

X-ray Image Segmentation using Multi-task Learning

Sejin Park¹, Woojin Jeong¹ and Young Shik Moon^{1*}

¹ Hanyang University - Ansan Campus, Department of Computer Science and Engineering
1271 Sa-3 Dong, Ansan-Si, Gyeonggi-do, Korea
[e-mail: gnoses@hanyang.ac.kr]
[e-mail: wjjeong86@sk.com]
[e-mail: ysmoon@visionlab.or.kr]

*Corresponding author: Young Shik Moon

*Received July 12, 2019; revised October 6, 2019; accepted November 11, 2019;
published March 31, 2020*

Abstract

The chest X-rays are a common way to diagnose lung cancer or pneumonia. In particular, the finding of a lung nodule is the most important problem in the early detection of lung cancer. Recently, a lot of automatic diagnosis algorithms have been studied to find the lung nodules missed by doctors. The algorithms are typically based on segmentation network like U-Net. However, the occurrence of false positives that similar to lung nodules present outside the lungs can severely degrade performance. In this study, we propose a multi-task learning method that simultaneously learns the lung region and nodule-labeled data based on the prior knowledge that lung nodules exist only in the lung. The proposed method significantly reduces false positives outside the lung and improves the recognition rate of lung nodules to 83.8 F1 score compared to 66.6 F1 score of single task learning with U-net model. The experimental results on the JSRT public dataset demonstrate the effectiveness of the proposed method compared with other baseline methods.

Keywords: image segmentation, convolutional neural network, lung nodule segmentation

1. Introduction

Accurate and robust lesion detection is a key component of an automated medical diagnosis system [1],[2],[3],[4],[5],[6],[7],[8]. Notable achievements in deep learning have benefited several research trials in medical image analysis, [9],[10],[11],[12],[13],[14],[15],[16] and the most recent major lesion detection algorithms are based on convolutional neural networks. In particular, semantic segmentation methods such as U-Net [17] allow for precise lesion detection with respect to intensity and shape variations. Although current state-of-the-art semantic segmentation networks such as ImageNet [18] demonstrate satisfactory performance with regard to common object detection, the task of detecting lung nodules in Chest X-rays remains a challenge.

Chest X-ray is a common way to diagnose lung cancer or pneumonia. A lung nodule is a relatively small focal density in the lung. The nodule most commonly represents a cancer, especially in older adults and smokers. In particular, the finding of a lung nodule is most important problem in early detection of lung cancer. Recently, a lot of automatic diagnosis algorithms have been studied to find the lung nodules missed by doctors. However, the occurrence of false positives present outside the lungs significantly degrades performance. To avoid detection of nodule-like blobs outside lung are very important and challenging task for improving system performance and stability.

To address the challenge, we propose a novel multi-task learning (MTL) method to utilize lung regions and nodule-labeled data; the method mimics the knowledge and experience of a clinical expert regarding disease and anatomy. In our study, we considered the relationship between nodule and lung as an example of knowledge and experience of such clinical expert. A lung nodule is a small round or oval-shaped growth in the lung and is less than 3 cm (approximately 1.2 inches) in diameter. [19] Accordingly, clinical experts concentrate on the disease-related part to diagnose and treat patients. They recognize that small faint nodular densities outside the lung are not actual nodules; nonetheless, a nodule in the lung is more distinctive. Our main contributions are as follows:

- To propose a novel multiple-objective segmentation network (MO-SegNet) for detecting lung nodules in chest X-rays. We demonstrate that the anatomical region data enhance lesion detection performance in the MTL model.
- To demonstrate the effect of multiple loss balancing and how it affects nodule detection performance significantly. In addition, we demonstrate that our segmentation network can detect other lesions in the chest, such as pleural effusion.

Fig. 1 shows the pipeline of proposed lung nodule detection system. The proposed pipeline is divided into training phase, test phase and post processing phase. In the training phase, the segmentation network is trained by receiving input images, lung region labels, and nodule labels at the same time. In the test phase, only the input image is taken as an input to produce a nodule segmentation output. In the post processing phase, the bounding box is created using the connected component labeling algorithm with the nodule segmentation output.

The rest of the paper is organized as follows. Section 2 describes the related work. The MTL method is proposed in Section 3. Section 4 gives out the experiment results. The conclusion is described in Section 5.

2. Related Work

2.1. Multitask Learning

Humans often infer facts from several related knowledge sources and learn relationships based on complex tasks. An understanding of one task is often expanded and affects the learning of other tasks. This is the motivation behind MTL, [20] which effectively increases the sample size and decreases the noise level of the training data by ignoring data-dependent noise and generalizing with an improved representation, learned by composing multiple tasks. A single model with multiple related objectives has a higher probability of obtaining a more generalized representation by averaging the noise patterns.

We can view MTL as a form of inductive transfer. [21] The inductive bias is provided by auxiliary tasks, which cause the model to prefer hypotheses that explain more than one task. This generally results in solutions that generalize better. Agrawal et al. [22] demonstrate multiple visual tasks, such as ego-motion and image classification, as forms of self-supervision for visual feature learning. Further, they provide an effective demonstration of learning visual representations from non-visual access to ego-motion information in a real-world setting, by training neural networks to predict the camera transformation between pairs of images.

2.2. Semantic Segmentation

Semantic segmentation refers to understanding an image at the pixel level. [23] demonstrates that fully convolutional networks can efficiently learn to make dense predictions for per-pixel tasks, such as semantic segmentation. Almost all semantic segmentation networks have an encoder–decoder architecture to capture location information of an object. [24], [15]. The encoder reduces spatial resolution with pooling layers and the decoder recovers the spatial resolution. There are usually shortcut connections, such as in residual networks, which are akin to U-Net.

In medical image domain, [16] proposed tongue image segmentation aims to extract the image object from tongue body. However they used traditional image processing algorithms such as image thresholding, gray projection and active contour model. [25] proposed to solve segmentation of skin lesion image based on convolutional neural network and adversarial networks.

Misra et al. [26] demonstrated that, in the computer vision task, one can exploit multiple related properties to improve performance using MTL. They proposed combined semantic segmentation and surface normal estimation networks. They also demonstrated that all layers from the first convolution to the last fully connected layer are shared by all tasks, and only the last layers are task-specific. In [21], MTL with multiple regression and classification for semantic segmentation is proposed. The authors suggested that MTL is beneficial for regularization and accuracy improvement. Furthermore, they proposed using multitask loss functions to simultaneously learn to balance various classification and regression losses using homoscedastic uncertainty.

2.3 Bounding Box Detection

Another object detection method is bounding box detection network such as Faster R-CNN [36], SSD [37], YOLO [38]. For MTL with bounding box detection network, we need simultaneous bounding box regression and pixel classification (segmentation) architecture, which has a much more complicated loss function. Those architectures makes it relatively difficult to balance the two tasks. The proposed method is more advantageous for learning two segmentation branches with one loss function, and converting the bounding box into a rectangular segmentation mask does not affect significantly the performance.

3. X-ray Image Segmentation using Multi-task Learning

3.1 Lung Nodule Detection in Chest X Ray

The occurrence of false positives that similar to lung nodules present outside the lungs can severely degrade performance.

To address these challenges, we utilize anatomical regions and nodule-labeled data for MTL, which mimics the prior knowledge of doctors regarding disease and anatomy. In our study, we considered the relationship between nodule and lung as an example of knowledge and experience of such clinical expert. A lung nodule is a small round or oval-shaped growth in the lung and is less than 3 cm (approximately 1.2 inches) in diameter. Accordingly, clinical experts concentrate on the disease-related part to diagnose and treat patients. They recognize that small faint nodular densities outside the lung are not actual nodules; nonetheless, a nodule in the lung is more distinctive.

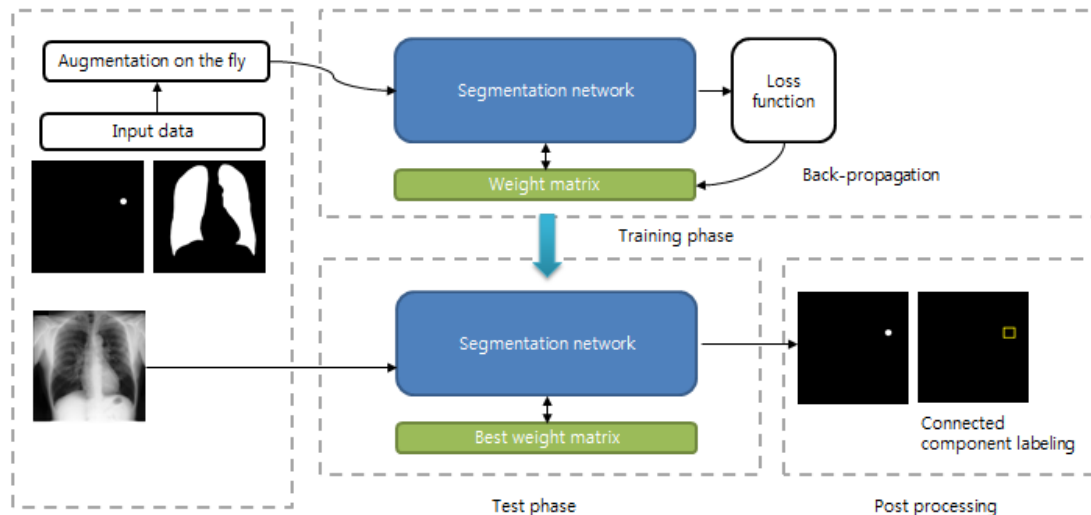


Fig. 1. The pipeline of proposed lung nodule detection system (Single task)

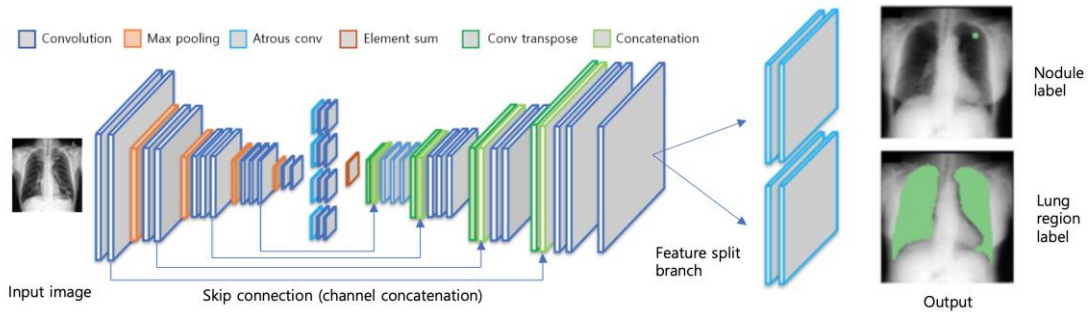


Fig. 2. Multi objective segmentation architecture for chest and lung nodule segmentation

3.2 Multi Task Learning For Image Segmentation

The most commonly used method to perform MTL in deep learning is parameter sharing of hidden layers, which is generally achieved by sharing hidden layers between all tasks. This type of parameter sharing significantly reduces the risk of overfitting. Ruder [27] demonstrated that the risk of overfitting the shared parameters is less than that of overfitting task-specific parameters. When we attempt to learn more tasks, our model has to search a more discriminative representation that captures all tasks. One more view of MTL is for auxiliary tasks. The goal of an auxiliary task in MTL is to enable the model to learn more helpful representations for main tasks. If the nodule detection problem is the main task, chest region segmentation is an auxiliary task that can generate representations that are more helpful. Similarly, the auxiliary task can be used to focus on parts of the image that a model might normally ignore. In our case, chest region segmentation results in ignoring nodules outside the lung. Fig. 2 illustrates an overview of our multiple-objective segmentation network architecture to perform chest and lung nodule segmentation simultaneously. Multiple source data is incorporated in feature split branch for MTL. Each branch share same U-Net network and split branch for lung region output and nodule region output. Brief network architecture is illustrated in Fig. 2. Our architecture has both chest and lung nodule segmentation networks, which share feature layers. The single-objective nodule segmentation network is based on the structure of the U-Net architecture, but with some modifications. The core of the U-Net architecture is an encoder–decoder scheme and it has lateral skip connections. In our modified architecture, the encoder is followed by an atrous spatial pyramid pooling (ASPP) layers to detect multiscaled objects. ASPP uses atrous (dilated) convolutions with parameters of different rates for arbitrary scale detection. [24] Atrous convolutions are special convolutions with a factor that expands the field of view. Fig. 3 shows details of ASPP module and other alternative network structures of semantic segmentation network. The factor expands the convolution filter according to the dilation rate and fills the empty spaces with zeroes. ASPP module will sharply increase the amount of parameters. However, rather than building deeper general convolutions to get larger FOV, Atrous convolution can have similar FOV with less depth. In creating a multi-scale network with a larger FOV, ASPP has relatively shallow network than composition of general convolutions. Moreover, sparse input data is more efficiently represented using as few components as possible, and thus, it generalizes satisfactorily. Therefore, varying the dilation rate of atrous convolution generates filters that detect objects at multiple scales. Using multiple parallel atrous convolution layers with different sampling rates, we can aggregate the multiscaled object detector in a single model. The decoder part has upsampling layers applied by a convolution transpose operation.

Furthermore, lateral skip connections pass the feature maps from lower layers of the contracting path to the analogous level in the expanding path. Details of the neural network layers are described in [Table 1](#).

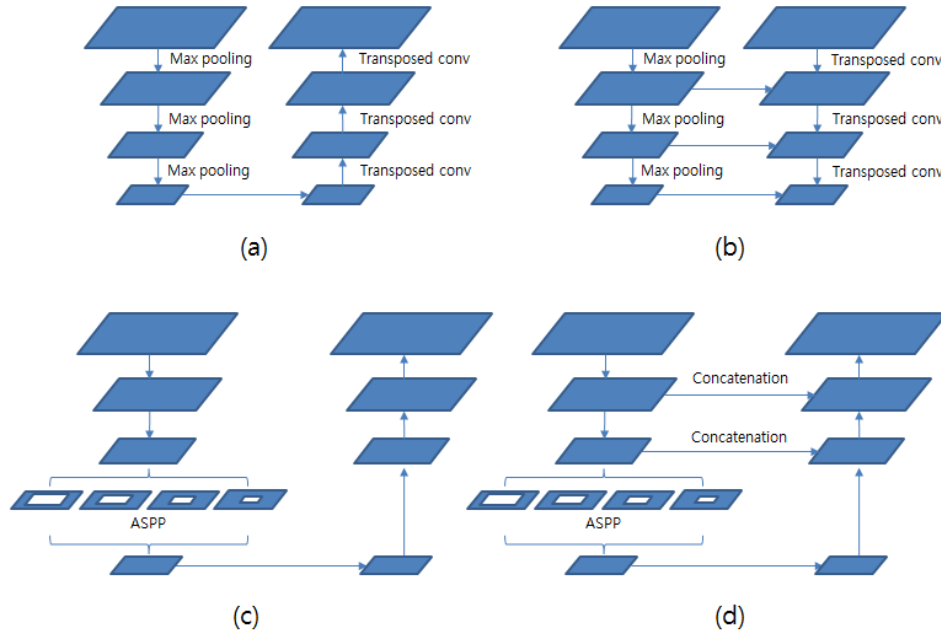


Fig. 3. Types of segmentation network architecture.

- (a) Deconvnet (b) U-Net (c) Deconvnet with Atrous pyramid pooling networker
(d) U-Net with Atrous pyramid pooling networker

Loss function for single objective function For the base segmentation model, we used adaptive weighted cross-entropy [28] exploited the class weighting parameter to manage the imbalanced size of each class. We let $w \in \mathbb{R}^k$ denote a weight vector with elements $w_k > 0$ defined over the range of class labels $k \in \{1, 2, \dots, K\}$. We then define the weighted cross-entropy as

$$\mathcal{L} = -w_c y_c \log p(x) \quad (1)$$

w is calculated by $(N - \sum r_n) / \sum r_n$ in every batch, where r_n denotes the pixel label count. Variables y_c and x are the true label and input image, respectively.

Data Augmentation In variable image settings, we may have a dataset of images captured under a limited set of conditions. However, our target images may have been captured across a variety of conditions. Based on our observations, chest X-rays differ in translation, orientation, randommargin, brightness, and have pattern noise from radiography equipment. For all observations and prior knowledge, we generate augmented data with random cropping, orientation, brightness, and with additional Gaussian noise and Poisson noise.

Optimizer We use the ADAM [29] optimizer with an initial learning rate of 0.001. The exponential decay rates for the first moment and second moment estimates are 0.9 and 0.999, respectively.

Table 1. Details of neural network architecture

No.	Layer Type	Bottom Layer	Neurons	Activation	Kernel Size	Norm
1	Conv	Input	16	RELU	3×3	GN
2	Conv	1	16	RELU	3×3	GN
3	Max Pooling	2	N/A	N/A	3×3 (stride 2×2)	GN
4	Conv	3	32	RELU	3×3	GN
5	Conv	4	32	RELU	3×3	GN
6	Max Pooling	5	N/A	N/A	3×3 (stride 2×2)	GN
7	Conv	6	64	RELU	3×3	GN
8	Conv	7	64	RELU	3×3	GN
9	Conv	8	64	RELU	3×3	GN
10	Max Pooling	9	N/A	N/A	3×3 (stride 2×2)	GN
11	Conv	10	128	RELU	3×3	GN
12	Conv	11	128	RELU	3×3	GN
13	Conv	12	128	RELU	3×3	GN
14	Max Pooling	13	N/A	N/A	3×3 (stride 2×2)	GN
15	A. Conv	14	128	RELU	3×3 (rate 3)	BN
16	Conv	15	128	RELU	1×1	GN
17	Conv	16	128	RELU	1×1	GN
18	A. Conv	14	128	RELU	3×3 (rate 5)	BN
19	Conv	18	128	RELU	1×1	GN
20	Conv	19	128	RELU	1×1	GN
21	A. Conv	14	128	RELU	3×3 (rate 7)	BN
22	Conv	21	128	RELU	1×1	GN
23	Conv	22	128	RELU	1×1	GN
24	A. Conv	14	128	RELU	3×3 (rate 9)	BN
25	Conv	24	128	RELU	1×1	GN
26	Conv	25	128	RELU	1×1	GN
27	Element Sum	17,20,23,26	128	N/A	N/A	N/A
28	T. Conv	27	128	RELU	3×3 (stride 2×2)	BN
29	Concat	28, 13	128	N/A	N/A	N/A
30	Conv	29	64	RELU	3×3	BN
31	Conv	30	64	RELU	3×3	BN
32	T. Conv	31	64	RELU	3×3 (stride 2×2)	BN
33	Concat	32, 9	64	N/A	N/A	N/A
34	Conv	33	32	RELU	3×3	BN
35	Conv	34	32	RELU	3×3	BN
36	T. Conv	35	32	RELU	3×3 (stride 2×2)	BN
37	Concat	36, 5	32	N/A	N/A	N/A
38	Conv	37	16	RELU	3×3	BN
39	Conv	38	16	RELU	3×3	BN
40	T. Conv	39	16	RELU	3×3 (stride 2×2)	BN
41	Concat	40, 2	16	N/A	N/A	N/A
42	Conv	41	16	RELU	3×3	BN

43	Conv	42	2	RELU	1×1	N/A
44	Conv	41	16	RELU	3×3	BN
45	Conv	44	2	RELU	1×1	N/A

Group Normalization We use group normalization (GN) [30] instead of batch normalization (BN). BN's error increases rapidly when the batch size becomes smaller; this is because of inaccurate batch statistics estimation. Our model uses one sample for each batch; thus, the training curve is observed to oscillate during every iteration. GN significantly mitigates such an oscillation. We have demonstrated that GN also enhances the final performance compared to Batch Normalization (BN) in Table 4, as we show in the section 4.

3.3 Weighting Multiple Objective Loss Function

We define two loss functions for \mathcal{L}_y and \mathcal{L}_z , namely a nodule segmentation loss function and a lung region segmentation loss function, respectively. We formulate the following function for balancing multiple-objective losses.

$$\begin{aligned} \mathcal{L}_{mtl} &= \lambda \mathcal{L}_y + (1 - \lambda) \mathcal{L}_z \\ \mathcal{L}_y &= -w_{n,c} y_{n,c} \log p(x) \\ \mathcal{L}_z &= -w_{l,c} y_{l,c} \log p(x) \end{aligned} \quad (2)$$

\mathcal{L}_y is weighted cross entropy with lung nodule label, \mathcal{L}_z is weighted cross with lung region. Moreover, we introduce a principled multiple loss weighting function that can be trained to balance segmentation loss in the case of large differences. Our method can learn to balance these weightings optimally. We demonstrate how loss weighting affects model performance. Moreover, we compare manually optimized fixed weighting and our loss weighting method. We observe that our loss weighting function is robust against the initial values selected for the parameters. At the end of training, the losses are weighted with the parameter λ to determine the balance between chest and lung nodule segmentation.

3.4 Implementation Details

We re-scale the X-ray images such that their shorter side is 1024 pixels. We used tensorflow 1.2.0 and python 3.6 for implementation. It took 1 day on training with single nVidia Titan-X GPU. On testing time, it took 500 msec for single X-ray image with preprocessing.

4. Experimental Results and Analysis

4.1 Experimental Dataset and Evaluation Metric

We applied the proposed method to a public dataset of chest X-rays (JSRT). [31] Segmentation performance was assessed by comparing the output of the proposed automated segmentation methods with the ground truth, in terms of F1-score and FROC (sensitivity and number of false positives per scan). We also compared our proposed method with the U-Net baseline network. For this experimental work, we used images from the public JSRT dataset,

which is the standard database of digital images with and without chest lung nodules, created by the Japanese Society of Radiological Technology (JSRT) in cooperation with the Japanese Radiological Society in 1988. It comprises 154 nodule and 93 non-nodule images as high-resolution (2048×2048 matrix size, 0.175-mm pixel size) chest X-ray images. It covers a wide density range (12 bit, 4096 grayscale). Images with chest lung nodules have annotations with the X- and Y-coordinates of the nodules.

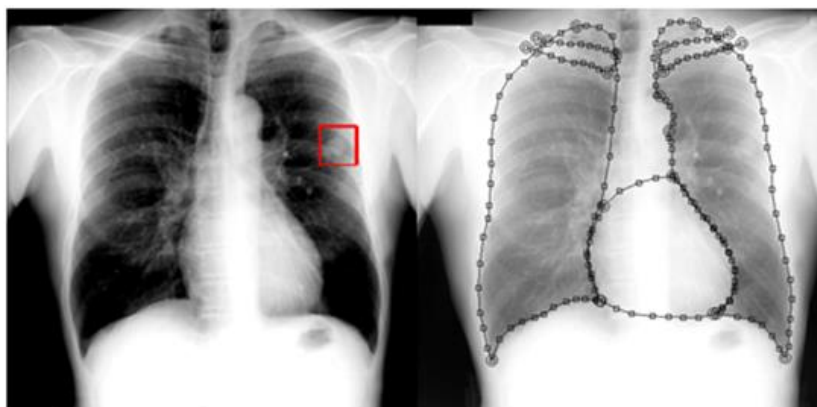


Fig. 4. JSRT dataset (left) and SCR dataset (right).

We used the segmentation in chest radiographs (SCR) database [32] for lung region labels. The SCR database has been established to facilitate comparative studies on the segmentation of the lung fields, heart, and clavicles in standard posterior–anterior chest radiographs. All chest radiographs of the SCR database were obtained from the JSRT database. Fig. 4 shows sample images from the JSRT and SCR datasets.

We evaluated the F1/Recall/Precision score by 5-fold cross-validation of positive samples. The 154 cases were divided into five subsets of similar size: four subsets were used for training and the remaining one was used for testing. Additional qualification testing was performed to verify the design and performance of the proposed segmentation network on other lesion data.

4.2 Evaluation Results

Table 2. Comparing accuracy of various data and CNN architectures

Task	Model	F1	Recall	Precision
Single	U-net	66.6±3.5 %	68.7±1.8 %	64.6 ± 2.9 %
Single	MO-SegNet	64.5±3.4 %	62.5±3.9 %	66.6 ± 3.1 %
Multiple	U-net	74.0±2.5 %	62.5±3.7 %	90.6 ± 2.1 %
Multiple	MO-SegNet	83.8±3.4 %	81.3±4.1 %	86.4 ± 3.8 %

Comparison of single and multiple objective segmentation method In Table 2, we list the results from the experiment with the effect of incorporating multiple-objective segmentation methods. In the table, task means whether single or multiple objective learning is used. Model

means whether U-net baseline model or proposed MO-SegNet is used. For U-net architecture, we reimplement U-net with feature split branch. We compare individual models of the baseline U-Net and our proposed architecture. This distinctly illustrates the benefit of multiple-objective segmentation networks, which obtain significantly better results than individual task models. The proposed method with a fixed balance loss attains a value of 83.8, 81.3, and 86.4 for F1-score, recall, and precision, respectively. Multiple task learning with both of U-net and MO-SegNet are outperformed single task learning. This means that the performance of MTL is higher regardless of model. Moreover, proposed MO-SegNet with MTL showed most higher performance.

Fig. 6 shows the false positive reduction cases outside lung by our model; **Fig. 6 (left)** shows the input image and mask of the pixels that are of nodule lesions, and **Fig. 6 (center)** and **Fig. 6 (right)** show the output of the baseline network and the proposed method, respectively. First 3 pictures of first line have nodules in groundtruth, so that it is X ray of cancer disease case. 3 pictures of second line have no nodules in groundtruth, so that it is X ray of healthy case. We can observe that the false positives (red circle) in **Fig. 6 (center)** is disappeared by the proposed method in **Fig. 6 (right)**. However detected nodule (yellow circle) is still visible in both center and right pictures in first line of cancer disease case.

Fig. 7 shows the lesions detected by our model; **Fig. 6 (left)** shows the input image and mask of the pixels that are of nodule lesions (yellow circle), and **Fig. 6 (center)** and **Fig. 6 (right)** show the output of the baseline network and the proposed method, respectively. We can observe that the false negative lesion **Fig. 6 (center)** is detected by (yellow circle) the proposed method **Fig. 6 (right)**.

Effect of weighting multiple objective loss We demonstrate the effect of weighting multiple-objective losses. Model performance is extremely sensitive to weight selection, as illustrated in **Fig. 5**. The training curve was measured by repeated learning 10 times while adjusting the weighting parameter. The change in F1 score was measured at every iteration. When the weighting parameter is 1.0 (only nodule segmentation works), it show very low performance. We select the best-performing value of loss weighting as 0.99, from the experiments in **Table 3**.

In the experiment, it was confirmed that a small difference of 1.0 and 0.99 makes a big difference in performance. This proves that the proposed MTL method shows a big performance improvement even if only a little effect. In the case of loss weighting between 0.95 and 0.99, there was a slight difference, but in the case of less than that, the performance dropped again. This shows that the performance of lung nodule segmentation may be reduced if too many tasks are performed by overbalanced weighting parameter to segment lung regions during multi task.

More detailed performance analysis We show our more detailed analysis on JSRT test set in **Table 5**. comparing to other state-of-the-art semantic segmentation methods of Deeplab V3+ [33] and Pyramid Scene Parsing Network (PSPNet) [34]. Our proposed architecture makes further improvement compared to the baseline. Finally, Deeplab V3+ and PSPNet shows 81.2 and 77.6 F1 score, which shows lower performance than ours. Two models are trained with only single task (nodule segmentation) to compare with original model architecture.

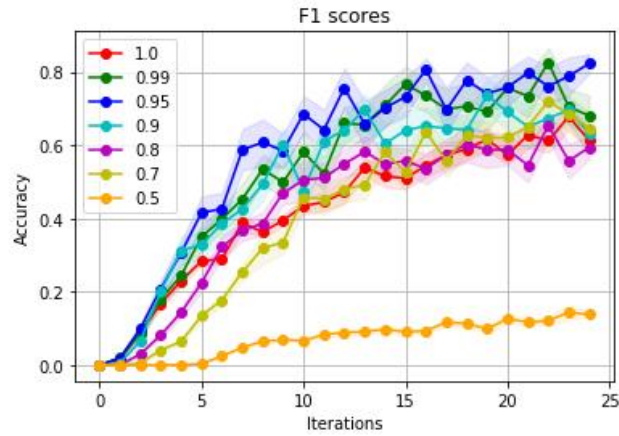


Fig. 5. Learning curve of varying weight balance

Table 3. Comparing accuracy of various data and CNN architectures

Weight balance	F1	Recall	Precision
1.0	76.0 %	62.5 %	96.9 %
0.99	83.8 %	81.3 %	86.5 %
0.95	82.5 %	79.8 %	85.4 %
0.9	78.5 %	68.7 %	91.6 %
0.8	71.2 %	75.8 %	67.1 %
0.7	75.1 %	68.7 %	82.8 %
0.5	14.7 %	81.2 %	8.1 %

Table 4. Ablation test for normalization method

Model	F1	Recall	Precision
U-net + BN	72.3 %	60.8 %	89.1 %
U-net + GN	74.0 %	62.5 %	90.6 %
MO-SegNet + BN	81.9 %	79.4 %	84.6 %
MO-SegNet + GN	83.8 %	81.3 %	86.4 %

Table 5. Comparison against state-of-the-art semantic segmentation methods

Model	F1	Recall	Precision
Deeplab V3+ [33]	81.2±3.9 %	80.1±2.9 %	82.3±4.5 %
Pyramid Scene Parsing Network (PSPNet) [34]	77.7 ± 4.8 %	75.2±5.3 %	80.3±6.9 %
MO-SegNet (proposed)	83.8±3.4 %	81.3±4.1 %	86.4 ± 3.8 %

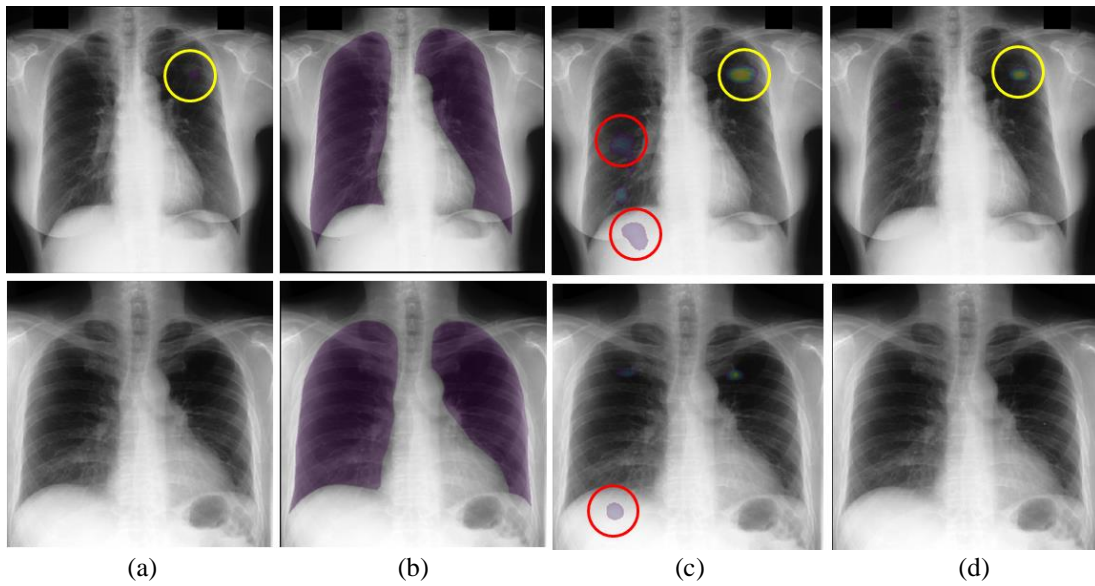


Fig. 6. Inference result for false positive reduction case outside lung.
 (a) Ground truth lesion (b) Ground truth lung region (c) Single data U-net (d) MO-SegNet

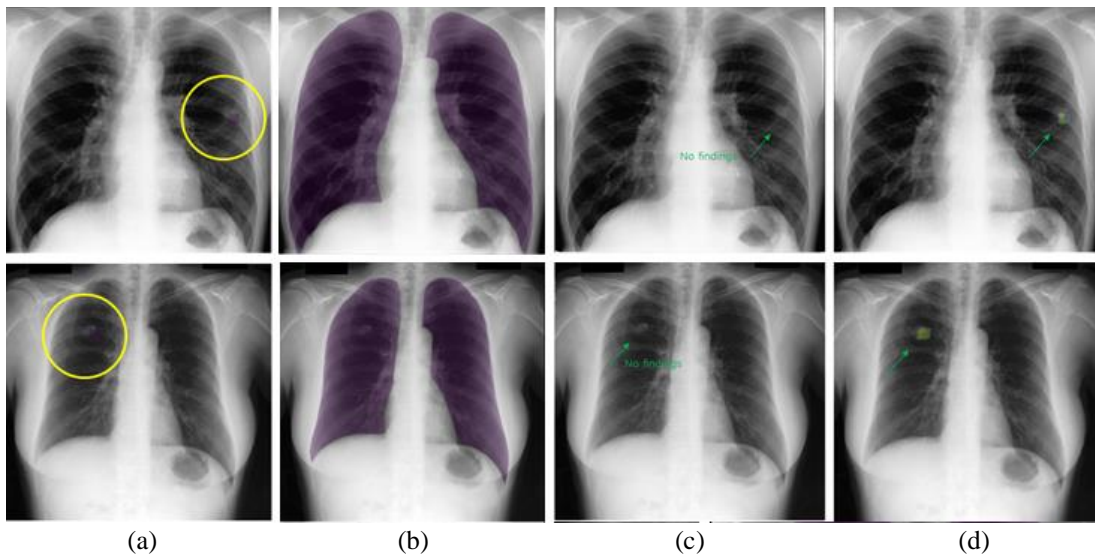


Fig. 7. Inference result of the proposed method.
 (a) Ground truth lesion (b) Ground truth lung region (c) Single data U-net (d) MO-SegNet

4.3 Extra Experiments

As an additional experimental result, we observed that the semantic segmentation network can correct errors in the ground truth labeling. An inaccurate bounding box label exists in the JSRT dataset, but our semantic segmentation network produces a more accurate bounding box for the nodule boundary. **Fig. 8** shows the output images for those cases. We also investigated the use of our segmentation network to detect other chest lesions, such as pleural effusion. Pleural effusion is the presence of water in the lungs, caused by the accumulation of excess fluid

between the layers of the pleura, outside the lungs. A horizontal water line is observed in the middle of the lung region. In that region, pleural effusion overlaps with bones and has no boundary on the other side. The nodule is shaped in an isolated region but pleural effusion is in the overlapping region, which makes detection more challenging. In spite of the difficulty, our segmentation network demonstrated particularly satisfactory performance: pleural effusion was well segmented using the proposed method, which is shown in Fig. 9. The segmentation performance was evaluated using the Dice coefficient. We used the NIH chest X-ray8 dataset [35] for pleural effusion lesion segmentation.

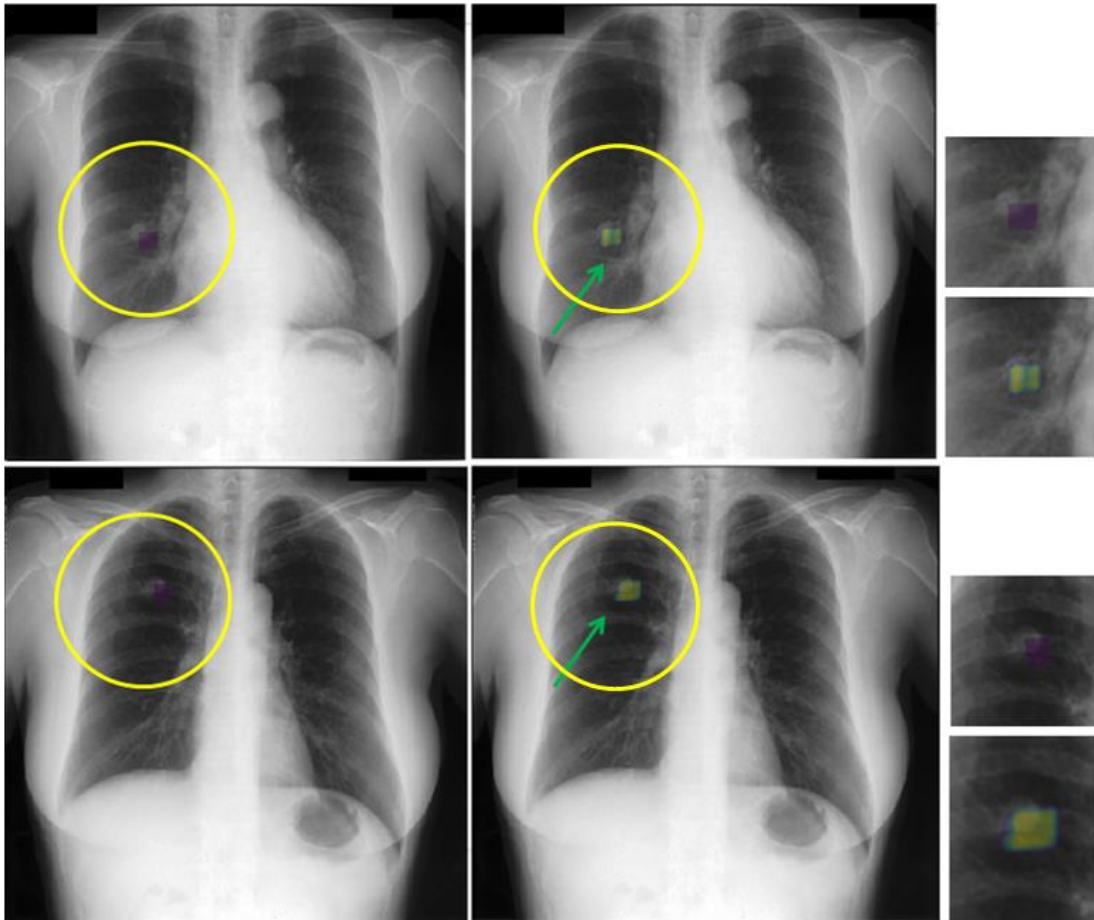


Fig. 8. Label correcting effect of semantic segmentation network
Some labels in groundtruth are miss-aligned to actual nodule boundary.

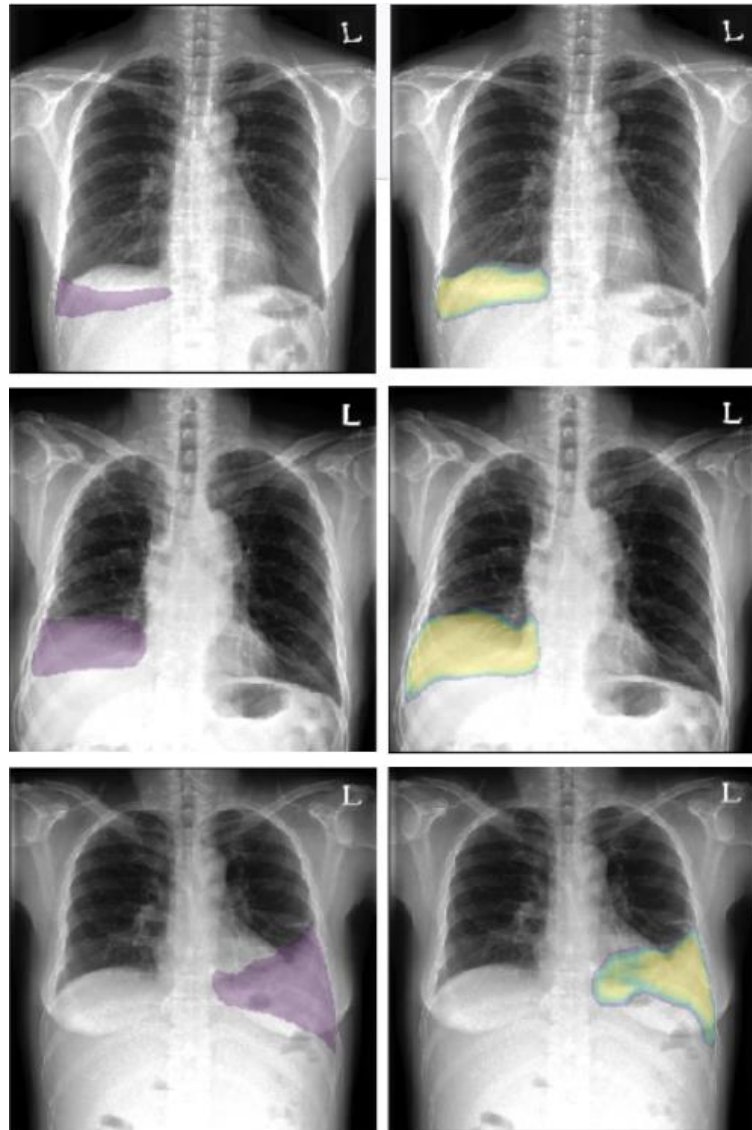


Fig. 9. Pleural effusion segmentation results
(Left) Groundtruth (Right) Segmentation output

5. Conclusion

In this paper, we proposed MO-SegNet, which is a novel multiple-objective segmentation network for lung nodule detection. We demonstrated that MTL significantly improves nodule detection performance. Our proposed method uses semantic segmentation and MTL networks for simultaneous segmentation of chest region and lung nodule lesion position. Moreover, we used multiple-objective loss balancing for unbiased multitask loss optimization. The experimental results demonstrated that our proposed method is not only a reliable technique for lung region segmentation and nodule detection but also improves the final performance of the nodule detection task. The performance of the proposed method was evaluated on the

public dataset from JSRT, using 5-fold cross-validation. The experimental results demonstrated that the performance of the proposed approach can compete with a conventional U-Net segmentation network in a single-objective manner.

One of the limitations of this research was the weight balance tuning. Additional research needs to be conducted for weight balance tuning, which is currently too expensive. There is a more convenient approach, which is to learn the optimal weights. One of the most reasonable solutions is adaptive parameter learning that can be undertaken to weight two multiple-objective loss functions.

References

- [1] S. Schalekamp, B. van Ginneken, E. Koedam, "Computer-aided detection improves detection of pulmonary nodules in chest radiographs beyond the support by bone-suppressed images," *Radiology*, vol. 272, no. 1, pp. 252-261, 2014. [Article \(CrossRefLink\)](#)
- [2] B. de Hoop, D. W. De Boo, H. A. Gietema, "Computer-aided detection of lung cancer on chest radiographs: effect on observer performance," *Radiology*, vol. 257 no. 2, pp. 532-540, 2010. [Article \(CrossRefLink\)](#)
- [3] J. Nam, S. Park, E. Hwang, J. Lee, K. Jin, K. Lim, T. H. Vu, J. Sohn, S. Hwang, J. Goo, "Development and Validation of Deep Learning-based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs," *Radiology*, vol. 290, no. 1, pp. 218-228, 2018. [Article \(CrossRefLink\)](#)
- [4] A. M. R. Schilham, B. van Ginneken, M. Loog, "A computer-aided diagnosis system for detection of lung nodules in chest radiographs with an evaluation on a public database," *Medical Image Analysis*, vol. 10, no. 2, pp. 247-258, 2006. [Article \(CrossRefLink\)](#)
- [5] R. C. Hardie, S. K. Rogers, T. Wilson, A. Rogers, "Performance analysis of a new computer aided detection system for identifying lung nodules on chest radiographs," *Medical Image Analysis*, vol. 12, no. 3, pp. 240-258, 2008. [Article \(CrossRefLink\)](#)
- [6] B. Gupta, M. Tiwari, S. Lamba, "Visibility improvement and mass segmentation of mammogram images using quantile separated histogram equalisation with local contrast enhancement," *CAAI Transactions on Intelligence Technology*, vol. 4, pp. 73-79, 2019. [Article \(CrossRefLink\)](#)
- [7] A. Knokher, R. Talwar, "Content-based Image Retrieval: Feature Extraction Techniques and Applications," *International Journal of Science, Engineering and Technology Research*, vol. 3, no. 5, 2014.
- [8] J. Song, Y. Guo, L. Gao, X. Li, A. Hanjalic, "From Deterministic to Generative: Multimodal Stochastic RNNs for Video Captioning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 10, pp. 3047-3058, 2019. [Article \(CrossRefLink\)](#)
- [9] Y. Zhu, Z. Chen, S. Zhao, H. Xie, "ACE-Net: Biomedical Image Segmentation with Augmented Contracting and Expansive Paths," in *Proc. of the 22th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 712-720, 2019. [Article \(CrossRefLink\)](#)
- [10] H. Xie, D. Yang, N. Sun, Z. Chen, Y. Zhang, "Automated pulmonary nodule detection in CT images using deep convolutional neural networks," *Pattern Recognition*, vol. 85, pp. 109-119, 2019. [Article \(CrossRefLink\)](#)
- [11] J. Song, X. Li, L. Gao, H. T. Shen, "Hierarchical LSTMs with Adaptive Attention for Visual Captioning," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, pp. 1-1, 2019. [Article \(CrossRefLink\)](#)
- [12] B.B. Oğul, P. Koşucu, A. Özçam, S.D. Kanik, "Lung Nodule Detection in X-Ray Images: A New Feature Set," in *Proc. of 6th European Conference of the International Federation for Medical and Biological Engineering*, vol. 45, pp. 150-155, 2014. [Article \(CrossRefLink\)](#)
- [13] S. Lia, Y. Gao, A. Oto, D. Shen, "Representation learning: a unified deep learning framework for automatic prostate MR segmentation," in *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 254-261, 2013. [Article \(CrossRefLink\)](#)

- [14] S. Zhai, Z. Cheng, Y. Wei, Z. Liang, Y. Chen, "Compressive sensing ghost imaging object detection using generative adversarial networks," *Optical Engineering*, vol. 58, no. 1. p. 15, 2019. [Article \(CrossRefLink\)](#)
- [15] B. Li, C. Tang, T. Zheng, Z. Lei, "Fully automated extraction of the fringe skeletons in dynamic electronic speckle pattern interferometry using a U-Net convolutional neural network," *Optical Engineering*, vol. 58, no. 1, pp. 105, 2019. [Article \(CrossRefLink\)](#)
- [16] W. Liu, J. Hu, Z. Li, Z. Zhang, Z. Ma, D. Zhang, "Tongue image segmentation via thresholding and gray projection," *KSII Transactions on Internet and Information Systems*, vol. 13, no. 2, pp. 945-961, 2019. [Article \(CrossRefLink\)](#)
- [17] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234-241, 2015. [Article \(CrossRefLink\)](#)
- [18] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *IJCV*, vol. 115, no. 3, pp. 211-252, 2015. [Article \(CrossRefLink\)](#)
- [19] A. Devaraj and B. van Ginneken and A. Nair, D. Baldwin, "Use of Volumetry for Lung Nodule Management: Theory and Practice," *Radiology*, vol. 284, no. 3, pp. 630-644, 2017. [Article \(CrossRefLink\)](#)
- [20] R. Caruana, "Multitask Learning," *Springer*, vol. 28, pp. 95-133, 1998. [Article \(CrossRefLink\)](#)
- [21] A. Kendall, Y. Gal, R. Cipolla, "Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7482-7491, 2018. [Article \(CrossRefLink\)](#)
- [22] P. Agrawal, J. Carreira, J. Malik, "Learning to See by Moving," in *Proc. of IEEE International Conference on Computer Vision (ICCV)*, 2015. [Article \(CrossRefLink\)](#)
- [23] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. [Article \(CrossRefLink\)](#)
- [24] L.C. Chen, G. Papandreou, F. Schroff, H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv:1706.05587*, 2017. [Article \(CrossRefLink\)](#)
- [25] N. Wang, Y. Peng, Y. Wang, M. Wang, "Skin lesion image segmentation based on adversarial networks," *KSII Transactions on Internet and Information Systems*, vol. 12, no. 6, pp. 2826-2840, 2018.
- [26] I. Misra, A. Shrivastava, A. Gupta, M. Hebert, "Cross-Stitch Networks for Multi-task Learning," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. [Article \(CrossRefLink\)](#)
- [27] S. Ruder, "An Overview of Multi-Task Learning in Deep Neural Networks," *arXiv preprint arXiv:1706.05098*, 2017. [Article \(CrossRefLink\)](#)
- [28] Y. Xu, Y. Li, Y. Wang, M. Liu, Y. Fan, M. Lai, I. Eric, C. Chang, "Gland instance segmentation using deep multichannel neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 12, pp. 2901-2912, 2017. [Article \(CrossRefLink\)](#)
- [29] D. P. Kingma, J. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. of International Conference on Learning Representations (ICLR)*, 2015. [Article \(CrossRefLink\)](#)
- [30] Y. W, K. He, "Group normalization," in *Proc. of European Conference on Computer Vision (ECCV)*, 2018. [Article \(CrossRefLink\)](#)
- [31] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists-detection of pulmonary nodules," *American Journal of Radiology*, vol. 174, pp. 71-74, 2000. [Article \(CrossRefLink\)](#)
- [32] B. van Ginneken, M. B. Stegmann, M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database," *Medical Image Analysis*, 2006.

- [33] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adan, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. of the European conference on computer vision (ECCV)*, pp. 801-818, 2018. [Article \(CrossRefLink\)](#)
- [34] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, "Pyramid scene parsing network," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2881-2890, 2017. [Article \(CrossRefLink\)](#)
- [35] X. Wang, Y. Peng, L. Lu, "Chestx-ray8 : hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2017. [Article \(CrossRefLink\)](#)
- [36] S. Ren, K. He, R. Girshick, J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017. [Article \(CrossRefLink\)](#)
- [37] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, A. C. Berg, "SSD: Single shot multibox detector," in *Proc. of European Conference on Computer Vision*, pp. 21-37, 2016. [Article \(CrossRefLink\)](#)
- [38] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 779-788, 2016. [Article \(CrossRefLink\)](#)



Sejin Park received the B.S. and M.S degree in the Department of Computer Science and Engineering from Ajou University, Korea, in 2002 and 2007, respectively. He is currently working towards PhD. Degree at the Department of Computer Science and Engineering from Hanyang University, Korea, From 2012. His research interests include computer vision and machine learning.
Email : gnoses@hanyang.ac.kr



Woojin Jeong received the B.S degree and PhD. degree in the Department of Computer Science and Engineering from Hanyang University, Korea, in 2012 and 2019. He is currently working as a data analyst at the SK Corporation C&C, from 2019. His research interests include computer vision and machine learning.
Email : wjjeong86@sk.com



Young Shik Moon received the B.S. and M.S. degrees in Electronics Engineering from Seoul National University and Korea Advanced Institute of Science and Technology, Korea, in 1980 and 1982, respectively, and PhD. degree in Electrical and Computer Engineering from the University of California at Irvine, CA, in 1990. From 1982 to 1985, he had been a researcher at the Electronics and Telecommunication Research Institute, Daejeon, Korea. In 1992, he joined the Department of Computer Science and Engineering at Hanyang University, Korea, as an Assistant Professor, and is currently a Professor. Dr. Moon served as General Chair of 2014 IEEE International Symposium on Consumer Electronics, and worked as the President of the Institute of Electronics and Information Engineer, Korea.
Email : ysmoon@hanyang.ac.kr