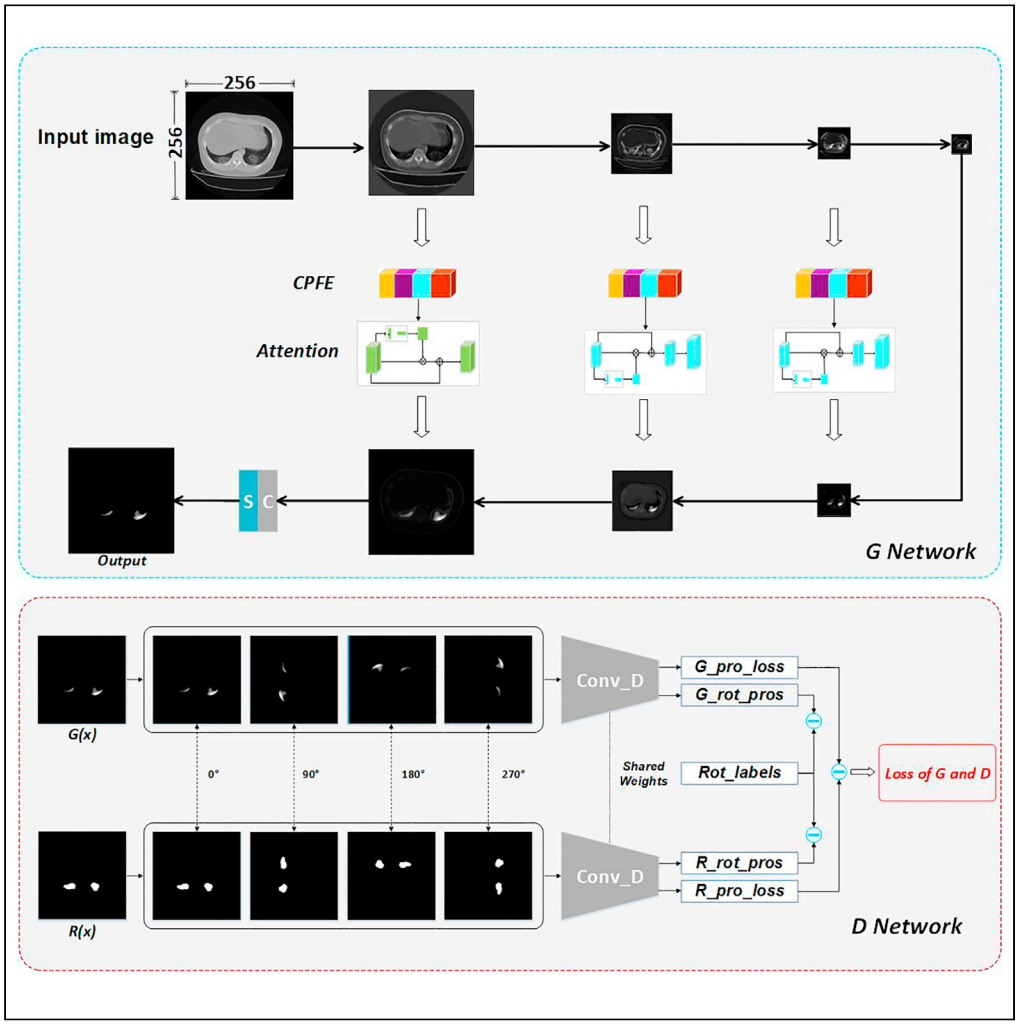
iScience

CellPress
OPEN ACCESS

Article

A general approach for automatic segmentation of pneumonia, pulmonary nodule, and tuberculosis in CT images



Lu Wang, He Zhou, Nan Xu, ..., Hainan Xu, Kexue Deng, Jiangdian Song

xuhn@sj-hospital.org (H.X.)
dengkexue-anhui@163.com (K.D.)
song.jd0910@gmail.com (J.S.)

Highlights

GSAL reduces the workload of manual annotation of lung lesions on CT images

The semantic dependencies in both the spatial and channel dimensions are decoded

The self-supervised rotation loss mitigates discriminator forgetting in GAN

Wang et al., iScience 26, 107005
July 21, 2023
https://doi.org/10.1016/j.isci.2023.107005

iScience

CellPress
OPEN ACCESS

Article

A general approach for automatic segmentation of pneumonia, pulmonary nodule, and tuberculosis in CT images

Lu Wang,[1,2,7] He Zhou,[2,7] Nan Xu,[2] Yuchan Liu,[3] Xiran Jiang,[4] Shu Li,[2] Chaolu Feng,[5] Hainan Xu,[6,*] Kexue Deng,[3,*] and Jiangdian Song[2,8,*]

SUMMARY

Proposing a general segmentation approach for lung lesions, including pulmonary nodules, pneumonia, and tuberculosis, in CT images will improve efficiency in radiology. However, the performance of generative adversarial networks is hampered by the limited availability of annotated samples and the catastrophic forgetting of the discriminator, whereas the universality of traditional morphology-based methods is insufficient for segmenting diverse lung lesions. A cascaded dual-attention network with a context-aware pyramid feature extraction module was designed to address these challenges. A self-supervised rotation loss was designed to mitigate discriminator forgetting. The proposed model achieved Dice coefficients of 70.92, 73.55, and 68.52% on multi-center pneumonia, lung nodule, and tuberculosis test datasets, respectively. No significant decrease in accuracy was observed ($p > 0.10$) when a small training sample size was used. The cyclic training of the discriminator was reduced with self-supervised rotation loss ($p < 0.01$). The proposed approach is promising for segmenting multiple lung lesion types in CT images.

INTRODUCTION

The automatic segmentation of lung lesions in CT images, including COVID-19 pneumonia and other lesions with similar CT findings, such as tuberculosis and pulmonary nodules, has always been challenging because of the various morphologies, intensities, and locations of lung lesions.[1] Traditional morphology-based segmentation methods, such as region-growing, level set, and graph cut algorithms, have been proven to automatically segment certain types of lung lesions.[2–4] However, leakage of the regions, as well as over- and under-segmentation limit the automatic segmentation of different types of lung lesions with diverse morphological variations. Emerging convolutional neural network (CNN)-based algorithms, particularly generative adversarial networks (GANs), have been demonstrated to be effective architectures for the automatic detection and segmentation of regions of interest (ROI) in medical images.[5–8] Continuous adversarial and collaborative training of the generator and discriminator enabled the GAN to detect lesion characteristics and produce synthetic images by directly extracting the ROIs of the lesions from the source CT images. Previous studies have demonstrated that GAN is an effective network for lung lesion segmentation. Zhang et al. proposed a conditional GAN model to segment the lung region and pneumonic lesions simultaneously using a pyramid-based GAN architecture.[9] ROI masks at different scales were used to process radiological images via a pyramid-based GAN to mimic images with better resolution, and an average dice coefficient (DC) of 54.1% with a standard deviation of 21.6% was obtained. In another study, a weakly supervised GAN architecture was proposed, consisting of a segmentation process producing a segmentation mask, a generator replacing the predicted lesion region with a generated region that resembles an uninfected area, and a discriminator to distinguish images of healthy synthetic regions from real uninfected regions.[10] By successfully deceiving the discriminator using synthetic images, the proposed architecture achieved an average DC of 70.3% for pneumonic lesion segmentation. Second, for lung nodules, a three-dimensional (3D) GAN-based data augmentation approach was proposed for bounding box-based 3D lesion detection.[11] In another study, two discriminators were simultaneously applied to classify real and synthetic nodule image pairs to improve the detection and segmentation accuracy of lung nodules. In addition, GAN-based data augmentation for image-style transfer has been used to synthesize training data for robust lung nodule segmentation.[12,13] In the aforementioned studies, a DC>80% was obtained with using augmented datasets from the Lung Image

[1]Department of Library, Shengjing Hospital of China Medical University, Shenyang, Liaoning 110004, China

[2]School of Health Management, China Medical University, Shenyang, Liaoning 110122, China

[3]Department of Radiology, The First Affiliated Hospital of University of Science and Technology of China (USTC), Division of Life Sciences and Medicine, USTC Hefei, Anhui 230036, China

[4]School of Intelligent Medicine, China Medical University, Shenyang, Liaoning 110122, China

[5]Key Laboratory of Intelligent Computing in Medical Image (MIIC), Ministry of Education, Shenyang, Liaoning 110169, China

[6]Department of Obstetrics and Gynecology, Pelvic Floor Disease Diagnosis and Treatment Center, Shengjing Hospital of China Medical University, Shenyang, Liaoning 110004, China

[7]These authors contributed equally

[8]Lead contact

*Correspondence: xuhn@sj-hospital.org (H.X.), dengkexue-anhui@163.com (K.D.), song.jd0910@gmail.com (J.S.)

https://doi.org/10.1016/j.isci.2023.107005

Table 1. Pneumonia lesion segmentation on the training, validation, and test datasets

DC	Training	Validation	Test	p value
M_100	77.63%	75.01%	74.93%	Ref.
M_70	76.96%	73.77%	72.19%	0.509
M_40	72.01%	71.33%	70.92%	0.220
M_10	63.20%	59.39%	58.55%	<0.01

M_100, M_70, M_40, M_10 represent the model trained on 100%, 70%, 40% and 10% training images.

Database Consortium image collection (LIDC-IDRI). By integrating multi-omics data, an end-to-end multi-conditional GAN was proposed for the multi-scale fusion of image and gene sequencing data to ensure the authenticity and quality of the synthesized lung nodule images.[14] Furthermore, although admission symptoms of tuberculosis are similar to those of pneumonia, it has been reported to be responsible for more than a million deaths per year based on the World Health Organization Global Tuberculosis Report. The morphological and intensity changes of tuberculosis in CT images are as diverse as those of lung nodules and pneumonia.[15] One of the previous studies applied nine different deep CNNs to tuberculosis detection on X-ray images.[16] In addition, an intensity affinity metric with clustering was proposed to select optimal thresholding for segmenting spatially diffuse and multi-focal radiotracer uptake of tuberculosis on PET images, with both sensitivity and specificity higher than 90%.[17] Based on U-net architecture, chest X-ray (CXR) modality-specific U-Nets was proposed for semantic segmentation of tuberculosis ROIs, and a DC of 75.5% was obtained using the proposed VGG16-CXR-U-Net architecture.[18] Recently, a GAN-based disentangle learning framework was proposed for effective tuberculosis area detection,[19] which obtained an intersection-over-union of 60.3% on CXR images.

Despite the performance of GAN in lung lesion segmentation tasks, the development of a general GAN-derived segmentation approach for the automatic segmentation of pneumonia, pulmonary nodules, and tuberculosis in CT images has three drawbacks.

First, various manually annotated training samples of lung lesions are required to achieve precise segmentation. Because there is no universal solution for the segmentation of diverse types of lung lesions, such as pneumonia, lung nodules, and tuberculosis, a large dataset of manually labeled images of specific lesion types is required in each segmentation study to guarantee data diversity. However, manual segmentation of ROIs in medical images is hampered by limitations such as time and labor resources, as well as inter-observer bias. For example, it takes approximately 30 s on average for an experienced radiologist to delineate the boundary of a solitary lung nodule with a diameter of 10 mm on one CT slice; this delineation procedure takes even longer for adhesion-type lung nodules or pneumonia cases with multiple diffuse lesions. In addition, most mature neural networks used for image analysis are trained on hundreds of thousands of images, such as ImageNet.[20] However, unlike natural images, manually annotated medical images are often difficult to obtain because of limited patient sources and privacy regulations in medical institutions. Although lung nodule images with manual delineations by radiologists from 1,018 patients were published in the LIDC-IDRI dataset[21] and CT images of COVID-19 cases with radiologists' annotations are available,[22,23] they are insufficient for current CNN networks with millions of parameters, particularly for the latest GANs.

Second, the information emphasized by the feature map layers obtained through step-by-step neural network convolution varies significantly. Specifically, the features of global context-aware semantics, which are relevant to precisely locating salient regions in images, are distributed in deeper layers. Conversely, shallow layer features are rich in spatial and local texture information, which is more appropriate for representing the details of the target regions.[24] Studies have demonstrated that the accuracy of saliency detection and segmentation benefits from the adaptive integration of local features in feature maps with corresponding global dependencies.[25,26] However, the process of encoding and decoding feature maps is always consecutively implemented without distinction in current GAN-based segmentation algorithms.[27,28] Therefore, a module that reorganizes and recognizes features related to the texture, shape, and location invariance of diverse lung lesions would be helpful for automatic segmentation.[29]

Finally, the alternating stochastic gradient descent training scheme in the GAN causes the discriminator network to depend on the generator network output.[30,31] Specifically, the discriminator is a classifier for

Table 2. DC of COVID-19 and non-COVID-19 pneumonia lesions on the test dataset

DC	COVID-19	non-COVID-19	p value
M_100	77.36%	72.68%	0.410
M_70	73.55%	70.03%	0.353
M_40	71.30%	67.11%	0.341
M_10	60.80%	57.17%	0.208

which the distribution of one class (i.e., fake samples) shifts as the generator changes during GAN training. In this case, the discriminator disregards the information learned from the previous mimic images produced by the generator, particularly in nonstationary training environments.[32] Training may become unstable or cyclic if the discriminator cannot recall previous classification boundaries or segmentations. This affects the learning of the target characteristics and eventually reduces the quality of the mimicked image.

Related works

The emerging development of attention networks has shown the potential to simultaneously decode both spatial and textural information in images, whereas the latest self-supervised loss in GAN has been proven to reduce the instability of the discriminator learning of GANs. In the field of computer vision, the attention mechanism focuses on discriminating image features and has proven to be an effective solution for object-tracking tasks.[33–35] Recently, attention networks have been applied to image analysis, such as image classification,[36] object detection,[37] and semantic segmentation.[25] Among these, channel and spatial attention modules are the two widely used modules in attention networks. The channel attention module emphasizes "what" informative part of an image should be the focus and the spatial attention module is responsible for "where" the informative part is located.

First, the channel attention module is necessary because the feature map contributions from each channel do not equally represent the object, with some channels representing the visual pattern of the target better than others, and vice-versa. Each channel map of high-level features can be regarded as a particular visual pattern, and different semantic patterns are associated with each other.[38] Therefore, the aim of the channel attention network in semantic segmentation is to optimize the channel-wise weights around the object to be segmented.[39] By exploiting the interdependencies between channel maps, interdependent feature maps can be emphasized to improve the feature representation of semantics related to the target regions.

The spatial attention module encodes a wider range of contextual information into local features, thereby enhancing their representation capabilities. It highlights the location of informative features of the target in an image so that the target location can be determined. Thus, it complements the channel attention.[40] Pooling is a type of spatial attention that selects the most probable attention region or the attention-weighted average of spatial features.[41,42] In addition to the basic forms of attention, variations in spatial attention, such as the stacked spatial attention network,[43] in which the later attention layer is based on an attention-based feature map modulated by the earlier one, have been proposed to help locate desired target regions. Recently, a dual-attention network that simultaneously applies spatial and channel attention networks to perform the same task has been proposed.[38,44]

However, because the generator network in GAN is constantly updated during training, the input of the discriminator network is changed accordingly. In this dynamic training situation, discriminator forgetting arises because the discriminator network is required to continuously extract new features to distinguish the updated fake sample input. Discriminator forgetting contributes to GAN instability.[32] To address this problem, a conditional GAN, which allows both the generator and discriminator to access the labeled data, and an augmented discriminator with supervised learning are feasible approaches for avoiding catastrophic forgetting.[45,46] Recently, an approach that integrates an auxiliary self-supervised loss in the discriminator to stimulate the generator was proposed, and discriminator collaboration in representation learning while competing for image generation was shown to be effective for natural image synthesis.[31]

Self-supervised loss is implemented by training the network on a pretext task, such as predicting a rotated angle or the relative location of an image patch, and then extracting the representations from

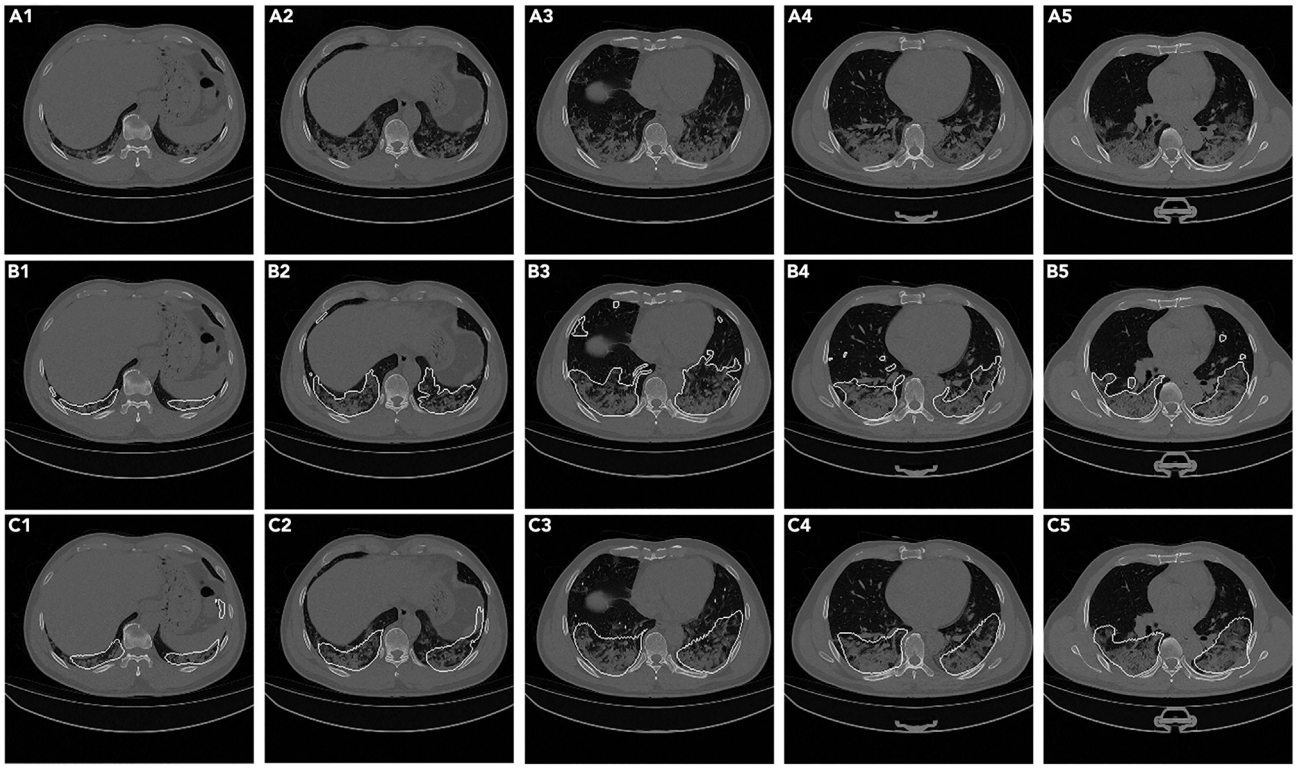


Figure 1. Segmentation of bilateral COVID-19 lesion

(A) Original images.

(B) Proposed model; and (C) Radiologists.

the resulting network. Studies on natural images have demonstrated that with the help of self-supervised loss, GAN retains generalizable representations in a nonstationary environment, which prevents the forgetting of classes in discriminator representations during training iterations.[47,48] However, studies using self-supervised rotation loss to enhance discriminator training in the field of medical image segmentation are rare.

In summary, developing a new architecture for automatic lung lesion segmentation in CT images to address the above-mentioned GAN challenges in this task is promising with the emerging attention network and auxiliary self-supervised loss. Therefore, we first proposed a self-supervised adversarial learning approach herein using an emerging dual-feature pyramid attention network module and auxiliary rotation loss for end-to-end lung lesion segmentation. Compared with previous studies, the main advantages of this study are that it has the potential to overcome the three main limitations of GAN in automatic lung lesion segmentation.

1) Reducing the workload of radiologists of manually annotating lung lesions on CT images for network training.
2) Decoding the semantic dependencies of lung lesion characteristics in both the spatial and channel dimensions in the shallow and deep convolution layers to enhance training efficiency.
3) Mitigating discriminator forgetting in GAN and preserving generalizable representations of lung lesion characteristics during nonstationary training.

The remainder of the paper is structured as follows. "results" section outlines the experiments performed to develop the model and comparison of the model is made with state-of-the-art methods. Followed by the "discussion", "conclusions", and "limitations of the study" of the paper. Finally, "STAR Methods" section describes the network architectures in detail.

Table 3. Lung nodule segmentation on the training, validation, and test datasets

DC	Training	Validation	Test	p value
M_100	82.39%	77.66%	76.30%	Ref.
M_70	80.55%	76.62%	76.23%	0.685
M_40	77.91%	75.63%	74.10%	0.566
M_10	76.10%	73.90%	73.55%	0.209

RESULTS

Table 1 presents the results of the pneumonia lesion segmentation. DCs of 77.63, 75.01, and 74.93% were obtained using all images (*M_100*) on the training, validation, and test datasets, respectively. Similarly, *M_40* (training dataset reduced to 40%) obtained DCs of 72.01, 71.33, and 70.92%, and *M_10* (training dataset reduced to 10%) obtained DCs of 63.20, 59.39, and 58.55%, respectively. A significant difference was found only when the training data were reduced to 10%, including 1,166 training images ($p < 0.01$). This finding indicates that when the training dataset of pneumonia images is reduced to several thousand images, the segmentation results of the *GSAL* model are similar to those obtained using tens of thousands of images, which significantly reduces the effort required by radiologists in manual delineation in future public health emergencies, such as pneumonia.

Table 2 lists the segmentation results of the COVID-19 and non-COVID-19 pneumonia lesions. The results showed that COVID-19 pneumonia segmentation was better than non-COVID-19 pneumonia lesion segmentation; however, no significant difference was found ($p > 0.2$) (see Table 2). The results demonstrate that the proposed *GSAL* model enables the segmentation of different types of pneumonia lesions, including COVID-19 pneumonia and other community-acquired pneumonia. This indicates the potential applicability of the *GSAL* model as a generic automatic segmentation algorithm for diverse lung lesions. Figure 1 shows the segmentation example of a COVID-19 lesion with the CT manifestation of bilateral, mixed ground-glass opacity (GGO) and consolidation, and Figures S1–S3 show those of other pneumonia subtypes: https://github.com/JD910/general_net_for_lesion_seg#supp_materials.

Table 3 lists the results of the lung nodule segmentation. DCs of 82.39, 77.66, and 76.30% were obtained using *M_100* for the training, validation, and test datasets, respectively. DCs of 76.10, 73.90, and 73.55% were obtained for *M_10*. No significant difference was found between the results for *M_100* and *M_10* ($p = 0.209$). The results show that for lung nodule segmentation, the proposed *GSAL* obtained an average DC of 75%, which is similar to the segmentation results reported in previous studies.

Table 4 lists the segmentation accuracy for each subtype of lung nodules on the test dataset. The solid, juxta-vascular, juxta-pleural, and GGO subtypes were manually identified by clinical experts. The results revealed that the segmentation accuracy for solid nodules and GGOs was better than that for adhesive-type lung nodules; however, the difference was not statistically significant ($p > 0.05$). Figure 2 illustrates the segmentation of a solid nodule, and Figures S4 and S5 present other nodule subtypes: https://github.com/JD910/general_net_for_lesion_seg#supp_materials.

For tuberculosis segmentation, the proposed model obtained DCs of 78.33, 70.90, and 73.30% on the training, validation, and test datasets, respectively, using *M_100* (see Table 5). No significant reduction in segmentation accuracy was found when using *M_10*, which included 330 images ($p = 0.089$), and DCs of 72.02, 65.37, and 68.52% were obtained for the three datasets. Figure 3 illustrates two examples of nodular tuberculosis segmentation. The results indicate that *GSAL* is a potentially effective approach to tuberculosis segmentation, although only hundreds of images were used for training. We speculate that the potential reason may be that tuberculosis lesions in this study tend to be more similar to the features of pulmonary nodules on CT images (see Figure 3).

A comparison of *GSAL* with the recently proposed general segmentation architecture nnU-net indicated that for lung nodule segmentation, nnU-net obtained significantly better DC than the proposed approach ($p < 0.05$). However, for pneumonia, with relatively fewer training samples, the proposed model obtained a significantly higher DC for the test dataset ($p < 0.05$). No significant difference was found between the two methods for tuberculosis segmentation ($p > 0.05$). The average time consumption from inputting an image

Table 4. DC of lung nodule subtypes of solid, juxta-vascular, juxta-pleural, and GGO on the test dataset.

DC	Juxta-pleural	Juxta-vascular	Solid	GGO
M_100	75.33%	75.20%	76.50%	77.33%
M_70	75.51%	74.10%	76.30%	77.03%
M_40	73.30%	72.58%	74.36%	74.22%
M_10	71.50%	71.79%	72.90%	73.94%

No statistically significant difference was found between the DCs ($p > 0.05$).

to outputting the generated image was 9.1 s for nnU-net and 0.2 s for *GSAL* in the test procedure. With *Model_100*, the time to train one epoch of the lung nodule dataset of *GSAL* was 30.2 min. Table 6 presents the detailed results of the segmentation accuracy.

Furthermore, a comparison of *GSAL* with the universal U-net was conducted (see Table 7). The results indicate that the proposed approach is superior to U-net ($p < 0.05$) in pneumonia segmentation on CT images, which is consistent with the results on nnU-net. For pulmonary nodules and tuberculosis, the proposed approach obtained a higher segmentation accuracy than U-net on the test dataset; however, no statistically significant difference was observed ($p > 0.05$). The results of the above comparison demonstrate that for pneumonia segmentation, the proposed approach is superior to U-net-derived approaches. For the other two types of lung lesions, *GSAL* is comparable to U-net-derived approaches. Although U-net has been proved to be an architecture for automatic segmentation, our experiments have demonstrated that the proposed *GSAL* is more appropriate for inflammatory lung lesion segmentation in CT images.

To demonstrate the improvement in the training efficiency obtained by the proposed model, an architecture without self-supervised rotation loss in the discriminator network was trained for comparison. Figure 4 presents the loss curves of the discriminator models in the two architectures at different training epochs (based on lung nodule images). The results show that the model with self-supervised loss converged faster ($p < 0.05$ on DC for epochs <60, *t-test*). In the ablation study without cascaded cascaded context-aware pyramid feature extraction (CPFE), the DCs were 61.55, 54.89, and 47.23% for pneumonia, pulmonary nodules, and tuberculosis, respectively, on the test dataset based on *M_10*, which were significantly lower than those of the proposed *GSAL* model ($p = 0.027$). In addition, when the attention modules were removed, the DCs of *M_10* on the test dataset decreased by 8.55, 6.98, and 11.29% for pneumonia, pulmonary nodules, and tuberculosis on the test dataset, respectively, which were significantly lower than those of the proposed *GSAL* model ($p = 0.003$). These findings confirm our hypothesis regarding the integration of an emerging attention network and self-supervised loss into a general lung lesion segmentation approach.

To verify the lung lesion detection precision of *GSAL*, the *TP*, *FP*, *FN*, *P*, *R*, and *F1*-score metrics of lung nodule and pneumonia segmentation on the test dataset were analyzed by two radiologists with more than five years of experience. The average accuracy obtained by the radiologists was reported to avoid potential inter-observer bias. The results in Table 8 and Table S1 show that when using the training dataset of thousands of images, the proposed *GSAL* achieved an *F1*-score of approximately 70% for both lung nodule and pneumonia lesion detection. However, when the training dataset of pneumonia was reduced to *M_10*, an *F1*-score of approximately 63% was obtained, which was significantly lower than that of *M_40* ($p < 0.05$).

DISCUSSION

In this study, we propose a general self-supervised adversarial learning architecture for end-to-end segmentation of pneumonia, lung nodules, and tuberculosis lesions and validate it on multicenter datasets. We demonstrated that the proposed model did not require massive numbers of manually labeled training samples, and no significant decrease in segmentation accuracy was observed when a large amount of training data was discarded, which would reduce the workload required for manual annotation in this field of study. Using the proposed cascaded CPFE and dual attention modules, diverse types and scales of lung lesions could be identified. High-level semantic interdependence and low-level textural features complemented lung lesion segmentation. Finally, the self-supervised rotation loss countered discriminator forgetting in the GAN, significantly improving the lung lesion segmentation efficiency ($p < 0.05$).

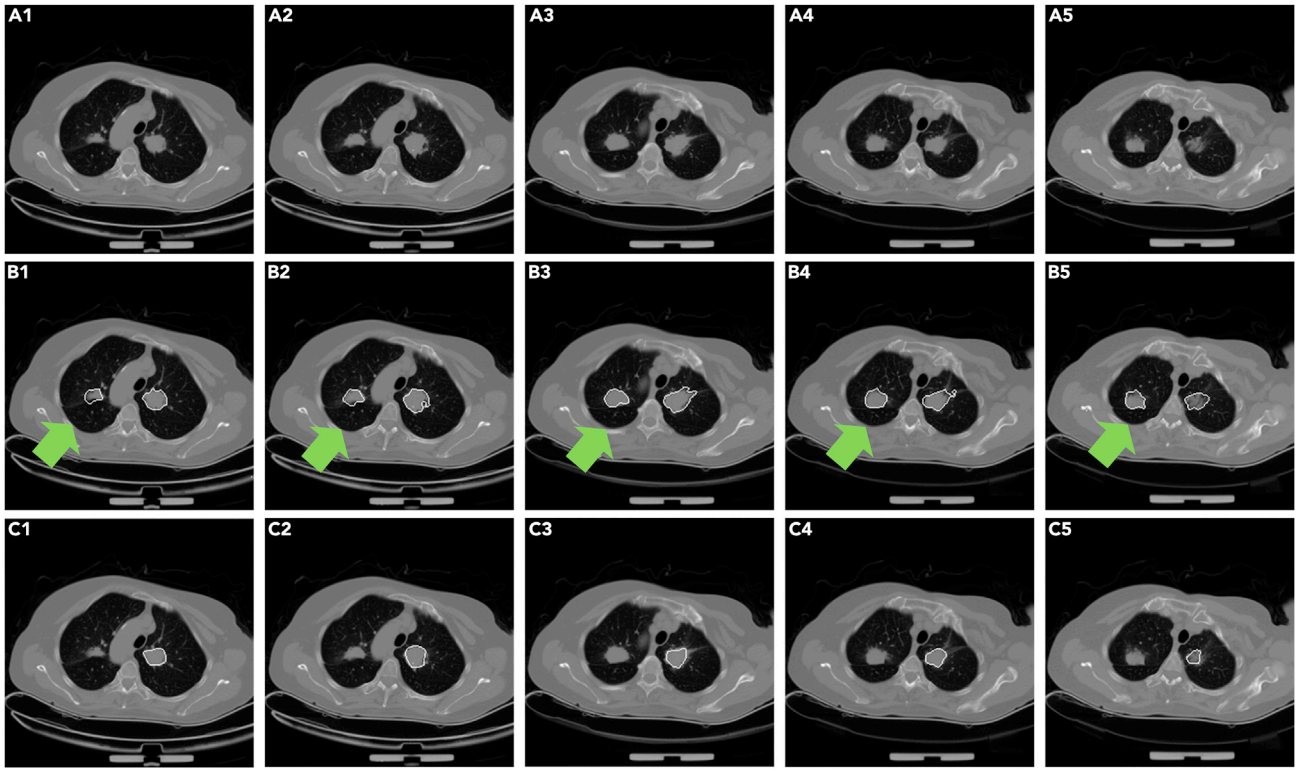


Figure 2. Solid nodule segmentation

(A) Original images; segmentation obtained by (B) Proposed model; and (C) Radiologists. One nodule on the right lung was not delineated by the radiologists (marked by green arrow), whereas the proposed *GSAL* model detected both nodules in the images.

Studies on pneumonia and pulmonary nodule segmentation in CT images have contributed substantially to the automatic segmentation of lung lesions.[28,49–55] However, previous studies have been conducted on specific lesion types, and general neural networks for segmenting these distinct lung diseases are rare. The lack of large, manually labeled training samples has hindered the development of accurate lung lesion segmentation methods. Furthermore, neural networks designed for the segmentation of specific lung lesion types, such as solid nodules, are unlikely to be directly suitable for the segmentation of other lung diseases, such as diffuse pneumonia, owing to their variations in structure, morphology, and intensity in CT images. Although transfer learning can achieve this goal to some extent,[56] a general segmentation network for different types of lung lesions will significantly reduce network design and manual labeling. This study demonstrated that the morphology and CT intensity of solid, GGO, adhesive-type lung nodules, and unilateral/bilateral GGO and multifocal consolidation of pneumonia, and tuberculosis characteristics could be identified using the proposed self-supervised adversarial learning architecture. Although opaque lesions of lung nodules (i.e., GGO) and pneumonia have similar intensity values on CT images, their varied morphologies and dimensions increase the difficulty of segmentation. The competitive detection accuracy (average *F1*-score: 77.1% and 70.3% in Table 8) of the *GSAL* demonstrates the superiority of the proposed architecture in detecting diverse lung opacities. Moreover, the proposed approach obtained the best segmentation accuracy for lung nodules, followed by tuberculosis; however, the DC decreased for pneumonia lesions using *M_10* (73.55, 68.52, and 58.55%,

Table 5. Tuberculosis segmentation on the training, validation, and test datasets

DC	Training	Validation	Test	p value
M_100	78.33%	70.90%	73.30%	Ref.
M_70	80.55%	76.62%	76.23%	0.583
M_40	77.91%	75.63%	74.10%	0.300
M_10	72.02%	65.37%	68.52%	0.139

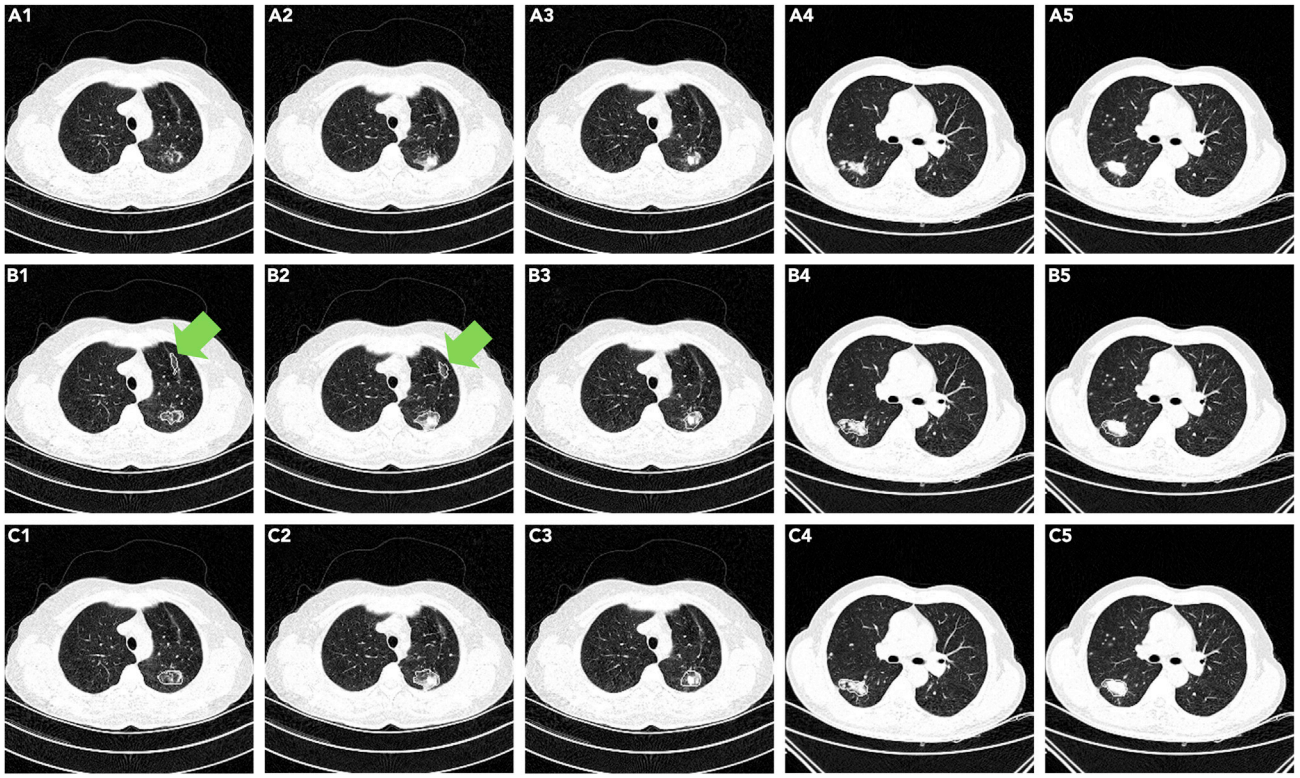


Figure 3. Tuberculosis segmentation

(A) Original images; segmentation obtained by (B) Proposed model; and (C) Radiologists. In 1-3, radiologists did not identify the lesion indicated by the arrow, and (4)-(5) indicates another patient with one nodular tuberculosis on the right lung.

respectively). This is because pneumonia is accompanied by a greater degree of diffusion than tuberculosis and pulmonary nodules in this study, which primarily consist of solid lesions, account for most of the training data. Therefore, the model performs better segmentation for solid lesions such as pulmonary nodules or nodular tuberculosis than pneumonia lesions.

In previous studies, the training sample size used to train neural networks on natural images reached a million samples, such as the large-scale visual recognition challenge subset of ImageNet.[57] However, for studies on lung lesion segmentation, the available annotated training samples consisted of approximately one thousand; these can be retrieved from the LIDC-IDRI dataset or other pneumonia datasets.[58,59] Owing to the limitations of the natural properties of medical data and the cost of manual labeling, it is impossible to accumulate as many training samples as it is for natural images, which means that current studies on medical image segmentation are limited by insufficient training samples. Therefore, the question of how to counter the reliance of neural networks on a large number of training samples has attracted considerable attention in medical image segmentation, particularly for lung lesions. This study proposes an effective way to reduce the training data to 3,000 images while achieving a segmentation DC of 70% on the test dataset

Table 6. Comparison of lung lesion segmentation between the proposed approach and the nnU-net on the test dataset

	Lung nodule		Pneumonia		Tuberculosis	
DC	Proposed	nnU-net	Proposed	nnU-net	Proposed	nnU-net
M_100	76.30%	85.30%	74.93%	70.20%	73.30%	74.69%
M_70	76.23%	82.60%	72.19%	68.39%	76.23%	74.33%
M_40	74.10%	80.90%	70.92%	63.58%	74.10%	71.08%
M_10	73.55%	78.02%	58.55%	50.41%	68.52%	64.97%

Table 7. Comparison of lung lesion segmentation between the proposed approach and traditional U-net on the test dataset

	Lung nodule		Pneumonia		Tuberculosis	
DC	Proposed	U-net	Proposed	U-net	Proposed	U-net
M_100	76.30%	72.66%	74.93%	65.30%	73.30%	71.11%
M_70	76.23%	75.30%	72.19%	55.63%	76.23%	75.96%
M_40	74.10%	71.03%	71.92%	51.39%	74.10%	73.28%
M_10	73.55%	68.33%	58.55%	51.96%	68.52%	70.36%

for pneumonia, lung nodule, and tuberculosis lesions, whereas previous studies required tens of thousands of training data. Specifically, the integrated workflow using a U-Net-like structure and dual attention module to design a self-supervised adversarial learning architecture provides a novel generator network for GAN with significantly less dependence on the size of the training dataset. Although U-Net and attention networks have been previously used for image segmentation,[60] and nnU-net achieved comparable accuracy in the present study, the advantage of our approach is the real-time and end-to-end exploitation of multi-receptive-field semantic interdependencies in both spatial and channel dimensions that adapt to the diversity of lung lesions. Each channel of a high-level feature map can be regarded as a class-specific response with different associated semantic responses.[61,62] Exploiting the interdependencies between channels emphasizes interdependent feature maps; thus, feature representation is more relevant to the specific semantics of lung lesions. Moreover, using the spatial attention module to evaluate pixel-wise correlation on spatial feature maps, specific details, such as the boundary and texture of lung lesions, are emphasized. Hu et, al.[63] combining convolutional filters should lead to an informative architecture by fusing spatial and channel-wise information within the appropriate receptive fields. Here, parallel spatial and channel-wise attention modules indicated that the high-level and low-level feature representations complement each other and could be integrated to generate mimic lung lesion segmentation images.

Another advantage of this study is that we reduced the discriminator forgetting that occurs in GAN training by introducing a self-supervised rotation loss for lung lesion segmentation. Although rotation loss values for salient region detection and segmentation in natural images have been reported,[31,64] related studies on medical images are limited. Motivated by the challenge of discriminator forgetting in GAN training, we added a mechanism to the discriminator that allows uninterrupted representation learning, thus rendering it independent of the generator status. Our results indicate that when coupled with the self-supervised rotation loss, the network learning of lung lesion representations is transferred across the generator and discriminator tasks; thus, the training efficiency is significantly improved compared with that of the traditional GAN ($p < 0.05$).

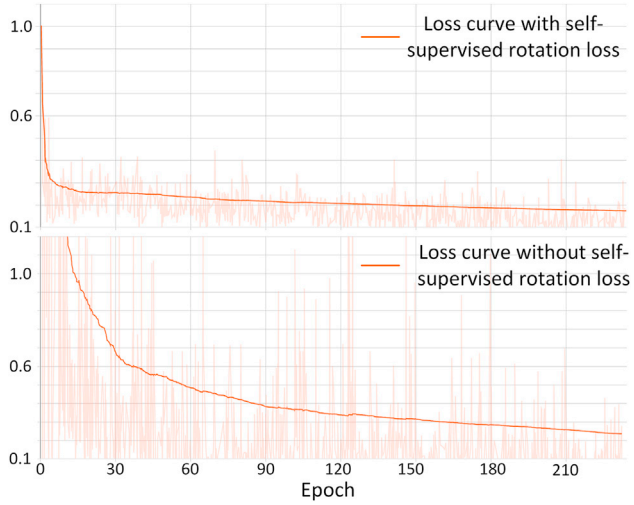


Figure 4. Loss curve of discriminator network training with and without the self-supervised rotation loss

Table 8. Segmentation of lesion detection by the radiologists (R1 and R2) on lung nodule and pneumonia images on the test dataset

	Lung nodule (*M_10*)						Pneumonia (*M_40*)					
	TP	*FP*	*FN*	*P*	*R*	*f1*	*TP*	*FP*	*FN*	*P*	*R*	*f1*
R1	602	512	28	54.9%	95.5%	69.7%	1269	730	139	63.5%	90.7%	74.7%
R2	618	498	12	55.5%	98.1%	70.9%	1320	649	79	67.0%	96.5%	79.1%
Avg	610	505	20	55.2%	96.8%	70.3%	1309.5	689.5	109	65.5%	93.6%	77.1%

Conclusions

We proved that the proposed *GSAL* architecture is promising as a general approach for automatic lung lesion segmentation of pulmonary nodules, pneumonia, and tuberculosis. Large numbers of annotated samples are unnecessary for *GSAL* training, thereby reducing the workload of radiologists in manual lesion delineation. The integration of cascaded context-aware pyramid feature extraction and a dual attention module enables the decoding of the semantic dependencies of lung lesion characteristics in both spatial and channel dimensions, which enhances training efficiency. Our study paves the way for a GAN-derived architecture for multiple lung lesion segmentation while mitigating discriminator forgetting using self-supervised rotation loss.

Limitations of the study

Our study was limited by the class imbalance in the sample of pulmonary nodules and pneumonia. For example, *M_10* for lung nodules and *M_40* for pneumonic lesions yielded similar results. Hence, further experiments should be performed to determine the most suitable amount of training data for each lung lesion type. Owing to the introduction of the rotated image into the discriminator, the computational cost of our model in space and time is higher than that of the traditional GAN. In the future, we will explore more cost-effective solutions to prevent discriminators from forgetting and optimizing representation learning. In addition, the model was developed on 2D CT images for lung lesion segmentation because self-supervised rotation loss was proposed for 2D natural images.[31] Future studies should focus on developing 3D segmentation models based on these modules. Finally, U-net and nnU-net were used for comparison because they were proposed as baseline segmentation model, whereas other recent segmentation models for specific human organs[65,66] were not used because this study aimed to propose a general segmentation model for multiple lung lesions.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND PARTICIPANTS DETAILS
- METHOD DETAILS
  - Generator network G
  - Discriminator network D
  - Experiment

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.107005.

ACKNOWLEDGMENTS

The National Natural Science Foundation of China (92259104, 82001904) supported this work.

AUTHOR CONTRIBUTIONS

L.W., H.Z., N.X., Y.L., X.J., S.L., C.F., H.X., K.D., and J.S.

Integrity guarantor of the study: J.S.

Study concepts and design: J.S.

Literature research: All authors.

Clinical studies: All authors.

Experimental studies/data analysis: All authors.

Statistical analyses: J.S.

Manuscript preparation: All authors.

Manuscript editing: All authors.

DECLARATION OF INTERESTS

The authors have no conflicts of interest to declare.

Received: October 14, 2022
Revised: April 27, 2023
Accepted: May 26, 2023
Published: May 30, 2023

REFERENCES

1. Song, J., Huang, S.-C., Kelly, B., Liao, G., Shi, J., Wu, N., Li, W., Liu, Z., Cui, L., Lungre, M., et al. (2022). Automatic lung nodule segmentation and intra-nodular heterogeneity image generation. IEEE J. Biomed. Health Inform. *26*, 2570–2581. https://doi.org/10.1109/JBHI.2021.3135647.

2. Song, J., Yang, C., Fan, L., Wang, K., Yang, F., Liu, S., and Tian, J. (2016). Lung lesion extraction using a toboggan based growing automatic segmentation approach. IEEE Trans. Med. Imaging *35*, 337–353. https://doi.org/10.1109/TMI.2015.2474119.

3. Hoogi, A., Beaulieu, C.F., Cunha, G.M., Heba, E., Sirlin, C.B., Napel, S., and Rubin, D.L. (2017). Adaptive local window for level set segmentation of CT and MRI liver lesions. Med. Image Anal. *37*, 46–55. https://doi.org/10.1016/j.media.2017.01.002.

4. Nardelli, P., Jimenez-Carretero, D., Bermejo-Pelaez, D., Washko, G.R., Rahaghi, F.N., Ledesma-Carbayo, M.J., and San Jose Estepar, R. (2018). Pulmonary artery-vein classification in CT images using deep learning. IEEE Trans. Med. Imaging *37*, 2428–2440. https://doi.org/10.1109/TMI.2018.2833385.

5. Ruan, Y., Li, D., Marshall, H., Miao, T., Cossetto, T., Chan, I., Daher, O., Accorsi, F., Goela, A., and Li, S. (2020). MB-FSGAN: joint segmentation and quantification of kidney tumor on CT by the multi-branch feature sharing generative adversarial network. Med. Image Anal. *64*, 101721. https://doi.org/10.1016/j.media.2020.101721.

6. Onishi, Y., Teramoto, A., Tsujimoto, M., Tsukamoto, T., Saito, K., Toyama, H., Imaizumi, K., and Fujita, H. (2020). Multiplanar analysis for pulmonary nodule classification in CT images using deep convolutional neural network and generative adversarial networks. Int. J. Comput. Assist. Radiol. Surg. *15*, 173–178. https://doi.org/10.1007/s11548-019-02092-z.

7. Qin, Y., Zheng, H., Huang, X., Yang, J., and Zhu, Y.M. (2019). Pulmonary nodule segmentation with CT sample synthesis using adversarial networks. Med. Phys. *46*, 1218–1229.

8. Valvano, G., Leo, A., and Tsaftaris, S.A. (2021). Learning to segment from scribbles using multi-scale adversarial attention gates. IEEE Trans. Med. Imaging *40*, 1990–2001.

9. Zhang, P., Zhong, Y., Deng, Y., Tang, X., and Li, X. (2020). CoSinGAN: learning COVID-19 infection segmentation from a single radiological image. Diagnostics *10*, 901. https://doi.org/10.3390/diagnostics10110901.

10. Wang, X., Deng, X., Fu, Q., Zhou, Q., Feng, J., Ma, H., Liu, W., and Zheng, C. (2020). A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT. IEEE Trans. Med. Imaging *39*, 2615–2625. https://doi.org/10.1109/TMI.2020.2995965.

11. Han, C., Kitamura, Y., Kudo, A., Ichinose, A., Rundo, L., Furukawa, Y., Umemoto, K., Li, Y., and Nakayama, H. (2019). Synthesizing diverse lung nodules wherever massively: 3D multi-conditional GAN-based CT image augmentation for object detection. Preprint at arXiv. https://doi.org/10.48550/arXiv.1906.04962.

12. Shi, H., Lu, J., and Zhou, Q. (2020). A Novel Data Augmentation Method Using Style-Based GAN for Robust Pulmonary Nodule Segmentation (IEEE)), pp. 2486–2491.

13. Gao, C., Clark, S., Furst, J., and Raicu, D. (2019). Augmenting LIDC Dataset Using 3D Generative Adversarial Networks to Improve Lung Nodule Detection (International Society for Optics and Photonics)), p. 109501K.

14. Xu, Z., Wang, X., Shin, H.C., Yang, D., Roth, H., Milletari, F., Zhang, L., and Xu, D. (2019). Correlation via Synthesis: End-To-End Nodule Image Generation and Radiogenomic Map Learning Based on Generative Adversarial Network.

15. World Health Organization (2013). Global Tuberculosis Report 2013 (World Health Organization).

16. Rahman, T., Khandakar, A., Kadir, M.A., Islam, K.R., Islam, K.F., Mazhar, R., Hamid, T., Islam, M.T., Kashem, S., Mahbub, Z.B., et al. (2020). Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization. IEEE Access *8*, 191586–191601.

17. Foster, B., Bagci, U., Ziyue, X., Dey, B., Luna, B., Bishai, W., Jain, S., and Mollura, D.J. (2014). Segmentation of PET images for computer-aided functional quantification of tuberculosis in small animal models. IEEE Trans. Biomed. Eng. *61*, 711–724. https://doi.org/10.1109/TBME.2013.2288258.

18. Rajaraman, S., Folio, L.R., Dimperio, J., Alderson, P.O., and Antani, S.K. (2021). Improved semantic segmentation of tuberculosis-consistent findings in chest X-rays using augmented training of modality-specific U-net models with weak localizations. Diagnostics *11*, 616. https://doi.org/10.3390/diagnostics11040616.

19. Han, L., Lyu, Y., Peng, C., and Zhou, S.K. (2022). GAN-Based Disentanglement Learning for Chest X-Ray Rib Suppression (Medical Image Analysis), 102369.

20. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A Large-Scale Hierarchical Image Database (Ieee)), pp. 248–255.

21. Armato, S.G., 3rd, McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., et al. (2011). The lung image database Consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Med. Phys. *38*, 915–931. https://doi.org/10.1118/1.3528204.

22. Ma, J., Wang, Y., An, X., Ge, C., Yu, Z., Chen, J., Zhu, Q., Dong, G., He, J., and He, Z. (2020). Towards efficient covid-19 ct annotation: a benchmark for lung and infection segmentation. Preprint at arXiv. https://doi.org/10.48550/arXiv.2004.12537.

23. Morozov, S., Andreychenko, A., Pavlov, N., Vladzymyrskyy, A., Ledikhova, N., Gombolevskiy, V., Blokhin, I.A., Gelezhe, P., Gonchar, A., and Chernina, V.Y. (2020). Mosmeddata: chest ct scans with covid-19 related findings dataset. Preprint at arXiv. https://doi.org/10.48550/arXiv.2005.06465.

24. Zhao, T., and Wu, X. (2019). Pyramid feature attention network for saliency detection. Preprint at arXiv, 3085–3094. https://doi.org/10.48550/arXiv.1903.00179.

25. Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., and Lu, H. (2019). Dual attention network for scene segmentation. Preprint at arXiv, 3146–3154. https://doi.org/10.48550/arXiv.1809.02983.

26. Gao, J., Wang, Q., and Yuan, Y. (2019). SCAR: spatial-/channel-wise attention regression networks for crowd counting. Neurocomputing *363*, 1–8.

27. Zhao, C., Han, J., Jia, Y., and Gou, F. (2018). Lung Nodule Detection via 3D U-Net and Contextual Convolutional Neural Network (IEEE)), pp. 356–361.

28. Wang, S., Zhou, M., Liu, Z., Liu, Z., Gu, D., Zang, Y., Dong, D., Gevaert, O., and Tian, J. (2017). Central focused convolutional neural networks: developing a data-driven model for lung nodule segmentation. Med. Image Anal. *40*, 172–183.

29. Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature pyramid networks for object detection. Preprint at arXiv, 2117–2125. https://doi.org/10.48550/arXiv.1612.03144.

30. Schonfeld, E., Schiele, B., and Khoreva, A. (2020). A u-net based discriminator for generative adversarial networks. Preprint at arXiv, 8207–8216. https://doi.org/10.48550/arXiv.2002.12655.

31. Chen, T., Zhai, X., Ritter, M., Lucic, M., and Houlsby, N. (2019). Self-supervised gans via auxiliary rotation loss. Preprint at arXiv, 12154–12163. https://doi.org/10.48550/arXiv.1811.11212.

32. Jaiswal, A., Moyer, D., Ver Steeg, G., AbdAlmageed, W., and Natarajan, P. (2020). Invariant representations through adversarial forgetting. Preprint at arXiv, 4272–4279. https://doi.org/10.48550/arXiv.1911.04060.

33. Wang, X., Girshick, R., Gupta, A., and He, K. (2018). Non-local neural networks. Preprint at arXiv, 7794–7803. https://doi.org/10.48550/arXiv.1711.07971.

34. Zhang, H., Goodfellow, I., Metaxas, D., and Odena, A. (2019). Self-attention generative adversarial networks (PMLR)), pp. 7354–7363.

35. Wang, Q., Teng, Z., Xing, J., Gao, J., Hu, W., and Maybank, S. (2018). Learning Attentions: Residual Attentional Siamese Network for High Performance Online Visual Tracking, pp. 4854–4863.

36. Woo, S., Park, J., Lee, J.-Y., and So Kweon, I. (2018). Cbam: convolutional block attention module. Preprint at arXiv, 3–19. https://doi.org/10.48550/arXiv.1807.06521.

37. Khan, F.S., van de Weijer, J., and Vanrell, M. (2012). Modulating shape features by color attention for object recognition. Int. J. Comput. Vis. *98*, 49–64.

38. Rahman, M.M., Fiaz, M., and Jung, S.K. (2020). Efficient visual tracking with stacked channel-spatial attention learning (IEEE Access).

39. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., and Fu, Y. (2018). Image super-resolution using very deep residual channel attention networks. Preprint at arXiv, 286–301. https://doi.org/10.48550/arXiv.1807.02758.

40. Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., and Tang, X. (2017). Residual attention network for image classification. Preprint at arXiv, 3156–3164. https://doi.org/10.48550/arXiv.1704.06904.

41. Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R., and Bengio, Y. (2015). Show, attend and tell: neural image caption generation with visual attention. Preprint at arXiv, 2048–2057. https://doi.org/10.48550/arXiv.1502.03044.

42. Zhu, Y., Groth, O., Bernstein, M., and Fei-Fei, L. (2016). Visual7w: grounded question answering in images. Preprint at arXiv, 4995–5004.

43. Xu, H., and Saenko, K. (2016). Ask, Attend and Answer: Exploring Question-Guided Spatial Attention for Visual Question Answering (Springer), pp. 451–466.

44. Chen, L., Zhang, H., Xiao, J., Nie, L., Shao, J., Liu, W., and Chua, T.-S. (2017). Sca-cnn: spatial and channel-wise attention in convolutional networks for image captioning. Preprint at arXiv, 5659–5667. https://doi.org/10.48550/arXiv.1611.05594.

45. Odena, A., Olah, C., and Shlens, J. (2016). Conditional image synthesis with auxiliary classifier GANs. Preprint at arXiv. https://doi.org/10.48550/arXiv.1610.09585.

46. Miyato, T., and Koyama, M. (2018). cGANs with projection discriminator. Preprint at arXiv. https://doi.org/10.48550/arXiv.1802.05637.

47. Doersch, C., Gupta, A., and Efros, A.A. (2015). Unsupervised visual representation learning by context prediction. Preprint at arXiv, 1422–1430. https://doi.org/10.48550/arXiv.1505.05192.

48. Gidaris, S., Singh, P., and Komodakis, N. (2018). Unsupervised representation learning by predicting image rotations. Preprint at arXiv. https://doi.org/10.48550/arXiv.1803.07728.

49. Ibrahim, D.A., Zebari, D.A., Mohammed, H.J., and Mohammed, M.A. (2022). Effective hybrid deep learning model for COVID-19 patterns identification using CT images. Expet Syst. *39*, e13010. https://doi.org/10.1111/exsy.13010.

50. Mahmoudi, R., Benameur, N., Mabrouk, R., Mohammed, M.A., Garcia-Zapirain, B., and Bedoui, M.H. (2022). A deep learning-based diagnosis system for COVID-19 detection and pneumonia screening using CT imaging. Appl Sci-Basel *12*, 4825. https://doi.org/10.3390/app12104825.

51. Mohammed, M.A., Al-Khateeb, B., Yousif, M., Mostafa, S.A., Kadry, S., Abdulkareem, K.H., and Garcia-Zapirain, B. (2022). Novel crow swarm optimization algorithm and selection approach for optimal deep learning COVID-19 diagnostic model. Comput. Intell. Neurosci. *2022*, 1307944. https://doi.org/10.1155/2022/1307944.

52. Shamim, S., Awan, M.J., Zain, A.M., Naseem, U., Mohammed, M.A., and Garcia-Zapirain, B. (2022). Automatic COVID-19 lung infection segmentation through modified unet model. J Healthc Eng *2022*. https://doi.org/10.1155/2022/6566982.

53. Allioui, H., Mohammed, M.A., Benameur, N., Al-Khateeb, B., Abdulkareem, K.H., Garcia-Zapirain, B., Damaševičius, R., and Maskeliūnas, R. (2022). A multi-agent deep reinforcement learning approach for enhancement of COVID-19 CT image segmentation. J. Personalized Med. *12*, 309. https://doi.org/10.3390/jpm12020309.

54. Abdulkareem, K.H., Mostafa, S.A., Al-Qudsy, Z.N., Mohammed, M.A., Al-Waisy, A.S., Kadry, S., Lee, J., and Nam, Y. (2022). Automated system for identifying COVID-19 infections in computed tomography images using deep learning models. J. Healthc. Eng. *2022*, 5329014. https://doi.org/10.1155/2022/5329014.

55. Shen, W., Zhou, M., Yang, F., Yang, C., and Tian, J. (2015). Multi-scale Convolutional Neural Networks for Lung Nodule Classification (Springer), pp. 588–599.

56. Chen, S., Ma, K., and Zheng, Y. (2019). Med3d: transfer learning for 3d medical image analysis. Preprint at arXiv. https://doi.org/10.48550/arXiv.1904.00625.

57. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015).

Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. *115*, 211–252.

58. Aresta, G., Jacobs, C., Araújo, T., Cunha, A., Ramos, I., van Ginneken, B., and Campilho, A. (2019). iW-Net: an automatic and minimalistic interactive lung nodule segmentation deep network. Sci. Rep. *9*, 11591. https://doi.org/10.1038/s41598-019-48004-8.

59. Shakir, H., Rasool Khan, T.M., and Rasheed, H. (2018). 3-D segmentation of lung nodules using hybrid level sets. Comput. Biol. Med. *96*, 214–226. https://doi.org/10.1016/j.compbiomed.2018.03.015.

60. Chen, X., Yao, L., and Zhang, Y. (2020). Residual attention u-net for automated multi-class segmentation of covid-19 chest ct images. Preprint at arXiv. https://doi.org/10.48550/arXiv.2004.05645.

61. Chen, X., Lin, L., Liang, D., Hu, H., Zhang, Q., Iwamoto, Y., Han, X.-H., Chen, Y.-W., Tong, R., and Wu, J. (2019). A Dual-Attention Dilated Residual Network for Liver Lesion Classification and Localization on CT Images (IEEE)), pp. 235–239.

62. Chen, X., Lin, L., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., Tong, R., and Wu, J. (2019). A Cascade Attention Network for Liver Lesion Classification in Weakly-Labeled Multi-phase Ct Images (Springer), pp. 129–138.

63. Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. Preprint at arXiv, 7132–7141. https://doi.org/10.48550/arXiv.1709.01507.

64. Jing, L., Yang, X., Liu, J., and Tian, Y. (2018). Self-supervised spatiotemporal feature learning via video rotation prediction. Preprint at arXiv. https://doi.org/10.48550/arXiv.1811.11387.

65. Liang, X., Li, N., Zhang, Z., Xiong, J., Zhou, S., and Xie, Y. (2021). Incorporating the hybrid deformable model for improving the performance of abdominal CT segmentation via multi-scale feature fusion network. Med. Image Anal. *73*, 102156. https://doi.org/10.1016/j.media.2021.102156.

66. Liu, J., Liu, H., Gong, S., Tang, Z., Xie, Y., Yin, H., and Niyoyita, J.P. (2021). Automated cardiac segmentation of cross-modal medical images using unsupervised multi-domain adaptation and spatial neural attention structure. Med. Image Anal. *72*, 102135. https://doi.org/10.1016/j.media.2021.102135.

67. Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). Pyramid scene parsing network. Preprint at arXiv, 2881–2890. https://doi.org/10.48550/arXiv.1612.01105.

68. Wang, H., Wang, L., Lee, E.H., Zheng, J., Zhang, W., Halabi, S., Liu, C., Deng, K., Song, J., and Yeom, K.W. (2021). Decoding COVID-19 pneumonia: comparison of deep learning and radiomics CT image signatures. Eur. J. Nucl. Med. Mol. Imag. *48*, 1478–1486. https://doi.org/10.1007/s00259-020-05075-4.

69. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., and Maier-Hein, K.H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods *18*, 203–211.

70. Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation (Springer), pp. 234–241.

71. Kingma, D.P., and Ba, J. (2014). Adam: a method for stochastic optimization. Preprint at arXiv. https://doi.org/10.48550/arXiv.1412.6980.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Lung nodule CT images	The Cancer Imaging Archive	https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=1966254
Pneumonia CT images	zenodo	https://github.com/JunMa11/COVID-19-CT-Seg-Benchmark
Code for lung lesion segmentation	This paper	https://github.com/JD910/general_net_for_lesion_seg
Software and algorithms		
Pytorch	PyTorch Foundation	https://pytorch.org/
Python	Python Software Foundation	https://www.python.org

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jiangdian Song (song.jd0910@gmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

CT images have been deposited at The Cancer Imaging Archive and zenodo, and are publicly available as of the date of publication. DOIs are listed in the key resources table.

All original code has been deposited at GitHub and is publicly available as of the date of publication. DOIs are listed in the key resources table.

Additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

The source code of this study is published at: https://github.com/JD910/general_net_for_lesion_seg.

EXPERIMENTAL MODEL AND PARTICIPANTS DETAILS

This study used 110 COVID-19 positive cases and 103 COVID-19 negative cases, and 20 COVID-19 positive cases from the open-access datasets. For lung nodule, 1018 patients with lung nodule from the LIDC-IDRI dataset were used. In addition, 1652 patients with lung nodules and 220 patients with tuberculosis images were enrolled in this study (1012 males, median age: 40 years). Appropriate Institutional Review Board approval was obtained, and the need for informed consent was waived in this retrospective study. Because studies in the field of lung lesion segmentation do not specifically calculate the required sample size, and the number of subjects included in this study is more than previous studies, therefore, no additional sample size calculation was performed. In addition, this study divided the dataset according to the consensus of deep learning with an 80:10:10 ratio to construct a training, validation, and test dataset (the influence of sex was not specifically considered according to previous studies in this field), respectively.

METHOD DETAILS

Figure S6 shows the proposed self-supervised adversarial learning (*GSAL*) architecture. A generator network G, which is composed of a U-Net-like structure that includes a cascaded context-aware pyramid feature extraction (CPFE) with a dual attention module, was constructed to first produce a lung lesion image $G(x)$ from the original CT image. A discriminator network comprising a series of convolution blocks with self-supervised auxiliary rotation loss, was subsequently employed to distinguish the generated image $G(x)$ from the ground truth image $R(x)$.

Generator network G

The generator network *G* proposed here comprises a U-Net-like structure that includes a cascaded CPFE for multi-receptive-field feature extraction and a dual attention module consisting of a spatial attention module and channel-wise attention module to enhance lung lesion representation.

U-Net-like structure

The U-Net-like structure is a feasible GAN generator for image segmentation.[27,30] From the contracting path of U-Net, we obtained feature maps for the layers with 64, 128, 256, and 512 filters (called G_64, G_128, G_256, and G_512, respectively), and the corresponding expansive path with deconvolution was used to output the nodule image. Subsequently, the feature maps of the U-Net-like structure were input into the cascaded CPFE and attention modules to consider the lung lesion semantic interdependencies in the spatial and channel dimensions.

Cascaded CPFE

A dual attention module was used in *GSAL* to enhance the high-level contextual features in the deep layers and low-level spatial textural features in the shallow layers. Specifically, a spatial attention module was used to enhance the texture details of lung lesions in a CT image, and a channel-wise attention module was applied to enhance the semantic context of a lung lesion. In addition, because stepwise feature maps only obtain the approximate regions of lesions, a CPFE network[24] was first used to obtain multiscale, multi-receptive-field features to represent lesions of various sizes. The cascaded structure sends the output of the CPFE module to the dual attention module such that the most appropriate scale and receptive field for lung lesion detection are recognized.

Figure S7 shows that the CPFE module comprises parallel dilated atrous convolutional layers at different dilation rates. It was employed on each of the three semantic feature map outputs using a U-Net-like structure to capture the contextual information of multiple receptive fields. There are distinct variations in the scale, shape, and location of lung lesions in CT images; hence, parallel dilated atrous convolution is suitable for recognizing lung lesions such that contextual information at different scales can be captured. The dilation rates were set to one, two, and three to extract scale- and shape-invariant features.[67] For each feature map, the outputs of the parallel dilated atrous convolution layers were combined via cross-channel concatenation (see Figure S7).

The dual-attention module enabled us to adaptively integrate similar features at any scale from a global perspective, thereby improving the detection and segmentation of lung lesions. The steps of the dual attention module are as follows. First, a spatial/channel-wise attention matrix is generated to model the correlation between any two channels or pixels of a feature map. Second, the attention matrix and transformation of the original feature map are combined using matrix multiplication. Third, the results of the multiplication and the original feature map are summed element-wise to obtain the final feature representations.

Spatial attention module

The spatial attention module encodes a wider range of contextual information into local features, thus enhancing their ability to represent the ROIs details. By evaluating the spatial correlation between any two pixels on the feature map, the spatial attention module focuses on the discrimination of structural and textural details in the ROI from those of the background, which helps discriminate features for lung lesion extraction.

Figure S8 shows that the low-level feature map derived from the CPFE module, $LF \in R^{C\times H\times W}$, is first reshaped to $S \in R^{C\times N}$, where $N = H \times W$. The transpose of S and S are then multiplied to obtain the correlation between any two pixels in the feature map. The result is passed through a softmax layer to calculate spatial attention feature map $SA \in R^{N\times N}$.

$$SA_{ji} \in \frac{exp\left(S_i^T \cdot S_j\right)}{\sum_{i=1}^{N} exp\left(S_i^T \cdot S_j\right)} \quad \text{(Equation 1)}$$

where S^T is the transpose of S and SA_{ji} denotes the impact of the pixel at the i^{th} position on pixel at the j^{th} position. A closer representation of these two positions contributes to a stronger correlation.

Next, matrix $S \in R^{C\times N}$ is multiplied by the transpose of SA; the result is then reshaped to $R^{C\times H\times W}$. Finally, we multiply it by a trainable scale parameter α and subsequently sum it element-wise with the original LF to obtain the final output: $SAM \in R^{C\times H\times W}$.

$$SAM = \alpha \times reshape\left(S \cdot SA^{T}\right) + LF \quad \text{(Equation 2)}$$

where SA^{T} denotes the transpose of SA, and $reshape$ outputs the feature map with dimensions $R^{C\times H\times W}$. α is a trainable weight.

Channel-wise attention module

The channel-wise attention module aims to exploit the interdependencies between high-level channel maps and improve the feature representation of specific semantics. First, the high-level feature maps output from the CPFE module, $HF \in R^{C'\times H'\times W'}$, are reshaped to $C \in R^{C'\times N'}$, where $N' = H' \times W'$. Then, C and the transpose of C are multiplied and passed through a softmax layer to output the channel attention feature map CA.

$$CA_{ji} \in \frac{exp\left(C_i \cdot C_j^{T}\right)}{\sum_{i=1}^{C} exp\left(C_i \cdot C_j^{T}\right)} \quad \text{(Equation 3)}$$

where $CA \in R^{C'\times C'}$, C^{T} is the transpose of C, and CA_{ji} denotes the impact of the i^{th} channel on the j^{th} channel.

Next, the transposes of CA and C are multiplied; the result is reshaped to $R^{C'\times H'\times W'}$. The final output of the network is obtained by multiplying the result of $CA^{T} \cdot C$ by a trainable scale parameter β and then computing the element-wise sum with input HF.

$$CAM = \beta \times reshape\left(CA^{T} \cdot C\right) + HF \quad \text{(Equation 4)}$$

where CAM is the output of the channel-wise attention module, and CA^{T} denotes the transpose of the channel attention feature map. β is a trainable weight.

Using the above spatial and channel-wise attention modules, the inter-channel and inter-pixel correlations of the feature maps were detected, and the contextual semantic and local texture of lesions were enhanced, respectively, thus improving the discriminability of the semantics of lung lesion features.

Discriminator network D

Previous studies have indicated that an auxiliary self-supervised rotation loss in the discriminator of a GAN retains generalizable representations in a nonstationary environment, and this prevents the catastrophic forgetting of classes in discriminator representations during training iterations.[47,48] In the proposed *GSAL*, an auxiliary rotation loss is introduced into discriminator network D to eliminate discriminator forgetting during model training. The generated images $G(x)$ and ground truth image $R(x)$ are first rotated by $r \in R$ ($R = \{0°, 90°, 180°, 270°\}$). The angles of rotation are then decoded as artificial labels to calculate the rotation loss of the corresponding input images. The details are as follows.

Using images obtained by rotating both $G(x)$ and $R(x)$ at four angles R, the rotated images are concatenated channel-wise and input into a series of convolution blocks (see Figure S9). Here, a batch size of eight and two graphics processing units were used. Thus, for each graphic processing unit, there were 16 images derived from $G(x)$ and 16 corresponding images derived from $R(x)$ that were input into the discriminator network. We define the input image as $Input()$.

$$\begin{cases} Input(G(x)) = Concat\left(Rot = \left(G(x)^{0°}, G(x)^{90°}, G(x)^{180°}, G(x)^{270°}\right)\right) \\ Input(R(x)) = Concat\left(Rot = \left(R(x)^{0°}, R(x)^{90°}, R(x)^{180°}, R(x)^{270°}\right)\right) \end{cases} \quad \text{(Equation 5)}$$

Figure S9 shows the proposed discriminator network comprises four successive convolution blocks called Conv_D1, Conv_D2, Conv_D3, and Conv_D4. Two fully connected networks are then used, one to output the discriminator loss of the image and the other to output the corresponding rotation loss.

$$\begin{cases} (G\_pro\_loss, G\_rot\_pros) = Conv\_D_{r\sim R}[Concat(Rot = r|G(x)^{r})] \\ (R\_pro\_loss, R\_rot\_pros) = Conv\_D_{r\sim R}[Concat(Rot = r|R(x)^{r})] \end{cases} \quad \text{(Equation 6)}$$

where *pro_loss* denotes the discriminator loss with a shape of [4, 1] for each input image and *rot_pros* represents the rotation probabilities of each input image with a shape of [4, 4]. In addition, $r \sim R$ is the rotation angle r selected from a set of possible rotations $R \in \{0°, 90°, 180°, 270°\}$. Image x rotated by r degrees is denoted as $(x)^r$. Furthermore, $Conv\_D(Rot|(x)^r)$ denotes the predicted distribution of the discriminator over the angles of rotation of the sample.

One-hot encoding is used to produce the artificial labels of the rotation angles, named *rot_labels* (see Figure S6), to evaluate the distance between the discriminator output and the rotation label. As Equation 7 describes, the one-hot encoded *rot_labels* form a [4, 4] matrix per input image, which the shape is the same as the *rot_pros* mentioned above. Cross entropy is used to calculate the distance between the *rot_labels* and *rot_pros*.

$$\begin{cases} G\_Rot\_loss = sum\left[binary\_cross\left(G\_rot\_pros, rot\_labels\right)\right] \\ R\_Rot\_loss = sum\left[binary\_cross\left(R\_rot\_pros, rot\_labels\right)\right] \end{cases} \quad \text{(Equation 7)}$$

where *G_Rot_loss* and *R_Rot_loss* denote the rotation loss of the generated image $G(x)$ and ground truth image $R(x)$ through the discriminator network, respectively. Moreover, *binary_cross* is the loss obtained by the function of *binary_cross_entropy_with_logits*. The sum of the loss of images in one batch is used in the proposed method.

For the backpropagation, the following loss, employed by the Wasserstein GAN with a gradient penalty, was used:

$$GAN\_loss = G\_pro\_loss - R\_pro\_loss + GP[G(x), R(x)] \quad \text{(Equation 8)}$$

where *GP* represents the gradient penalty in the Wasserstein GAN with a gradient penalty. The final loss of the proposed discriminator network is the weighted sum of the above loss and the rotation loss of $R(x)$, expressed as follows:

$$D_{loss} = GAN\_loss + \gamma \cdot R\_Rot\_loss \quad \text{(Equation 9)}$$

where γ denotes the weight of rotation loss of ground truth image $R(x)$. For the loss of the proposed generator network, we used the summed loss of the $G(x)$ image obtained by the discriminator network and the weighted rotation loss of $G(x)$, as follows.

$$G_{loss} = G\_pro\_loss + \delta \cdot G\_Rot\_loss \quad \text{(Equation 10)}$$

where δ is the weight of the rotation loss of generated image $G(x)$.

Experiment

To evaluate the proposed *GSAL* model for lung lesion segmentation, pneumonia, lung nodule, and tuberculosis images from multiple centers were used in this study.

For the pneumonia lesion segmentation, two open-access datasets, which included both COVID-19 pneumonia and other viral-infected community-acquired pneumonia with similar CT signs to COVID-19, with manual delineation from radiologists were used. A total of 14,260 pneumonia images were collected from Wang et al.'s study,[68] including 110 COVID-19 positive cases and 103 COVID-19 negative cases. In addition, 20 COVID-19 positive cases were collected from the COVID-19-CT-Seg-Benchmark,[22] and 1,540 images with manual annotations were collected. Images were resampled to 256 × 256 pixels, and only images with over 1,000 pneumonia pixels manually delineated by radiologists were included in our study to prevent under-segmentation caused by training ROIs that were too small. All the included cases were randomly divided at an 80:10:10 ratio to construct a training, validation, and test dataset, respectively.

To evaluate lung nodule segmentation performance, we used the LIDC-IDRI dataset and CT images collected from three institutions participating in this study. All the images from the LIDC-IDRI database were manually delineated by one to four radiologists. For the remaining datasets, all the nodules were manually segmented and reviewed by at least two experts, and the ITK-Snap software was used to segment the lung nodules slice-by-slice. Images were resampled to 256 × 256 pixels, and all lung nodules with a maximum diameter greater than 3 mm after resampling were included in the study. Finally, 63,080 CT images that included at least one lung nodule were included. All the included lung nodules were randomly divided into the training, validation, and test dataset with the same ratio.

A total of 3361 tuberculosis CT images of 220 patients were collected from two participant units, and all tuberculosis areas were manually delineated by the local radiologists with over five years of radiology experience using ITK-SNAP. To ensure generality, this study included nodular tuberculosis, secondary pulmonary tuberculosis, hematogenous pulmonary tuberculosis, and other pulmonary tuberculosis subtypes with different signs in CT images. The training, validation, and test datasets were constructed with the same ratio after resampling all the images to 256 × 256.

The radiologists' segmentations intersection for each lung lesion is denoted as $R(x)$, which is used as the gold standard for the segmentation in this study, and the predicted lung lesion image is denoted as $G(x)$.

To verify the superiority of the proposed model independent of training sample size, three experiments with reduced sample sizes were also conducted. *M_100*, *M_70*, *M_40*, and *M_10*, are used to refer to experiments in which 100, 70, 40, and 10% of the images were randomly selected to construct the training, validation, and test dataset to build and evaluate the *GSAL* model, respectively. In addition, to demonstrate the cascaded CPFE, the attention module, and the self-supervised rotation loss proposed for lung lesion segmentation, an ablation study was conducted to remove the above modules for lung lesion segmentation, and the comparison of segmentation results between the two was then performed.

The DC was used in this study to evaluate the automatic segmentation of the lung lesions accuracy. The DC was calculated by comparing the overlapping area between the manual segmentation of the ROI of lung lesions on the original CT images and the generated ROI of lung lesions on the mimic images.

$$DC_{R,G} = (2|R \cap G| / (|R| + |G|)) * 100\% \quad \text{(Equation 11)}$$

where $DC_{R,G}$ is the DC of the real ROI R and the generated ROI G, and R and G indicate the two clustered nodule volumes (i.e., of the real and generated lung lesion images). The segmentation results obtained by *M_100*, *M_70*, *M_40*, and *M_10*, were compared with the gold standard $R(x)$ to calculate the DC. The statistics of the segmentation accuracy were acquired by comparing the generated lesions and the corresponding manually delineated lesions on the CT images.

A state-of-the-art nnU-Net method, considered a promising general medical image segmentation architecture, was compared with the proposed model in lung lesion segmentation.[69] Segmentation accuracy and time consumption were calculated for comparison. In addition, the current segmentation networks for medical images are mostly designed using U-net as a baseline framework.[70] Therefore, as a representative of the widely used segmentation networks to date, U-net was also used for comparative analysis with the proposed algorithm.

In addition, to demonstrate the lung lesion detection accuracy of the proposed *GSAL*, two in-house well-trained radiologists reviewed all the images produced in the test dataset. True positive (*TP*), false positive (*FP*), and false-negative (*FN*) were manually determined, and the precision (*P*), recall (*R*), and *f*1 measurement, which is the harmonic mean of precision and recall, were used to evaluate the performance. These metrics are calculated as follows.

$$\begin{cases} P = TP/(TP+FP) \\ R = TP/(TP+FN) \\ f1 = (2 \cdot P \cdot R)/(P+R) \end{cases} \quad \text{(Equation 12)}$$

The PyTorch platform was used to implement the algorithms. All the training, validation, and tests were performed on a computer with two 24 GB GeForce RTX3090. A pre-experiment was conducted using 1,000 randomly selected CT images to determine the best hyper-parameters for the experimental conditions. A batch size of eight was used; the number of training iterations was set to 200; the Adam optimizer[71] was used with an initial learning rate of 1e−4; the parameters for the Adam optimizer were set to $\beta 1 = 0.9$ and $\beta 2 = 0.99$; the size of the input CT image was set to 256 × 256. *GSAL* parameters were 15.17M for the generator and 2.93M for the discriminator. Data augmentation including translating, rotating, and scaling of the CT images was performed on the training datasets. The initialization of the trainable scale parameters, α in the SAM and β in the CAM, was 1.0. For the rotation loss, the weight γ =1.0 for the ground truth image for D_{loss}, and $\delta = 0.5$ for the generated image for G_{loss}. The hyper-parameters of nnU-Net are defined in Isensee et al.'s study.[69]