

Received 13 October 2023; revised 4 March 2024; accepted 7 May 2024.
Date of publication 9 May 2024; date of current version 17 May 2024.

Digital Object Identifier 10.1109/JTEHM.2024.3399261

Weakly-Supervised Segmentation-Based Quantitative Characterization of Pulmonary Cavity Lesions in CT Scans

WENYU XING¹, YANPING YANG², YANAN ZHOU², TAO JIANG³, YIFANG LI¹,
YUANLIN SONG^{2,4}, DONGNI HOU⁴, AND DEAN TA^{1,3}, (Senior Member, IEEE)

¹Institute of Biomedical Engineering and Technology, Academy for Engineering and Technology, Fudan University, Shanghai 200433, China

²Shanghai Institute of Infectious Disease and Biosecurity, Fudan University, Shanghai 200032, China

³Department of Biomedical Engineering, School of Information Science and Technology, Fudan University, Shanghai 200438, China

⁴Department of Pulmonary and Critical Care Medicine, Zhongshan Hospital, Fudan University, Shanghai 200032, China

(Wenyu Xing and Yanping Yang contributed equally to this work.) CORRESPONDING AUTHORS: D. TA (tda@fudan.edu.cn)
AND D. HOU (hou.dongni@zs-hospital.sh.cn).

This work was supported in part by Shanghai Municipal Science and Technology Major Project ZD2021CY001; in part by the Science and Technology Commission of Shanghai Municipality under Grant 20DZ2261200, Grant 220Z11901000, and Grant 220XD1401200; in part by Shanghai Municipal Key Clinical Specialty under Grant shslczdc02201; and in part by China Post-Doctoral Science Foundation under Grant 2023M740712.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Zhongshan Hospital under Application No. B2021-183R.

ABSTRACT **Objective:** Pulmonary cavity lesion is one of the commonly seen lesions in lung caused by a variety of malignant and non-malignant diseases. Diagnosis of a cavity lesion is commonly based on accurate recognition of the typical morphological characteristics. A deep learning-based model to automatically detect, segment, and quantify the region of cavity lesion on CT scans has potential in clinical diagnosis, monitoring, and treatment efficacy assessment. **Methods:** A weakly-supervised deep learning-based method named CSA²-ResNet was proposed to quantitatively characterize cavity lesions in this paper. The lung parenchyma was firstly segmented using a pretrained 2D segmentation model, and then the output with or without cavity lesions was fed into the developed deep neural network containing hybrid attention modules. Next, the visualized lesion was generated from the activation region of the classification network using gradient-weighted class activation mapping, and image processing was applied for post-processing to obtain the expected segmentation results of cavity lesions. Finally, the automatic characteristic measurement of cavity lesions (e.g., area and thickness) was developed and verified. **Results:** the proposed weakly-supervised segmentation method achieved an accuracy, precision, specificity, recall, and F1-score of 98.48%, 96.80%, 97.20%, 100%, and 98.36%, respectively. There is a significant improvement ($P < 0.05$) compared to other methods. Quantitative characterization of morphology also obtained good analysis effects. **Conclusions:** The proposed easily-trained and high-performance deep learning model provides a fast and effective way for the diagnosis and dynamic monitoring of pulmonary cavity lesions in clinic.

INDEX TERMS Pulmonary cavity lesion, weakly-supervised segmentation, CSA²-ResNet, Grad-CAM, quantitative characterization.

Clinical and Translational Impact Statement—This model used artificial intelligence to achieve the detection and quantitative analysis of pulmonary cavity lesions in CT scans. The morphological features revealed in experiments can be utilized as potential indicators for diagnosis and dynamic monitoring of patients with cavity lesions.

I. INTRODUCTION

PULMONARY cavity is a common type of lesions which refers to gas-containing space seen as a lucency or

low-attenuation area, within a nodule, mass, or area of parenchymal consolidation [1]. It is caused by malignancies, suppurative or caseous infections, and ischemic parenchymal

necrosis [2]. Pathologically, it mainly results from necrosis and liquefaction of lung parenchyma, followed by the entry of air through bronchial drainage [3]. On CT imaging, pulmonary cavities appear as ring-shaped shadows within the parenchymal region [1].

In clinical practice, manifestations of pulmonary cavitory lesions in CT images, such as location, morphology, size, and margin features are of crucial significance for differentiating benign or malignant disease, and inferring etiology of the lesion. Currently, clinicians use manual measurement and qualitative morphological features to evaluate cavity lesions [4], which is time-consuming, subjective, and difficult to compare between serial CT scans and between different cases. Therefore, computer aided automatic segmentation and characteristics quantification of cavity lesions is expected to provide new opportunities for faster and more accurate clinical evaluation, which is essential in differential diagnosis, treatment efficacy assessment, and prognostication. Furthermore, it will also facilitate large scale longitudinal clinical researches and epidemiological investigations to gain deeper insights into the pathological mechanisms and outcome patterns.

Previously, supervised deep learning models has been widely used in CT scan image segmentation tasks. For example, Xing et al. [5] designed a novel segmentation model by combining convolution and multilayer perceptron (MLP) modules to extract and reconstruct the bilateral parenchyma in CT scans. Zhao et al. [6] integrated a three-dimensional V-Net with a shape deformation module implemented using a spatial transformer network to segment lung parenchyma in CT scans. Hu et al. [7] proposed a parallel deep learning algorithm with multi-attention mechanism and densely connected convolutional networks to segment the tumor in CT slice images. Chaganti et al. [8] proposed a convolutional neural network (CNN) model based on the pretrained DenseNet model for the segmentation of CT patterns of viral pneumonia. Although these methods achieve good segmentation results, they usually require obtaining mask labels of the target area in advance for training, which is labor-intensive and inefficient. What's more, it is noteworthy that compared to the relatively simple task of lung parenchyma segmentation, the edges of pulmonary cavities are complex and diverse in shape, making it difficult to achieve accurate mask label extraction through traditional image processing or manual delineation. Therefore, supervised segmentation models may not be well-suited for the segmentation task of cavity lesions.

To address such issues, weakly-supervised semantic segmentation has been proposed [9], [10]. It primarily relies on image classification and activation map generation to obtain segmentation results of regions of interest (ROI). Compared to directly obtaining segmentation labels, using only image-level labels makes the entire task simpler. It no longer requires pre-obtaining segmentation labels to achieve target segmentation in complex backgrounds, significantly improving the efficiency and feasibility of segmentation tasks, especially for complex objects.

Researchers previously introduced the weakly-supervised idea in detecting COVID-19 infection lesions, enabling precise segmentation of the lesions. For example, Li et al. [11] proposed a classifier-augmented generative adversarial network, which achieved weakly supervised localization of COVID-19 lung lesion in CT slice images. Yang et al. [12] proposed a weakly-supervised method based on generative adversarial network (GAN) with image-level labels only for COVID-19 lesion localization. Sun et al. [13] employed a Local Self-Coherence Mechanism to accomplish label propagation based on lesion area labeling characteristics and achieved the weakly supervised segmentation of COVID-19 infection on CT images. Xie et al. [9] propose a weakly-supervised segmentation method based on dense regression activation maps with attention neural network module, which achieved lesion segmentation in CT scans. Lu et al. [14] proposed a novel weakly supervised inpainting-based learning method, in which only bounding box labels are required for accurate lesion segmentation.

However, for pulmonary cavity lesions in CT scans, more effective segmentation methods are required for clinical application. In this paper, considering the current research status and existing problems, we proposed a novel weakly-supervised deep learning framework-based pulmonary cavity lesion automatic segmentation model, which was composed of ROI extraction, image classification, and lesion segmentation. Based on this method, we also proposed automatic characteristic measurement methods and application scenarios of lesion morphology features. The main contributions are summarized as follows: (1) the proposed weakly-supervised model only requires image-level labels, which are easier and efficient to obtain than pixel-level labels; (2) the CSA²-ResNet classification model employed a hybrid attention mechanism into ResNet backbone model to enhance sensitivity and directionality towards lesions, which was beneficial for subsequent weakly-supervised lesion localization; (3) the accurate quantitative analysis of cavitory lesions based on precise segmentation results is beneficial for clinical diagnosis, assessment, and dynamic monitoring.

II. METHODS

A. PATIENTS

We retrospectively enrolled 35 adult patients aged 18 ~ 90 years old and admitted to Zhongshan Hospital Fudan University from July 2016 to October 2021. These patients had pulmonary cavitory lesions in chest CT scans, and the lesions were diagnosed as lung tuberculosis ($n = 5$), malignant tumor ($n = 10$), aspergilloma ($n = 5$), or bacterial abscess ($n = 5$). The patients underwent chest CT scans at different times points during the disease course. A total of 64 CT scans were obtained. All CT slice images were first segmented to isolate the lung parenchyma regions, and then 523 CT slice images were randomly selected for the 5-fold cross validation experiments. Among these selected images, 248 of them had cavity lesions, while 275 did not.

The CT scans were performed in a heat-first position at the end of the patient’s inhalation with breath holding. CT scans were collected using SOMATON Emotion 16 scanner (SIEMENS Healthcare) in spiral scanning mode and stored in Medical Digital Imaging and Communication (DICOM) format. No contrast agents were administered during the scanning process. All patients provided informed consent and approved by the Ethics Committee of Zhongshan Hospital (Number: B2021-183R).

B. CAVITY LESION SEGMENTATION

Figure 1 shows the schematic diagram of proposed weakly-supervised segmentation model for pulmonary cavity lesions. We first segment the lung parenchyma region using our previously proposed CM-SegNet segmentation model, and fed it into the self-designed CNN model (i.e., CSA²-ResNet) to predict the probability of presence of cavity lesion. Then the lesion was localized by the activation region of the classification network, which is generated by Gradient-weighted Class Activation Mapping (Grad-CAM) algorithm. Finally, by a simple post-processing, the expected automatic segmentation results of cavity lesions are obtained.

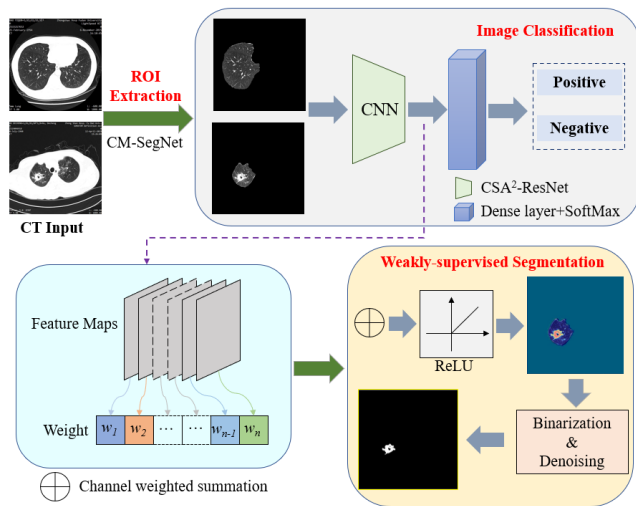


FIGURE 1. Schematic diagram of weakly-supervised segmentation model for cavity lesions.

1) ROI EXTRACTION

As the pulmonary cavity lesions are located within the pulmonary parenchyma, the key to the weakly supervised segmentation is ensure the accuracy of determining presence of cavity lesions in pulmonary parenchyma. Accurate segmentation of the pulmonary parenchyma in CT images is of paramount importance. Previously, we proposed a deep learning-based automatic segmentation approach named CM-SegNet [5] for 2D medical images by combining convolution and multilayer perceptron, which have good performance in CT image segmentation. In this study, before analyzing pulmonary cavity lesions, a pre-trained CM-SegNet model

was used to preprocess the input CT images to obtain the lung parenchyma ROI region. This preprocess step is to reduce the complexity of subsequent research and improve accuracy of lesion segmentation.

2) SUPERVISED DEEP LEARNING-BASED IMAGE CLASSIFICATION

As a classic convolutional neural network model, ResNet effectively alleviates the problem of gradient vanishing in deep networks by introducing residual connections, which improves the classification performance and plays an important role in computer vision tasks. However, based on CNN structures, ResNet is a local feature extraction method which only included the correlation between pixels of long distance is overlooked in ResNet model. In the task of recognizing cavity lesions, such long-distance correlations are of great importance because the morphology of one lesion is closely related with other lesions within a certain patient. Additionally, traditional residual structures usually fail to highlight the importance of certain crucial channels and prevent the model from focusing on the hollow lesion region due to the same weight allocated to different channels. To address these issues, a new hybrid-attention ResNet classification model referred to convolutional block attention module (CBAM) [18] and called CSA²-ResNet (Figure 2), was proposed by integrating channel and spatial attention modules to the classic ResNet model, achieving accurate identification of cavity lesions in lung parenchyma.

① ResNet-style base network

ResNet model is a typical CNN architecture composed of Identity and Conv residual modules, which utilizes recursive convolution operations to down-sample the input images, increase channel dimensions, and achieve high-dimensional feature extraction.

②Hybrid-attention-based ResNet-style network

To improve the performance of the base CNN models, hybrid channel and spatial attention modules were utilized in our proposed CSA²-ResNet classification model, as shown in Figure 2. The hybrid attention modules increase the directionality of feature analysis and improve the representation of interests.

Channel attention module was composed of squeeze and excitation [15]. Assuming the input feature map is $U \in R^{H \times W \times C}$, $U = \{u_1, u_2, u_3, \dots, u_c\}$, where $u_i \in R^{H \times W}$ represents a single channel, and H , W , and C represent the dimensions of the feature map in terms of height, width, and channels, respectively. The squeeze operation [Eq. (1)] utilizes global average-pooling in both the H and W dimensions to compress the feature maps of each channel. This encodes the spatial features of the entire space on a single channel into global features, resulting in $1 \times 1 \times C$ outputs used for channel-wise weight learning.

$$F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_j^W u_c(i, j) \tag{1}$$

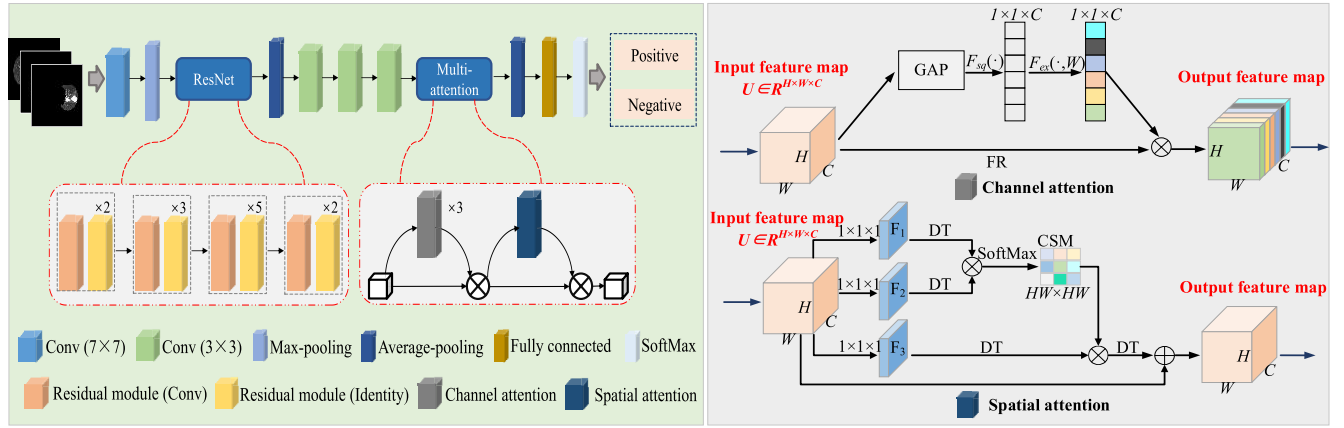


FIGURE 2. Schematic diagram of CSA²-ResNet classification model. GAP: global average pooling, FR: feature recalibration, DT: dimension transformation, CSM: correlation strength matrix.

In order to better capture the global feature information extracted by the squeeze operation, the model employs excitation operations to analyze the dependencies between channels. Two fully connected layers are primarily used to learn the non-linear cross-correlation between channels and the weights for each channel. The learned channel weights are then normalized using Sigmoid function, as shown in Eq. (2), to ensure that the weights range from 0 to 1 and sum up to 1. Finally, the weights of each channel learned through the excitation operation are multiplied with the original channels of the input feature map, achieving channel-wise feature recalibration (FR) of the original features. The resulting output of this operation serves as input for the next layer.

$$F_{ex}(z, W) = \sigma(W_2 \delta(W_1 z_c)) \quad (2)$$

where W_1 and W_2 represent fully connected layers. δ and σ represent ReLU and Sigmoid activation function, respectively. Z_c represents the global spatial information extracted by squeeze operation.

Spatial attention module captures the long-range dependencies of spatial information and plays a crucial role in focusing the model on the target region [16]. The process is shown in Eq. (3).

$$y_i = \frac{1}{\sum_j f(x_i, x_j)} \sum_j f(x_i, x_j) g(x_j) \quad (3)$$

where x and y represent the input and output, respectively. i and j represent the spatial positions of the input feature map. x_i is a feature vector with the same dimension as the input x . $f(\cdot)$ is a function that computes the similarity between i and j . $g(\cdot)$ represents the representation of the feature map at position j . The output y is obtained after normalization.

For the detailed calculation process of spatial attention, we firstly introduced a convolution layer with kernel size of $1 \times 1 \times 1$ to linearly map the input feature map and reduce its dimension, resulting in three features: F_1 , F_2 , and F_3 . Next, the dimensions of the three features,

in terms of height and width, are merged using a dimension transformation (DT) operation into $(HW) \times C$. Then we used matrix multiplication operation to obtain the correlation strength matrix (CSM) between any two points in F_1 and F_2 , with a size of $(HW) \times (HW)$, and normalized the results using a SoftMax function to obtain attention coefficients of each pixel with respect to other pixels. The higher the similarity between two-point features, the larger the response coefficient. The attention coefficients are used to perform weighted fusion with the F_3 feature matrix and the results were dimensionally transformed into a feature map with size of $H \times W \times C$. After a residual operation with the input feature map, the spatially fused feature map is obtained, which serves as the input for the next layer of the model.

Finally, we used a Global Average Pooling (GAP) layer to extract the global feature f_{Glo} from the output of the convolution layers. Identification of the presence of lung cavity lesions were then achieved by employing fully connected layers and a SoftMax classifier.

3) WEAKLY-SUPERVISED SEGMENTATION

To improve the accuracy of cavity lesion segmentation, we introduced a weakly-supervised algorithm based on Grad-CAM [17]. Firstly, we calculated the weights of each channel in the final convolution layer with respect to the image category. Then the weight vector was linearly weighted with the corresponding channel of the feature map to obtain a 2D activation map. Finally, a heatmap that is significantly positively correlated with the cavity lesion was outputted through the ReLU activation function, resulting in a visualization effect like attention mechanism. This enables discrimination of the location of cavity lesion in CT slice images.

After obtaining the visualized activation feature map of cavity lesions, the image processing algorithm (*i.e.*, adaptive threshold binarization, hole filling, and the connected-domain labeling) are employed to further eliminate other

surrounding infected lesions and achieve accurate extraction of the binary mask for cavity lesions.

C. CHARACTERISTICS ANALYSIS OF CAVITY LESIONS

Quantitative feature analysis of cavity lesions can help clinical observation of dynamic changes and identify the benign and malignant. To address the problems of objectivity, inaccuracy, and low efficiency in manual comparison, we proposed an automated characteristic analysis method based on weakly-supervised precise segmentation for analyzing the lesion features in CT images [1], [18], such as area, thickness, etc.

1) MEASUREMENT OF THE AREA OF CAVITY LESION

The hole filling in morphological reconstruction and the connected-domain labeling are used to extract the binary mask of the entire lesion region (ELR) and the single cavity region (SCR). Additionally, automated calculations of area and proportion are implemented as shown in Eq. (4). Based on the above results, it is possible to automatically capture the largest lesion slice and dynamically observe the development of cavity lesions over different time periods, as illustrated in Figure 3.

$$R_{ELR} = \frac{S_{ELR}}{S_P}, R_{SCR} = \frac{S_{SCR}}{S_{ELR}} \quad (4)$$

where S_{ELR} , S_{SCR} , and S_P represent the area of the entire lesion region, single hole region, and unilateral lung parenchymal region where the lesion is located, respectively. R_{ELR} and R_{SCR} represents the proportion of entire lesions in the parenchyma and single hole in the entire lesion, respectively.

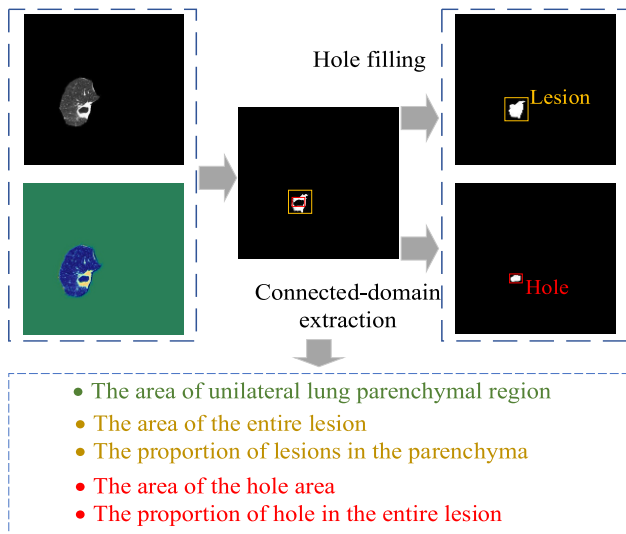


FIGURE 3. Schematic diagram of area and proportion measurement of cavity lesions.

2) MEASUREMENT OF THE THICKNESS OF THE HOLE WALL

The centroid (O) of separate hole was firstly obtained in the entire cavity lesion. With the centroid as the vertex, we make

outward spreading rays that will intersect with the edge of the cavity lesion at two points (i.e., P_1 and P_2), and the distance (L_n) between P_1 and P_2 is the thickness of the cavity lesion' wall in a single direction. After then, using a rotation angle of 1 degree and performing the same measurement operation, after 360 rounds (i.e., $N = 360$), the wall thickness of the cavity lesion in each direction can be obtained [Eq. (5)]. Based on the above results, it is convenient to achieve the statistic analysis of various features about the wall thickness of the cavity lesion in CT images, such as maximum, minimum, average, variance, etc. The schematic diagram of thickness measurement of the wall of cavity lesions is represented in Figure 4.

$$L_n = \sqrt{(x_{P_1} - x_{P_2})^2 + (y_{P_1} - y_{P_2})^2}, \quad n \in \{Z \mid (0, 360)\} \quad (5)$$

where P_1 and P_2 are the two intersection points between the spreading ray and the cavity lesion's inner and outer ring edges. x and y are the horizontal and vertical coordinates of intersection point, respectively.

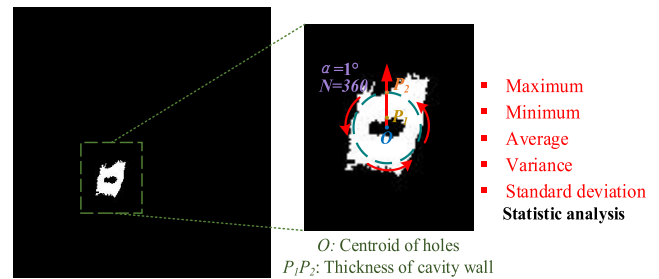


FIGURE 4. Schematic diagram of thickness measurement of the wall of cavity lesions.

D. EVALUATION OF WEAKLY-SUPERVISED SEGMENTATION MODEL

The two independent stages of our proposed weakly-supervised segmentation model were both evaluated using the "Accuracy" index, including ACC_1 and ACC_2 , as shown in Eqs. (6) and (7), respectively.

$$ACC_1 = \frac{TP' + TN'}{TP' + TN' + FP' + FN'} \quad (6)$$

$$ACC_2 = \frac{TP''}{TP'' + FP''} \quad (7)$$

where ACC_1 and ACC_2 represent the accuracy of the CSA²-ResNet classification model and the independent segmentation model, respectively. TP' , FP' , TN' , and FN' represent the number of correctly and incorrectly identified images with cavity lesions and without cavity lesions, respectively. TP'' and FP'' represent the number of correctly and incorrectly segmented cavity lesions. Because the edges of most cavity lesions are blurred, it is difficult to obtain the ground truth lesion, the segmentation correctness judgments are finally judged by two experienced pulmonary physicians

from Zhongshan Hospital Fudan University (≥ 6 years in reading lung CT scans). Meanwhile, the performance of total weakly-supervised segmentation model was quantitatively evaluated by five indexes including accuracy, precision, specificity, recall, and *F1* score.

III. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed method has been tested on the data set collected from the 35 patients who were admitted to the Zhongshan Hospital Fudan University and have at least one cavity lesion on chest CT scans. A total of 523 CT slice images randomly selected from these pneumonia patients. 248 (47.42%) of them had cavity lesions, while 275 (52.58%) were normal. Meanwhile, 5-fold cross validation experiments were utilized to train and test the proposed method to ensure the reliability of the experimental results.

Deep learning models and image processing algorithms involved in this paper were performed using Pytorch framework and Matlab 2022b, respectively. All experiments were run on a workstation with a CPU: Intel Xeon Gold 6248R, RAM: 256G, and GPU: Tesla V100.

A. RESULTS OF WEAKLY-SUPERVISED SEGMENTATION MODEL

All CT slice images were firstly segmented to obtain the ROI of lung parenchyma. Then the lung parenchyma regions were used to the subsequent training and testing process of the proposed weakly-supervised segmentation model with the hyper-parameters of epoch, batch size, learning rate, optimizer, and loss function of 20, 4, 0.0001, AdamW, and cross entropy loss function, respectively. The accuracy of CSA²-ResNet-based classification and Grad-CAM-based segmentation models were $100\pm 0\%$ and $96.8\pm 2.3\%$, respectively. Figure 5 shows the process of cavity lesion segmentation task. Furthermore, we selected 100 CT slice images with clear lesion edges and invited experienced clinicians to provide the accurate ground truth. Then, five indicators [*i.e.*, Dice coefficient (DC), Jaccard similarity coefficient (JSC), Hausdorff distance (HD), average surface distance (ASD), and conformity coefficient (CC)] were employed to quantitatively evaluate the segmented results, as shown in Table 1. These experimental results demonstrate that the proposed classification model CSA²-ResNet with channel and spatial attention has good feature directionality, and the final segmentation results are precise.

TABLE 1. Performance evaluation of segmented results.

Index	DC	JCS	HD	ASD	CC
Value	0.96	0.93	5.74	0.55	0.92
	\pm	\pm	\pm	\pm	\pm
	0.02	0.04	2.25	0.21	0.05

Note: The data was summarized in the form of mean value \pm standard deviation.

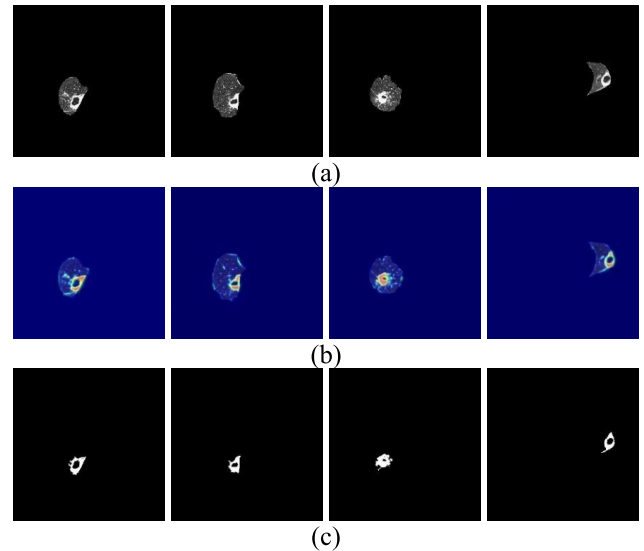


FIGURE 5. Examples of pulmonary cavity lesions and experimental results of the proposed models. (a) Segmented lung parenchyma images; (b) Category activation hot map; (c) Segmented cavity lesion.

To evaluate the entire weakly-supervised segmentation model, five indexes [*i.e.*, accuracy (ACC), precision (PRE), specificity (SPE), recall (REC), and F1-score (F1)] were employed with the results of $98.48\pm 1.09\%$, $96.80\pm 2.28\%$, $97.20\pm 1.97\%$, $100\pm 0\%$, and $98.36\pm 1.17\%$, respectively (Table 2). The results demonstrate that CT images of lung parenchyma with or without cavity lesions all have good analytical results.

TABLE 2. Performance evaluation of total segmentation model.

Index	Value (%)					A \pm SD
	1	2	3	4	5	
ACC	98.10	99.05	98.10	97.14	100	98.48 \pm 1.09
PRE	96.00	98.00	96.00	94.00	100	96.80 \pm 2.28
SPE	96.49	98.21	96.49	94.83	100	97.20 \pm 1.97
REC	100	100	100	100	100	100 \pm 0
F1	97.96	98.99	97.96	96.91	100	98.36 \pm 1.17

Note: The data was summarized in the form of mean value \pm standard deviation.

B. COMPARISON WITH PREVIOUSLY RELATED MODELS

1) COMPARISON WITH ResNet-BASED WEAKLY SUPERVISED SEGMENTATION MODEL

We compared our proposed model with the original ResNet-based model in this task. The segmentation results of cavities, as shown in Figure 6, indicate that the proposed model outperforms the original ResNet-based model in handling details due to its clear directionality of attention mechanism. Meanwhile, the quantitative comparison

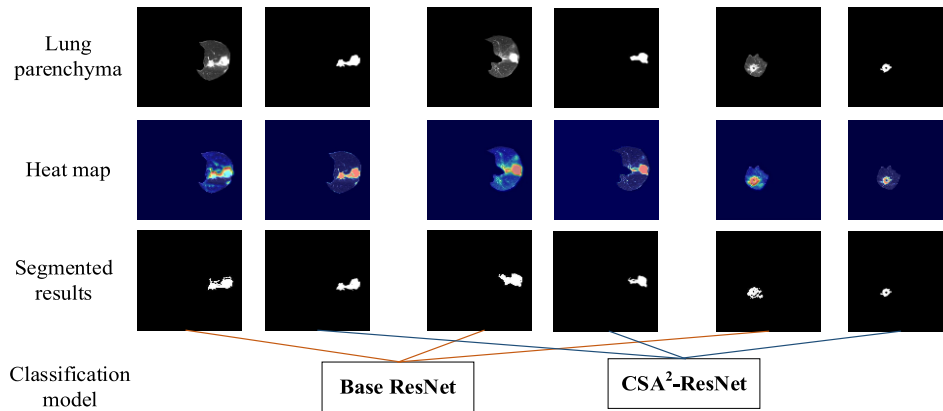


FIGURE 6. Segmentation results of weakly-supervised segmentation models using CSA²-ResNet and base ResNet.

results are shown in Table 3 and 4. For stepwise comparison (*i.e.*, Table 3), the proposed CSA²-ResNet-based segmentation model significantly improved the accuracy improvements of 5.8% and 17.0% at the classification and segmentation stages when compared with base ResNet, respectively (*all P* < 0.001). In full model comparison (*i.e.*, comprehensive evaluation of two-stage models, Table 4), compared with ResNet-based segmentation model, the proposed CSA²-ResNet-based segmentation model performed significantly better regarding all indexes, with the gain of 23.24%, 22.80%, 20.84%, 26.00%, and 24.63% in accuracy, precision, specificity, recall, and F1 score, respectively (*all P* < 0.001).

TABLE 3. Accuracy of CSA²-ResNet and ResNet-based weakly-supervised segmentation models at classification and segmentation Stages.

Task	Base ResNet	CSA ² -ResNet	Gain	<i>P</i> -value
Classification	94.2%	100%	5.8%	<0.001
Segmentation*	79.8%	96.8%	17.0%	<0.001

* Grad-CAM-based segmentation models was used after classification of CSA²-ResNet or ResNet classification.

2) COMPARISON WITH OTHER CNN AND TRANSFORMER-BASED WEAKLY SUPERVISED SEGMENTATION MODEL

The base CSA²-ResNet classification model in the proposed weakly-supervised model adopts the form of “CNN + Attention” to achieve visualization of cavity lesions in CT scans. In order to verify the superiority of our model in the experimental task of this paper, we compared it with CNN-based models (*e.g.*, AlexNet, Vgg-16, GoogLeNet, DenseNet-201, and Inception-V3) and Transformer-based models (*e.g.*, Swin Transformer, Vision Transformer, Tokens-To-Token ViT, Transformer in Transformer, and Twins-PCPVT). All models were trained and testing using the same hyper-parameters, and the segmentation results were obtained by Grad-CAM.

TABLE 4. Comparison of the weakly-supervised segmentation models using CSA²-ResNet and base ResNet.

Index	Backbone model		Gain	<i>P</i> -value
	Base ResNet	CSA ² -ResNet		
ACC	75.24%	98.48%	23.24%	<0.001
PRE	74.00%	96.80%	22.80%	<0.001
SPE	76.36%	97.20%	20.84%	<0.001
REC	74.00%	100%	26.00%	<0.001
F1	74.00%	98.36%	24.36%	<0.001

The performance comparison of two stages (*i.e.*, classification and segmentation) using accuracy and dice coefficient are shown in Figure 7, which demonstrate that our proposed model has the optimal segmentation performance.

C. AUTOMATIC MEASUREMENT AND APPLICATION OF MORPHOLOGICAL CHARACTERISTICS

1) AREA

Based on the precise segmentation results, the specific area and proportion of each lesion relative to the entire lung parenchyma can be obtained through automated calculations, approximately following a normal distribution curve. Simultaneously, the method permits automated and accurate measurement of lesion and inner cavity area, enabling morphological feature analysis using the slice which contains the largest lesion. Taking one patient as an example, CT slices containing lesions were identified from Layers 16 to 23. The area of lesion in each slice and its proportion to the whole lung parenchyma were shown in Figure 8. The largest lesion cross-section was identified in CT slice 21, where the area of the cavity region and its proportion to the entire lesion were determined as 374 pixels and 11.17%, respectively.

In addition, based on the automatic capture of the largest lesion CT cross-section and the quantitative analysis of the

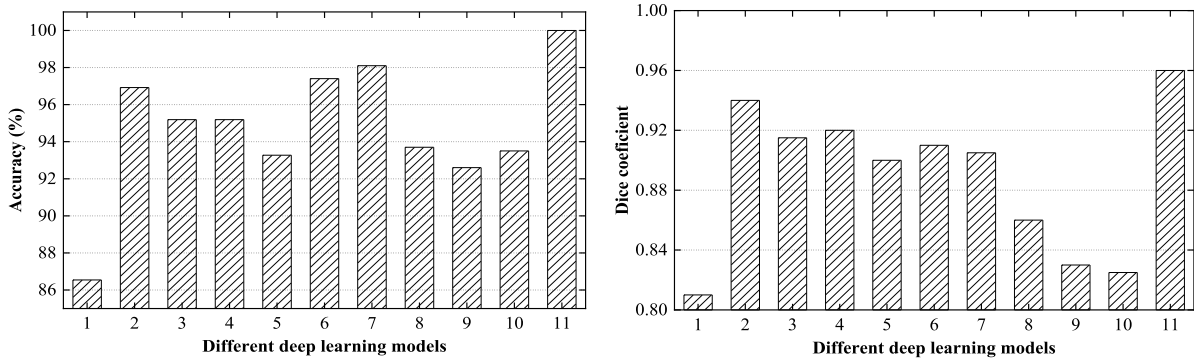


FIGURE 7. Comparison of different models' performance. 1-Swin Transformer, 2-Vision Transformer, 3-Tokens-To-Token ViT, 4-Transformer in Transformer, 5-Twins-PCPVT, 6-AlexNet, 7-Vgg-16, 8-GoogLeNet, 9-DenseNet-201, 10-Inception-V3, 11-Ours.

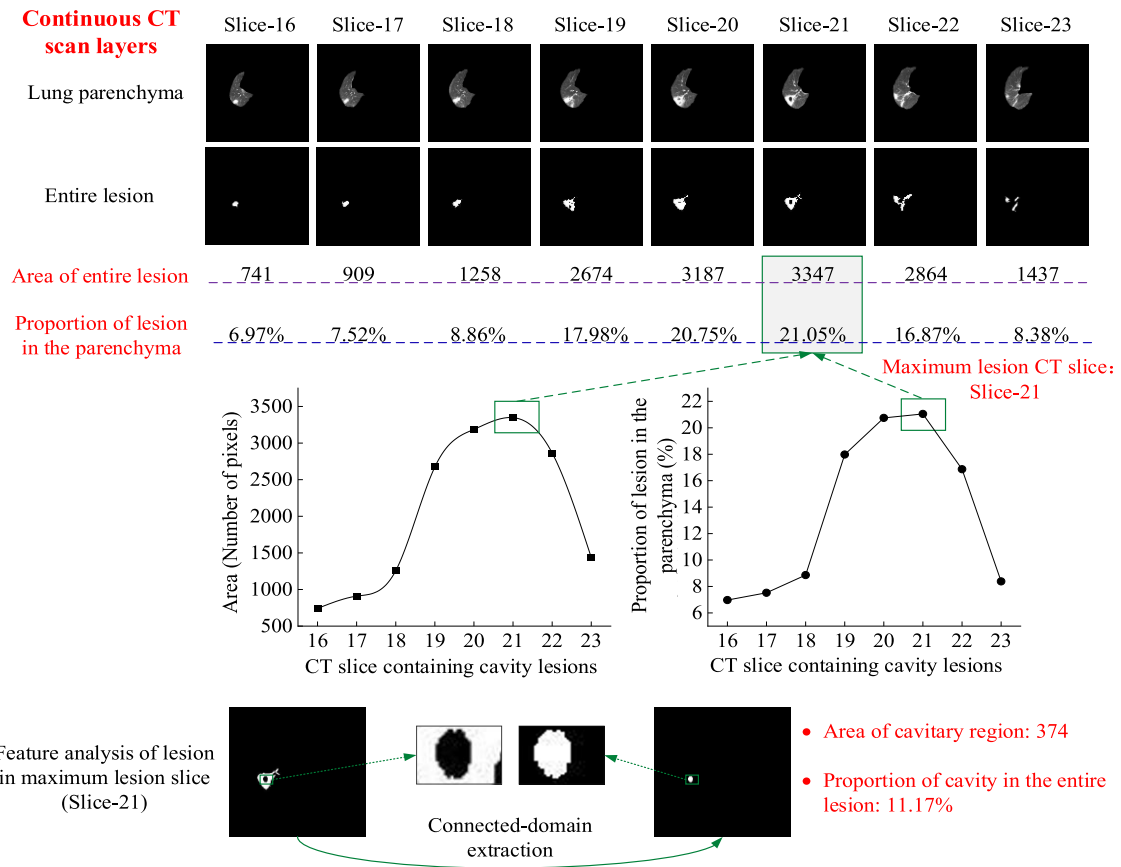


FIGURE 8. An example of segmentation and quantitative analysis of a cavity lesion in a serial of continuous CT slices.

cavity lesion region, dynamic monitoring and comparison of the same lesion in a patient at different treatment time points can be achieved. Taking one patient as an example, the presentation images and change curves of cavity lesions at three different time points (*i.e.*, 2017.08.18, 2017.10.11, and 2017.12.01) are shown in Figure 9. Quantified parameters of the cavity lesions suggest that the area and relative proportion of entire lesion and inner cavitory region significantly decreased after clinical treatment (entire lesion: 4165 pixels (16.72%) to 970 pixels (3.06%); inner cavitory region:

1601 pixels (38.44%) to 38 pixels (3.92%)), indicating improvement in disease condition. These results demonstrate the potential application value of the proposed method for pulmonary cavitory lesion segmentation and quantitative analysis in the clinical diagnosis and follow-up.

2) THICKNESS

The distribution of chest wall thickness is also an important indicator for clinical analysis. Thus, we conducted a quantitative analysis on the wall thickness of cavity lesions

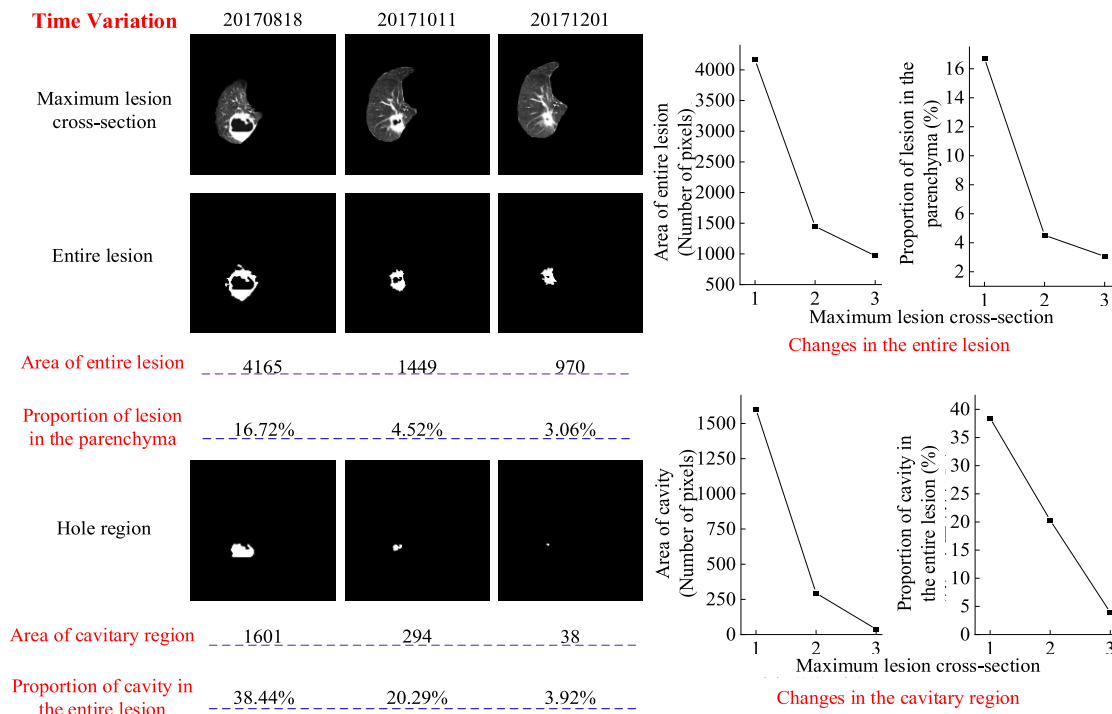


FIGURE 9. An example of analyzing dynamic changes of the same cavity lesion at different time points. 1, 2, and 3 represent three different CT acquisition time points: 20170818, 20171011, and 20171201, respectively.

in CT images. Taking one CT image as an example, we extracted the centroid of the cavity and calculated the thickness of the cavity walls in various directions within a 360-degree range. Subsequently, we performed a histogram-based statistical analysis on the collected thickness data to identify the primary range of lesion wall thickness that the range of wall thickness in this case is mainly concentrated in the 5-6 pixel. Furthermore, we successfully computed key statistical features including maximum, minimum, average, standard deviation, and variance and their values are 19.00 pixels, 2.03 pixels, 7.98 pixels, 14.75 pixel², and 3.84 pixels, providing accurate metrics for characterizing the morphology of cavity lesions. This comprehensive analysis not only enhances our understanding of the wall thickness characteristics of these lesions but also ensures a solid technical foundation for further research in this area. The main process of statistical analysis of cavity wall thickness in CT scans is shown in Figure 10.

IV. DISCUSSION

To facilitate automatic segmentation and quantitative characterization of pulmonary cavity lesions, we designed an effective weakly-supervised deep learning-based algorithm and tested it by five-fold cross validation. The experimental results show the proposed model achieved high accuracy in segmentation task, and thus provide opportunities for fast and precise quantitative characterization of cavity lesions.

The purpose of this study was to develop a model for automated characterization of cavity lesions in CT imaging

which can assist in the diagnosis and treatment of pneumonia diseases. Thus, our primary task in this study is to segment of cavity lesions automatically and precisely. Considering the complex manifestations of cavity lesions (e.g., diverse morphologies and blurry edges), classical supervised segmentation model is limited due to the difficulty in obtaining pixel-level mask labels. In contrast, image- or patient- level labels are often accessible for weakly-supervised segmentation algorithm based on binary classification. Therefore, it is crucial to design a pre-classification model that is suitable for small samples with lightweight model and of high accuracy is needed.

Due to the small sample size and the relatively simple classification task, we adopted the classic convolutional neural network (i.e., ResNet) as backbone and added the hybrid attention modules to it. The hybrid attention helps the classification model to focus on the cavity lesion regions [19]. Among them, spatial attention focuses on the spatial position in the feature map, improving the accuracy of the model in locating the target object in the image. Meanwhile, channel attention focuses on the channel dimensions in the feature map, which helps the model better select and utilize feature channels that are beneficial to the task in the image, and improves overall performance. The effects of hybrid attention were demonstrated in both qualitative and quantitative experimental results (e.g., Figure 6, Table 3, and Table 4), where the proposed model improved classification and segmentation accuracy by 5.8% and 17.0% (all $P < 0.001$), respectively, compared with the original ResNet model.

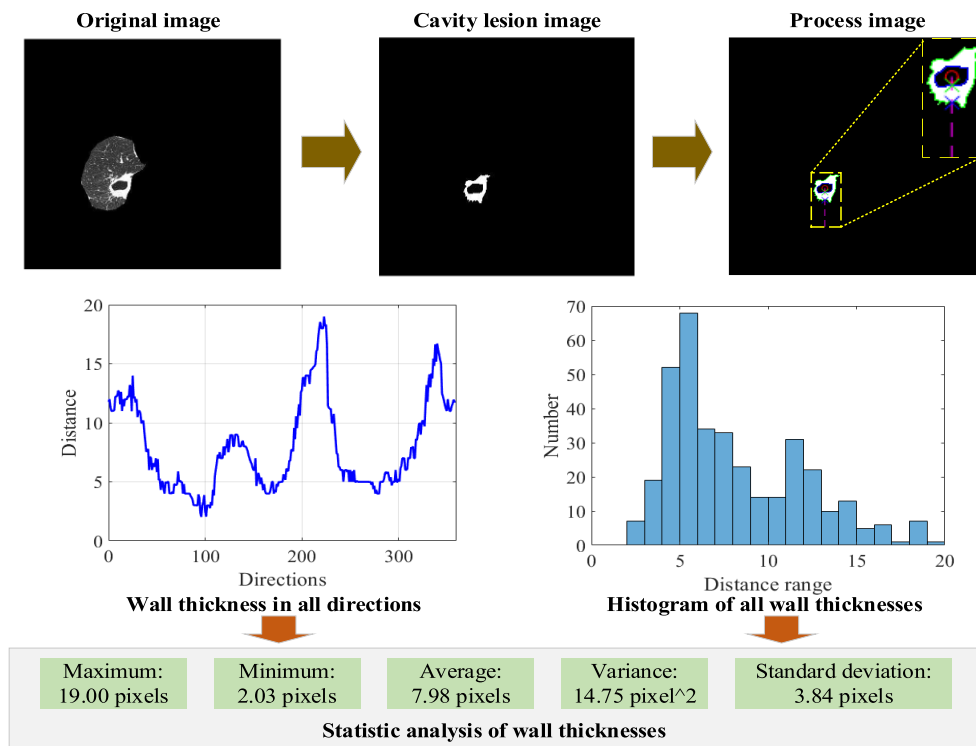


FIGURE 10. An example of statistical analysis of wall thickness of cavity lesions in CT scans.

Of course, there have been many other deep learning models based on CNN or Transformer, most of which achieved good classification performance. Notably, our CSA²-ResNet model outperforms several CNN and Transformer models in preciseness, as shown in Figure 7. The high efficiency of the proposed CSA²-ResNet model is attributable to its simpler structure compared to Transformer models, which is suitable for small sample sizes and simple classification tasks. Additionally, by introducing attention mechanisms into the CNN-based ResNet models, our model also achieved better lesion-focused effects closing to Transformer-based model than other classical CNN-based deep learning models.

In addition to segmentation as the foundation, we achieved capture, analysis, and monitor of the cavity lesion for automatically measuring morphological features (*e.g.*, area and thickness). Owing to the precise segmentation by our model, feasible and intuitive quantitative analysis of single and serial CT scans (*e.g.*, area of entire lesion and single hole region, proportion of lesion in the parenchyma and hole in the entire lesion) is possible as presented in Figure 8 and Figure 9. Meanwhile, by extracting the centroid of the lesion and measuring by rotation, a clear statistical distribution of the thickness of cavity lesion (Figure 10) has also been obtained. The above automated measurement of cavity lesion features reduced errors caused by subjective and objective factors in manual measurement, which has important value for clinical diagnosis and guidance of treatment. Furthermore, the proposed model is also applicable to segment and quantify

other lesion types, such as lung nodules in CT images, thyroid tumors in ultrasound images, brain lesions in MRI images, and so on.

Limitations of This Study: There are still several limitations in this study. First, this paper mainly focuses on capturing and analyzing the section with the largest lesion. The 3D morphology and the correlation between upper and lower layers are also important for diagnosis, demanding for 3D segmentation in future studies. Second, the CT scan dataset used in this study was collected from a single hospital. Due to the potential impacts of different devices and acquisition environments on segmentation results, the generalizability of our model needs further verification in multi-center dataset. Third, our work mainly focused on segmentation task, while more efforts are needed to use morphological and texture features to differentiate the benign/malignant nature and etiology of the lesions in the future.

V. CONCLUSION

In this paper, we proposed a weakly-supervised deep-learning segmentation framework for pulmonary cavity lesions in CT slice images with only image-level labels. It achieved good classification and precise segmentation performance. Meanwhile, we applied the segmentation results to quantitative characterizing cavity lesions, which is essential in automatic assessment and dynamic monitoring. Therefore, our proposed methods have great potential in facilitating accurate diagnosis, observation of clinical course, and evaluation of treatment efficacy.

REFERENCES

- [1] A. P. Parkar and P. Kandiah, "Differential diagnosis of cavitory lung lesions," *J. Belg. Soc. Radiol.*, vol. 100, no. 1, pp. 1–8, Nov. 2016.
- [2] K. H. Mortensen, J. L. Babar, and A. Balan, "Multidetector CT of pulmonary cavitation: Filling in the holes," *Clin. Radiol.*, vol. 70, no. 4, pp. 446–456, Apr. 2015.
- [3] N. R. Kim and J. Han, "Pathologic review of cystic and cavitory lung diseases," *Korean J. Pathol.*, vol. 46, no. 5, pp. 407–414, 2012.
- [4] K. Gafoor et al., "Cavitory lung diseases: A clinical-radiologic algorithmic approach," *Chest*, vol. 153, pp. 1443–1465, Jun. 2018.
- [5] W. Xing et al., "CM-SegNet: A deep learning-based automatic segmentation approach for medical images by combining convolution and multilayer perceptron," *Comput. Biol. Med.*, vol. 147, Aug. 2022, Art. no. 105797.
- [6] C. Zhao et al., "Lung segmentation and automatic detection of COVID-19 using radiomic features from chest CT images," *Pattern Recognit.*, vol. 119, Nov. 2021, Art. no. 108071.
- [7] H. Hu, Q. Li, Y. Zhao, and Y. Zhang, "Parallel deep learning algorithms with hybrid attention mechanism for image segmentation of lung tumors," *IEEE Trans. Ind. Informat.*, vol. 17, no. 4, pp. 2880–2889, Apr. 2021.
- [8] S. Chaganti et al., "Automated quantification of CT patterns associated with COVID-19 from chest CT," *Radiol. Artif. Intell.*, vol. 2, no. 4, Jul. 2020, Art. no. e200048.
- [9] W. Xie, C. Jacobs, J.-P. Charbonnier, and B. van Ginneken, "Dense regression activation maps for lesion segmentation in CT scans of COVID-19 patients," *Med. Image Anal.*, vol. 86, May 2023, Art. no. 102771.
- [10] K. Yuan et al., "A multi-strategy contrastive learning framework for weakly supervised semantic segmentation," *Pattern Recognit.*, vol. 137, May 2023, Art. no. 109298.
- [11] X. Li et al., "CAGAN: Classifier-augmented generative adversarial networks for weakly-supervised COVID-19 lung lesion localisation," *IET Comput. Vis.*, vol. 18, no. 1, pp. 1–14, Feb. 2024.
- [12] Z. Yang, L. Zhao, S. Wu, and C. Y. Chen, "Lung lesion localization of COVID-19 from chest CT image: A novel weakly supervised learning method," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 6, pp. 1864–1872, Jun. 2021.
- [13] W. Sun, X. Feng, J. Liu, and H. Ma, "Weakly supervised segmentation of COVID-19 infection with local lesion coherence on CT images," *Biomed. Signal Process. Control*, vol. 79, Jan. 2023, Art. no. 104099.
- [14] F. Lu et al., "A weakly supervised inpainting-based learning method for lung CT image segmentation," *Pattern Recognit.*, vol. 144, Dec. 2023, Art. no. 109861.
- [15] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.
- [16] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 7794–7803.
- [17] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.
- [18] X. Li, Y. Zhou, P. Du, G. Lang, M. Xu, and W. Wu, "A deep learning system that generates quantitative CT reports for diagnosing pulmonary tuberculosis," *Int. J. Speech Technol.*, vol. 51, no. 6, pp. 4082–4093, Jun. 2021.
- [19] K. Zhang, B. Tang, L. Deng, and X. Liu, "A hybrid attention improved ResNet based fault diagnosis method of wind turbines gearbox," *Measurement*, vol. 179, Jul. 2021, Art. no. 109491.

• • •