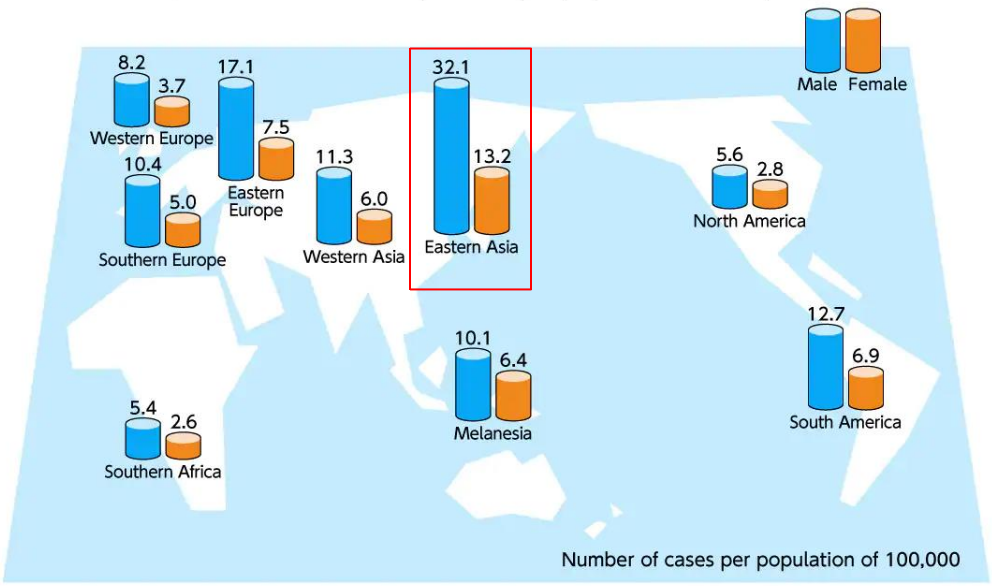
Gastric cancer has remained a burdensome disease in East Asian countries.

■ Gastric cancer in East Asian

- The percentage of people who develop gastric cancer is known to be high in East Asian compared to other regions.

- Gastritis (infection with the Helicobacter pylori) increases the risk of gastric cancer.

Prevalence of stomach cancer by region (2018) [1]
(Estimated number of patients per population of 100,000)



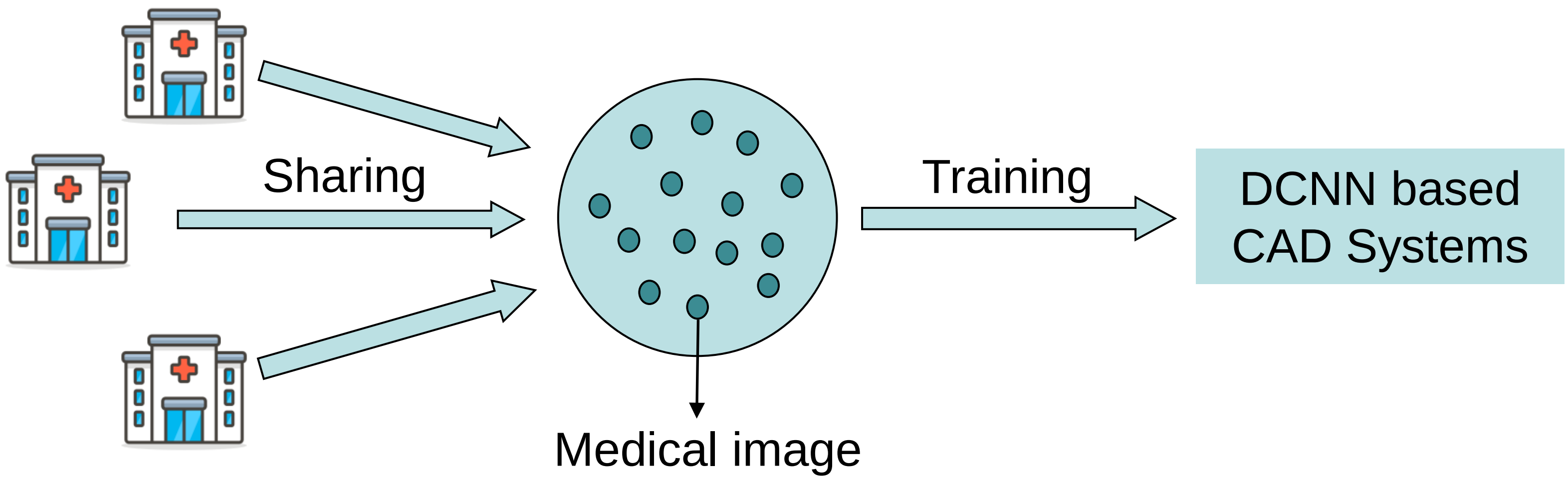
■ Challenges

- ・DCNN based gastritis CAD (Computer-aided diagnosis) systems are important.
- ・Gastric image sharing between different clinical facilities are needed.

Target of Our Study and Challenges

The sharing of medical images is a primary method for building high-accuracy DCNN based CAD systems [2].

■ Medical Image Sharing



■ Problems

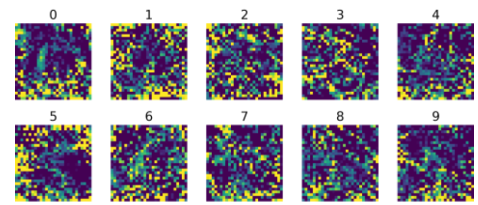
- The large size of medical image datasets make the sharing inefficient.
- The medical images often include the private information of patients.

Practical methods are needed to solve these problems in medical image sharing.

[2] P. Rajpurkar et al., "Mura: Large dataset for abnormality detection in musculoskeletal radiographs," in Proc. Medical Imaging with Deep Learning, 2018.

Solution Strategy

■ Dataset Distillation [3]



MNIST [4]
10 images
93.8% accuracy

■ Dataset reduction

- Can distill the valid information based on gradient descent.

CIFAR10 [5]
100 images
54.0% accuracy

■ Anonymity

- Can generate images with different data distributions.

Dataset Distillation can generate distilled anonymous images.

- Large size of the medical dataset.
- Including the private information.

Dataset distillation may overcome these problems.

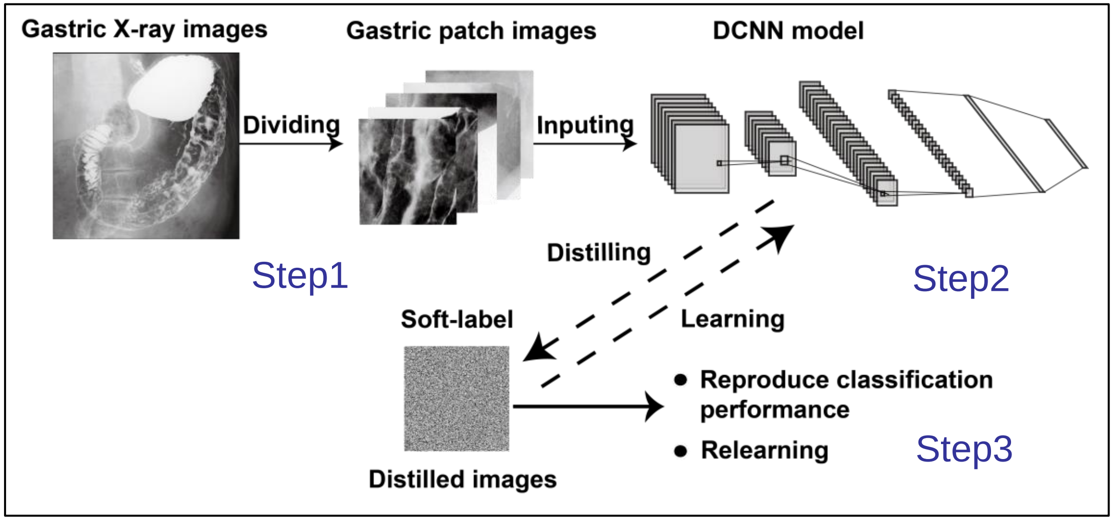
[3] T. Wang et al., "Dataset distillation," *arXiv preprint arXiv:1811.10959*, 2018.
[4] Y. Lecun et al., *Proc. of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
[5] A. Krizhevsky et al., "Learning multiple layers of features from tiny images ," *Citeseer*, 2009.

Overview of the Proposed Method

■ Aim of our study

- Distill valid information of a whole gastric X-ray image dataset into several images.

- Anonymonize the gastric X-ray images.



■ Proposed method

Soft-label anonymous gastric X-ray image distillation.

Novelty

Distilling information of different patches.

Prevent the overfitting of distilling process.

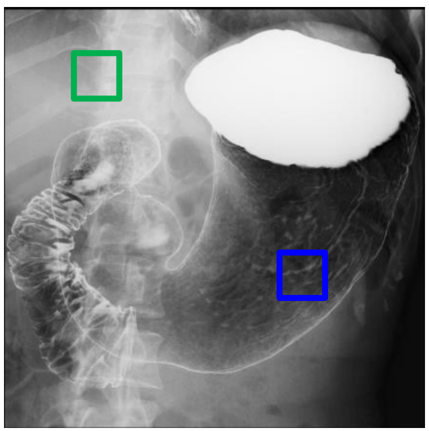
We can generate anonymous soft-label gastric images with the proposed method.

Details of Proposed Method|Step 1

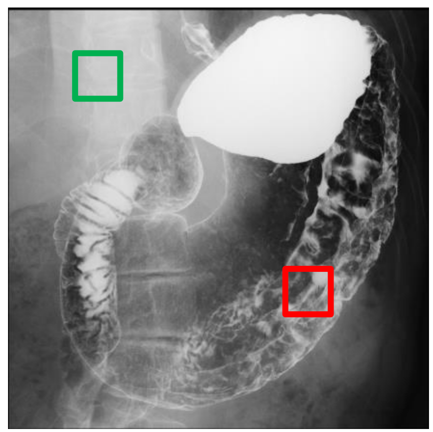
■ Characteristics of gastric X-ray images

- High resolution (ex. 2,048 × 2,048).
- Differences of gastritis and non-gastritis are described in local regions.

■ Patch-based gastric X-ray image labeling



Non-gastritis image (Negative)



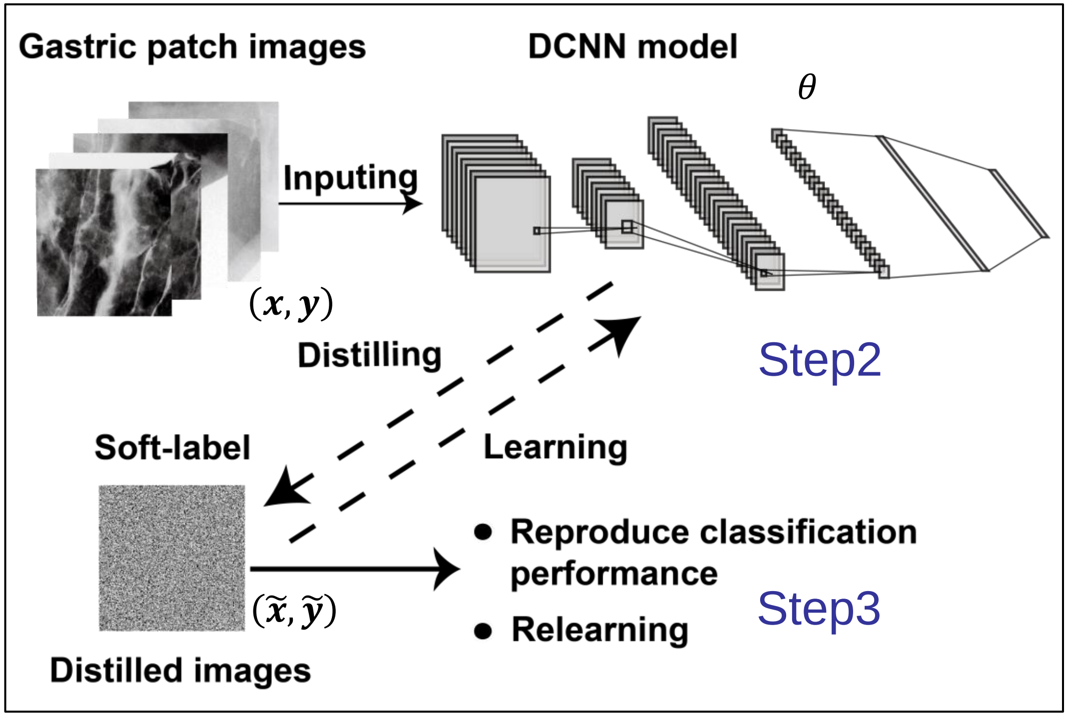
Gastritis Image (Positive)

Irrelevant patch (I)
Negative patch (N)
Positive patch (P)

Divided patches are used for soft-label patch image distillation of the next step.

Details of Proposed Method|Steps 2 & 3

■ Soft-label gastric patch image distillation



- Compute updated weights with the distilled data.

$$\theta_{i+1} \leftarrow \theta_i - \tilde{\alpha}\nabla_{\theta_i}\ell\left(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \theta_i\right)$$

- Evaluate the objective function with the derived new weights.

$$\begin{aligned}\tilde{\mathbf{x}}^*, \tilde{\mathbf{y}}^*, \tilde{\alpha}^* &= \arg\min \mathcal{L}\left(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\alpha}; \theta_i\right)\\ &= \arg\min \ell\left(\mathbf{x}, \mathbf{y}, \theta_{i+1}\right)\\ &= \arg\min \ell\left(\mathbf{x}, \mathbf{y}, \theta_i - \tilde{\alpha}\nabla_{\theta_i}\ell\left(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \theta_i\right)\right)\end{aligned}$$

- Update distilled data.

$$\tilde{\mathbf{x}} \leftarrow \tilde{\mathbf{x}} - \alpha\nabla_{\tilde{\mathbf{x}}}\mathcal{L}$$
$$\tilde{\mathbf{y}} \leftarrow \tilde{\mathbf{y}} - \alpha\nabla_{\tilde{\mathbf{y}}}\mathcal{L}$$
$$\tilde{\alpha} \leftarrow \tilde{\alpha} - \alpha\nabla_{\tilde{\alpha}}\mathcal{L}$$

l : loss function α : learning rate
$\tilde{\alpha}$: optimized learning rate

■ Full gastric X-ray image classification

- Evaluate the full gastric X-ray image classification performance.

$$y^{\text{test}} = \begin{cases} 1 & \text{if } \frac{\mathrm{Num}(\mathcal{P})}{\mathrm{Num}(\mathcal{P})+\mathrm{Num}(\mathcal{N})} \geq \epsilon \\ 0 & \text{otherwise} \end{cases}$$

Dataset and Evaluation Methods

■ Dataset

- X-ray images: gray-scale, 2,048×2,048 pixels
- Number of images: 815 images [training data 200, test data 615]
- Patch size, sliding interval: 299×299 pixels, 50 pixels
- Ground Truth: results of diagnostic report

■ Evaluation methods

$$\text{Sensitivity (Sen)} = \frac{TP}{TP+FN} \qquad \text{Specificity (Spe)} = \frac{TN}{TN+FP}$$

$$\text{Harmonic mean (HM)} = \frac{2 \times \text{sensitivity} \times \text{specificity}}{\text{sensitivity} + \text{specificity}}$$

TP: true positive TN: true negative FP: false positive FN: false negative

Experimental Conditions

■ Experimental settings

- Model, framework: ResNet18, PyTorch
- Loss function: cross-entropy loss
- The number of distilled images, distill epochs and steps: 3 (I, N, P), 3, 3
- Soft-label initial values: one-hot values of the original labels
- Distilled images: the best images in the training phase
- Ex.1: comparison with the ordinary DCNN
- Ex.2: comparison with the hard-label distillation

■ Comparative methods

- CM1: three ordinary ResNet18 models (3,000, 6,000, 9,000)
- CM2: hard-label distillation

Classification Performance Evaluation

■ Results of Ex.1 (Compare with the ordinary DCNN)

Method	Sen	Spe	HM
Proposed method (3)	0.886	**0.869**	**0.877**
ResNet18 (9000)	0.814	0.832	0.823
ResNet18 (6000)	0.907	0.760	0.827
ResNet18 (3000)	**0.914**	0.669	0.773

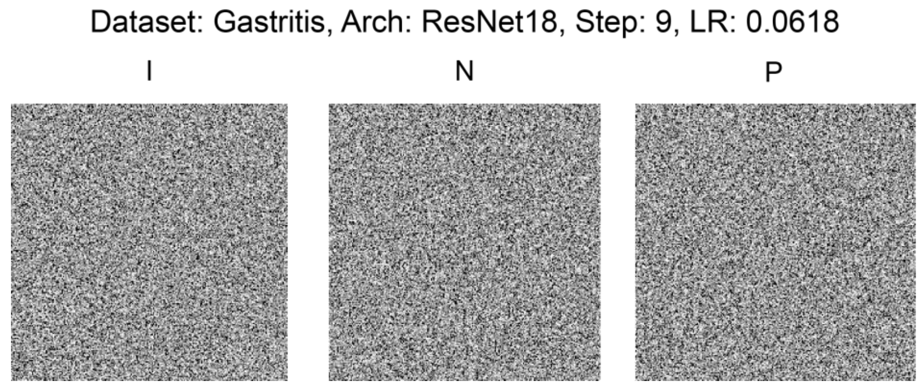
■ Results of Ex.2 (Compare with the hard-label distillation)

Method	Sen	Spe	HM
Proposed method (3)	**0.886**	0.869	**0.877**
Hard-label Distillation (3)	0.829	**0.884**	0.856

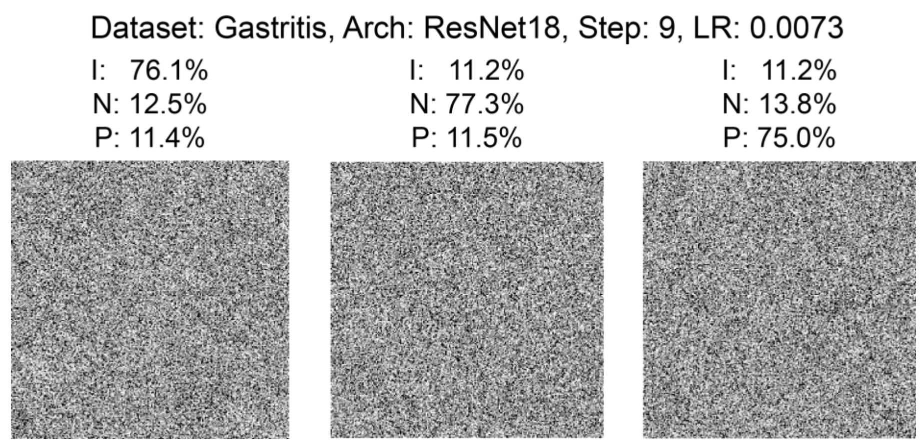
Our method realizes high performance with only three distilled images.

Anonymization Effect Evaluation

■ Distilled hard-label images



■ Distilled soft-label images



The distillation methods can completely anonymize the gastric X-ray images.

Conclusion and Future Works

■ Contributions

Effectiveness

- Achievement of a high performance with only three distilled images.
- Completely anonymized the medical images.

Significance

- Improve the efficiency and security of the medical data sharing.

■ Future works

- Transfer learning with the distilled images.
- Improve the efficiency of the distillation algorithm.

Contact Inforamtion

If you have any questions, please contact me.

Email: guang@lmd.ist.hokudai.ac.jp