



OPEN

## A deep learning based dual encoder–decoder framework for anatomical structure segmentation in chest X-ray images

Ihsan Ullah<sup>1,7</sup>, Farman Ali<sup>2,7</sup>, Babar Shah<sup>3</sup>, Shaker El-Sappagh<sup>4,5</sup>, Tamer Abuhmed<sup>6</sup> & Sang Hyun Park<sup>1</sup>✉

Automated multi-organ segmentation plays an essential part in the computer-aided diagnostic (CAD) of chest X-ray fluoroscopy. However, developing a CAD system for the anatomical structure segmentation remains challenging due to several indistinct structures, variations in the anatomical structure shape among different individuals, the presence of medical tools, such as pacemakers and catheters, and various artifacts in the chest radiographic images. In this paper, we propose a robust deep learning segmentation framework for the anatomical structure in chest radiographs that utilizes a dual encoder–decoder convolutional neural network (CNN). The first network in the dual encoder–decoder structure effectively utilizes a pre-trained VGG19 as an encoder for the segmentation task. The pre-trained encoder output is fed into the squeeze-and-excitation (SE) to boost the network's representation power, which enables it to perform dynamic channel-wise feature calibrations. The calibrated features are efficiently passed into the first decoder to generate the mask. We integrated the generated mask with the input image and passed it through a second encoder–decoder network with the recurrent residual blocks and an attention the gate module to capture the additional contextual features and improve the segmentation of the smaller regions. Three public chest X-ray datasets are used to evaluate the proposed method for multi-organs segmentation, such as the heart, lungs, and clavicles, and single-organ segmentation, which include only lungs. The results from the experiment show that our proposed technique outperformed the existing multi-class and single-class segmentation methods.

Chest X-rays are mostly used to examine the chest's anatomical structures, such as the lungs, heart, and clavicles for various pulmonary and cardiac disorders. In the healthcare industry, 3.6 billion chest X-rays are performed each year to evaluate the patients' health conditions<sup>1</sup>. With the increasing amount of chest X-rays, computer-aided diagnostic (CAD) systems can play a significant role to identify chest diseases. Significant research has recently been conducted in CAD systems to facilitate pulmonologists/radiologists to assess chest radiographs. The CAD considers a range of analytical tasks that demand precise segmentation of the anatomical structures in the chest radiography images. These tasks include disease prediction and various size measurements in the chest radiographs, such as determining the existence of pulmonary nodules<sup>2</sup> or lung disease<sup>3</sup> using lung field segmentation. Furthermore, cardiomegaly<sup>4</sup> can be predicted using heart segmentation in chest radiographs. The

<sup>1</sup>Department of Robotics and Mechatronics Engineering, Daegu Gyeonbuk Institute of Science and Engineering (DGIST), Daegu 42988, South Korea. <sup>2</sup>Department of Computer Science and Engineering, School of Convergence, College of Computing and Informatics, Sungkyunkwan University, Seoul 03063, South Korea. <sup>3</sup>College of Technological Innovation, Zayed University, Dubai 19282, United Arab Emirates. <sup>4</sup>Faculty of Computer Science and Engineering, Galala University, Suez 435611, Egypt. <sup>5</sup>Information Systems Department, Faculty of Computers and Artificial Intelligence, Benha University, Banha 13518, Egypt. <sup>6</sup>Department of Computer Science and Engineering, College of Computing and Informatics, Sungkyunkwan University, Suwon 16419, South Korea. <sup>7</sup>These authors contributed equally: Ihsan Ullah and Farman Ali. ✉email: shpark13135@dgist.ac.kr

information about the clavicle position can be used to overcome false positive results and identify the lesions behind a clavicle more consistently.

However, evaluating chest radiographic images is still challenging for indistinct and overlapping body structure borders, such as the lungs, heart, and clavicles. Furthermore, anatomical structures vary in shapes and sizes depending on the patient's gender, age, and physique. Also, the presence of medical equipment, which includes pacemakers and guidewires, and various artifacts in chest radiography make the segmentation of anatomical structures problematic. Despite these difficulties, significant development has occurred in recent years regarding the improvement of the segmentation methods<sup>4–7</sup>. The researchers have been able to test various segmentation approaches that employ the Japanese Society of Radiological Technology (JSRT) dataset<sup>8</sup>. These approaches include rule-based systems, such as the thresholding of intensity<sup>9</sup>, edge detection-based methods<sup>10,11</sup>, hybrid models<sup>12,13</sup>, and landmark-based models<sup>14,15</sup>. However, these methods require optimal parameter settings for good performance, and they frequently fail when the anatomical structures overlap.

Due to the excellent performance of the CNNs in segmentation tasks, which are notably in biomedical imaging, the CNN-based approaches<sup>5,16</sup> have quickly gained popularity. Moreover, deep learning methods, such as U-Net have substantially improved segmentation in medical applications, which include vascular segmentation<sup>17</sup>, catheter segmentation in X-rays and realistic images<sup>18–20</sup>, and the lungs segmentation in chest radiography<sup>21</sup>. However, the interest region in the medical images can have similar appearances, which makes it difficult to segment them using U-Net<sup>22</sup>. Deeper CNN structures with more layers are preferred for better feature representation to achieve a higher segmentation performance. Therefore, we utilized a dual CNN architecture with more layers to achieve a better feature representation for the anatomical structure segmentation.

This paper presents a novel segmentation framework that uses a dual encoder–decoder CNN structure with a pre-trained network weights for the chest anatomical structure segmentation in the X-ray fluoroscopic images. The first encoder–decoder network predicts the initial mask using the VGG19 pre-trained network as an encoder. The initially predicted mask is coupled with the input image, and it is fed into a second encoder–decoder network, which comprises of the recurrent residual convolutional layers and an attention gate module to predict the final improved segmentation mask and capture the additional contextual features. The proposed framework is evaluated using three standard benchmark datasets. The JSRT dataset is utilized for three-class segmentation, which contains the data about heart, clavicles, and lungs. The MCCXR and SCXR datasets are employed for single-class segmentation, which includes the lungs only. The proposed method outperforms the existing state-of-the-art (SOTA) CNN architectures for the anatomical structures segmentation in the chest X-rays. The following are the significant contributions of this study.

- We propose a novel dual encoder–decoder architecture to effectively segment the anatomical structures in chest X-ray images. The proposed method can accurately segment both the more prominent structures (lungs and heart) and the smaller structures (clavicles) in chest X-rays.
- Instead of previous methods requiring post-processing<sup>23</sup> or multi-stage<sup>5</sup>, we employed an end-to-end CNN architecture with efficient training and high accuracy to improve the model sensitivity to foreground pixels without requiring complex heuristics.
- The proposed method incorporates Attention Gating Modules (AGMs) to allow the model to focus on the regions of interest while maintaining the spatial resolution and improving the quality of the feature maps.
- We extensively evaluated three public datasets for the anatomical structure segmentation, that includes multiclass segmentation and single-class segmentation. The experimental results show that our proposed method outperforms the previous SOTA methods on single class anatomical structure segmentation as well as multiclass anatomical structure segmentation.

The rest of the manuscript is arranged as follows. “[Related work](#)” presents the prior studies about the anatomical structure segmentation for the X-ray fluoroscopic images. “[Proposed method](#)” describes the proposed method in detail. We evaluate the proposed method extensively in “[Experiments and results](#)”. “[Discussion](#)” provides a discussion about the proposed framework. Finally, “[Conclusion](#)” presents the conclusion of the proposed method.

## Related work

In this section, we present the recent advancements in the anatomical structure segmentation in three domains, which include the conventional, shallow learning, and deep learning methods.

**Conventional methods.** The conventional feature-based methods employ heuristics algorithms relying on low-level image features. For example, Cheng et al.<sup>24</sup> locate lung areas by evaluating the Horizontal and Vertical profiles (HVP) of a chest X-ray image. After that, the lungs boundaries are detected using static global intensity thresholding and smoothing operations. Armato et al.<sup>25</sup> first construct a range of threshold values using a histogram analysis instead of utilizing the threshold of static global, which they later apply global thresholding to distinguish the initial lung area, and then use local thresholding to refine the initial segmented lung region. Nevertheless, iterative thresholding makes the process computationally expensive. Furthermore, these techniques<sup>24,25</sup> produce lower performances when large deformation occurs in the chest region.

Also, these methods are sensitive to image intensities. It is challenging to define a proper threshold value for the accurate segmentation in the X-ray images due to poor contrast<sup>24,25</sup>. Li et al.<sup>10</sup> employed the first derivative of the HVP to define the initial boundary of the lung area, which was succeeded by an adjustment of iterative boundary and edge tracing. However, this method fails to segment the complicated anatomical structures, such as mediastinum and hemidiaphragm. Iakovidis et al.<sup>11</sup> utilized a similar approach<sup>10</sup> to identify the lung boundaries in chest X-ray images using curve interpolation<sup>26</sup>. Xu et al.<sup>27</sup> introduced an improved method to

identify the anatomical structures in chest X-ray images by performing a gradient analysis using the structural relationships among the anatomic landmark positions. After that, the polynomial functions are utilized to smoothen the boundaries of the anatomical structures. Ahmad et al.<sup>28</sup> employed c-mean clustering (CMC) to improve lung borders. However, these techniques do not need any prior understanding of lung anatomy. Several approaches<sup>10,11,27,28</sup> are still noise-sensitive, and the results of pre-processing noise filtering with a Gaussian are contingent on the filter size and threshold.

**Shallow learning methods.** The shallow learning techniques are considerably sensitive to the feature extraction process. However, the major limitation of these approaches is to select the suitable attributes. Several shallow learning-based frameworks for anatomical structure classification are described in the prior literature. To identify the lung area, Gray et al.<sup>29</sup> used a multi-layer neural network (MLP) with the k Nearest Neighbor (kNN) and the Linear Discriminant Analysis (LDA) classifiers, which were trained using a variety of local textures, grey levels, and local difference based features. Tsujii et al.<sup>30</sup> used a Hybrid Adaptive Neural Network (HANN), which includes features, such as adjacent pixel locations, normalized intensity, and histogram equalized entropy. The hidden layers in the HANN are dynamically settled, and thus they can be trained efficiently without over-fitting. Nevertheless, these techniques perform poorly with similarly intensity regions, low contrast, and tilted images.

The Markov random field (MRF) model represents a flexible and successful approach to obtain spatial and texture-relevant information. An MRF was also used by Vittitoe et al.<sup>31</sup> to blend spatial and textural information using the potential functions that were parameterized with a probability distribution and the recursive conditional approach to classify each pixel of the chest X-ray fluoroscopic images optimally for the respective anatomic class. Furthermore, Ginneken et al.<sup>8</sup> combined the rule-based approach with a shallow learning-based method, whereas the rule-based method uses prior information about the lung region. Also, the pixel classifier uses these attributes, the image intensities, and the entropy measurements to classify the X-ray image to their various categories, such as the lung regions and the non-lung regions. Shi et al.<sup>32</sup> suggested an unsupervised technique to segment the lung regions. The authors used a CMC-based method with Gaussian kernels, and they achieved good results via the CMC's flexible and robust mathematical modeling. Nevertheless, the CMC methods fail in medical image segmentation due to noise, artifacts, and illumination variations.

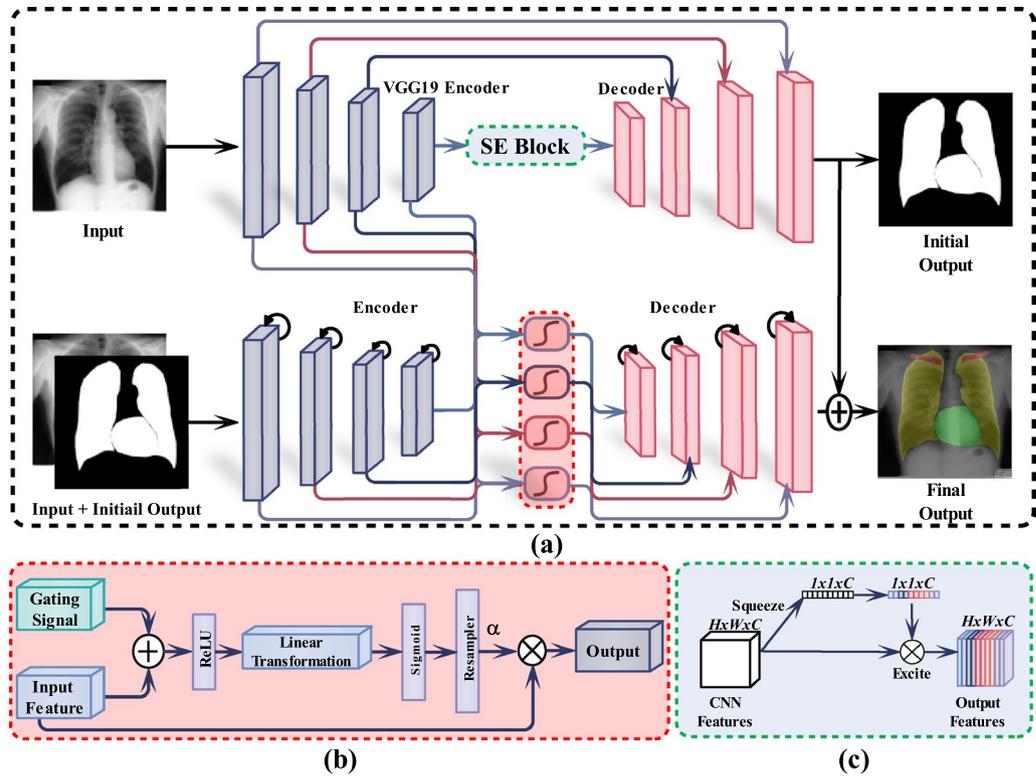
**Deep learning methods.** Deep learning-based approaches are contrasted to shallow learning techniques, leverage dynamics, and hierarchical feature representations with several levels of abstraction. Deep learning methods significantly outperform the shallow learning techniques in many applications, but the most significant impediment to the practical deployment of deep learning is a lack of labeled data for training. In recent years the CNNs have been employed in medical field for diagnosing chest problems<sup>33</sup>. Kalinovsky et al.<sup>34</sup> recently adopted the four-layered encoder–decoder architecture, which is called SegNet<sup>35</sup>, for the lung area segmentation. Novikov et al.<sup>16</sup> adopted U-Net<sup>22</sup> for multi-class segmentation in chest radiographs. Furthermore, Mittal et al.<sup>36</sup> proposed LF-SegNet, a modified U-Net-based approach that incorporates a normalization mechanism and improves the up-sampling strategy for segmentation of lung fields in chest radiographs. However, these U-Net-based methods introduce outliers and holes inside the targeted structures, which were solved using a level-set method post-processing step.

Recently, amore complex CNN method such as Mask R-CNN<sup>37</sup> is proposed by Wang et al.<sup>38</sup> to segment and identify the lung field, heart, and clavicles in the chest X-ray images. However, this approach is computationally costly due to the excessive region proposals. Furthermore, Hwang et al.<sup>5</sup> proposed a two-stage cascade network training approach, in which a network is trained using input chest X-rays in the first stage, and then both the input chest X-ray image and the output from the learned model in the first stage are fed into the second network. However, these approaches were not trained in an end to end manner. In addition, Peng et al.<sup>39</sup> proposed a two-stage approach for the lung segmentation, where in the first step a Deep Belief Network and K-Nearest Neighbor is utilized to segment the lungs and then an improved principal curve and machine learning method is used for the refinement of the segmentation result. However, these types of techniques are computationally costly and necessitate a post-processing step to improve the output. In contrast, the proposed dual encoder–decoder method can be trained in the end-to-end manner. As a result, post-processing and pre-processing are not required.

## Proposed method

Figure 1 shows the overview of the proposed framework. The proposed method contains two encoders and decoders. The image is first given to a pre-trained VGG19 encoder which consists of highly connected convolutional and fully connected layers for better feature extraction. The extracted features from a pre-trained VGG19 encoder are then given to the SE block which pass the most relevant information to the first decoder for the generation of the initial binary segmentation mask. Furthermore, the initial binary segmentation mask and input image are coupled and fed to the second encoder to ensure that the produced initial binary segmentation mask can still be improved. The output features of the second encoder are passed through the AGMs, improving the quality of the feature maps. The AGMs focus on the regions of interest while maintaining the spatial resolution of the feature maps. The feature maps from both encoders are upsampled and passed to the second decoder which generate the final improved segmentation output. Moreover, the pre-trained VGG19 network can be replaced with other networks such as MobileNet<sup>40</sup>, ResNet<sup>41</sup>, DenseNet<sup>42</sup> or EfficientNet<sup>43</sup>. Furthermore, incorporating DenseNet<sup>42</sup> in our proposed framework obtained the best results.

**Dual encoder–decoder CNN architecture (DED-CNN).** *Encoders.* The first Deep Learning (DL) network employed a VGG19 encoder with successive convolution and max-pooling layers. The number of filters is doubled after the max-pooling layer, and the process is repeated four times. The fully connected layer is



**Figure 1.** (a) The overall architecture of the proposed anatomical structures segmentation in the X-ray images, and (b) illustrates the attention-gating module in the red dotted lines. The attention coefficients computed in the AGM are used to scale the input features. The gating signal in the AGMs uses both the activations and the contextual information to identify the spatial regions. (c) The Squeeze-and-Excite (SE) block.

substituted with a single convolutional layer that serves as a network bottleneck, which separates the encoder and decoder. The Rectified Linear Unit (ReLU) activation function is used to introduce nonlinearity into the model. The Squeeze-and-Excite (SE)<sup>44</sup> block, which is shown in Fig. 1c, is then applied to enhance the feature map’s quality.

The second DL encoder is built from scratch. We utilize the prior segmentation mask estimated by the first encoder to guide the second encoder towards the instance of interest. The second network can learn appropriate boundaries of anatomical structures in chest X-ray images by emphasizing features derived from relatively coarse segmentation outputs. Furthermore the second encoder incorporates a recurrent residual convolutional layer (R2CL) in each step to increase the model’s capacity for integrating the context information. R2CL helps to develop a more deeper model and ensure better and stronger feature representation accumulation with respect to different time-steps. The R2CL is comprised of  $3 \times 3$  convolutions, which is repeated twice with recurrent connections. When the R2CL is employed, the number of feature maps increases, which results in a size reduction of about 50%. Each convolution operation is followed by batch normalization (BN)<sup>45</sup>. The BN regularizes the model while reducing the internal covariant shift. Concatenation is used in the R2U-Net architecture to map the features from the encoding path to the decoding path. The residual network<sup>46</sup> performs better when the R2CL operations are conducted in discrete time steps. As illustrated in Fig. 1a, we assume that  $x_l$  is the input of the  $l$ th layer of the R2CL block. Where the pixel is localized at  $(i, j)$  in the input on the  $k$ th feature map. Furthermore, the network output  $O_l$  is  $(t)$  at time step  $t$ . The output can be mathematically represented by the equation that is given below.

$$O_{ijk}^l(t) = (W_k^c)^T \times x_l^{c(i,j)}(t) + (W_k^r)^T \times x_l^{r(i,j)}(t-1) + b_k, \tag{1}$$

wherein  $x_l^{c(i,j)}$  and  $x_l^{r(i,j)}$  are the standard convolutional layer inputs as well as the input for  $l$ th residual convolutional layer (RCL).  $W_k^c$  and  $W_k^r$  are the weights of the forward convolutional layer and RCL of the  $k$ th feature map, and  $b_k$  is used as the bias.  $O_{ijk}^l$  is given to the standard ReLU, which can be mathematically expressed using the equation that is given below.

$$F(x_l, w_l) = f(O_{ijk}^l(t)) = 0, O_{ijk}^l(t) \tag{2}$$

Final outputs  $x(l+1)$  of the recurrent convolutional unit is given to the residual unit, which can be formulated mathematically with the equation that is given below.

$$x_{l+1} = x_l + F(x_l, w_l). \tag{3}$$

The inputs of R2CL are represented by  $x_l$ . The outputs  $F(x_l, w_l)$  are used in the down-sampling and up-sampling layers in the encoder and decoder path of the proposed network. The successive sub-sampling or up-sampling layers use the final output  $x_{l+1}$  as the input. The basic forward convolutional unit is shown in Fig. 2a, and the structure of the recurrent residual convolutional layer unit (R2CL) is shown in Fig. 2b. We deepen the R2CL block, which each comprise of two RCL units, to extract the features of the deeper layers. The RCL expansion strategy is represented in Fig. 2c, which is where the RCL is expanded to two-time steps ( $T = 2$ ).

**Decoders.** Each decoder block in the first DL network performs up-sampling on the given feature map via the transposed convolution, which effectively increases the size of the feature maps. Afterwards, the features maps of the encoder skip-connections are concatenated with the output feature maps. The skip-connections from the first encoder are used in the first decoder. Nevertheless, the skip-connections from both encoders are used in the second decoder, which allow the model to maintain spatial resolution and improve the quality maps of the output feature. In each step of the decoding path, the output of the R2CL unit from the preceding layer is up-sampled by the R2-Unit. The number of feature maps is reduced to half and size is doubled with each up-sampling process. The size of the feature map is restored to its original size of the input image at the last layer of the decoder. Furthermore, the BN is used in the up-sampling process to increase the proposed network's stability and speed up its convergence during the training phase. The output of the BN is fed to the Attention Gating Modules (AGMs).

The second network uses the AGMs to optimize the previous encoder's output features, which then merge them with the corresponding features in the decoder. Furthermore, the AGMs allow the model to emphasize the high-quality feature representations while preserving the spatial resolution and boosting the quality of the resulting feature maps. The additive attention map employed in AGMs is described in Fig. 1b. The attention values are computed for each pixel of all the input features  $x_i^l$ . To determine the attention regions in the feature map, the gating vector  $g_i$  is applied to each pixel  $i$ . The following is the additive equation.

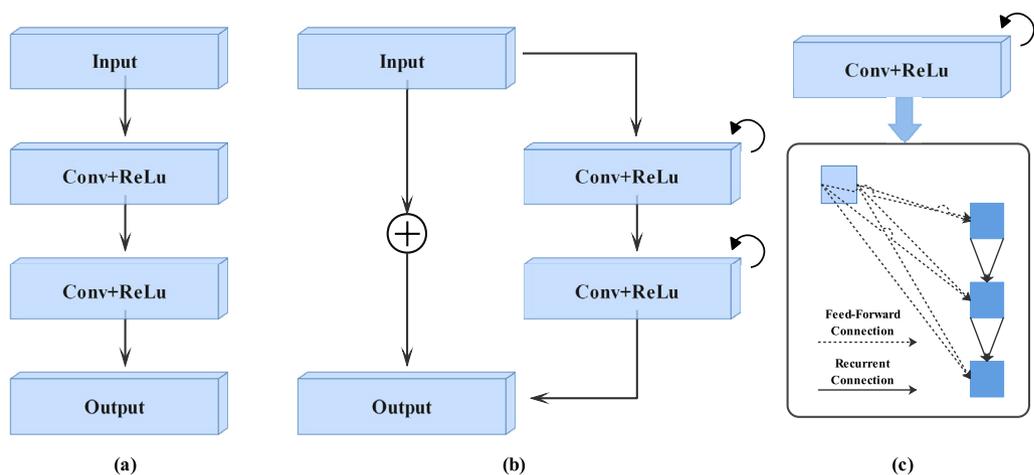
$$\alpha_i^l = \sigma_2(\psi^T(\sigma_1(W_x^T x_i^l + W_g^T g_i + b_g)) + b_\psi), \quad (4)$$

wherein  $\sigma_1$  and  $\sigma_2$  represent the ReLU and sigmoid activation. Furthermore, the weights of a linear transformation are represented as  $W_x$  and  $W_g$ , and the biases are  $b_g$  and  $b_\psi$ . A linear transformation is used to minimize the number of learnable parameters and the computational complexity of the AGMs. In addition, we adjust the input features to match the size of the gating signal. To prevent the aliasing issue, grid resampling of the attention coefficients is performed using tri-linear interpolation. Our network incorporates the AGMs to emphasize the essential features to segment the clavicles, heart, and lungs. The coarse-scale information is used in the gating to disentangle the unnecessary and noisy responses in the skip connections, which boosts the model's performance with the foreground pixel prediction. The output of the AGMs is the combination of the input feature map and the attention coefficient, which can be represented mathematically by the equation that is given below.

$$g_i = x_i^l \times \alpha_i^l. \quad (5)$$

Finally, the output is predicted using a sigmoid activation function.

The proposed framework was implemented using the Keras and TensorFlow libraries, and the models were trained from scratch. The model was initialized with random values during the training step. Adam was used to optimize the model with an initial learning rate (LR) of 1e-5 and a batch size of 8.



**Figure 2.** Convolutional and recurrent convolutional layers for different variants, which include (a) the forward convolutional layer, (b) the recurrent residual convolutional layer (R2CL), and (c) depicting the unfolded recurrent convolutional layer for  $T = 2$ .

## Experiments and results

In this section, the datasets that were utilized to conduct the experiments are extensively discoursed, and the implementation procedure is described. In addition, the performance of the proposed framework to segment the anatomical structures in the chest fluoroscopic images is evaluated, and the results are discussed.

**Datasets.** We performed several experiments on three different datasets: MCCXR dataset<sup>47</sup>, the JSRT dataset<sup>48</sup>, and the SCXR dataset<sup>47</sup>, as shown in Fig. 3. The JSRT dataset contains 247 posterior-anterior (PA) chest radiographs with a resolution of  $2048 \times 2048$  pixels and their ground-truth segmentation<sup>8</sup> with a resolution of  $1024 \times 1024$  pixels. The MCCXR<sup>47</sup> comprises of 138 fluoroscopic images with a resolution of  $4020 \times 4892$ , which contains 80 normal subjects and 58 abnormal subjects, including tuberculosis, effusions, and military patterns gathered from Montgomery County's tuberculosis screening program. The SCXR dataset<sup>47</sup> was collected by Shenzhen No. 3 Hospital in Shenzhen, Guangdong province, China. The SCXR dataset<sup>47</sup> contain 326 normal X-rays images and 336 abnormal X-rays that show various manifestations of tuberculosis. Further, we have used NIH Chest X-ray dataset<sup>49</sup> which contain both PA and AP views in cross-dataset evaluation. We have performed several experiments in cross-dataset evaluation setup on different datasets to validate the robustness and effectiveness of the proposed framework.

The JSRT dataset is initially split into five subsets. The five-folds cross-validation is then carried out with one fold being dedicated to testing, which includes 20% of all the available chest X-ray data. The remaining four folds are divided into training, which includes 90% the chest X-ray data, and the validation, which includes 10% of the chest X-ray data. For the SCXR and MCCXR datasets, we performed three fold cross-validation. For each fold, the total number of epochs was set to 100. An NVIDIA 1080Ti GPU was used to train each model. We compared the proposed model to the recent state of the art techniques, i.e., U-Net<sup>22</sup>, Linknet<sup>50</sup>, PSPNet<sup>51</sup>, Seg-Net<sup>52</sup>, Adaptive Scan<sup>53</sup>, X-Net+<sup>54</sup>, RX-Net+<sup>54</sup>, TVC<sup>55</sup>, TMI<sup>16</sup>, and SCIA<sup>56</sup>. For fair comparison, we normalized the image size as  $256 \times 256$  images and then trained the proposed model and the comparison methods.

**Performance metrics.** We utilized the dice<sup>57</sup>, Intersection over Union (IoU)<sup>58</sup>, and the Hausdorff distance (HD)<sup>59</sup> similarity coefficients to evaluate the proposed method and compare it to the SOTA approaches. For the pairwise comparison of the binary segmentation of the foreground with the ground truth, the DSC is a frequently used overlap metric. In formal terms, it is written using the equation that is given below.

$$Dice = \frac{2 \times (A \cap B)}{A + B}, \quad (6)$$

where the ground truth is A, and the predicted mask is B. The Dice coefficient varies from 0 to 1, and 1 indicates a complete overlap. In addition, we also utilized the Intersection over Union (IoU) metric to quantify the percent overlap between the target mask and our predicted output.

$$IoU = \frac{A \cap B}{|A \cup B|}. \quad (7)$$

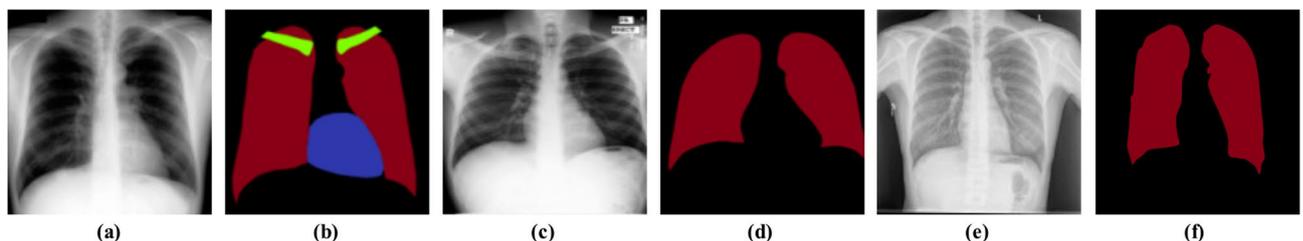
The HD is a distance measurement between two sets of points. HD is the maximum of the distances between any segmentation point and the nearest ground truth point.

$$HD = \max(d(A, B), d(B, A)), \quad (8)$$

where A and B is the finite point set, the function  $d(A, B)$  is referred to as the directed HD from A to B, and  $d(B, A)$  is the distance from B to A.

## Experimental results

This section presents the quantitative and qualitative results of the proposed method using three different datasets. In the quantitative results, the proposed framework performance of the multiclass segmentation and the single-class segmentation is compared with the different approaches using the JSRT dataset and the MCCXR and SCXR datasets. In the qualitative results, the visual results of the proposed method are demonstrated.



**Figure 3.** Input images and labels from three datasets: (a) radiographic X-ray image of the chest from JSRT dataset, (b) JSRT dataset labels of the lungs, clavicles and heart, (c) X-ray image of the chest from MCCXR dataset, (d) MCCXR dataset lungs label image, (e) chest X-ray image of SCXR dataset, and (f) SCXR dataset lungs label image.

**Quantitative results.** *JSRT benchmark.* Table 1 presents the comparison of the proposed work in terms of the segmentation performance with the existing approaches i.e X-Net+, RX-Net+, TVC, SCIA and TMI that are applied for all the anatomical structures of the X-ray images from the JSRT dataset, such as the heart, lungs, and clavicles. As we can see in Table 1, the proposed method outperformed the majority of the existing techniques. Our proposed method obtained the most significant results in terms of the IoU and dice score for the clavicles and heart segmentation. Additionally, our proposed method is comparably good for the lung segmentation. The RX-Net+ reported the lowest performance for the clavicles segmentation, which is 0.859 and 0.924 for the IoU and dice score, respectively. Another variant of the same method X-Net+ achieved slightly better results. However, this method uses an image size of  $1024 \times 1024$ , which is computationally expansive and challenging to train a model. The proposed method also yielded higher dice and IoU scores than SCIA for the clavicles and heart and comparable dice and IoU scores for the lungs. Although SCIA achieved slightly better results on the lungs than ours, SCIA employed a level set approach to post-process the output segmentation to remove the artifacts, which is computationally costly. In contrast, no post-processing is performed in the proposed method for the improvement of the output segmentation.

Table 2 presents the results of the comparison models trained on a single-class dataset. Similar to the results on the multi-class, DED-CNN outperformed the RX-Net+ and TMI. DED-CNN achieved 0.868 IoU for clavicle segmentation, whereas the TMI had the lowest IoU of 0.833. Also, the RX-Net+ showed a slightly worse IoU than our proposed model by achieving a 0.86 IoU. The proposed method obtained a 0.907 IoU for heart segmentation, whereas the RX-Net+ and the TMI scored 0.889 and 0.869 IoU, respectively. The RX-Net+ and the proposed DED-CNN achieved a similar IoU score of 0.955. However, the proposed method outperformed the TMI for lung segmentation. In terms of the HD, the proposed DED-CNN outperforms both the RX-Net+ and TMI for the heart and lungs by achieving 37.186 and 37.419 of HD, respectively. For the clavicles, the RX-Net+ achieved a lower HD than our proposed method.

*MCCXR benchmark.* We also compared our proposed model in terms of the lung segmentation with the existing methods for the MCCXR dataset, which are presented in Table 3. The proposed DED-CNN outperforms the other existing approaches. The proposed DED-CNN attained a dice score of 97.67%, and an IoU of 95.48%. The

Methods	Clavicles				Lungs				Heart			
	IoU	IoU-FR	Dice	Dice-FR	IoU	IoU-FR	Dice	Dice-FR	IoU	IoU-FR	Dice	Dice-FR
Human <sup>8</sup>	–	0.896	–	0.945	–	0.946	–	0.972	–	0.878	–	0.935
X-Net+ <sup>54</sup>	0.848	<b>0.874</b>	–	0.933	0.951	0.956	–	0.978	0.881	0.884	–	0.938
RX-Net+ <sup>54</sup>	0.838	0.859	–	0.924	0.947	0.948	–	0.973	0.876	0.876	–	0.934
TVC <sup>55</sup>	–	–	–	–	–	0.951	–	0.975	–	0.893	–	0.943
SCIA <sup>56</sup>	0.863	–	0.926	–	<b>0.959</b>	–	<b>0.979</b>	–	0.899	–	0.947	–
TMI <sup>16</sup>	0.833	–	<b>0.929</b>	–	0.95	–	0.974	–	0.882	–	0.937	–
DED-CNN	0.86	–	0.909	–	0.954	–	0.974	–	0.906	–	0.949	–
DED-CNN*	0.868	–	0.91	–	0.955	–	0.976	–	<b>0.907</b>	–	<b>0.95</b>	–

**Table 1.** Comparative evaluation of the proposed method segmentation performance with the existing SOTA algorithms was reported on the JSRT dataset. The performance of the models was trained using full resolution (FR) (i.e.,  $2048 \times 2048$ ) and reported with IoU-FR and Dice-FR. The proposed method (\*) represents the segmentation results of the model trained with the augmented dataset. Significant values are in bold.

Methods	Organs	IoU					Hausdorff distance				
		Mean	Std	Median	Min	Max	Mean	Std	Median	Min	Max
TMI <sup>16</sup>	Clavicles	0.833	0.833	0.843	0.639	0.905	22.39	10.721	20.18	6.031	72.764
	Heart	0.869	0.869	0.894	0.511	0.955	50.245	32.378	40.464	13.867	195.971
	Lungs	0.951	0.951	0.957	0.842	0.972	56.795	40.207	42.804	13.176	229.373
RX-Net+ <sup>54</sup>	Clavicles	0.86	0.86	0.874	0.661	0.929	20.047	10.479	17.202	7.066	70.187
	Heart	0.889	0.889	0.905	0.738	0.961	38.007	18.074	34.35	13.15	99.823
	Lungs	<b>0.955</b>	0.955	0.961	0.889	0.976	45.146	30.254	36.291	11.338	167.7
DED-CNN	Clavicles	<b>0.868</b>	0.868	0.873	0.706	0.935	38.694	3.275	38.849	31.118	46.342
	Heart	<b>0.907</b>	0.907	0.916	0.767	0.961	<b>37.186</b>	3.557	37.2	30.044	45.587
	Lungs	<b>0.955</b>	0.955	0.958	0.881	0.972	<b>37.419</b>	3.535	37.464	28.347	44.961

**Table 2.** An overview of the results based on the IoU and the Hausdorff distance for each architecture and three organs, which include the clavicles, heart, and lungs. The bolded scores are the best-performing networks on the JSRT dataset according to each column's metric.

Methods	Dice	IoU
Fuzzy C-mean <sup>60</sup>	0.9580	0.9350
Atlas-based NR <sup>12</sup>	0.9600	0.9410
Deformation-tolerant <sup>61</sup>	0.9230	0.8620
Adaptive scan <sup>53</sup>	0.9592	0.9216
U-NET <sup>22</sup>	0.9547	0.9256
SegNet <sup>52</sup>	0.8418	0.8914
Linknet <sup>50</sup>	0.9687	0.9397
PSPNet <sup>51</sup>	0.3889	0.2443
DED-CNN	<b>0.9767</b>	<b>0.9548</b>

**Table 3.** Analysis of the lung segmentation results in the terms of the dice score and the IoU score using the MCCXR dataset. The bolded values demonstrate that the proposed approach performed better in terms of the metric presented in each column.

proposed model achieved higher scores than the baseline UNet by 2.2% and 2.92% in terms of the dice score and IoU. In addition, our model also achieved a higher score when it was compared with the PSPNet model.

*SCXR benchmark.* The proposed method is also applied for lung segmentation on the SCXR dataset. As shown in Table 4, we achieved better results than the existing methods. As we can see in Table 4, the proposed DED-CNN outperforms the other methods by achieving a dice score and an IoU of 97.67% and 95.48%, respectively. These results show that the proposed method outperformed the baseline UNet by 2.2% dice score and 2.92% IoU. Moreover, PSPNet showed the lowest performance when it was compared to the achieved results from the proposed method.

To further confirm the significance of the proposed DED-CNN, we compared it with the existing deep learning methods in the literature using three standard benchmark datasets. We avoided the traditional methods, because our main aim was to compare the proposed model with the deep learning based approaches for lung segmentation. Table 5 presents the comparison results of the proposed method and the existing deep CNN architectures for lung segmentation. As we can see in Table 5, the proposed method achieved dice scores of 97.60%, 97.67%, and 95.68%, and it achieved IoU scores of 95.50%, 95.48%, and 91.84% for the JSRT, MCCXR, and SCXR

Methods	Dice	IoU
U-NET <sup>22</sup>	0.9532	0.9119
Linknet <sup>50</sup>	0.9491	0.9044
PSPNet <sup>51</sup>	0.6685	0.5420
FPN <sup>62</sup>	0.9023	0.8328
DED-CNN	<b>0.9568</b>	<b>0.9184</b>

**Table 4.** Comparison of the proposed method with the other existing models in terms of lung segmentation using the SCXR database. The bolded values demonstrate that the proposed method performs better in terms of the metric presented in each column.

Methods	Datasets	Method description	Evaluation	
			Dice	IoU
Human <sup>8</sup>	JSRT	Annotated by humans	0.9720	0.9460
	JSRT		0.9740	0.9220
Ronneberge et al. <sup>22</sup>	MCCXR	U-Net:encoder-decoder CNN	0.9547	0.9256
	SCXR		0.9532	0.9119
Suza et al. <sup>23</sup>	MCCXR	AlexNet for patch classification and ResNet18 based reconstruction	0.9400	0.8800
Novikov et al. <sup>16</sup>	JSRT	FCN Inverted-Net	0.9685	0.9490
Lei et al. <sup>56</sup>	JSRT	GAN based segmentation	0.9747	0.9508
DED-CNN	JSRT	Dual encoder-decoder structure	<b>0.9760</b>	<b>0.9550</b>
	MCCXR		<b>0.9767</b>	<b>0.9548</b>
	SCXR		<b>0.9568</b>	<b>0.9184</b>

**Table 5.** Detailed comparison of the proposed method with the existing models using three different datasets. The bolded values represent the best performances.

datasets, respectively. We compared these results with the U-Net model, which indicate that the proposed model increases the dice score by +0.2%, +2.2%, and +0.36%, and IoU score by +3.3%, +2.92%, and +0.65% for the JSRT, MCCCXR, and SCXR datasets, respectively. For the JSRT dataset, the proposed DED-CNN surpasses the dice score by +0.4% and the JI score by +0.9% when it is compared to the gold standard<sup>8</sup>. These results indicate that the proposed method outperformed the recently published segmentation approaches in terms of the dice and IoU scores using the three datasets.

**Qualitative results.** To confirm the qualitative performance and further assess the efficiency of the proposed method, the visual results of the different anatomical structures segmentation are demonstrated in Fig. 4. The class activation maps (CAMs) strategy was initially adopted to interpret classification models; however, Vinogradova et al.<sup>63</sup> used the CAMs for the interpretation of semantic segmentation networks. We analyzed the feature's activation here and employed CAMs to generate the heat maps to visualize the most activated areas in the X-ray images<sup>64</sup>. In addition, the last layer of the trained model is retrieved to construct the CAMs. By taking the weighted sum of the feature maps with their associated weights, we obtained a map, which is denoted as  $C_m$ , of the most important features, which is used to identify the image pixel as the lung region. The map  $C_m$  can be formally stated using the equation that is given below.

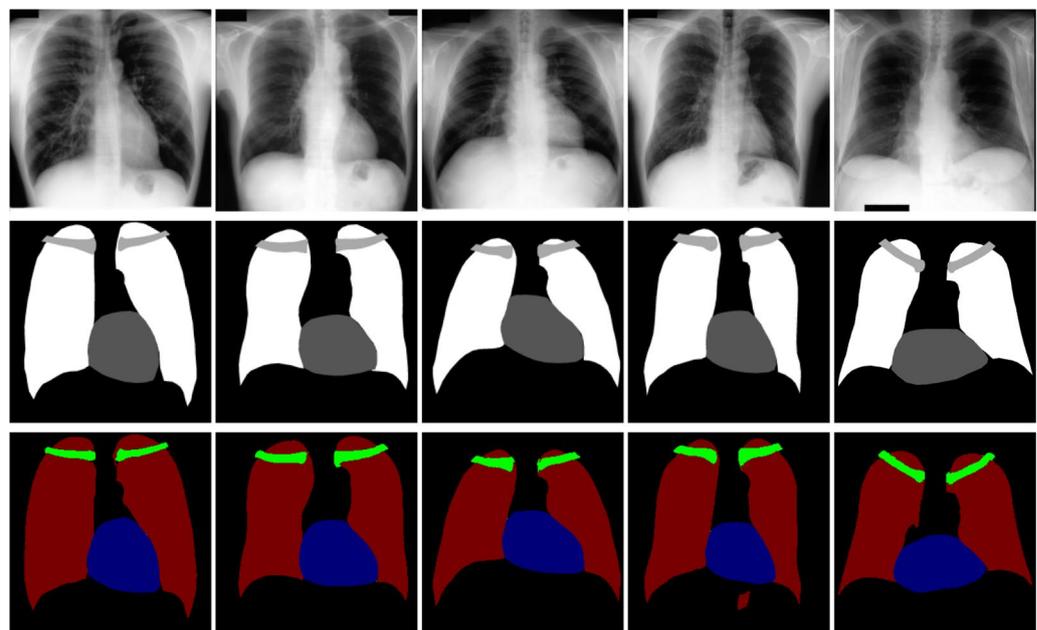
$$C_m = \sum_k w_{c,k} f_k, \quad (9)$$

where  $f_k$  is the last feature map, and  $w_{c,k}$  is weight of the final classification layer. We identify the essential features, which are utilized by the model to predict the lung region and by up-scaling the map  $C_m$  to the image's dimensions and overlaying the input image. Figure 5 shows various examples of the lung segmentation task using the CAMs. As shown in Fig. 5, the CAMs that are employed in our proposed method produced more activation on specific lung regions, whereas the U-Net CAMs have some outliers that were triggered on the right lung's bottom side. In addition, the Linknet CAMs have less activation on the left lung when it was compared to the proposed CAMs.

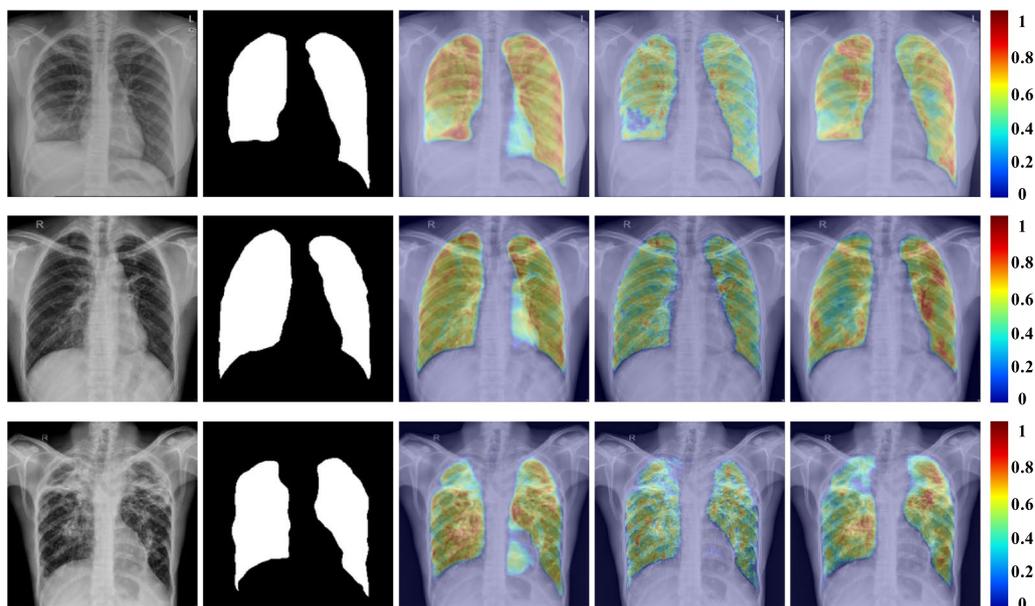
## Discussion

The popularity of the CADs among physicians and healthcare professionals is increasing daily. Therefore, it has become vital to provide a stable and reliable framework for chest radiographs that can be constantly evaluated and used for possible medical diagnostics. This study provides a new method for autonomous multi-organ and lung area segmentation in the chest X-rays. The proposed method incorporates a dual encoder–decoder CNN architecture with an initial encoder, which is a pre-trained VGG19 network. In this model, the attention gates and recurrent residual convolutional blocks are utilized instead of the regular convolutional blocks in the second network. The extensive experimental evaluation affirmed that the proposed dual encoder–decoder CNN architectures produced the best results for anatomical structure segmentation compared to the state-of-art methods.

The results showed in Fig. 4 confirmed the effectiveness of our proposed method in regards to segmenting several organs in the chest X-ray fluoroscopic radiographs. Furthermore, the results provided in Fig. 5 demonstrate that our technique can segment the heart, clavicles, and lungs in severely aberrant structures, which makes it suitable for clinic diagnostics. Furthermore, we performed lung segmentation generalization across datasets in



**Figure 4.** Qualitative results of our proposed method on a test sample from the JSRT dataset. From top to bottom: inputs, groundtruth labels, and the segmentation prediction.



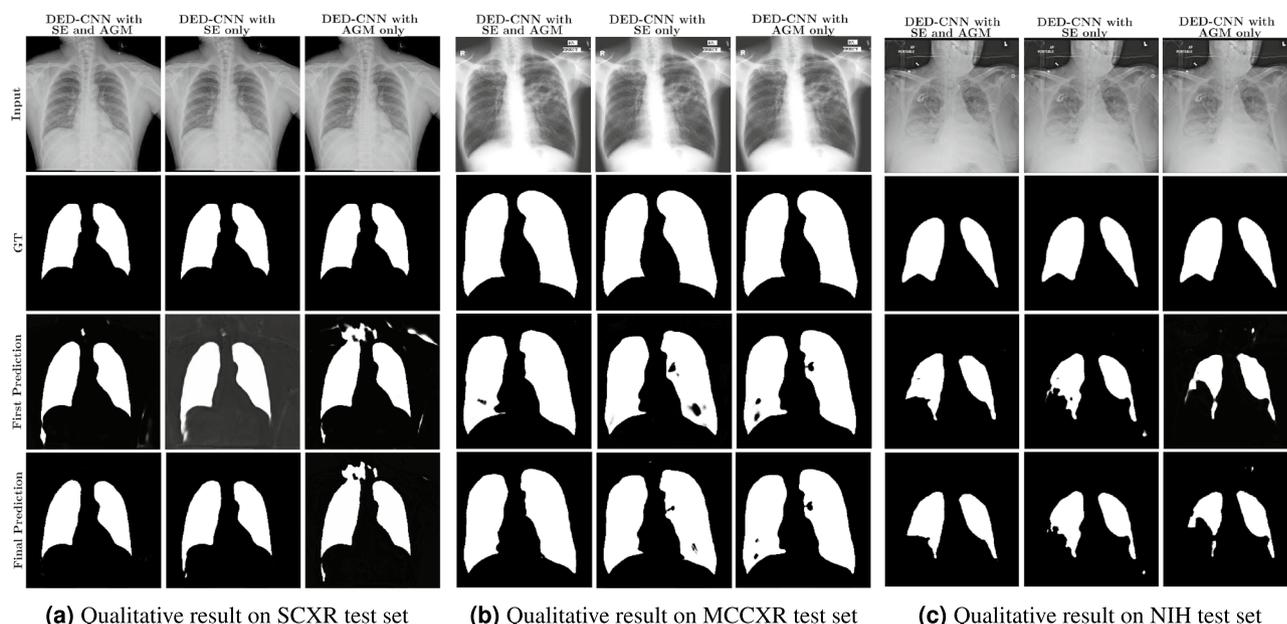
**Figure 5.** Class activation map (CAM) for the lung region. From left to right: inputs, groundtruth labels, U-Net CAM, Linknet CAM, and the proposed method CAM.

order to evaluate all possible dataset combinations. Table 6 shows the dice and IoU scores for the proposed DED-CNN. The performance of the DED-CNN is higher to that of the model that was trained using only AGM or SE blocks on test datasets. Furthermore, the Fig. 6 demonstrate that the qualitative results are consistent with our quantitative results presented in Table 6 for cross dataset testing configuration. Moreover, the performance of the proposed method with various pre-trained models is shown in Table 7. When we employed the DenseNet-121 as a pre-trained model in our proposed approach, we obtained the best dice and IoU score in cross-dataset settings compared to VGG-19, ReseNet-50, MobileNetv2 and EfficientNet pre-trained models.

Our proposed approach has notable advantages over SCIA and TMI since it is an “end-to-end” method and has robust accuracy performances. There is no pipeline that necessitates a significant amount of image pre and post-processing or registration, which leads to a noticeable decrease in performance time. Additionally, by utilizing the first network output in the second network, where we combine the input image and output segmentation map to guide the second network, our approach effectively lowers outliers and misclassification. In contrast to SCIA and TMI, the proposed method makes use of a pre-trained model, maximizing its generalizability. Additionally, using a pre-trained model is advantageous when there is limited training data.

Methods	Train dataset	Test dataset	Dice	IoU
DED-CNN	JSRT	SCXR	<b>0.9525</b>	<b>0.9108</b>
DED-CNN w/o SE	JSRT	SCXR	0.9381	0.8855
DED-CNN w/o AGM	JSRT	SCXR	0.9399	0.8869
Network-1	JSRT	SCXR	0.9416	0.8913
Network-2	JSRT	SCXR	0.9480	0.9014
DED-CNN	SCXR	MCCXR	<b>0.9597</b>	<b>0.9228</b>
DED-CNN w/o SE	SCXR	MCCXR	0.9478	0.9013
DED-CNN w/o AGM	SCXR	MCCXR	0.9550	0.9142
Network-1	SCXR	MCCXR	0.9495	0.9068
Network-2	SCXR	MCCXR	0.9559	0.9167
DED-CNN	SCXR+MCCXR	NIH	<b>0.9168</b>	<b>0.8470</b>
DED-CNN w/o AGM	SCXR+MCCXR	NIH	0.9148	0.8436
Network-1	SCXR+MCCXR	NIH	0.9051	0.8383
Network-2	SCXR+MCCXR	NIH	0.9081	0.8433

**Table 6.** Detailed comparison of the proposed method in the cross-dataset setting. The proposed method includes two encoder-decoder networks, referred to as Network-1 and Network-2, respectively. The bolded values represent the best performances.



**Figure 6.** Qualitative results of the proposed DED-CNN in the cross-dataset setting where all the models are trained on different dataset and tested on different dataset. From left to right: the first row is the input image, the second row represent the ground truths, the third row is the prediction result of the first encode-decoder model in DED-CNN framework and the last row represent the results of the second encoder-decoder model in the DED-CNN framework. All the models are trained and test in cross dataset configuration.

Methods	Pre-trained model	Train dataset	Test dataset	Dice	IoU
DED-CNN	VGG-19 <sup>65</sup>	JSRT	SCXR	0.9525	0.9108
	ResNet-50 <sup>41</sup>			0.9540	0.9160
	DenseNet-121 <sup>42</sup>			<b>0.9565</b>	<b>0.9166</b>
	MobileNet-v2 <sup>40</sup>			0.9491	0.9073
	EfficientNet <sup>43</sup>			0.9508	0.9096
DED-CNN	VGG-19 <sup>65</sup>	SCXR	MCCXR	0.9597	0.9228
	ResNet-50 <sup>41</sup>			0.9599	0.9236
	DenseNet-121 <sup>42</sup>			<b>0.9610</b>	<b>0.9297</b>
	MobileNet-v2 <sup>40</sup>			0.9423	0.9183
	EfficientNet <sup>43</sup>			0.9537	0.9213
DED-CNN	VGG-19 <sup>65</sup>	SCXR+MCCXR	NIH	0.9168	0.8470
	ResNet-50 <sup>41</sup>			0.9204	0.8530
	DenseNet-121 <sup>42</sup>			<b>0.9334</b>	<b>0.8752</b>
	MobileNet-v2 <sup>40</sup>			0.9287	0.8672
	EfficientNet <sup>43</sup>			0.9213	0.8557

**Table 7.** Cross-dataset qualitative comparisons of the component of the proposed approach. The values in bold represent the best performances.

Despite the outstanding performance of the proposed model in various settings, we illustrated some flaws and constraints that should be carefully considered. First and foremost, the proposed system only accomplishes the segmentation task. Second, this approach is computationally expensive because it uses a dual encoder and decoder. We will integrate additional downstream tasks in future studies, such as pneumonia classification to provide a more comprehensive pipeline for the CAD of chest radiographs. In addition, we also aim to increase the speed of the proposed method by reducing the model parameters using pruning.

### Conclusion

In this study, we presented a deep learning-based framework that can effectively identify anatomical structures, which include the lungs, heart, and clavicles, in chest X-ray images. A dual encoder-decoder network is notably employed, which can iteratively refine the output of the first network by fusing it with the input image and passing it through the second network to identify the anatomical structures in the X-ray images. Furthermore,

the first encoder–decoder incorporates the use of a pre-trained VGG19 network, which allows the proposed framework to be efficiently trained using limited datasets. In addition, it also allows the proposed model to extract essential features to enhance the anatomical structure segmentation. In the second encoder–decoder, we integrated the input image with a segmentation mask to guide the network and focus on the essential features and avoid the outliers, which permits the proposed network to segment the anatomical structures effectively. Furthermore, instead of the standard convolutional layers, we employed the R2CL and AGMs, which enable the proposed framework to focus on the regions of interest simultaneously and improve the feature maps. The proposed method accurately extracts information from the first encoder–decoder network and integrates it with the second encoder–decoder network to provide precise segmentation of the anatomical structures, which will assist physicians with diagnosing various pulmonary and cardiac diseases.

In the future, the performance of the anatomical structures segmentation framework will be enhanced by utilizing synthetically generated samples using generative adversarial networks. Additionally, downstream tasks, such as pneumonia and covid classification, will be included to provide a more comprehensive approach for the CADs of chest radiographs. Finally, a more sophisticated segmentation and classification method will be investigated for more accurate and reliable anatomical structure segmentation and diseases classification.

## Data availability

The JSRT dataset used in this study is published by the Japanese Society of Radiology Technology (JSRT) and is accessible at <http://db.jsrt.or.jp/eng.php>, The Montgomery dataset (MCCXR) used in this study is published by the U.S. National Institute of Health and is accessible at <https://academictorrents.com/details/ac786f74878a5775c81d490b23842fd4736bfe33>. The Shenzhen dataset (SCXR) images are available on <https://www.kaggle.com/datasets/raddar/tuberculosis-chest-xrays-shenzhen> and Shenzhen dataset (SCXR) masks are accessible at <https://www.kaggle.com/datasets/yoctoman/shcxr-lung-mask>. The NIH dataset used in this work is publicly available on [https://github.com/rsummers11/CADLab/tree/master/Lung\\_Segmentation\\_XLSor](https://github.com/rsummers11/CADLab/tree/master/Lung_Segmentation_XLSor).

Received: 31 May 2022; Accepted: 9 January 2023

Published online: 16 January 2023

## References

1. UNSCEAR. *Report of the United Nations Scientific Committee on the Effects of Atomic Radiation. General Assembly 56. Session (10–18 July 2008). Official Records: 63. Session, Suppl. No. 46 (a/63/46)*. Technical Report, United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) (2008).
2. Kakeda, S. *et al.* Improved detection of lung nodules on chest radiographs using a commercial computer-aided diagnosis system. *Am. J. Roentgenol.* **182**, 505–510 (2004).
3. Wang, J., Li, F. & Li, Q. Automated segmentation of lungs with severe interstitial lung disease in CT. *Med. Phys.* **36**, 4592–4599 (2009).
4. Nakamori, N., Doi, K., Sabeti, V. & MacMahon, H. Image feature analysis and computer-aided diagnosis in digital radiography: Automated analysis of sizes of heart and lung in chest images. *Med. Phys.* **17**, 342–350 (1990).
5. Hwang, S. & Park, S. Accurate lung segmentation via network-wise training of convolutional networks. in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. 92–99 (Springer, 2017).
6. Ibragimov, B., Likar, B., Pernuš, F. & Vrtovec, T. Accurate landmark-based segmentation by incorporating landmark misdetections. in *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. 1072–1075 (IEEE, 2016).
7. Yang, W. *et al.* Lung field segmentation in chest radiographs from boundary maps by a structured edge detector. *IEEE J. Biomed. Health Inform.* **22**, 842–851 (2017).
8. Van Ginneken, B., Stegmann, M. B. & Loog, M. Segmentation of anatomical structures in chest radiographs using supervised methods: A comparative study on a public database. *Med. Image Anal.* **10**, 19–40 (2006).
9. Dawoud, A. Lung segmentation in chest radiographs by fusing shape information in iterative thresholding. *IET Comput. Vis.* **5**, 185–190 (2011).
10. Li, L., Zheng, Y., Kallergi, M. & Clark, R. A. Improved method for automatic identification of lung regions on chest radiographs. *Acad. Radiol.* **8**, 629–638 (2001).
11. Iakovidis, D. K. & Papamichalis, G. Automatic segmentation of the lung fields in portable chest radiographs based on Bézier interpolation of salient control points. in *2008 IEEE International Workshop on Imaging Systems and Techniques*. 82–87 (IEEE, 2008).
12. Candemir, S. *et al.* Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE Trans. Med. Imaging* **33**, 577–590 (2013).
13. Lee, J.-S., Wu, H.-H. & Yuan, M.-Z. Lung segmentation for chest radiograph by using adaptive active shape models. *Biomed. Eng. Appl. Basis Commun.* **22**, 149–156 (2010).
14. Shao, Y. *et al.* Hierarchical lung field segmentation with joint shape and appearance sparse learning. *IEEE Trans. Med. Imaging* **33**, 1761–1780 (2014).
15. Iakovidis, D. K. & Savelonas, M. Active shape model aided by selective thresholding for lung field segmentation in chest radiographs. in *2009 9th International Conference on Information Technology and Applications in Biomedicine*. 1–4 (IEEE, 2009).
16. Novikov, A. A. *et al.* Fully convolutional architectures for multiclass segmentation in chest radiographs. *IEEE Trans. Med. Imaging* **37**, 1865–1876 (2018).
17. Fu, H., Xu, Y., Lin, S., Wong, D. W. K. & Liu, J. Deepvessel: Retinal vessel segmentation via deep learning and conditional random field. in *International Conference on Medical Image Computing and Computer-assisted Intervention*. 132–139 (Springer, 2016).
18. Ullah, I., Chikontwe, P., Choi, H., Yoon, C.-H. & Park, S. H. Synthesize and segment: Towards improved catheter segmentation via adversarial augmentation. *Appl. Sci.* **11**, 1638 (2021).
19. Ullah, I., Chikontwe, P. & Park, S. H. Catheter synthesis in X-ray fluoroscopy with generative adversarial networks. in *International Workshop on Predictive Intelligence In Medicine*. 125–133 (Springer, 2019).
20. Ullah, I., Chikontwe, P. & Park, S. H. Real-time tracking of guidewire robot tips using deep convolutional neural networks on successive localized frames. *IEEE Access* **7**, 159743–159753 (2019).
21. Gao, J., Jiang, Q., Zhou, B. & Chen, D. Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: An overview. *Math. Biosci. Eng.* **16**, 6536–6561 (2019).
22. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 234–241 (Springer, 2015).

23. Souza, J. C. *et al.* An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks. *Comput. Methods Programs Biomed.* **177**, 285–296 (2019).
24. Cheng, D. & Goldberg, M. An algorithm for segmenting chest radiographs. in *Visual Communications and Image Processing'88: Third in a Series*. Vol. 1001. 261–268 (International Society for Optics and Photonics, 1988).
25. Armato, S. G. III, Giger, M. L. & MacMahon, H. Automated lung segmentation in digitized posteroanterior chest radiographs. *Acad. Radiol.* **5**, 245–255 (1998).
26. Bartels, R. H., Beatty, J. C. & Barsky, B. A. *An Introduction to Splines for Use in Computer Graphics and Geometric Modeling* (Morgan Kaufmann, 1995).
27. Xu, X.-W. & Doi, K. Image feature analysis for computer-aided diagnosis: Detection of right and left hemidiaphragm edges and delineation of lung field in chest radiographs. *Med. Phys.* **23**, 1613–1624 (1996).
28. Ahmad, W. S. H. M. W., Zaki, W. M. D. W. & Fauzi, M. F. A. Lung segmentation on standard and mobile chest radiographs using oriented Gaussian derivatives filter. *Biomed. Eng. Online* **14**, 1–26 (2015).
29. McNitt-Gray, M. F., Sayre, J. W., Huang, H. & Razavi, M. Pattern classification approach to segmentation of chest radiographs. in *Medical Imaging 1993: Image Processing*. Vol. 1898. 160–170 (International Society for Optics and Photonics, 1993).
30. Tsujii, O., Freedman, M. T. & Mun, S. K. Automated segmentation of anatomic regions in chest radiographs using an adaptive-sized hybrid neural network. *Med. Phys.* **25**, 998–1007 (1998).
31. Vittitoe, N. F., Vargas-Voracek, R. & Floyd, C. E. Jr. Identification of lung regions in chest radiographs using Markov random field modeling. *Med. Phys.* **25**, 976–985 (1998).
32. Shi, Z. *et al.* Lung segmentation in chest radiographs by means of Gaussian kernel-based fcm with spatial constraints. in *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*. Vol. 3. 428–432 (IEEE, 2009).
33. DeGrave, A. J., Janizek, J. D. & Lee, S.-I. Ai for radiographic Covid-19 detection selects shortcuts over signal. *Nat. Mach. Intell.* 1–10 (2021).
34. Kalinovsky, A. & Kovalev, V. *Lung Image Segmentation Using Deep Learning Methods and Convolutional Neural Networks*. (2016).
35. Badrinarayanan, V., Handa, A. & Cipolla, R. Segnet: A deep convolutional encoder–decoder architecture for robust semantic pixel-wise labelling. arXiv preprint [arXiv:1505.07293](https://arxiv.org/abs/1505.07293) (2015).
36. Mittal, A., Hooda, R. & Sofat, S. Lf-segnet: A fully convolutional encoder–decoder network for segmenting lung fields from chest radiographs. *Wirel. Pers. Commun.* **101**, 511–529 (2018).
37. He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask r-cnn. in *Proceedings of the IEEE International Conference on Computer Vision*. 2961–2969 (2017).
38. Wang, J., Li, Z., Jiang, R. & Xie, Z. Instance segmentation of anatomical structures in chest radiographs. in *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. 441–446 (IEEE, 2019).
39. Peng, T., Xu, T. C., Wang, Y. & Li, F. Deep belief network and closed polygonal line for lung segmentation in chest radiographs. *Comput. J.* (2020).
40. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4510–4520 (2018).
41. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 770–778 (2016).
42. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4700–4708 (2017).
43. Tan, M. & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. in *International Conference on Machine Learning*. 6105–6114 (PMLR, 2019).
44. Hu, J., Shen, L. & Sun, G. Squeeze-and-excitation networks. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 7132–7141 (2018).
45. Ioffe, S. & Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. in *International Conference on Machine Learning*. 448–456 (PMLR, 2015).
46. Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T. M. & Asari, V. K. Improved inception-residual convolutional neural network for object recognition. *Neural Comput. Appl.* **32**, 279–293 (2020).
47. Jaeger, S. *et al.* Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quant. Imaging Med. Surg.* **4**, 475 (2014).
48. Shiraishi, J. *et al.* Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *Am. J. Roentgenol.* **174**, 71–74 (2000).
49. Tang, Y.-B., Tang, Y.-X., Xiao, J. & Summers, R. M. Xlsor: A robust and accurate lung segmentor on chest X-rays using criss-cross attention and customized radiorealistic abnormalities generation. in *International Conference on Medical Imaging with Deep Learning*. 457–467 (PMLR, 2019).
50. Chaurasia, A. & Culurciello, E. Linknet: Exploiting encoder representations for efficient semantic segmentation. in *2017 IEEE Visual Communications and Image Processing (VCIP)*. 1–4 (IEEE, 2017).
51. Zhao, H., Shi, J., Qi, X., Wang, X. & Jia, J. Pyramid scene parsing network. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2881–2890 (2017).
52. Badrinarayanan, V., Kendall, A. & Cipolla, R. Segnet: A deep convolutional encoder–decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 2481–2495 (2017).
53. Chandra, T. B., Verma, K., Jain, D. & Netam, S. S. Segmented lung boundary correction in chest radiograph using context-aware adaptive scan algorithm. in *Advances in Biomedical Engineering and Technology*. 263–275 (Springer, 2021).
54. Gomez, O., Mesejo, P., Ibanez, O., Valsecchi, A. & Cordon, O. Deep architectures for high-resolution multi-organ chest X-ray image segmentation. *Neural Comput. Appl.* **32**, 15949–15963 (2020).
55. Bi, L., Feng, D. & Kim, J. Dual-path adversarial learning for fully convolutional network (FCN)-based medical image segmentation. *Vis. Comput.* **34**, 1043–1052 (2018).
56. Wang, C. Segmentation of multiple structures in chest radiographs using multi-task fully convolutional networks. in *Scandinavian Conference on Image Analysis*. 282–289 (Springer, 2017).
57. Sorensen, T. A. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on danish commons. *Biol. Skar.* **5**, 1–34 (1948).
58. Jaccard, P. The distribution of the flora in the alpine zone. 1. *New Phytol.* **11**, 37–50 (1912).
59. Beauchemin, M., Thomson, K. P. & Edwards, G. On the Hausdorff distance used for the evaluation of segmentation results. *Can. J. Remote Sens.* **24**, 3–8 (1998).
60. Jangam, E. & Rao, A. Segmentation of lungs from chest X rays using firefly optimized fuzzy c-means and level set algorithm. in *International Conference on Recent Trends in Image Processing and Pattern Recognition*. 303–311 (Springer, 2018).
61. Bosdelekidis, V. & Ioakeimidis, N. S. Lung field segmentation in chest X-rays: A deformation-tolerant procedure based on the approximation of rib cage seed points. *Appl. Sci.* **10**, 6264 (2020).
62. Liu, M. & Yin, H. Feature pyramid encoding network for real-time semantic segmentation. arXiv preprint [arXiv:1909.08599](https://arxiv.org/abs/1909.08599) (2019).
63. Vinogradova, K., Dibrov, A. & Myers, G. Towards interpretable semantic segmentation via gradient-weighted class activation mapping (student abstract). *Proc. AAAI Conf. Artif. Intell.* **34**, 13943–13944 (2020).

64. Selvaraju, R. R. *et al.* Grad-cam: Visual explanations from deep networks via gradient-based localization. in *Proceedings of the IEEE International Conference on Computer Vision*. 618–626 (2017).
65. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014).

### Acknowledgements

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (MSIT) (No. 2019R1C1C1008727), and by the grant of the medical device technology development program funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea) (20006006). This research work was also supported by the Cluster grant R20143 of Zayed University

### Author contributions

I.U. contributed to the conceptualization of this study, conducted experiments, wrote the original draft, and revised the manuscript, S.H.P. and F.A. contributed to the experimental plan, supervised the work and revised the manuscript, B.S, S.S. and T.A. contributed to the evaluation of the developed technique, analysis of results and revised the manuscript. All authors reviewed the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to S.H.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023, corrected publication 2023