Article

DEHA-Net: Dual-Encoder based Hard Attention Network with Adaptive ROI Mechanism for Lung Nodule Segmentation

Muhammad Usman [1,*] and Yeong-Gil Shin [1]

[1] Department of Computer Science and Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, South Korea
* Correspondence: ussman@snu.ac.kr

Abstract: Measuring pulmonary nodules accurately can help with early diagnosis of lung cancer, which can improve a patient's chances of survival. Many methods for segmenting nodules have been developed, but they all rely on input from radiologists in the form of a 3D volume of interest (VOI) or use a constant region of interest (ROI) and only consider the presence of nodule within the given VOI. These approaches limit the networks' ability to detect nodules outside the VOI and can also include unnecessary structures in the VOI, leading to potentially inaccurate segmentation. In this work, we propose a novel approach for 3D lung nodule segmentation by using 2D region of interest (ROI) inputted from radiologist or computer-aided detection (CADe) system. Particularly, we design a dual-encoder-based hard attention network (DEHA-Net) which incorporates the full slice of thoracic computed tomography scan along with the ROI mask to produce an accurate segmentation mask of lung nodule in the given slice. The proposed architecture exploits the adaptive region of interest (A-ROI) algorithm to automatically investigates the penetration of lung nodule into surrounding slices while eliminating the need to drawing separate ROIs in each slice. Further, the framework performs the multi-view analysis, i.e., in sagittal and coronal views, to improve the segmentation performance. The proposed scheme has been rigorously evaluated on the lung image database consortium and image database resource initiative (LIDC/IDRI) dataset and an extensive analysis of results have been performed. The quantitative analysis shows that the proposed method is not only improves the existing state-of-the-art in term of dice score but also, significantly robust against the different types, shape and dimensions of lung nodules.

Keywords: lung nodule segmentation; 3D segmentation; dual-encoder-based CNN; hard attention

1. Introduction

Lung cancer is the most deadly form of cancer, and early detection is crucial for potentially life-saving treatment. Accurate quantification of pulmonary nodules, which may be associated with various conditions but are often indicative of lung cancer, is essential for continuous monitoring of lung nodule volume in order to assess the malignancy and predict the likelihood of lung cancer [1][2]. However, manual segmentation of nodules, a necessary step in calculating their volume, is a laborious and time-consuming process that can also introduce variability between and within observers [3].

Computer-aided diagnosis (CAD) systems can greatly enhance the productivity of radiologists by assisting them to overcome the challenges associated with the manual segmentation of pulmonary nodules. CAD system consists of two subsystems, i.e., computer-aided detection (CADe) and computer-aided diagnosis (CADx). The CADe component aims to distinguish between nodules and other structures, such as tissues and blood vessels. The CADx component then evaluates the detected nodules and determines whether they are benign or malignant tumors. The primary goal of these CAD systems is to improve the accuracy and efficiency of cancer diagnosis by radiologists. They are designed to assist in decision making by providing additional information and reducing the time needed to interpret CT images. This work is focused to designed the CADx system for accurate

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2022 doi:10.20944/preprints202212.0532.v1

segmentation of lung nodule which is a challenging task due to variable shapes, different sizes, and complicated surrounding tissues in the lung region. To address this issue, various automatic segmentation frameworks for nodule quantification have been devised, which incorporates the traditional image processing-based methods as well as the deep learning-based approaches [4]. However, the significant heterogeneity of lung nodules, particularly, the shape and contrast variations of lung nodules hinders the development of robust nodule segmentation framework. These variations, both within and between nodules, as well as the visual similarity between nodules and their surrounding non-nodule tissue, necessitate the use of a 3-D volume of interest (VOI) as input in order to accurately estimate the shape of the nodule. Figure 1 demonstrates the intra-nodule and inter-nodule variations, showcasing the diversity between the shapes of different nodules and the variations present within individual nodules. Providing a 3-D VOI is quite time consuming and laborious job as radiologist has to specify the region of interest into each slice containing the nodule. A few studies have resolved this issue by utilizing the fix ROI for all the slices, this approach requires only one ROI input from user which significantly reduces the time and hassle; however, employing fix ROI adds redundant non-nodular region in the input ROI which leads to the poor segmentation performance.

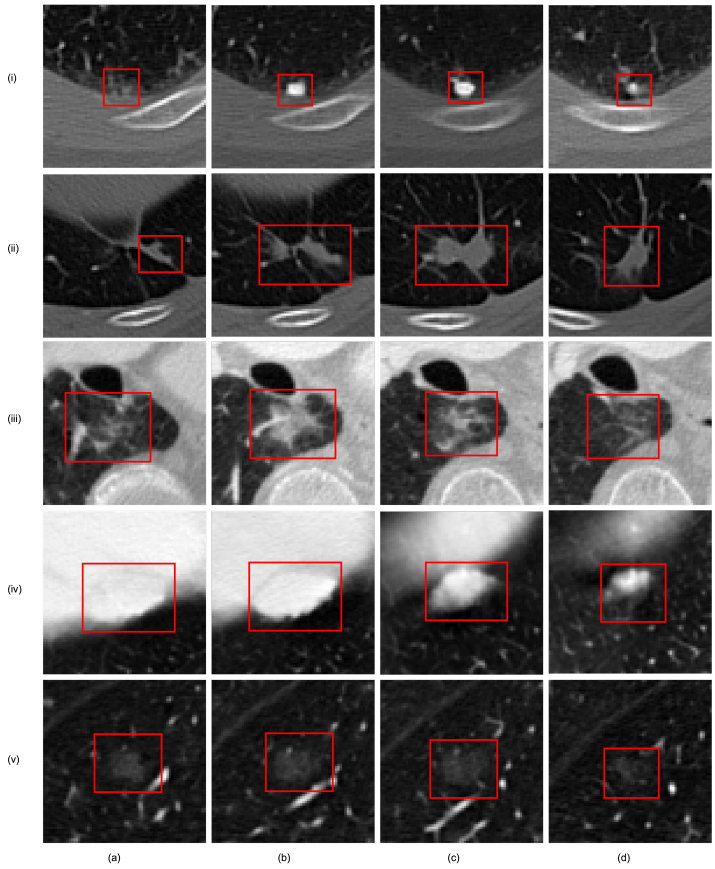

Figure 1. Illustration different type of the lung nodule. The intra-nodule diversity can be noticed from column (a) to (e), whereas inter-nodule variation is demonstrated from the row (i) to (v).

To address the issues related to the use of 3-D VOIs as input and fix ROIs, in our previously work [5], we proposed a novel approach of dynamic ROIs for accurate volumetric segmentation of pulmonary nodules. To determine the dynamic ROIs, we proposed adaptive region of interest (A-ROI) algorithm that utilizes a single 2-D ROI provided by radiologists [5] to estimate the dynamic ROIs in the surrounding slices. This approach begins by segmenting the nodule in the provided ROI first by employing residual-UNet, and then utilizes the segmentation mask to determine the ROIs for the surrounding slices

 doi:10.20944/preprints202212.0532.v1

to further extend the nodule segmentation in both directions. Concretely, the A-ROI algorithm dynamically adjust the position and size of bounding box for the adjacent slices to investigate the penetration of nodule in the other slices. The technique demonstrated exceptional performance and outperformed the previous state-of-the-art method. However, the previous approach required cropping the ROI, which can cause problems if the ROIs are too smaller or larger than the normalized dimensions which are used to input into the network. Similarly, the mask obtained after inference must be resized to match the original cropped ROI size, introducing error when interpolation is used to obtain the target dimensions. To address these issues, we propose a dual-encoder-based architecture that takes two inputs: the original slices and the ROI mask, eliminating the need to rescale the ROIs before and after inference. The A-ROI algorithm is then used to further produce ROI masks for surrounding slices for which ROIs are not provided. Specifically, the A-ROI algorithm is applied along the axial plane to provide an initial estimation of nodule shape, which is then used to automatically extract a 3-D VOI from the scan. The extracted VOI is further utilized to create coronal and sagittal views of the nodule, and slices from these views are analyzed using two separate dual-encoder-based architectures. Finally a consensus module is employed to ensemble the three predictions obtained from axial, coronal and sagittal view models. A number of experiements have been performed on LIDC dataset [6] to demonstrate the effectiveness of proposed technique in terms of overall performance and robustness to the type and size variations of lung nodules.

2. Related work

Accurate assessment of lung nodules is essential for evaluating their potential malignancy and likelihood of being indicative of lung cancer. Subsequently, an extensive efforts have been made by numerous researchers to devise an efficient nodule segmentation framework to assist the radiologists. These studies can be classified into two categories, i.e., conventional image processing-based methods and advanced deep learning-based techniques [4].

Jamshid *et al.* [7] proposed a framework which segmented the nodule by employing region growing techniques, such contrast-based region growing and fuzzy connectivity region growing, within a volumetric mask created using a local adaptive segmentation algorithm that distinguishes between foreground and background regions within a specified window size. While the algorithm demonstrated good performance for isolated nodules, it was unable to effectively segment attached nodules. By utilizing geodesic impact zones in a multi-threshold picture representation, Stefano *et al.* [8] offer a user-interactive algorithm that meets the fusion-segregation criterion based on both gray-level similarity and object shape. They extending their work in [9], by eliminating the need for user interaction. Correction is now performed based on a 3D local shape analysis, allowing for the refinement of an initial nodule segmentation to distinguish possible vessels from the nodule itself, without requiring any input from the user. Rendon *et al.* [10] used morphological and threshold approaches to get rid of extraneous structures from the ROI that was given. The last step was to use a support vector machine (SVM) to categorize each pixel in the discovered space.

Classical image processing techniques have had difficulty achieving accurate and robust volumetric nodule segmentation. In contrast, recent deep learning-based methods have made wast inroads into many medical imaging applications such as image segmentation applications [11,12], including lung nodule segmentation[13]. The introduction of the U-NET [14] architecture for medical image segmentation, in particular, has greatly enhanced performance in this field. As a result, there has been an increased focus on using deep learning for lung nodule segmentation. In [15], Tyagi et al. proposed 3D conditional generative adversarial network (GAN) for lung nodule segmentation. They utilized the UNet architecture as the backbone of GAN. Their employed simple classification network as discriminator which incorporated spatial squeeze and channel excitation module to differentiating between truth and fake segmentation. Similarly, Wang *et al.* [16] devel-

 doi:10.20944/preprints202212.0532.v1

oped a method for nodule segmentation called Central Focused Convolutional Neural Networks (CF-CNN). This approach involves using a volumetric patch centered around the voxel of interest as input to the model. In addition, the team [17] also published a multi-view CNN that can perform nodule segmentation using input from different views (axial, coronal, and sagittal) of the same voxel. One potential limitation of this method is that the patch extraction process is the same for all nodules, which could lead to incorrect segmentation if the nodule is larger than the size of the patch. By using skip connections in the encoder and decoder paths, Tong et al. enhanced the performance of U-NET for nodule segmentation, however the model was only intended for 2D segmentation. By using skip connections in the encoder and decoder paths, Tong *et al.* [18] enhanced the performance of U-NET for nodule segmentation, however the model was only intended for 2D segmentation. Hancock *et al.* [19] put forth a variation on the standard level set picture segmentation technique in which, as opposed to being manually created, the velocity function is instead learnt from data using regression machine learning techniques. They reported slightly improved performance when they applied this segmentation approach to the segmentation of lung nodules. Chen et al. [20] proposed end-to-end multi-task learning framework which consists of joint classification and multi-channel segmentation networks. Both networks utilized the same latent representation learned by the common encoder branch which improved the lung segmentation performance. The study also incorporated enhanced version of patches by using OTSU and SLIC methods. To extract local characteristics and detailed contextual information from lung nodules, Liu *et al.* [21] used a residual block-based dual-path network, which significantly improved performance. They also employed a fixed VOI, which restricts the search for the nodule and hence lowers 3D segmentation performance. To avoid this issue, Chen et al. [22] proposed fast multi-crop guided attention (FMGA) network for lung nodule segmentation by incorporating 2D and 3D cropped ROIs. They applied the greedy search algorithm to explore the penetration of lung nodule into surrounding slices. Their framework also exploited a customised loss function which enable the network to focus on improve the segmentation of nodule borders. The results demonstrated the robustness of proposed framework; however, the scheme failed to improve the state-of-the-art in term of overall dice score.

In our previous work [5], we addressed the limitations of a fixed volume of interest (VOI) by introducing the concept of an adaptive 2D region of interest (ROI) in each slice, which significantly improves the ability to utilize deep learning. Most notably, cropped ROIs were fed to the deep residual U-Net [23], which demonstrated promising performance along with several limitations. Particularly, due to the heterogeneity of lung nodules, numerous variations in dimensions are possible, which makes it impossible to find the optimal input dimensions for the network. Subsequently, the cropped ROI has to be severally resized, either by up-sampling or down-sampling the ROI, which affects the performance of the proposed framework. One possible alternative is to train various models with different input dimensions. However, this comes with an immense increase in the computational cost, which hinders the solution's real-time clinical applications. For instance, Zhang et al. [24] proposed multi-scale segmentation squeeze-and-excitation UNet with a conditional random field to segment the nodule in the given volume of interest. They extract VOIs at four different scales and trained four different networks and finally applied conditional random field to merge the four predictions. Their framework increased the computational complexity and only covers four scales defined according to dimensions available in the given dataset which is insufficient to cover the possible diversity of size of lung nodule in real-time clinical applications. To overcome this issue, in this work, we proposed dual-encoder-based architecture, which incorporates the ROI mask to input as hard attention, which enables the framework to avoid the pre- and post-inference resizing and leads to performance improvement.

 doi:10.20944/preprints202212.0532.v1

3. Materials and Methods

3.1. Dataset

In this work, we used he Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) Database [6,25], which is the largest publicly available resource of lung CT scans. which is developed to facilitate the development of computer-aided systems for evaluating lung nodule identification, categorization, and quantification, the LIDC-IDRI has gathered a sizable collection of thoracic CT scans. It comprises 1,018 diagnostics and screening thoracic CT images for lung cancer from 1,010 individuals with annotated lesions. Each thoracic CT scan has undergone a two-phase annotation process which is performed by four qualified radiologists. As in earlier research [26], [27], [5], we only considered nodules with a minimum diameter of 3 mm and annotations from all four radiologists. The ground truth border for pulmonary nodule segmentation was created using a 50% consensus criterion [28] due to the variability among the four radiologists, and a Python module named pyLIDC was employed. Total of 893 nodules from LIDC dataset are selected and randomly distributed 40%, 5% and 55% set which are used as training, validation and test set, respectively.

3.2. Data Preprocessing

The LIDC collection includes scans obtained from numerous locations and scanners. Consequently, it has a variety of pixel spacing and slice thickness. These variables are crucial for nodule appearance. In particular, slice thickness has a significant impact on the coronal and sagittal views. Slice thickness in the majority of LIDC scans, which spans from 0.45 mm to 5.0 mm, is greater than pixel spacing. Therefore, to enhance the visibility of nodules in all three views, the slice thickness was reduced to corresponding pixel spacing of scan by up-sampling the scan in z-axis. The pixel spacing has remained unchanged as it was less than one for each scan, producing an axial view of the nodules in reasonable resolution.

We also normalized the intensity values, ranging from 0 to 1, by using the window center and window width tag from corresponding DICOM files. The normalization can be defined as follows:

$$I_n = \frac{I - W_{Min}}{W_{Max} - W_{Min}}, \tag{1}$$

$$W_{Min} = WC - WW/2, \tag{2}$$

$$W_{Max} = WC + WW/2, \tag{3}$$

where, I, I_n, WC and WW represent the original, normalized image, window center and window width, respectively.

In contract to the studies [16–18,29], that produced the training samples by employing the constant margin scheme, in this work utilized ROI with random margins on each side as in [5]. To train our Dual-encoder-based architecture, we generated the ROI masks by using the ground truth nodule masks. To enforce our model to learn about the absence of lung nodule in the given slice, we also included two non-nodular slices from both sides of nodule.

3.3. Dual-Encoder based Hard Attention Network with adaptive ROI Mechanisms

The proposed framework utilized the novel Dual-Encoder based Network with Hard-Attention (*DEHA*) with adaptive ROI (A-ROI) Mechanisms. The overall framework has been illustrated in Figure 2. First stage, the 2D ROI, which can be source from manual input from radiologist or computer aided diagnosis (CADe) system, is provided to Dual-Encoder based Hard Attention Network (*DEHA*-Net) along the axial axis. It applies the A-ROI algorithm to generate the ROIs for remaining surrounding slices which enables the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2022 doi:10.20944/preprints202212.0532.v1

investigation of nodule along axial view to reconstruct the 3D mask of nodule. At second stage, the 3D mask constructed after the axial analysis is exploited to generate the ROIs along the sagittal and coronal views. Then we apply the proposed *DEHA*-Net along sagittal and coronal view with predefined ROIs generated from the 3D mask obtained at first stage. Finally, a consensus module has been utilized to produce the final 3D segmentation mask of nodule. It is important to note that in the whole pipeline, there is no resizing module which eliminates the issues associated to re-scaling of given input and network output. This enables our network to achieve improved performance and make it more robust to size variations of various types of nodules. The following subsection describe the details of the proposed *DEHA*-Net and A-ROI algorithm.

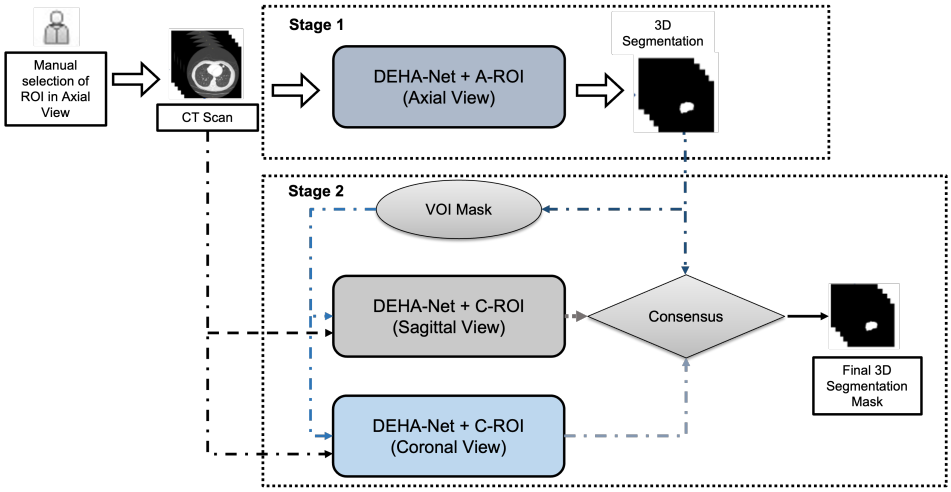

Figure 2. The proposed framework consists of two stages. In the first stage, the user or CADe system provides the ROI along the axial axis, and the *DEHA*-Net (dual-encoder-based convolutional neural network with self-hard attention) and adaptive ROI algorithm are used determine the ROIs in the surrounding slices to perform the 3D segmentation. At second stage, the sagittal and coronal views are created to segment the nodule and finally three segmentation predictions are fed into consensus module to produce the final 3D segmentation mask.

3.3.1. Dual-Encoder based Hard-Attention Network

Lung nodules vary in shape and dimensions, making impossible to set a suitable input dimension for network. To overcome this issue, we designed dual-encoder-based hard attention network (*DEHA*-Net) which incorporates two inputs, i.e., slice containing the nodule and ROI provided mask, to accurately segment the nodule in the given slice. Particularly, ROI mask provides the hard attention which enable network to focus on only the provided region of interest. The proposed *DEHA*-Net consists of two encoders and one decoder branch as demonstrated in Figure 3. Each encoder is connected to decoder with residual connections originated from four different levels. Both encoders are having identical architecture, consisting of four levels. At *nth* level, there is a convolution layer of $32 \times n^2$ filters and kernel size of 3x3 followed by rectified linear activation to add non linearity. After relu there is a batch normalisation layer and then a max-pooling layer which compresses the information. These five layers make a single level of encoder.

First encoder extracts features from images of the CT scan while the second encoder enforces the hard attention learned from the ROI mask of the nodule. Its main purpose is to focus the network towards the location of the nodule and segment it. Decoders output segmentation mask of nodule for current mask and ROI for next and previous slice. Similarly, the decoder consists of four levels and each level consists of concatenation layer followed by an convolution layer. After that rectified linear is applied for non-linearity followed by batch normalisation layer and finally these features are upsampled. In the last level of the decoder, the upsampling layer is replaced by a convolution layer of a single

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2022 doi:10.20944/preprints202212.0532.v1

filter with softmax activation. Each concatenation layer of the decoder concatenates features from each level of both encoders and previous decoder level to pass into the proceeding layers.

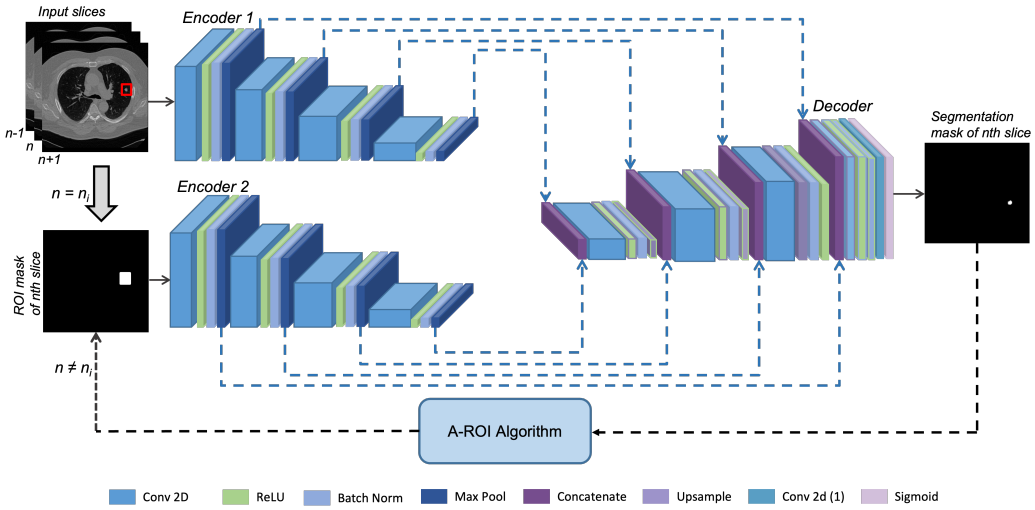

Figure 3. Illustration of proposed dual-encoder-based hard attention network (DEHA-Net) architecture.

3.4. Adaptive ROI Algorithm

The adaptive ROI (A-ROI) algorithm was proposed in [5] which enables the network to investigate nodule presence in the surrounding slices without having ROIs from user. Concretely, A-ROI algorithm exploits the segmented mask of nodule in current (*nth*) slice generated by the network to estimate the ROI for the next slices (i.e., $n \pm 1$). In this work, we employed A-ROI algorithm to complement the proposed *DEHA*-Net to perform the 3D segmentation of lung nodule. A-ROI utilizes a hyper parameter $R_T \in (0, 1)$ to moderate the margins around the nodule in the generated ROI masks.

The full impact of the A-ROI algorithm has been demonstrated in Figure 4. Two constant ROIs have been shown in first row in red and blue which remain fixed throughout all the slices, one with tight margins which failed to cover the nodule in the surrounding slices while other constant ROI is with wider margins that adds redundant area which confuses the network. Whereas in the second row dynamic ROI, produced by A-ROI algorithm, has been shown. The column shows the different slices, (a) represent the slice where first ROI is provided by user and (b)-(f) demonstrate the adjacent slices.

The proposed framework of generating the 3D segmentation mask of lung nodule along the axial view has been describe in algorithm 1. The algorithm 1 illustrated the steps followed to generate the 3D segmentation mask by investigating nodule penetration along axial axis. The provided ROI by radiologist or CADe system in $n_i th$ slice is represented by RoI_{n_i} is used as RoI_n initiate the segmentation. The normalized slice, I_n and provided ROI is fed to *DEHA*-Net which is denoted by Θ. Later the segmentation mask of nodule generated by *DEHA*-Net is inputted to A-ROI algorithm to produce the ROI mask for the next slice. The next slice can be in any direction, i.e., forward or backward. The same cycle is repeated until the next ROI mask becomes blank.

3.5. Ensembling Mechanism

The propose framework utilizes the consensus module to ensemble the segmentation results obtained from axial, sagittal and coronal axis. The consensus value, E_i, of i_{th} voxel is calculated as follows:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2022 doi:10.20944/preprints202212.0532.v1

Algorithm 1 : The algorithmic steps followed in propose framework for nodule investigation along axial view.

$n = n_i, RoI_n = RoI_{n_i}$
while $\sum RoI_n > 0$ **do**
    $Seg_n = \Theta(I_n, RoI_n)$
    $n \leftarrow n \pm 1$
    $RoI_n \leftarrow AROI(Seg_n, R_T)$
end while

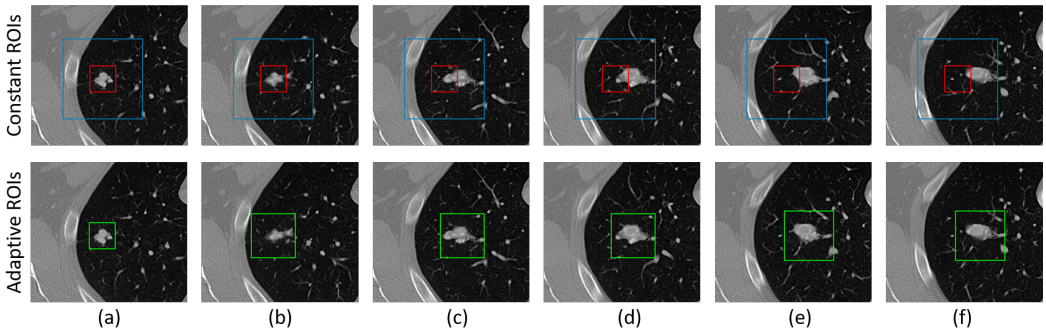

Figure 4. Constant and adaptive regions of interest (ROIs) have been depicted in a series of slices containing a nodule. The blue and red buonding boxes represent constant ROIs, while the green boxes represent adaptive ROIs. (Figure Credit: Usman *et al.*[5])

$$E_i = thr\left(\sum_{j=1}^{K} S_{ij}, \tau\right) \tag{4}$$

$$thr(g, \tau) = \begin{cases} 1, & if\ g \geq \tau \\ 0, & Otherwise \end{cases} \tag{5}$$

where, S_{ij} represents the prediction of ith from jth model and K denotes the number of models which in our case are three, i.e., axial, sagittal and coronal. τ is the threshold which is determined on validation set.

4. Experimental Setup and Implementation Details

4.1. Loss function

To train the proposed *DEHA*-Net, we utilized the dice similarity coefficient (DSC) [30] loss which can be defined as follows:

$$\mathcal{L}_{\text{DSC}} = \frac{1}{N}\sum_{i=1}^{N}\left[1 - \frac{2 * \Theta(I_i, RoI_i) \cap S_{g_i}}{\Theta(I_i, RoI_i) + S_{g_i}}\right] \tag{6}$$

Where, Θ, S_{g_i} and N represent the model, groud truth segmentation mask and number of samples in the training set, respectively. We use the stochastic gradient descent (SGD) to train our network.

4.2. Implementation Details and Training Strategy

We Keras [31] framework for implementing the proposed DEHA-Net and emplyed Equation 6 by SGD scheme to minimize the error. Model is trained on Nvidia Tesla V100 Tensor core gpu with 12,821 images sized 512×512 and batch size of 8. Training was initiated from random weights and with We begin training the model with random weights and with an initial learning rate of 0.001 and first and second momentum of 0.9 for the decay of the learning rate. We use early stopping with patience of 10 epochs to avoid overfitting.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2022 doi:10.20944/preprints202212.0532.v1

Table 1. The quantitative results of our proposed scheme along with previously published studies in terms of mean ± standard deviation of all quantitative measures used in this study. The best performance indicated in bold.

Authors, year	DSC (%)	SEN (%)	PPV (%)
Liu *et al.*, 2019 [21]	81.58 ± 11.05	87.30 ± 14.30	79.71 ± 13.59
Cao *et al.* 2020 [33]	82.74 ± 10.20	89.35 ± 11.79	79.64 ± 13.34
Usman *et al.*, 2020 [5]	87.55 ± 10.58	91.62 ± 8.47	88.24 ± 9.52
Maqsood *et al.*, 2021 [34]	81 ± -	-	-
Zhang *et al.*, 2022 [24]	85.1 ± 7.10	82.7 ± 10.8	**90 ± 10.7**
Tyagi *et al.*, 2022 [15]	80.74 ± -	85.46 ± -	80.56 ± -
Chen *et al.*, 2022 [22]	81.32 ± -	**92.33 ± -**	74.78 ± -
Zhou *et al.* 2022 [35]	86.75 ± 10.58	89.07 ± 8.31	83.26 ± 10.21
Our Method 2022	**87.91 ± 6.27**	90.84 ± 8.22	89.56 ± 10.07

4.3. Performance Measures

We three evaluation parameters to rigorously evaluate the performance of proposed framework. We use following evaluation parameters to evaluate the perform of our proposed method.

- **Dice Similarity Coefficient:** We use the dice similarity coefficient (DSC) [16,32] which measures the degree of overlap between the ground truth mask and predicted mask. The DSC values ranges from 0 to 1, while 1 and 0 indicate complete overlap and no overlap, respectively. It can be defined as follows:

$$DSC = \frac{2 * Y' \cap Y}{Y' \cup Y} \tag{7}$$

  Where Y' and Y are the predicted segmentation mask and reference segment mask, respectively.
- **Sensitivity:** To measure the pixel classification performance proposed framework, we used the sensitivity (SEN) which can be defined as follows:

$$SEN = \frac{Y' \cap Y}{Y} \tag{8}$$

- **Positive Predictive Value (PPV):** To measure the correctness of the segmentation area produced by the proposed framework, we used positive predictive value (PPV) which can defined as follows:

$$PPV = \frac{Y' \cap Y}{Y'} \tag{9}$$

5. Results and Discussion

5.1. Overall Performance Analysis

The proposed framework has been evaluated on all the parameters described in Section 4.3 to analyze and compare the overall performance with previously published studies. Table 1 summarises the results achieved by our framework on test set along with the reported performance of existing studies. It demonstrates that the proposed framework outperforms the existing methods in term of dice score and most importantly, our technique is having lowest standard deviation which depicts its robustness against the various types and size variations of lung nodules. Particularly, comparison with our previous work [5] where we utilized the cropped ROI input, this approach offers an improved performance with lower standard deviation. Such excellent performance can

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2022 doi:10.20944/preprints202212.0532.v1

Table 2. Mean dice score for various type of nodules from the LIDC-IDRI testing set.

Characteristics	Characteristics score					
	1	2	3	4	5	6
Calcification	-	-	85.99 [18]	91.25 [42]	85.98 [27]	87.77 [405]
Internal structure	87.98 [487]	78.04 [3]	-	84.13 [2]	-	-
Lobulation	91.07 [201]	86.09 [164]	84.79 [78]	85.08 [31]	87.54 [18]	-
Malignancy	89.18 [39]	87.76 [114]	79.45 [163]	89.14 [98]	91.02 [78]	-
Margin	92.08 [9]	89.81 [37]	79.25 [78]	82.99 [232]	92.97 [136]	-
Sphericity		88.77 [38]	83.22 [153]	91.61 [218]	90.24 [83]	-
Speculation	92.42 [257]	82.69 [165]	85.17 [32]	80.39 [14]	83.56 [24]	-
Subtlety	80.3 [4]	88.96 [22]	82.88 [131]	91.99 [238]	86.03 [97]	-
Texture	80.47 [11]	85.73 [18]	87.1 [26]	82.27 [107]	90.17 [330]	-

be attributed to the incorporation of ROI mask into our dual-encoder based architecture which eliminates the necessity to crop and normalize the input slice. It also shows the effectiveness of A-ROI algorithm which has been incorporated in proposed scheme to estimate the ROIs masks for the surrounding slices of given input slice.

5.2. Robustness Analysis

The LIDC dataset includes annotations that describe various characteristics of nodules, such as their subtlety, internal structure, calcification, sphericity, margin, lobulation, speculation, texture, and malignancy. These characteristics represent different levels of difficulty in detecting the boundaries of nodules. To evaluate the effectiveness of our method, we divided the test data into groups based on each characteristic and analyzed the results for each group. Table 2 presents the dice scores for each group, which demonstrates that our framework performs quite similar on each group and promising results can be obtained on all types of lung nodules. This can be attributed to the hard attention which enables the proposed DEHA-Net to focus on only given ROI region while keeping the surrounding information which better help to distinguish the nodule.

Further to illustrate the robustness of proposed method, the histogram of the distribution of dice scores on the test set of the LIDC dataset has been shown in Figure 5. The majority of test instances have a score of over 85%, which demonstrates the strong performance of our proposed method.

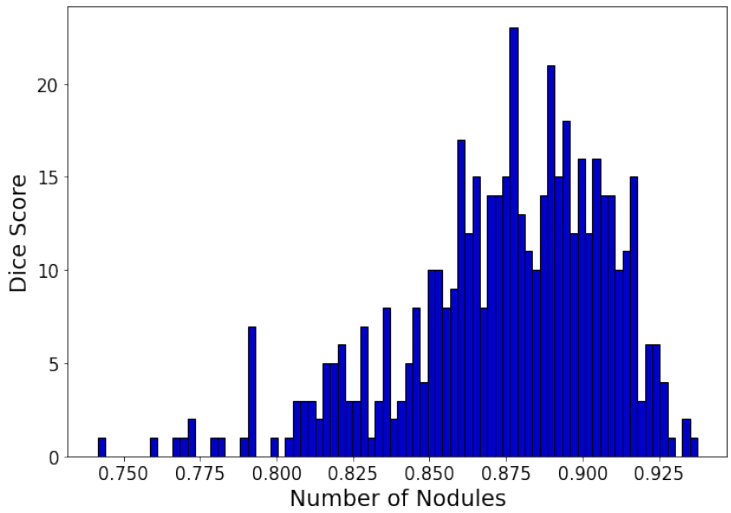

Figure 5. Dice similarity score distribution obtained on the LIDC testing set.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2022 doi:10.20944/preprints202212.0532.v1

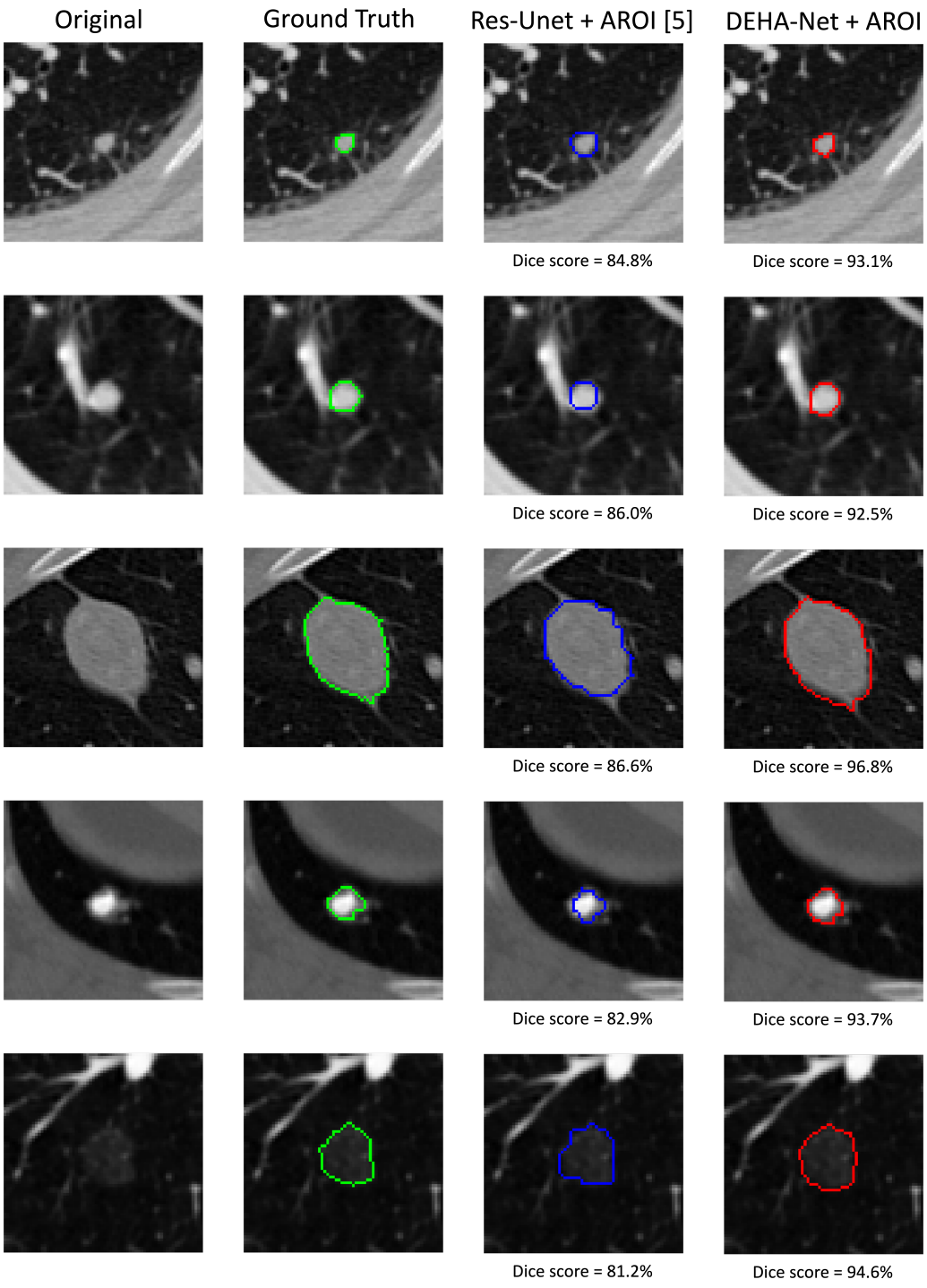

Figure 6. The visual results comparison of previous cropped slices-based approach with Res-UNet and full slice input-based approach with DEHA-Net.

5.3. Qualitative Analysis

To elaborate the difference in the performance of this work and previously published work in [5], we perform the visual analysis of results. Figure 6 shows the visual results with axial views on randomly selected nodules of difference size and types. The results demonstrate that the incorporation of hard attention by the ROI mask in the model has significantly improved the segmentation performance. It can also be noticed that the resizing of cropped slice disturb the boundary of segmented nodule which is crucial to

determine the exact dimensions of nodule to calculate the malignancy level. The proposed DEHA-Net enable our framework to utilize the full slice without severally losing the minor details of from the given input image which are crucial to perform the accurate segmentation of lung nodule.

6. Conclusions

In this work, we propose a dual-encoder-based hard attention network (DEHA-Net) which utilizes the adaptive region of interest (A-ROI) algorithm for accurate 3D lung nodule segmentation. The proposed framework takes 2D slice along with the region of interest, which covers the nodule area, to produce the 3D segmentation of the nodule. Concretely, the proposed model, with the help of A-ROI algorithm, automatically investigates the presence of nodules in the surrounding slices, which eliminates the requirement of providing 3D volume of interest by the radiologist. Most importantly, in contrast to the previous studies in which cropped patch of the given slice is inputted to the network, the proposed DEHA-Net exploits the entire slice to learn the meaningful feature, which assists to better distinguishing between the nodule and non-nodular voxels. An extensive evalaution of proposed framework has been performed on the lung image database consortium and image database resource initiative (LIDC/IDRI) dataset, which is the largest publicly available dataset. The quantitative and qualitative results are presented and analyzed, which demonstrate that the technique shows excellent performance by outperforming the existing state-of-the-art method in terms of dice similarity score. Furthermore, results depict that the framework is significantly robust to the various types and sizes of nodules. Future plans include the improvement of the framework by reducing its computational complexity to further optimize its performance in terms of execution time.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "Conceptualization, X.X. and Y.Y.; methodology, X.X.; software, X.X.; validation, X.X., Y.Y. and Z.Z.; formal analysis, X.X.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing—original draft preparation, X.X.; writing—review and editing, X.X.; visualization, X.X.; supervision, X.X.; project administration, X.X.; funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript.", please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: Not applicable.

Institutional Review Board Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mozley, P.D.; Bendtsen, C.; Zhao, B.; Schwartz, L.H.; Thorn, M.; Rong, Y.; Zhang, L.; Perrone, A.; Korn, R.; Buckler, A.J. Measurement of tumor volumes improves RECIST-based response assessments in advanced lung cancer. *Translational oncology* **2012**, *5*, 19.
2. Devaraj, A.; van Ginneken, B.; Nair, A.; Baldwin, D. Use of volumetry for lung nodule management: theory and practice. *Radiology* **2017**, *284*, 630–644.
3. Moltz, J.H.; Bornemann, L.; Kuhnigk, J.M.; Dicken, V.; Peitgen, E.; Meier, S.; Bolte, H.; Fabel, M.; Bauknecht, H.C.; Hittinger, M.; et al. Advanced segmentation techniques for lung nodules, liver metastases, and enlarged lymph nodes in CT scans. *IEEE Journal of selected topics in signal processing* **2009**, *3*, 122–134.
4. Wu, J.; Qian, T. A survey of pulmonary nodule detection, segmentation and classification in computed tomography with deep learning techniques. *Journal of Medical Artificial Intelligence* **2019**, *2*.
5. Usman, M.; Lee, B.D.; Byon, S.S.; Kim, S.H.; Lee, B.i.; Shin, Y.G. Volumetric lung nodule segmentation using adaptive roi with multi-view residual learning. *Scientific Reports* **2020**, *10*, 1–15.
6. Armato III, S.G.; 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. The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Medical physics* **2011**, *38*, 915–931.
7. Dehmeshki, J.; Amin, H.; Valdivieso, M.; Ye, X. Segmentation of pulmonary nodules in thoracic CT scans: a region growing approach. *IEEE transactions on medical imaging* **2008**, *27*, 467–480.

8. Diciotti, S.; Picozzi, G.; Falchini, M.; Mascalchi, M.; Villari, N.; Valli, G. 3-D segmentation algorithm of small lung nodules in spiral CT images. *IEEE transactions on Information Technology in Biomedicine* **2008**, *12*, 7–19.
9. Diciotti, S.; Lombardo, S.; Falchini, M.; Picozzi, G.; Mascalchi, M. Automated segmentation refinement of small lung nodules in CT scans by local shape analysis. *IEEE Transactions on Biomedical Engineering* **2011**, *58*, 3418–3428.
10. Rendon-Gonzalez, E.; Ponomaryov, V. Automatic Lung nodule segmentation and classification in CT images based on SVM. In Proceedings of the 2016 9th International Kharkiv Symposium on Physics and Engineering of Microwaves, Millimeter and Submillimeter Waves (MSMW). IEEE, 2016, pp. 1–4.
11. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Medical image analysis* **2017**, *42*, 60–88.
12. Guo, Y.; Liu, Y.; Georgiou, T.; Lew, M.S. A review of semantic segmentation using deep neural networks. *International journal of multimedia information retrieval* **2018**, *7*, 87–93.
13. Rocha, J.; Cunha, A.; Mendonça, A.M. Comparison of Conventional and Deep Learning Based Methods for Pulmonary Nodule Segmentation in CT Images. In Proceedings of the EPIA Conference on Artificial Intelligence. Springer, 2019, pp. 361–371.
14. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.
15. Tyagi, S.; Talbar, S.N. CSE-GAN: A 3D conditional generative adversarial network with concurrent squeeze-and-excitation blocks for lung nodule segmentation. *Computers in Biology and Medicine* **2022**, *147*, 105781.
16. Wang, S.; Zhou, M.; Liu, Z.; Liu, Z.; Gu, D.; Zang, Y.; Dong, D.; Gevaert, O.; Tian, J. Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation. *Medical image analysis* **2017**, *40*, 172–183.
17. Wang, S.; Zhou, M.; Gevaert, O.; Tang, Z.; Dong, D.; Liu, Z.; Tian, J. A multi-view deep convolutional neural networks for lung nodule segmentation. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2017, pp. 1752–1755.
18. Tong, G.; Li, Y.; Chen, H.; Zhang, Q.; Jiang, H. Improved U-NET network for pulmonary nodules segmentation. *Optik* **2018**, *174*, 460–469.
19. Hancock, M.C.; Magnan, J.F. Lung nodule segmentation via level set machine learning. *arXiv preprint arXiv:1910.03191* **2019**.
20. Chen, W.; Wang, Q.; Yang, D.; Zhang, X.; Liu, C.; Li, Y. End-to-End Multi-Task Learning for Lung Nodule Segmentation and Diagnosis. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021, pp. 6710–6717.
21. Liu, H.; Cao, H.; Song, E.; Ma, G.; Xu, X.; Jin, R.; Jin, Y.; Hung, C.C. A cascaded dual-pathway residual network for lung nodule segmentation in CT images. *Physica Medica* **2019**, *63*, 112–121.
22. Chen, Q.; Xie, W.; Zhou, P.; Zheng, C.; Wu, D. Multi-Crop Convolutional Neural Networks for Fast Lung Nodule Segmentation. *IEEE Transactions on Emerging Topics in Computational Intelligence* **2021**.
23. Zhang, Z.; Liu, Q.; Wang, Y. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters* **2018**, *15*, 749–753.
24. Zhang, B.; Qi, S.; Wu, Y.; Pan, X.; Yao, Y.; Qian, W.; Guan, Y. Multi-Scale Segmentation Squeeze-and-Excitation UNet with Conditional Random Field for Segmenting Lung Tumor from CT Images. *Computer Methods and Programs in Biomedicine* **2022**, p. 106946.
25. McNitt-Gray, M.F.; Armato III, S.G.; Meyer, C.R.; Reeves, A.P.; McLennan, G.; Pais, R.C.; Freymann, J.; Brown, M.S.; Engelmann, R.M.; Bland, P.H.; et al. The Lung Image Database Consortium (LIDC) data collection process for nodule detection and annotation. *Academic radiology* **2007**, *14*, 1464–1474.
26. Feng, X.; Yang, J.; Laine, A.F.; Angelini, E.D. Discriminative localization in CNNs for weakly-supervised segmentation of pulmonary nodules. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2017, pp. 568–576.
27. Wu, B.; Zhou, Z.; Wang, J.; Wang, Y. Joint learning for pulmonary nodule segmentation, attributes and malignancy prediction. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018, pp. 1109–1113.
28. Kubota, T.; Jerebko, A.K.; Dewan, M.; Salganicoff, M.; Krishnan, A. Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models. *Medical Image Analysis* **2011**, *15*, 133–154.
29. Amorim, P.H.; de Moraes, T.F.; da Silva, J.V.; Pedrini, H. Lung Nodule Segmentation Based on Convolutional Neural Networks Using Multi-orientation and Patchwise Mechanisms. In Proceedings of the ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing. Springer, 2019, pp. 286–295.
30. Zou, K.H.; Warfield, S.K.; Bharatha, A.; Tempany, C.M.; Kaus, M.R.; Haker, S.J.; Wells III, W.M.; Jolesz, F.A.; Kikinis, R. Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. *Academic radiology* **2004**, *11*, 178–189.
31. Chollet, F.; et al. Keras. https://github.com/fchollet/keras, 2015.
32. Jung, J.; Hong, H.; Goo, J.M. Ground-glass nodule segmentation in chest CT images using asymmetric multi-phase deformable model and pulmonary vessel removal. *Computers in biology and medicine* **2018**, *92*, 128–138.
33. Cao, H.; Liu, H.; Song, E.; Hung, C.C.; Ma, G.; Xu, X.; Jin, R.; Lu, J. Dual-branch residual network for lung nodule segmentation. *Applied Soft Computing* **2020**, *86*, 105934.
34. Maqsood, M.; Yasmin, S.; Mehmood, I.; Bukhari, M.; Kim, M. An efficient DA-net architecture for lung nodule segmentation. *Mathematics* **2021**, *9*, 1457.
35. Zhou, Z.; Gou, F.; Tan, Y.; Wu, J. A cascaded multi-stage framework for automatic detection and segmentation of pulmonary nodules in developing countries. *IEEE Journal of Biomedical and Health Informatics* **2022**, *26*, 5619–5630.