XLSor: A Robust and Accurate Lung Segmentor on Chest X-Rays Using Criss-Cross Attention and Customized Radiorealistic Abnormalities Generation

You-Bao Tang[*1] YOUBAO.TANG@NIH.GOV
Yu-Xing Tang[*1] YUXING.TANG@NIH.GOV
Jing Xiao[2] XIAOJING661@PINGAN.COM.CN
Ronald M. Summers[1] RMS@NIH.GOV
[1] *Imaging Biomarkers and Computer-Aided Diagnosis Laboratory, Radiology and Imaging Sciences, National Institutes of Health Clinical Center, Bethesda, MD 20892-1182, USA*
[2] *Ping An Insurance Company of China, Shenzhen, 510852, China*

Abstract

This paper proposes a novel framework for lung segmentation in chest X-rays. It consists of two key contributions, a criss-cross attention based segmentation network and radiorealistic chest X-ray image synthesis (*i.e.* a synthesized radiograph that appears anatomically realistic) for data augmentation. The criss-cross attention modules capture rich global contextual information in both horizontal and vertical directions for all the pixels thus facilitating accurate lung segmentation. To reduce the manual annotation burden and to train a robust lung segmentor that can be adapted to pathological lungs with hazy lung boundaries, an image-to-image translation module is employed to synthesize radiorealistic abnormal CXRs from the source of normal ones for data augmentation. The lung masks of synthetic abnormal CXRs are propagated from the segmentation results of their normal counterparts, and then serve as pseudo masks for robust segmentor training. In addition, we annotate 100 CXRs with lung masks on a more challenging NIH Chest X-ray dataset containing both posterioranterior and anteroposterior views for evaluation. Extensive experiments validate the robustness and effectiveness of the proposed framework. The code and data can be found from https://github.com/rsummers11/CADLab/tree/master/Lung_Segmentation_XLSor.

Keywords: Lung segmentation, chest X-ray, criss-cross attention, radiorealistic data augmentation

1. Introduction

Lung diseases and disorders are one of the leading causes of death and hospitalization throughout the world. According to the American Lung Association, lung cancer is the number one cancer killer of both women and men in the United States, and more than 33 million Americans are facing a chronic lung disease. The chest radiograph (chest X-ray, or CXR) is one of the most requested radiologic examination for pulmonary diseases such as lung cancer, chronic obstructive pulmonary disease (COPD), pneumonia, tuberculosis, etc. There are huge demands on developing computer-aided diagnosis/detection (CADx/CADe) methods to assist radiologists and other physicians in reading and comprehending chest X-ray images (Shin et al., 2016; Wang et al., 2017, 2018b; Tang et al., 2018c), given the fact that there is a shortage of experienced radiologists, especially in developing

[*] Contributed equally

countries. Precise segmentation of lung fields can provide rich structural information such as shape irregularity, size measurement and total lung volume, which further facilitates subsequent stages of automated diagnosis (*e.g.*, disease pattern recognition, segmentation and quantization) to assess certain serious clinical conditions.

Over the past decades, automated segmentation of lung boundaries in CXR has received substantial attention in the literature (Candemir et al., 2014; Dai et al., 2017) but still remained a challenging problem (El-Baz et al., 2016). Previous work mainly adopted hand-crafted features to design rule-based systems (Li et al., 2001), active shape/appearance models (Xu et al., 2012), or their hybrid methods (Candemir et al., 2014) to segment the lung boundaries. These approaches rely on the test CXR images being well modeled by the existing training images but they may fail on a different distribution or population. Recently, deep learning based methods (e.g. fully convolutional neural networks (FCN) (Shelhamer et al., 2017)) have achieved great successes in biomedical image segmentation (Chen et al., 2018; Tang et al., 2019a; Cai et al., 2018; Tang et al., 2018a) and other medical image analysis tasks (Tang et al., 2019d,c,b, 2018b; Jin et al., 2018; Yan et al., 2018, 2019). The FCN-based methods are intrinsically limited to local receptive fields and insufficient contextual information due to the fixed geometric structures of the convolution. These limitations impose unfavorable effects in segmenting boundaries around less clear lung regions caused by pathological conditions or poor image quality (*e.g.*, low contrast, costophrenic angle clipped off, bad positioning of the patient). Structure correcting adversarial network (SCAN) (Dai et al., 2017) incorporates FCN and adversarial learning (Goodfellow et al., 2014) to segment organs (lungs and heart) in CXRs. SCAN imposes regularization based on the physiological (global) structures by using a critic network that discriminates between the ground truth annotations from the segmentation masks generated by the FCN.

In order to capture richer global contextual information for robust and accurate lung segmentation, we make use of a criss-cross attention (CCA) module (Huang et al., 2018b) to aggregate long-range pixel-wise contextual information in both horizontal and vertical directions. Further dense contextual information can be achieved by stacking more CCA modules recurrently to cover all the pixels. In addition, since publicly available datasets only contain small numbers of lung masks and they are mainly for normal lungs and lungs with subtle findings or unique pathology in an posterioranterior view (*e.g.*, small nodules within the lung field in the JSRT database (Shiraishi et al., 2000), CXRs with tuberculosis presented in the Montgomery database (Jaeger et al., 2014)), it is insufficient to directly use these datasets for training a powerful lung segmentor that can be adapted to pathological lungs with hazy lung boundaries (*e.g.*, large masses, pneumonias, effusions, etc.) for both posterioranterior (PA) and anteroposterior (AP) views. Furthermore, it is very time consuming and tedious for radiologists to manually annotate lung masks, especially on CXRs with abnormalities/pathologies in lung regions (or the so-called abnormal CXRs in this paper). Therefore, we use an image-to-image translation method (Huang et al., 2018a) to synthesize radiorealistic (*i.e.* a synthesized radiograph that appears anatomically realistic) abnormal CXRs from the source of normal ones for data augmentation and mask propagation. The lung masks of synthetic abnormal CXRs are transferred from their normal counterpart and then used as pseudo masks for segmentor retraining.

The proposed framework **XLSor** (*i.e.* **X**-ray **L**ung **S**egment**or**) takes advantage of radiorealistic synthesized abnormal CXRs and pseudo masks, without requiring paired normal and abnormal CXRs from the same patient (which is infeasible in reality), as well as the criss-cross attention module to generate robust and accurate lung segmentation. We annotate 100 lung masks on a more

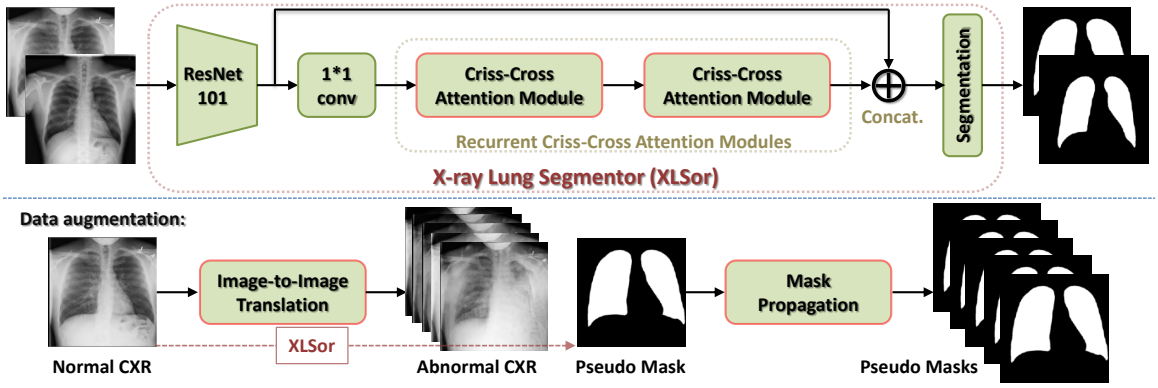


Figure 1: Framework of the proposed X-ray lung segmentor (XLSor).

challenging NIH Chest X-ray dataset (Wang et al., 2017) containing both PA and AP views for evaluation. Extensive experiments on different datasets validate the robustness and effectiveness of the proposed framework.

2. Methodology

2.1. XLSor Framework Overview

The overall XLSor framework is shown in Figure 1. Given a training set R with ground-truth masks, an initial lung segmentor is trained (see details in Sec. 2.2). Then, for an auxiliary external set, an image-to-image method MUNIT (Huang et al., 2018a) is used to synthesize abnormal CXRs from normal ones, so as to augment the training data and pseudo mask annotations (mask of normal CXR is obtained using the initial lung segmentor and propagated to its synthesized abnormal CXRs, see details in Sec. 2.3). The initial lung segmentor is updated using R along with the augmented dataset A with pseudo masks.

2.2. Criss-Cross Attention based Network for Lung Segmentation

In preliminary experiments, we trained a U-Net model (Ronneberger et al., 2015), a widely used model in many applications of medical image segmentation, for lung segmentation. When testing it on the unseen abnormal CXRs, the segmentations are not very promising. That is because the features are extracted from local receptive fields and cannot well capture sufficient contextual information of lungs in U-Net. However, the rich and global contextual information of lungs and their surrounding regions is very important for lung segmentation.

Criss-cross Network (CCNet) (Huang et al., 2018b) achieved state-of-the-art performance in semantic segmentation based on a novel criss-cross attention (CCA) module. Inspired by this, we employ CCA to build a robust and accurate lung segmentor (named XLSor) on chest X-rays. The XLSor is constructed with a fully convolutional network and two CCA modules to capture long-range contextual information (see Figure 1 top). Specifically, we replace the last two down-sampling layers in the ImageNet pre-trained ResNet-101 (He et al., 2016) with dilated convolution operation (Chen et al., 2015), resulting in an output stride of 8. The CCA module collects contextual information in horizontal and vertical directions to enhance pixel-wise representative capability.

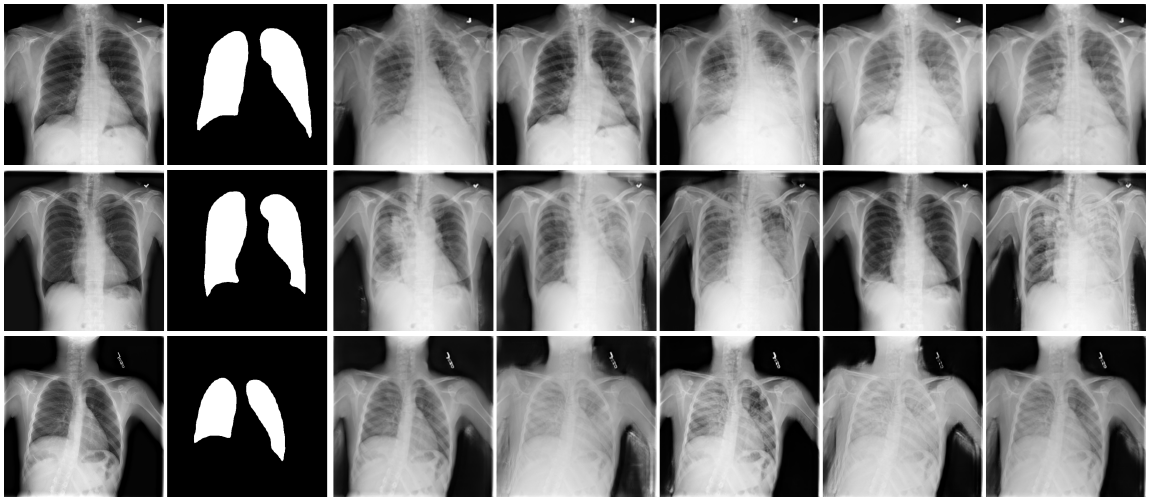
Recurrent criss-cross attention module can capture dense contextual information from all pixels by stacking two CCA modules with shared weights. CCA shares the similar idea of capturing global contextual information as the non-local neural network (Wang et al., 2018a) but with much higher computational efficiency. Please refer to (Huang et al., 2018b) for more details about the CCA module. Therefore, the CCA based XLSor can generate clear lung boundaries for more accurate lung segmentation by considering the richer and global contextual information.

The mean square error loss function and the SGD with momentum of 0.9 and weight decay of 0.0005 are used to optimize the XLSor. The initial learning rate is 0.02 and updated using a poly learning rate policy where the initial learning rate is multiplied by $1-(\frac{iter}{max\_iter})^{0.9}$, where $iter$ is the number of current iterations and max_iter is the total number of iterations. The batch size is set as 4. The size of the input CXR is 512×512.

2.3. Data Augmentation via Abnormal Chest X-Ray Pairs Construction

As discussed in Sec. 1, it is insufficient to train a robust lung segmentor using the existing datasets and mask annotations. A simple solution is to enrich the training data, which has been widely used in deep learning. The traditional data augmentation means is to use a combination of affine transformations to manipulate the training data, *e.g.*, shifting, zooming in/out, rotation, flipping, etc, so as to generate new duplicate images for each input image. The contextual information in these generated images do not change very much. To solve these problems, we propose a data augmentation strategy using an image-to-image translation method (Huang et al., 2018a) to construct a large number of abnormal chest X-ray pairs without involving any human intervention, based on which a powerful model can be learned for robust and accurate lung segmentation on different challenging CXRs.

To construct the pairs of abnormal CXR and its corresponding lung masks, there are two straightforward ways. One is to convert the abnormal CXRs into normal ones, and then compute the lung masks which serve as the ground truths for the abnormal CXRs. The other one is to convert the normal CXRs into abnormal ones, and then the lung masks segmented on the normal CXRs are considered as the ground truths of the abnormal ones. Here, we prefer the second way, since the lung regions in real normal CXRs are determined while the ones could be different for various generated normal CXRs in the first way. For the image-to-image translation task, *i.e.* from normal CXRs to abnormal ones, a state-of-the-art method, *i.e.* MUNIT (Huang et al., 2018a), is utilized in this work. MUNIT assumes that the image representation can be decomposed into a content code that is domain-invariant, and a style code that captures domain-specific properties. To translate an image to another domain, MUNIT recombines its content code with a random style code sampled from the style space of the target domain. Please refer to (Huang et al., 2018a) for more details about MUNIT. In this work, we first train the MUNIT model using the default parameter configuration and the NIH chest X-Ray dataset (Wang et al., 2017), from which 5,000 normal CXRs and 5,000 abnormal CXRs are randomly selected for training. Then, given a normal CXR (see Figure 2(a)), we use the trained MUNIT model to generate (or synthesize) a number of abnormal CXRs (see Figure 2(c)-(g)) by combining the content code of the normal CXR and different random style codes learned from the domain of abnormal CXRs. From Figure 2(c)-(g), we can see that the generated abnormal CXRs are radiorealistic. We also notice that the shape of lungs are distorted slightly in the generated abnormal CXRs sometimes. Therefore, the generated abnormalities are customized using the style codes and visually radiorealistic. At last, we use the initial XLSor model trained from the publicly available datasets to obtain the lung masks (see Figure 2(b)) of the given normal



(a) (b) (c) (d) (e) (f) (g)

Figure 2: Three examples (rows) of the constructed abnormal CXR pairs. Given an unseen normal CXR (a), XLSor outputs a lung segmentation that is binarized with a threshold of 0.5 to get the lung mask (b) and MUNIT generates different abnormal CXRs (c-g). The lung mask (b) and the synthesized abnormal CXRs (c-g) form the constructed abnormal CXR pairs.

CXR, which are also considered as the pseudo masks of the generated abnormal CXRs (*i.e.* mask propagation) to form the constructed abnormal CXR pairs (see Figure 1 bottom) for further training the XLSor model. We also iteratively conducted above processes and found that it is not helpful because the normal CXRs are easy to segment and the pseudo masks are good enough at the first iteration.

3. Experiments

3.1. Datasets and Evaluation Criteria

We evaluate the lung segmentation performance of the proposed XLSor using two publicly available datasets, *i.e.* JSRT (Shiraishi et al., 2000) and Montgomery (Jaeger et al., 2014), and our own annotated dataset (named NIH). **JSRT** contains 247 CXRs, among which 154 have lung nodules and 93 have no lung nodule. **Montgomery** contains 138 CXRs, including 80 normal patients and 58 patients with manifested tuberculosis (TB). Both datasets provide pixel-wise lung mask annotations. We notice that the abnormal lung regions in these two datasets are mild. Only using such datasets for evaluation cannot well demonstrate the effectiveness and generalizability of the methods, since diseases can occasionally cause severe damages to the lungs. Therefore, we manually annotate the lung masks of 100 abnormal CXRs with various severity of lung diseases, which are selected from the NIH Chest X-Ray dataset (Wang et al., 2017) by excluding the samples used for MUNIT training. Here, we name the manually labeled set as **NIH**.

JSRT and Montgomery datasets are combined and randomly split into three subsets for both normal and abnormal CXRs, *i.e.* training (70%), validation (10%) and testing (20%). Specifically, the validation and testing sets include 37 and 78 CXRs, respectively. The remaining 280 CXRs serve as a training set for model training. The validation set is used for model selection, and the testing set and the NIH dataset are used for performance evaluation. Five criteria, *i.e.* volumetric similarity (VS), averaged Hausdorff distance (AVD), Dice similarity coefficient (DICE), precision (PRE) and recall (REC) scores, are calculated pixel-wisely by a publicly available segmentation evaluation tool (Taha and Hanbury, 2015) with threshold of 0.5 and used to evaluate the quantitative segmentation performance.

3.2. Quantitative Results

In this work, U-Net (Ronneberger et al., 2015) is applied for performance comparisons to demonstrate the effectiveness of the criss-cross attention based XLSor. To validate the usefulness of adding the augmented samples for lung segmentation, we first use the proposed data augmentation strategy to generate four augmented training sets, denoted as A^1, A^2, A^3 and A^4, respectively. Here, A^1 contains 600 constructed pairs including 100 normal pairs and 500 abnormal pairs where five abnormal CXRs are synthesized from each normal CXR using MUNIT (Huang et al., 2018a). A^i $(i = 2,3,4)$ contains all samples in A^{i-1} and another new 600 constructed pairs. We then train the XLSor and U-Net models for lung segmentation using six different training settings, *i.e.* only using the real public training set (denoted R), using the real public training set and any augmented set A^i $i = 1,2,3,4$ (denoted $R+A^i$), and only using the augmented set A^4. To validate the effectiveness of CCA for segmentation performance improvement, we also train the XLSor model without CCA modules (denoted XLSor$^-$) and the U-Net model with CCA modules (denoted U-Net$^+$) using R and $R+A^4$. In each training setting, the same traditional data augmentation techniques (*e.g.*, scaling and flipping) are adopted. Finally, the five criteria are used to evaluate the performance of lung segmentation on the public testing set and NIH dataset, whose results are reported in Table 1.

From Table 1, we can see that 1) the proposed XLSor gets better results than U-Net on both the simple public testing set and the difficult NIH dataset. Especially, the performance of XLSor$_R$ is much better than the one of U-Net$_R$ on the NIH dataset (*e.g.*, improving the Dice score about 12%), meaning that the proposed XLSor is able to work much better than U-Net on the unseen CXRs whose data distribution is much different from the training data. This demonstrates that the proposed XLSor based on the criss-cross attention module can well learn the global contextual information of lung regions and strong discriminative features to distinguish the lung regions from their surrounding structures regardless of the CXRs' properties. 2) When adding the augmented samples for model training, the performance is improved, *i.e.* XLSor$_{R+A^i}$ (or U-Net$_{R+A^i}$) gets better results than XLSor$_R$ (or U-Net$_R$), suggesting the effectiveness of our data augmentation technique for lung segmentation performance improvement. Through experiments, we find that the performance remains stable when adding more augmented samples than A^4. 3) When only using the augmented samples for model training, both XLSor and U-Net still get very promising performance on the public testing set and the NIH dataset (see the results of XLSor$_{A^4}$ and U-Net$_{A^4}$ in Table 1), suggesting that the generated abnormal CXRs are radiorealistic and the pseudo lung masks effectively supervise the learning processes for lung segmentation. 4) The results by all models are quite similar in the public testing set, that is because the testing CXRs are all (near-)normal and the lung segmentation task is relatively easy. 5) U-Net obtains worse performance on NIH dataset than the

Table 1: Lung segmentation results on the **public testing set** and **NIH dataset** using the proposed XLSor and U-Net with different training settings. Results showing mean with standard deviation. ↑: the larger the better. ↓: the smaller the better.

Method	REC↑	PRE↑	DICE↑	AVD↓	VS↑
Public testing set					
XLSor$_{R}$	0.973±0.02	**0.979±0.02**	**0.976±0.01**	0.149±0.51	**0.992±0.01**
XLSor$_{R+A^1}$	0.973±0.02	**0.979±0.02**	**0.976±0.01**	0.152±0.52	0.991±0.01
XLSor$_{R+A^2}$	**0.974±0.02**	0.978±0.02	**0.976±0.01**	**0.117±0.31**	0.991±0.01
XLSor$_{R+A^3}$	0.972±0.02	**0.979±0.02**	**0.976±0.01**	0.126±0.33	0.991±0.01
XLSor$_{R+A^4}$	**0.974±0.02**	0.977±0.02	**0.976±0.01**	0.146±0.44	0.991±0.01
XLSor$_{A^4}$	0.965±0.03	**0.979±0.02**	0.972±0.02	0.162±0.36	0.989±0.01
XLSor$^{-}_{R}$	0.973±0.02	0.978±0.02	0.975±0.01	0.151±0.53	0.991±0.01
XLSor$^{-}_{R+A^4}$	0.972±0.02	0.978±0.02	0.976±0.01	0.148±0.47	0.991±0.01
U-Net$_{R}$	0.976±0.02	0.968±0.03	0.972±0.02	0.198±0.56	0.988±0.02
U-Net$_{R+A^1}$	0.973±0.02	0.976±0.02	0.974±0.01	0.162±0.54	**0.990±0.01**
U-Net$_{R+A^2}$	**0.977±0.02**	0.973±0.02	**0.975±0.01**	0.135±0.41	0.989±0.01
U-Net$_{R+A^3}$	0.976±0.02	0.975±0.02	**0.975±0.01**	**0.131±0.34**	**0.990±0.01**
U-Net$_{R+A^4}$	0.973±0.02	**0.978±0.01**	**0.975±0.01**	0.152±0.46	**0.990±0.01**
U-Net$_{A^4}$	0.967±0.02	0.975±0.01	0.971±0.01	0.164±0.37	0.989±0.01
U-Net$^{+}_{R}$	0.976±0.02	0.970±0.03	0.972±0.02	0.191±0.54	0.988±0.02
U-Net$^{+}_{R+A^4}$	0.975±0.02	0.977±0.01	0.975±0.01	0.130±0.33	0.990±0.01
NIH dataset					
XLSor$_{R}$	0.966±0.02	0.927±0.09	0.943±0.05	0.669±1.64	0.966±0.05
XLSor$_{R+A^1}$	0.958±0.03	0.973±0.02	0.965±0.02	0.172±0.26	0.985±0.01
XLSor$_{R+A^2}$	0.962±0.02	0.980±0.01	0.971±0.01	0.097±0.08	0.989±0.01
XLSor$_{R+A^3}$	0.967±0.02	0.978±0.02	0.973±0.01	0.089±0.07	0.990±0.01
XLSor$_{R+A^4}$	**0.974±0.01**	0.976±0.01	**0.975±0.01**	**0.078±0.06**	**0.993±0.01**
XLSor$_{A^4}$	0.964±0.02	**0.983±0.01**	0.973±0.01	0.098±0.13	0.988±0.01
XLSor$^{-}_{R}$	0.965±0.03	0.902±0.10	0.929±0.06	0.952±1.81	0.955±0.06
XLSor$^{-}_{R+A^4}$	0.965±0.02	0.981±0.01	0.967±0.01	0.093±0.10	0.990±0.01
U-Net$_{R}$	0.938±0.07	0.761±0.20	0.823±0.16	5.231±9.02	0.869±0.15
U-Net$_{R+A^1}$	0.926±0.05	**0.960±0.03**	0.942±0.03	0.832±1.29	0.971±0.02
U-Net$_{R+A^2}$	0.947±0.04	0.950±0.04	0.948±0.03	0.500±1.03	0.981±0.02
U-Net$_{R+A^3}$	0.950±0.03	0.954±0.03	0.951±0.02	0.393±0.58	0.983±0.02
U-Net$_{R+A^4}$	0.943±0.04	0.958±0.03	0.950±0.03	0.454±0.73	0.982±0.02
U-Net$_{A^4}$	**0.952±0.03**	0.959±0.03	**0.955±0.02**	**0.315±0.47**	**0.983±0.02**
U-Net$^{+}_{R}$	0.929±0.07	0.804±0.20	0.842±0.14	4.782±8.05	0.895±0.14
U-Net$^{+}_{R+A^4}$	0.956±0.03	0.969±0.02	0.962±0.02	0.262±0.54	0.985±0.02

public testing set, meaning that the CXRs in the NIH dataset are more complex and difficult than the ones in the public testing set. But XLSor can get comparable and good results on both datasets, sug-

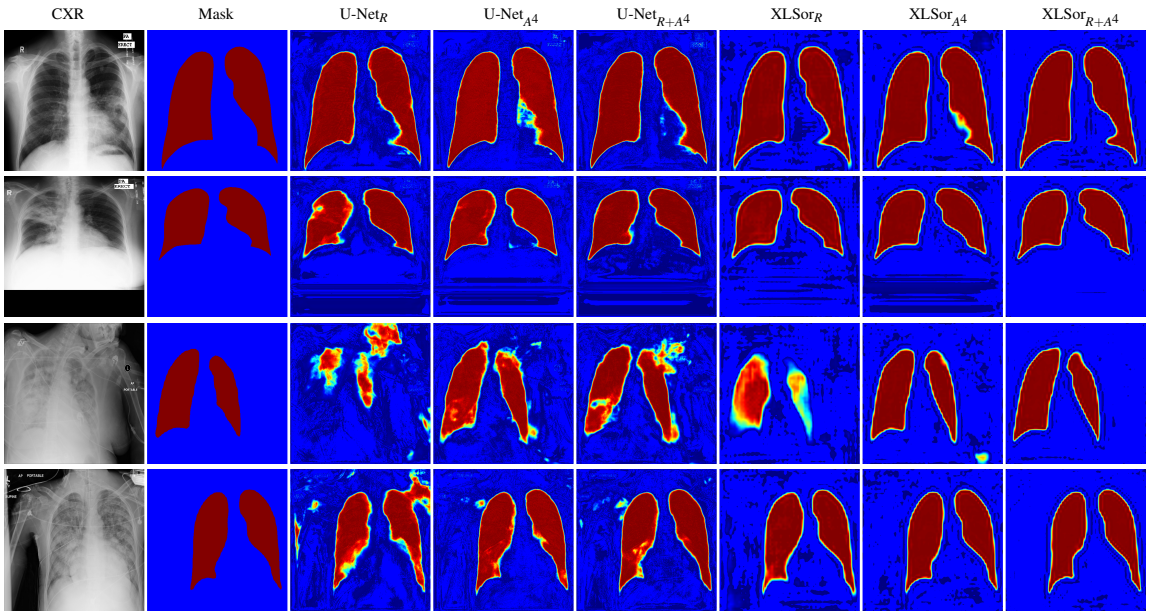


Figure 3: Four examples (rows) of lung segmentation results produced by XLSor and U-Net trained using R, A^4 and $R+A^4$. Here, the results are given as the probability maps directly outputted by the models, which can be binarized with a threshold of 0.5 to get the binary lung masks for performance evaluation. The first two rows are from the public testing set and the last two rows are from the NIH dataset. To better visualize the differences between lung segmentation results and ground truths, we colorize them with pseudo-colors. Better viewed in color.

gesting that the proposed XLSor is robust and powerful for lung segmentation in different scenarios. 6) XLSor/U-Net$^+$ achieves better results than XLSor$^-$/U-Net (especially, on the NIH dataset), suggesting that using CCA modules can make the model learn the global contextual information of lung regions better and extract more powerful discriminative features for performance improvement. All results quantitatively demonstrate the effectiveness and generalizability of the proposed XLSor for lung segmentation on various CXRs.

3.3. Qualitative Results

Figure 3 shows four qualitative lung segmentation results produced by the models (*i.e.* XLSor and U-Net) trained with the following settings: R, A^4 and $R+A^4$. Compared with U-Net, the lung segmentation results produced by the proposed XLSor are much closer to the ground truths in various challenging scenarios. To be specific, 1) the proposed XLSor not only highlights the correct lung regions clearly, but also well suppresses the probabilities of background regions, so as to produce the segmentation results with higher contrast between lung regions and background than U-Net. 2) With the help of the criss-cross attention module that considers sufficient contextual information, the proposed XLSor is able to output the lung segmentations with clear boundaries

and consistent probabilities, even when the model is trained and tested on CXRs with different distribution of abnormalities. 3) With the augmented samples for training, the qualities of lung segmentations are improved. These intuitively demonstrate the effectiveness of the proposed XLSor and the usefulness of the proposed data augmentation strategy for lung segmentation on chest X-rays.

4. Conclusions and Future Work

In this paper, we propose a robust and accurate lung segmentor based on a criss-cross attention network and a customized radiorealistic abnormalities generation technique for data augmentation. Experiments showed that the proposed framework was able to capture rich contextual information from both original and radiorealistic synthesized CXRs to adapt to more challenging images, resulting in much better segmentation, especially in unseen abnormal CXRs. Future work includes segmenting more organs and integrating with more downstream tasks such as disease classification and detection to provide comprehensive and accurate computer-aided detection on CXR images, *e.g.*, performing segmentation and classification simultaneously by training different MUNIT models for individual diseases and using them to generate abnormalities accordingly in categories.

Acknowledgments

This research was supported by the Intramural Research Program of the National Institutes of Health Clinical Center and by the Ping An Insurance Company through a Cooperative Research and Development Agreement. We thank Nvidia for GPU card donation.

References

J. Cai, Y. Tang, L. Lu, A. P Harrison, K. Yan, J. Xiao, L. Yang, and R. M. Summers. Accurate weakly-supervised deep lesion segmentation using large-scale clinical annotations: Slice-propagated 3D mask generation from 2D RECIST. In *MICCAI*, 2018.

S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE TMI*, 33(2):577–590, 2014.

L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, and D. Rueckert. Drinet for medical image segmentation. *IEEE TMI*, 37(11):2453–2462, 2018.

L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Semantic image segmentation with deep convolutional nets and fully connected CRFs. In *ICLR*, 2015.

W. Dai, J. Doyle, X. Liang, H. Zhang, N. Dong, Y. Li, and E. P. Xing. Scan: Structure correcting adversarial network for chest x-rays organ segmentation. *arXiv preprint arXiv:1703.08770*, 2017.

A. El-Baz, X. Jiang, and J.S. Suri. *Biomedical Image Segmentation: Advances and Trends*. CRC Press, Taylor & Francis Group, 2016. ISBN 9781482258554.

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *NeurIPS*, 2014.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

X. Huang, M.Y. Liu, S. Belongie, and J. Kautz. Multimodal unsupervised image-to-image translation. In *ECCV*, 2018a.

Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu. CCNet: Criss-cross attention for semantic segmentation. *arXiv preprint arXiv:1811.11721*, 2018b.

S. Jaeger, S. Candemir, Y. X. Antani, S.and Wang, P. X. Lu, and G. Thoma. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *QIMS*, 4(6):475–477, 2014.

D. Jin, Z. Xu, Y. Tang, A. P. Harrison, and D. J. Mollura. CT-realistic lung nodule simulation from 3D conditional generative adversarial networks for robust lung segmentation. In *MICCAI*, 2018.

L. Li, Y. Zheng, M. Kallergi, and R. A. Clark. Improved method for automatic identification of lung regions on chest radiographs. *AR*, 8(7):629 – 638, 2001.

O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015.

E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *IEEE TPAMI*, 39(4):640–651, 2017.

H. C. Shin, K. Roberts, L. Lu, D. Demner-Fushman, J. Yao, and R. M. Summers. Learning to read chest X-rays: Recurrent neural cascade model for automated image annotation. In *CVPR*, 2016.

J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi. Development of a digital image database for chest radiographs with and without a lung nodule. *AJR*, 174(1):71–74, 2000.

A. A. Taha and A. Hanbury. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC MI*, 15(1):29, 2015.

Y. Tang, J. Cai, L. Lu, A. P. Harrison, K. Yan, J. Xiao, L. Yang, and R. M. Summers. CT image enhancement using stacked generative adversarial networks and transfer learning for lesion segmentation improvement. In *MLMI*, 2018a.

Y. Tang, A. P. Harrison, M. Bagheri, J. Xiao, and R. M. Summers. Semi-automatic RECIST labeling on CT scans with cascaded convolutional neural networks. In *MICCAI*, 2018b.

Y. Tang, X. Wang, A. P. Harrison, L. Lu, J. Xiao, and R. M. Summers. Attention-guided curriculum learning for weakly supervised classification and localization of thoracic diseases on chest radiographs. In *MLMI*, 2018c.

Y. B. Tang, S. Oh, Y. X. Tang, J. Xiao, and R. M. Summers. CT-realistic data augmentation using generative adversarial network for robust lymph node segmentation. In *Medical Imaging: CAD*, 2019a.

Y. B. Tang, K. Yan, Y. X. Tang, J. Liu, J. Xiao, and R. M. Summers. ULDor: A universal lesion detector for CT scans with pseudo masks and hard negative example mining. In *ISBI*, 2019b.

Y. X. Tang, Y. B. Tang, M. Han, J. Xiao, and R. M. Summers. Abnormal chest X-ray identification with generative adversarial one-class classifier. In *ISBI*, 2019c.

Y. X. Tang, Y. B. Tang, M. Han, J. Xiao, and R. M. Summers. Deep adversarial one-class learning for normal and abnormal chest radiograph classification. In *Medical Imaging: CAD*, 2019d.

X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers. ChestX-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *CVPR*, 2017.

X. Wang, R. Girshick, A. Gupta, and K. He. Non-local neural networks. In *CVPR*, 2018a.

X. Wang, Y. Peng, L. Lu, Z. Lu, and R. M. Summers. Tienet: Text-image embedding network for common thorax disease classification and reporting in chest X-rays. In *CVPR*, 2018b.

T. Xu, M. Mandal, R. Long, I. Cheng, and A. Basu. An edge-region force guided active shape approach for automatic lung field detection in chest radiographs. *CMIG*, 36(6):452 – 463, 2012.

K. Yan, X. Wang, L. Lu, L. Zhang, A. P. Harrison, M. Bagheri, and R. M. Summers. Deep lesion graphs in the wild: relationship learning and organization of significant radiology image findings in a diverse large-scale lesion database. In *CVPR*, 2018.

K. Yan, Y. Peng, Z. Lu, and R. M. Summers. Fine-grained lesion annotation in CT images with knowledge mined from radiology reports. In *CVPR*, 2019.