Detection and Classification of Lung Abnormalities by Use of Convolutional Neural Network (CNN) and Regions with CNN Features (R-CNN)

Shoji Kido
Division of Information Science and Engineering
Graduate School of Science and Technology for Innovation
Yamaguchi University
16-1, Tokiwadai 2-chome, Ube, Yamaguchi, 755-8611
kido.ai@yamaguchi-u.ac.jp

Yasusi Hirano
Division of Information Science and Engineering
Graduate School of Science and Technology for Innovation
Yamaguchi University
16-1, Tokiwadai 2-chome, Ube, Yamaguchi, 755-8611
yhirano@yamaguchi-u.ac.jp

Noriaki Hashimoto
Division of Information Science and Engineering
Graduate School of Science and Technology for Innovation
Yamaguchi University
16-1, Tokiwadai 2-chome, Ube, Yamaguchi, 755-8611
n_hashi@yamaguchi-u.ac.jp

***Abstract*— Image-based computer-aided diagnosis (CADx) algorithm by use of convolutional neural network (CNN) does not necessarily require an image-feature extractor. Therefore, image-based CADx is powerful compared with feature-based CADx that requires the image-feature extractor for differential diagnosis of lung abnormalities such as lung nodules and diffuse lung diseases. We have also developed an image-based computer-aided detection (CADe) algorithm by use of regions with CNN features (R-CNN) for detection of lung abnormalities. We evaluated the performance of image-based CADx by use of CNN and that of image-based CADe by use of R-CNN for various kinds of lung abnormalities such as lung nodules and diffuse lung diseases.**

Keywords—computer-aided diagnosis (CAD), deep learning, convolutional neural network (CNN), regions with CNN features (R-CNN)

I. Introduction

In the diagnosis of radiological images, detection and classification processes are important. Lung diseases such as lung nodules and diffuse lung diseases have wide variations. Lung nodules include malignant and benign nodules. Also, diffuse lung diseases spread whole lungs widely, and include many causes such as infections and tumors. They include many clinical important diseases. For assisting radiologists' diagnosis, computer-aided diagnosis (CAD) systems include two types of CAD algorithms such as a computer-aided detection (CADe) that detect abnormal lesion, and a computer-aided diagnosis (CADx) that differentiate abnormal lesion into benign or malignant. In previous CAD algorithms, we used image features that could detect and classify lung abnormalities such as lung nodules or diffuse lung disease patterns. These image features are useful for computer-aided classification on lung diseases. However, to define such image features is a difficult task due to complicated image patterns of lung diseases. Deep learning technique has dramatically improved the state-of-the art in pattern recognition in the fields of speech and vison. Moreover, convolutional neural network (CNN) has brought about breakthrough in pattern recognition of images including medical images. In usual CAD algorithms, designing an image-feature extractor is important. However, this task is difficult. On the other hand, a CAD algorithm by use of CNN does not necessarily require the image-feature extractor. In this study, we have developed an image-based CADx for differential diagnosis of lung abnormalities such as lung nodules and diffuse lung diseases by use of CNN. CNN shows high performance for classification of natural images. Therefore, many researchers have studied about differential diagnosis of lung abnormalities such as lung nodules and diffuse lung diseases [1]. On the other hand, a method of regions with CNN features (R-CNN) was proposed for detection of objects such as persons, animals on natural images [2]. Therefore, we have developed an image-based CADe for detection of lung abnormalities such as lung nodules and diffuse lung diseases by use of R-CNN.

II. Image Cases

A. *Lung nodules*

We used 163 patients with lung nodule cases on CT images. We used two-dimensional images of nodules. All nodules were histologically proven. In the malignant nodules, adenocarcinomas and squamous cell carcinomas were included. In the benign nodules, hamartomas, tuberculomas, and organizing pneumonias were included.

B. *Diffuse lung diseases*

We used 372 patients with or without diffuse lung abnormalities on CT images. We used two-dimensional images. These image cases included 56 consolidation (CON), 56 honeycombing (HON), 121 ground-glass opacity (GGO), 167 emphysema (EMP), and 55 normal (NOR).

978-1-5386-2615-3/18/$31.00

C. *Image-based CADx by use of CNN*

We used a CNN as a non-feature extraction based approach. In this study, we used relatively small number of image cases. Therefore, we used the pre-trained CNN model “AlexNet” which has been trained on the ImageNet dataset, which has 1,000 object categories and 1.2 million training images [3, 4]. We also used data augmentation for supplementing a small number of image data. In the first step, image data were augmented by 8 times with rotation and reflection. For lung nodule cases, we used 163 nodules, and we obtained 1,304 nodules with data augmentation. For diffuse lung disease cases, we used 9,635 patches (715 CON, 1,051 HON, 1,886 GGO, 3,474 EMP, and 2,509 NOR) from 372 patients, and we obtained 77,080 patches with data augmentation. In all cases, same number of input images for benign or malignant nodules or all patterns of diffuse lung diseases were randomly selected. The images were divided into training and validation sets. 30% of images from each set were selected for the training data and the remainder, 70%, for the validation data. Randomization was performed to avoid biasing the results. The CNN model processed the training and test sets. In the first, features of training images were extracted using CNN. Next, these features were used to train a multiclass support vector machine (SVM). Finally, the multiclass SVM classifier evaluated validation set.

D. *Image-based CADe by use of R-CNN*

R-CNN is an object detection framework, which uses a CNN to classify image regions within an image. In our previous study, we used a sliding window for classification of diffuse lung diseases on CT images [5]. Instead of using a sliding window, the R-CNN detector only deals with those regions that are likely to contain object. R-CNN creates bounding boxes, or region proposals, using a selective search. Once the proposals are created, R-CNN warps the region to a standard square size and passes it through to a CNN model. On the final layer of the CNN, R-CNN adds a SVM that classify this is the target objet such as a lung nodule or a region with diffuse lung disease pattern (Fig.1). We marked bounding boxes of abnormal lesions on CT images for training data. We used a leave-one-out method for evaluation. Therefore, we selected one case for validation, and the other cases were used for training. R-CNN was trained with marked abnormal lesions, and it marked bounding boxes of abnormal lesions on the test image.

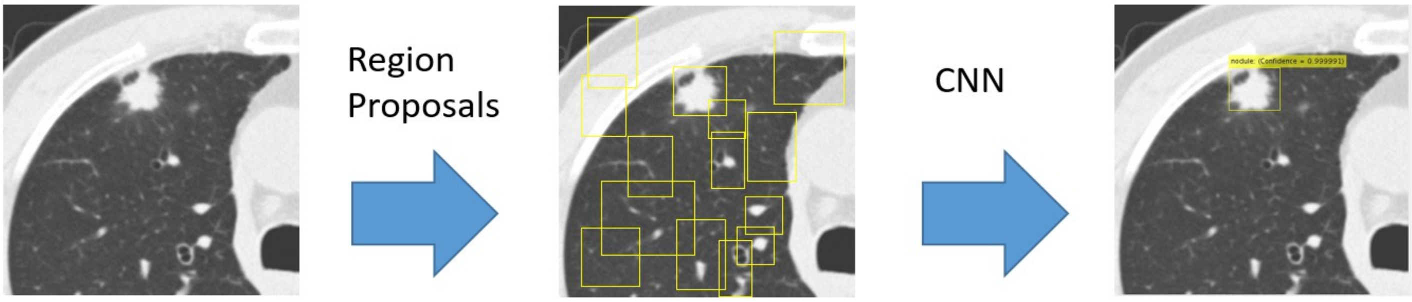

Fig.1 Scheme of an image-based CADe for lung nodule detction using R-CNN

III. RESULTS

A. *CADx results for lung nodules and diffuse lung diseases*

For the classification of benign and malignant lung nodules, the mean accuracy was 95.2% without data augmentation (TABLE □), and 99.4% with data augmentation (TABLE □). For five patterns of diffuse lung diseases (CON, EMP, GGO, HON, and NOR), the mean accuracy was 81.1% without data augmentation (TABLE □), and 84.7% with data augmentation (TABLE □).

TABLE I. NODULE CLASSIFICTION WITHOUT DATA AUGMENTATION

	Benign	*Malignant*
Benign	19	2
Malignant	0	21

TABLE II. NODULE CLASSIFICTION WITH DATA AUGMENTATION

	Benign	*Malignant*
Benign	164	2
Malignant	0	166

TABLE III. DLD CLASSIFICTION WITHOUT DATA AUGMENTATION

	CON	*EMP*	*GGO*	*HON*	*NOR*
CON	203	0	8	2	2
EMP	0	165	9	6	35
GGO	5	3	180	17	10
HON	4	22	120	168	1
NOR	0	42	17	0	156

TABLE IV. DLD CLASSIFICTION WITH DATA AUGMENTATION

	CON	*EMP*	*GGO*	*HON*	*NOR*
CON	1650	10	32	18	6
EMP	0	1358	35	85	238
GGO	40	44	1422	111	99
HON	36	86	115	1478	1
NOR	7	285	67	0	1357

978-1-5386-2615-3/18/$31.00 ©2018 IEEE

B. CADe results for lung nodules and diffuse lung diseases

For the detection of lung nodules, CADe correctly detected lung nodules attached to the chest wall and mediastinum, and it could detect various kinds of nodules such as nodule with air-bronchogram, and nodule with ground-glass opacity (Fig.2). However, it could not detect some kinds of nodules, because the number of training nodules was small (Fig.3). For the detection of diffuse lung disease patterns, we evaluated some kinds of patterns such as consolidation and honeycombing. After training with such opacities, we evaluated the test cases. In our results, diffuse lung disease patterns were well detected (Fig.4).

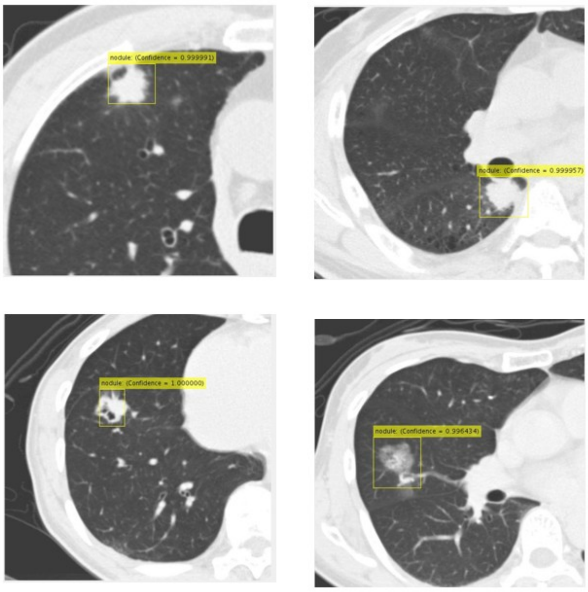

Fig.2 Succeceful cases of lug nodule dection by CADe using R-CNN

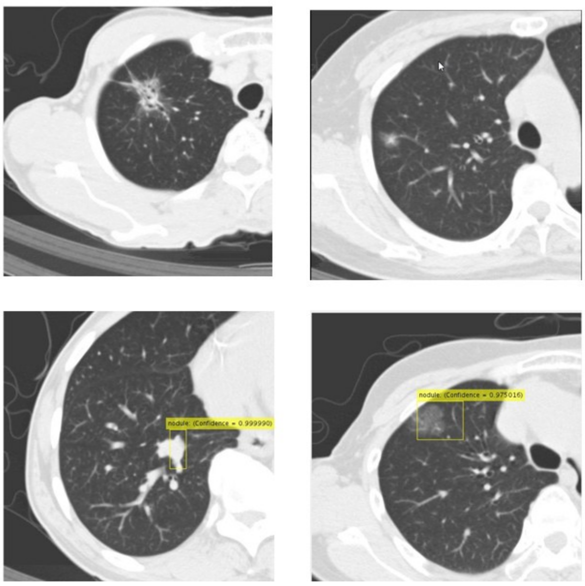

Fig.3 Failedl cases of lug nodule dection by CADe using R-CNN

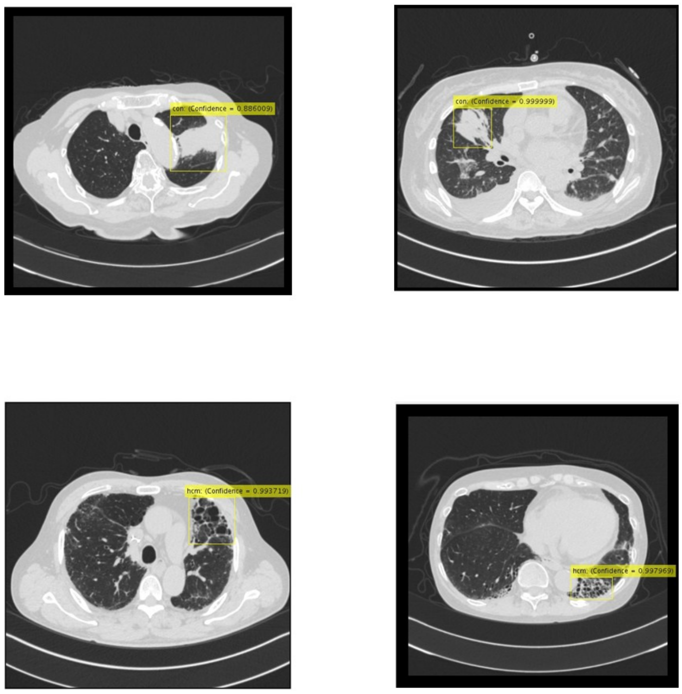

Fig.4 Detection of diffuse lung abnormalites by CADe using R-CNN. Upper two are consolidaion cases, and lower two are honeycombing cases.

IV. CONCLUSION

Image based CADx by use of CNN can differentiate various kinds of lung abnormalities such as lung nodules and diffuse lung diseases. Moreover, image-based CADe by use of R-CNN can detect also such lung abnormalities. In usual CAD algorithms that use image features, we have to extract image features by use of many kinds of image processing algorithms. However, it is difficult task to define various kinds of lung abnormalities properly. The CAD algorithms that use CNN or R-CNN do not necessarily require image-feature extractors. Therefore, both of image based CADx by use of CNN and image-based CADe by use of R-CNN are useful for radiologists' diagnosis for lung abnormalities such as lung nodules and diffuse lung diseases.

ACKNOWLEDGMENT

This research is supported in part by research grants of Grant-in-Aid for Scientific Research on Innovative areas, MEXT, Japanese Government.

REFERENCES

[1] Kido S, Hirano Y, Hashimoto N: Computer-aided classification of pulmonary diseases: Feature extraction based method versus non-feature extraction based method. Proceedings of IWAIT, 1-3, 2017..

[2] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation," Proceedings of the IEEE conference on computer vision and pattern recognition. 2014. 2012.

[3] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, June 2009.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems. pp.1106-1114, 2012.

[5] Murakami K, Kido S, Hashimoto N, et al. “Computer-aided classification of diffuse lung disease patterns using convolutional neural network,” Proceedings of the Computer Assisted Radiology and Surgery, S138-S139, 2017.