Structure Correction for Robust Volume Segmentation in Presence of Tumors

Pranjal Sahu, Yiyuan Zhao, Parmeet Bhatia, Luca Bogoni, Anna Jerebko, and Hong Qin

***Abstract*—CNN based lung segmentation models in absence of diverse training dataset fail to segment lung volumes in presence of severe pathologies such as large masses, scars, and tumors. To rectify this problem, we propose a multi-stage algorithm for lung volume segmentation from CT scans. The algorithm uses a 3D CNN in the first stage to obtain a coarse segmentation of the left and right lungs. In the second stage, shape correction is performed on the segmentation mask using a 3D structure correction CNN. A novel data augmentation strategy is adopted to train a 3D CNN which helps in incorporating global shape prior. Finally, the shape corrected segmentation mask is up-sampled and refined using a parallel flood-fill operation. The proposed multi-stage algorithm is robust in the presence of large nodules/tumors and does not require labeled segmentation masks for entire pathological lung volume for training. Through extensive experiments conducted on publicly available datasets such as NSCLC, LUNA, and LOLA11 we demonstrate that the proposed approach improves the recall of large juxtapleural tumor voxels by at least 15% over state-of-the-art models without sacrificing segmentation accuracy in case of normal lungs. The proposed method also meets the requirement of CAD software by performing segmentation within 5 seconds which is significantly faster than present methods.**

***Index Terms*—Lung volume segmentation, tumors, CNN.**

I. INTRODUCTION

A. Problem Formulation and Prior Work

LUNG VOLUME segmentation is a primary step in any lung computer aided diagnosis (CAD) system for nodule detection from CT scans [1]–[3]. Convolution Neural Networks (CNN) based segmentation models have shown huge success in performing segmentation and models such as U-Net, 3D U-Net, V-Net [4]–[7] have become ubiquitous. However, one major limitation of these methods is that during the training phase they require labelled data for pathological cases as well. A CNN model trained entirely on normal lung volumes will fail to segment the abnormality as shown in Fig. 1. The primary reason for such failures is that CNN based segmentation algorithms such as U-Net, V-Net are trained using cross entropy or dice score based loss functions and have no understanding of the shape characteristics of the object to be segmented [8] leading to failures in above mentioned cases. Typical solution for this problem is to collect human annotated segmentation masks for pathological cases and adding it in the training set. However, obtaining such masks from expert radiologists is very labour intensive and not cost effective.

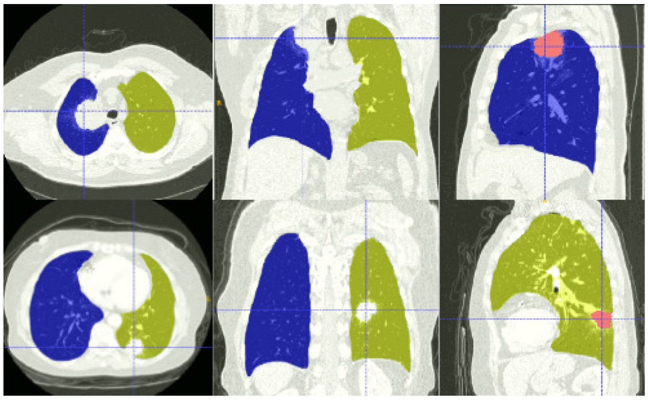


Fig. 1. Figure illustrating two pathological lung CT volumes where a large tumor is not captured by the segmentation algorithm (U-Net) due to similar intensity values as the non-lung volume. Each row shows the axial, coronal and sagittal sections of the pathological CT volume. Red colored region in coronal section highlights the mis-segmented tumor.

Classical lung segmentation methods typically rely on CT intensity to perform segmentation and such methods also fail to segment cases with severe pathological changes (such as mass, scars, and large tumors). There have been attempts in the past to segment such cases using shape prior for example in [9]–[17]. These methods relied on techniques such as sparse PCA [12], active contour [15], graph cut [16], atlas-driven [17], [18] etc. These methods are not readily parallelizable and have slow execution time [19]. This has also been demonstrated in [10], where authors provided running time comparisons for a number of state-of-the-art lung segmentation methods [12]–[15]. The mean running time for these methods is in the range of 189–586 seconds. A segmentation algorithm which alone takes more than 3 minutes is unsuitable for a typical commercial nodule detection software that usually needs to complete the entire nodule detection pipeline within 5 minutes [20].

Manuscript received December 16, 2019; revised April 30, 2020 and June 8, 2020; accepted June 16, 2020. Date of publication June 22, 2020; date of current version April 5, 2021. This work was supported by the Siemens Healthineers and National Science Foundation under Grants NSFIIS-1715985 and IIS-1812606. *(Corresponding author: Hong Qin.)*

Pranjal Sahu and Hong Qin are with the Department of Computer Science, Stony Brook University, Stony Brook, NY 11794 USA (e-mail: psahu@cs.stonybrook.edu; qin@cs.stonybrook.edu).

Yiyuan Zhao, Parmeet Bhatia, Luca Bogoni, and Anna Jerebko are with the Siemens Healthineers, Malvern, PA 19355 USA (e-mail: yiyuan.zhao@siemens-healthineers.com; parmeet.bhatia@siemens-healthineers.com; luca.bogoni@siemens-healthineers.com; anna.jerebko@siemens-healthineers.com).

Digital Object Identifier 10.1109/JBHI.2020.3004296

Example of segmentation failures as shown in Fig. 1 could lead to loss of nodules and tumors in the CAD system. This problem is particularly severe in cases where the nodule or tumor is wall attached. Ad-hoc measures such as smoothing, binary closing operation, curvature based methods etc. have limitations as they cannot capture wall attached large nodule/tumors (juxtapleural tumors) [21]–[23]. A data driven method which could seamlessly integrate with a CNN based segmentation model is needed.

Apart from the general purpose segmentation CNN such as U-Net, there have been attempts to design CNN architectures specific to lung volume segmentation [24]–[27]. For example, in [24], authors proposed a dilated convolution based lung segmentation method to reduce computational requirement. Their work showed promising results for few ILD cases (Interstitial Lung Disease), however due to absence of global shape prior their method fails to perform satisfactorily in presence of large tumors. In [25], authors proposed a progressive growing dense V-Net CNN for lung volume segmentation. In their work they tried to solve the problem of slow speed of inference and large memory requirement for 3D CNN based lung segmentation network when applied on lung CT volumes. They proposed to obtain the segmentation output by refining it in progression so that the number of filters operating at full resolution in a CNN could be reduced. In [26], authors proposed a progressive multi-path scheme called progressive holistically-nested networks (P-HNN). Here, again the major idea is to progressively refine the segmentation output and to use the VGG-16 [28] as the base network for doing segmentation. Similarly in [27], authors proposed a multi-resolution CNN for lung volume segmentation where they showed that incorporating both local and global features results in more accurate segmentation. They demonstrated that current approaches of performing slice based segmentation is sub-optimal and showed how a 2D slice based segmentation model sacrifices global context for high resolution details while a low resolution volumetric model sacrifices high resolution detail for global contexts. They also showed how different image sampling strategies can influence the resolution, GPU memory requirement and volumetric context of the segmentation model. A recurring theme across these models has been to perform segmentation in progression so as to reduce computation requirement. However, despite employing novel architectures the common limitation with the above mentioned CNN methods is the requirement of training data for pathological cases.

B. Our Contribution

Taking into account the above mentioned problems of data availability, efficiency and robustness, we propose a multi-stage CNN based segmentation algorithm which incorporates the shape information of lung volume into the segmentation pipeline using a 3D structure correction CNN. It is known that the shape properties of volume or image resides in the low frequency component and hence by only focusing on low resolution input the shape characteristics can be captured [27], [29]. The proposed segmentation framework is divided into three stages where the first two stages work on the low resolution CT volume to segment and correct the shape deformation caused due to the presence of tumors. A novel training strategy is adopted to train the structure correction 3D CNN using only normal lung volumes. The 3D CNN learns the shape characteristics of lung volume with the help of corrupted and non-corrupted segmentation masks, nullifying the role CT intensity plays during the training of a segmentation network. By performing the training in two stages, the network is forced to learn shape characteristics. Due to shape correction the ground truth mask annotation for entire pathological lung volume is not needed and a model trained using only normal volumes is able to capture tumor regions in pathological volumes. Finally the shape corrected low resolution volume is up-sampled and refined to obtain the final volume segmentation. For the refinement of the shape corrected segmentation mask a parallel flood-fill method is proposed. The flood-fill method acts as a substitute for the costly convolution operation at full resolution and improves the average surface distance. The end product of this pipeline is a fast and robust lung segmentation method which finishes within 5 seconds. We demonstrate the advantages of this model by analyzing its performance on different use cases. Three datasets have been used for this purpose namely LUNA [30], NSCLC [31] and LOLA11 [32].

In Section II we will describe the three stages of the proposed method. Training and dataset related details are mentioned in Section III. The results and comparison of the proposed method with other state-of-the-art lung segmentation methods are discussed in Section IV.

II. OUR METHOD

Designing a segmentation method which can work across a variety of CT volumes is a challenging task. CT volumes come in a range of x, y, z resolution. For example in LUNA [30] challenge dataset the x, y and z resolution are in range [0.48–0.97] mm, [0.48–0.97] mm and [0.5–2.50] mm respectively. Having a higher z-resolution results into higher number of CT slices. A naive 3D CNN architecture which works on the full resolution is not feasible to be trained on a single GPU due to memory constraints [27]. To solve this problem we propose a stage wise segmentation algorithm which significantly reduces the computation requirement and works across all sizes of input CT. In the following subsections we describe each stage of the algorithm. The entire segmentation pipeline is shown in Fig. 2.

A. Stage 1: Coarse Lung Shape Initialization

We use V-Net [6] architecture as our Stage 1 lung segmentation CNN. V-Net is a 3D CNN based volume segmentation network. In the proposed method we used a V-Net CNN with an input tensor size of $128 \times 128 \times 128$. All the CT scans are therefore interpolated using linear interpolation to a voxel spacing such that the interpolated volume has $128 \times 128 \times 128$ voxels. Output of the first stage V-Net is used to get the low resolution segmentation masks for trachea, left and right lungs. The training of the V-Net is done using the segmentation masks provided in the LUNA dataset [30]. The details of the V-Net training are discussed in Section III.

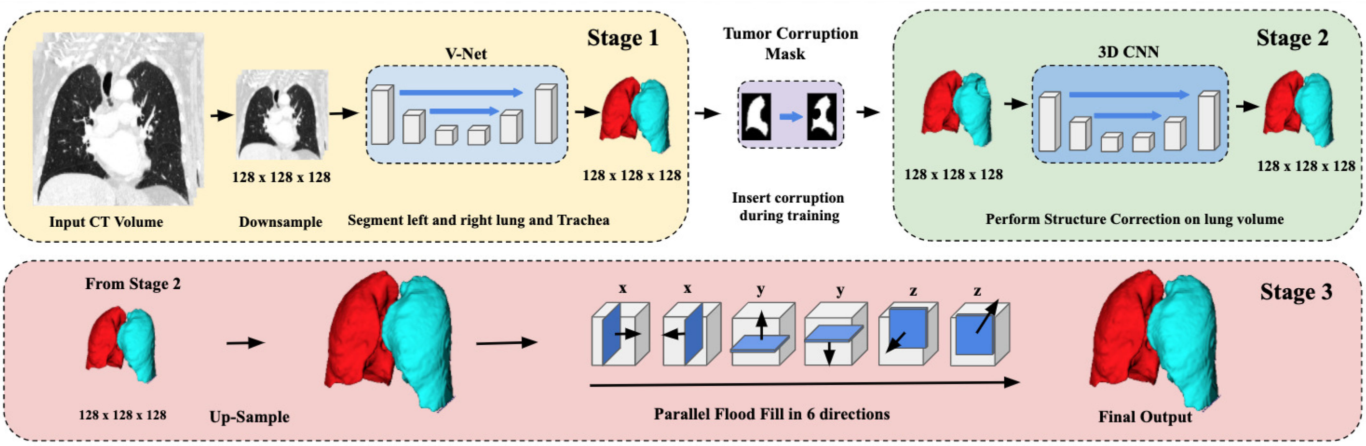


Fig. 2. Pipeline of proposed method for segmenting the lung region from CT volume. Structure correction is done in Stage 2 for the entire lung volume mask. Parallel flood fill operation is applied to refine the up-sampled mask.

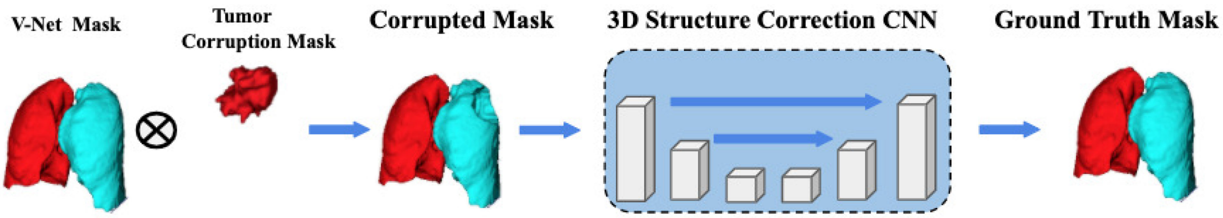


Fig. 3. Figure illustrating the functioning of structure correcting 3D CNN for lung mask. 3D CNN is trained using artificially generated data. The output of V-Net is multiplied with a tumor corruption mask and fed into the 3D CNN which is trained to reconstruct the ground truth mask.

The output of the first stage V-Net is not robust to the presence of tumors and hence requires structural correction. The left and right lung masks are then passed to the Stage 2 for the shape correction using a 3D CNN. Next we describe the second stage of the segmentation pipeline where correction in the shape of the lung masks is performed with the help of a 3D CNN.

B. Stage 2: 3D CNN for Shape Correction

The major component of the second stage of the segmentation pipeline is the 3D structure correction CNN. The segmentation mask for left and right lung obtained from Stage 1 are used in this stage. Training of the network in this stage is performed in a novel way so that the lung shape prior is encoded in the 3D structure correction CNN. To force the 3D CNN to learn the shape characteristics we introduce structural corruption in the masks obtained from stage 1 V-Net. The structural corruption is achieved by dropping a tumor shaped region at a randomly selected location from the V-Net mask, see Fig. 3. We call this the tumor corruption mask. The 3D binary corruption mask represents a large nodule or tumor. LUNA dataset takes CT samples from the LIDC/IDRI dataset [33] which comes with the nodule annotations [34].

We use the publicly available PyLIDC tool [35] for extracting the binary nodule masks from the XML annotation present in the LIDC/IDRI dataset. Each nodule is annotated by 2–4 radiologists and we take each annotation mask as different for increasing the number of samples. However, the size of these nodules is smaller compared to the size of tumors. Therefore, the nodule shapes obtained from the LIDC dataset are enlarged by doing nearest neighbour interpolation to make them similar to size of tumors (10–50 mm diameter). The binary nodule mask are enlarged separately and the actual CT volume from which it is extracted is not involved in this process. Example of the tumor corruption masks are shown in Fig. 4.

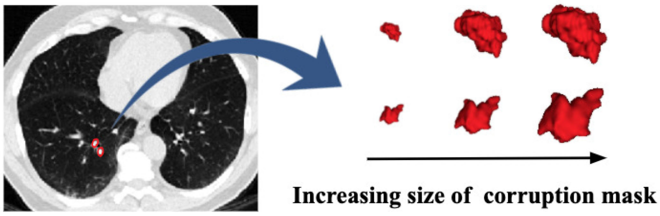


Fig. 4. Samples of nodule shapes extracted from LIDC nodule annotations used for introducing tumor corruption in V-Net output mask.

Here, we make an assumption that the shape characteristics of tumors are similar to an enlarged nodule. The location where the corruption is introduced is randomly selected from the edge

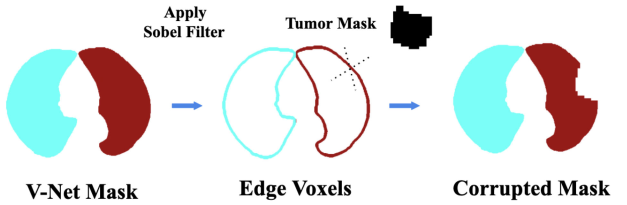


Fig. 5. Output of sobel filter gives the candidate locations to insert the corruption. A voxel is randomly selected from the candidate voxels.

voxels. Sobel filter is performed along the axial cross-section to obtain edge voxels. A randomly selected edge voxel then acts as the center of mass of the tumor corruption mask to be inserted. The V-Net mask overlapping with the tumor mask is erased while the region not overlapping with the tumor mask is retained as shown in Fig. 5.

The corrupted V-Net mask and the ground truth mask then act as the training data for the structure correction 3D CNN. The 3D CNN is then trained with the corrupted mask as input and ground truth mask as output. This results into 3D CNN learning the shape characteristics of lung volume. The details of the structure correction CNN architecture are discussed in Section III. Output of the structure correction is shape corrected segmentation mask. The resulting mask is then passed to stage 3 for obtaining the full resolution segmentation mask. Next we describe the up-sampling stage.

C. Stage 3: Up-Sampling

In this stage, the shape corrected segmentation mask is first up-sampled to the same size as input CT by performing nearest neighbour interpolation. The up-sampled shape corrected segmentation mask includes the region having large nodules/tumors. However, the output lacks fine boundary details. To obtain crisp segmentation boundary and to reduce the average surface distance we perform a parallel flood fill operation. The lung region in the body is well contained and has a distinct CT range. We utilize this property to perform a modified flood-filling of the lung volume starting from the up-sampled segmentation mask. This step produces mask which is tightly attached to the lung boundary. In the modified flood fill, the flood index moves in both directions across x, y and z axes, as shown in Stage 3 of Fig. 2. The benefit of doing flood-fill in this manner is that the algorithm can be parallelized easily. Before performing flood fill a maximum boundary volume is obtained by performing the dilation operation on the input up-sampled mask. The dilated mask gives an upper bound of region to be flooded which is necessary to avoid flooding into the colon region. The flood filling operation is implemented using Numba parallel processing library. The pseudo code for the modified flood fill is shown in Algorithm 1. The cutoff value for lung region C in rescaled CT is -500 HU while the mask is dilated by 2 voxels. Example of enhancement of the up-sampled mask after applying the flood-fill operation is shown in Fig. 6. Quantitative improvement in the segmentation mask due to the flood-fill operation is discussed in Section IV-B.

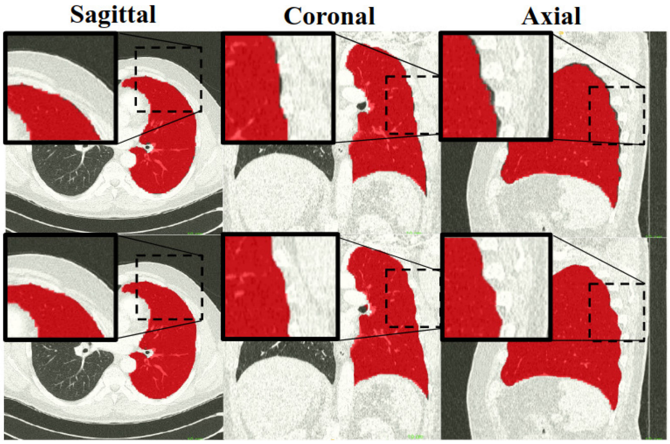


Fig. 6. Figure illustrating the result after the flood-fill operation on left lung mask. Top row shows the up-sampled segmentation mask which is not tightly attached to the lung boundary. After the parallel flood-fill operation the segmentation mask is tightly attached to the wall.

Algorithm 1: Parallel Flood-Fill.

```
Flood Fill($V, M, C, D, u_x, u_y, u_z$)
Where $V$ is CT volume, $M$ is up-sampled
segmentation binary mask, $C$ is the cutoff CT value for
lung region, $D$ is the dilated up-sampled mask,
$u_x, u_y, u_z$ are the volume dimensions of the lung CT.
    for $i = 0$ to $u_y$ do
      for $j = 0$ to $u_z$ do
        for $k = 0$ to $u_x$ do
          if $M[i,j,k] == 1$ & $D[i,j,k+1] == 1$ &
          $V[i,j,k+1] < C$ then
            $M[i,j,k+1] = 1$
          end if
        end for
      end for
    end for
    {Repeat the above block with $k = u_x$ to 0, i.e.
    opposite direction and perform the same operation
    for the other two axis i.e. $y$ and $z$ by iterating over
    $y$ and $z$ in the inner loop. }
    return $M$
```

III. Dataset and Training

In this section we describe the dataset, CNN architectures, loss function and the hyper-parameters used for training deep networks.

A. LUNA Dataset for Training

The first stage V-Net CNN is trained using the CT volumes and segmentation masks provided with LUNA dataset. LUNA dataset takes the CT scans from the publicly available LIDC dataset [33] and is available in 10 subsets. We used subsets 0 to 6 for training, subset 7 for validation and subset 8 and 9 as the testing sets. This results in total 519, 86 and 171 CT volumes for training, validation and testing splits respectively. The dataset

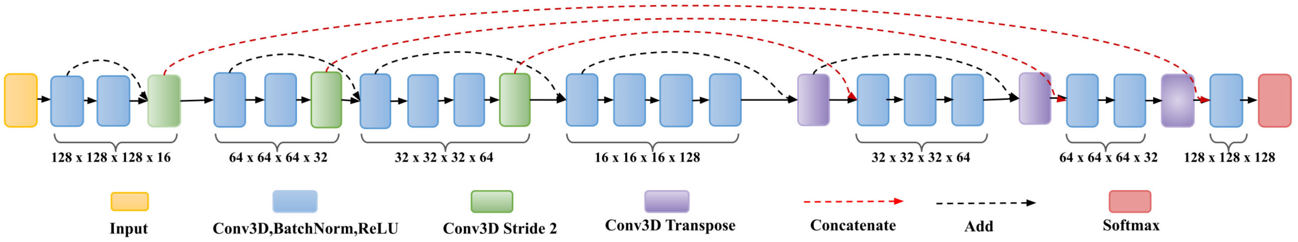


Fig. 7. V-Net architecture used in the proposed model. It has skip connections in the form of both addition and concatenation [6].

provided segmentation masks identifies each voxel as either background, left lung, right lung or trachea. It should be noted here that the segmentation masks in LUNA are computed using a sophisticated automated algorithm mentioned in [18] and are not human annotated. The method in [18] while segmenting lung CT takes care of error cases such as small nodules, sarcoidosis, pneumonia etc. Also the study conducted by authors in [18] shows that their proposed method is able to obtain a mean intersection over union of 0.95 ± 0.05 in comparison to 0.96 ± 0.04 obtained by a second human observer on LIDC dataset.

Standard dicom CT re-scaling is performed following the equation:

$$HU = IV * slope + \text{intercept} \tag{1}$$

where, HU is Hounsfield Unit, IV is intensity value and slope and intercept obtained from the dicom tags. After re-scaling value of 1024 is added. A cutoff value of 0 and 2000 is taken to ignore non-lung regions. Further all the voxels are divided by 2000 before feeding to network so that values lie in range 0 to 1.

B. Stage 1 V-Net Training Details

The architecture for stage 1 V-Net is shown in Fig. 7. It is trained with a learning rate of 0.0001 using Adam Optimizer [36] and loss function based on dice score. The loss function L for segmentation is defined in Equation (2):

$$L = 1 - \frac{1}{C}\sum_{c=1}^{C}\frac{2\sum_{i}^{S} p_{i,c} * g_{i,c}}{\sum_{i}^{S} p_{i,c} + \sum_{i}^{S} g_{i,c}}, \tag{2}$$

where, C is the number of segmentation classes, S is the number of voxels in the output i.e. $128 \times 128 \times 128$, $p_{i,c}$ is the value of tensor after applying softmax at location i for class c, Similarly, $g_{i,c} \in \{0, 1\}$ is the ground truth value at location i for class c. For stage 1 V-Net, refer Fig. 2, the number of classes C is 4 (left lung, right lung, trachea and background). The V-Net model is trained for a maximum of 20 epochs and the model with the least validation set loss is selected for getting results from the testing set.

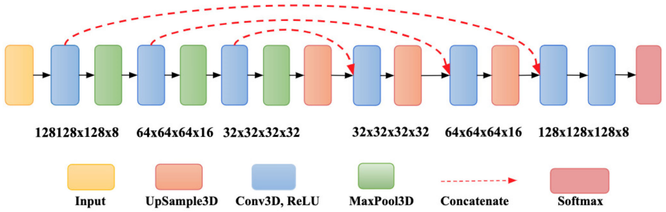


Fig. 8. Architecture of 3D CNN used in Fig. 2 for structure correction. Below each layer is the tensor size. Skip connection is used here in the form of concatenation.

C. Structure Correction CNN Training Details

The 3D structure correction CNN model is trained with a learning rate of 0.0001 using Adam Optimizer and the categorical cross entropy loss (CE). The output of 3D CNN is a tensor of size SxJ where S is the number of voxels in the input volume i.e. $128 \times 128 \times 128$, J is the number of classes i.e. 3 (left lung, right lung and background). The loss function CE used for training the parameters of 3D CNN becomes:

$$CE = -\frac{1}{S}\sum_{i=0}^{S}\sum_{j=0}^{J-1} y_{ij} \cdot \log\left(\hat{y}_{ij}\right) + \left(1 - y_{ij}\right) \cdot \log\left(1 - \hat{y}_{ij}\right) \tag{3}$$

where, $y_{ij} \in \{0, 1\}$ is the ground truth mask value, $\hat{y}_{ij}$ is the output of the softmax layer of 3D structure correction CNN at voxel i and y_{ij} is 1 if the voxel i belongs to class j or else 0. The structure correction 3D CNN is a standard encoder and decoder architecture with the skip connections between them as shown in Fig. 8. The skip connection helps in retaining the detail information from the input to the network's output. Also to keep the 3D CNN light weight we use the minimum number of filters needed to achieve the minimum attainable validation set loss. Increasing the filters above the chosen values does not decrease validation loss further. This results in a very computationally efficient architecture which does not add too much additional cost to the entire segmentation pipeline. For training the 3D CNN, artificial data is generated by introducing the corruption in the mask obtained from V-Net. The V-Net masks are augmented by 25 times by inserting tumor corruption at randomly selected edge voxel. Tumor corruption is introduced with a probability of 0.8 which means that roughly 80% of the augmented samples are corrupted and 20% samples remain the same. A total of 153 nodule annotations are taken from the LIDC dataset to be enlarged as tumor corruption mask out of which randomly selected 80% nodules are used for training split and 20% for validation split. Tumor corruption masks are rotated by 90° with equal probability along the three axes (x, y and z) to augment the tumor mask dataset. We train three configuration of 3D structure correction CNN which differ in the magnification factor of the tumor masks (10–20 mm, 10–30 mm, 10–50 mm). The impact of tumor magnification factor is discussed in Section V-A. The 3D CNN is trained for a maximum of 50 epochs and the model with least validation set loss is used for doing inference on the test set. Training of the models is done using a Titan 2080 Ti GPU on a linux machine with i9-9900 k CPU and 64 GB RAM. The model is implemented using the Keras Library with Tensorflow backend. Training of the Stage 1 V-net model takes around 2.5 hours, while the structure correction CNN takes around 4.7 hours.

D. Comparison CNN Training Details

We compare the proposed method against three strong baselines for medical image segmentation namely U-Net (using 2D CT slices), V-Net and 3D U-Net (using volume) along with three other publicly available state-of-the-art 2D slices based lung segmentation models namely P-HNN [26], SegNet [37] and Dilated CNN [24]. In addition, we also compare a GAN (Generative Adversarial Network) [38] based lung segmentation model proposed in [39]. All the models in comparison are trained using the LUNA dataset following the same protocol used for training the proposed model i.e. LUNA subset 0–6 is used for training and subset 7 is used for validation. U-Net, V-Net. 3D U-Net, SegNet and Dilated CNN are trained with both dice and categorical cross-entropy loss function using their publicly available implementation while P-HNN is trained following the official implementation. The model Multi-instance GAN model proposed in [39] is implemented following the details mentioned in the paper as the official implementation is not publicly available. V-Net and 3D U-Net are trained with an input volume of size 128 × 128 × 128 and the results obtained from it is then resized to original CT dimension. All the other 2D slices based models are trained with input size of 512 × 512 except for P-HNN which takes input of size 400 × 400 and resizes the result again to 512 × 512 size. The model weights with least validation loss is used for obtaining the results on the test datasets. For comparison purposes, we also obtained already trained P-HNN model weights from the authors.

IV. RESULTS AND COMPARISON

In this section we share the results of the experiments which demonstrate that the proposed algorithm while being fast is also robust to pathological changes in the lung volume such as large nodules and tumors. Three datasets have been used for testing the performance namely NSCLC [31], LUNA [30] and LOLA11 [32]. NSCLC dataset is used to demonstrate the performance in segmenting the lung volumes having large nodules/tumors. LUNA is used to demonstrate the performance while segmenting normal lungs and the speed of execution. And finally, LOLA11 is used to show that the proposed algorithm also performs satisfactorily in presence of other pathological changes such as scoliosis, emyphysema etc.

TABLE I
COMPARISON OF RECALL OF TUMOR VOXELS ON NSCLC DATASET (MEAN ± STD)

	NSCLC split		
	Juxtapleural Tumors	Rest	Total
Nodule Count	47	94	141
U-Net (d) [4]	0.541±0.255	0.888±0.184	0.772±0.266
V-Net (d) [6]	0.566±0.294	0.92±0.158	0.802±0.271
3D U-Net (d) [5]	0.422±0.253	0.822±0.231	0.689±0.304
SegNet (d) [37]	0.422±0.234	0.794±0.261	0.67±0.307
Dilated CNN(d) [24]	0.457±0.247	0.825±0.232	0.702±0.294
U-Net (ce) [4]	0.506±0.256	0.856±0.208	0.739±0.279
V-Net (ce) [6]	0.461±0.278	0.88±0.201	0.741±0.303
3D U-Net (ce) [5]	0.422±0.259	0.821±0.227	0.688±0.304
SegNet (ce) [37]	0.454±0.253	0.836±0.226	0.709±0.296
Dilated CNN(ce) [24]	0.562±0.272	0.892±0.183	0.782±0.267
P-HNN (Luna) [26]	0.723±0.246	0.937±0.145	0.866±0.211
P-HNN (Public) [26]	0.756±0.224	0.954±0.11	0.888±0.183
Multi-instance GAN [39]	0.516±0.268	0.888±0.188	0.764±0.28
Our (w/o Flood,10-20)	0.741±0.29	0.942±0.127	0.875±0.219
Our (w Flood,10-20)	0.751±0.287	0.946±0.123	0.881±0.215
Our (w/o Flood,10-30)	0.813±0.226	0.941±0.123	0.898±0.176
Our (w Flood,10-30)	0.823±0.223	0.955±0.109	0.911±0.168
Our (w/o Flood,10-50)	0.864±0.189	0.944±0.123	0.918±0.153
Our (w Flood,10-50)	**0.869**±0.187	**0.956**±0.108	**0.922**±0.15

(d) is dice loss and and (ce) is cross-entropy loss.

A. NSCLC Dataset Results for Tumor Voxel Recall

We first test the effectiveness of the proposed segmentation algorithm in preserving large nodules and tumors. For this purpose we performed the evaluation using the Non-Small Cell Lung Cancer (NSCLC) dataset [31]. The dataset comprises lung CT scans with the tumor annotation done by radiation oncologist. A total of 141 CT volumes are present in NSCLC with each volume having exactly one annotated tumor mask. The NSCLC dataset provides the segmentation mask only for the tumor regions and therefore the metric used for testing the segmentation performance is Tumor Voxel Recall as shown in Equation (4):

$$\textit{Tumor Voxel Recall (TVR)} = \frac{\sum_{i \in S^T} M^i}{\sum_{i \in S^T} 1}, \qquad (4)$$

where, $M = \{M_i \in \{0,1\}, i \in S^T\}$ is the predicted segmentation mask at voxel index i and S^T is the set of tumor voxel indices obtained from ground truth annotation. While calculating Tumor Voxel Recall, both left and right lung mask are taken while the trachea region is not considered. Results of this experiment is shown in Table I. To better appreciate the problem, we show the results of segmentation for cases with large juxtapleural nodules/tumors (>16 mm diameter) separately in Table I. Qualitative results for this experiment are shown in Fig. 9 where along with the segmentation masks, tumor voxel recall has also been shown. More qualitative results are shown in the supplementary material.

B. LUNA Test Set Results

LUNA dataset is devoid of pathologies such as tumors and only comprises smaller nodules. Therefore, we use LUNA test

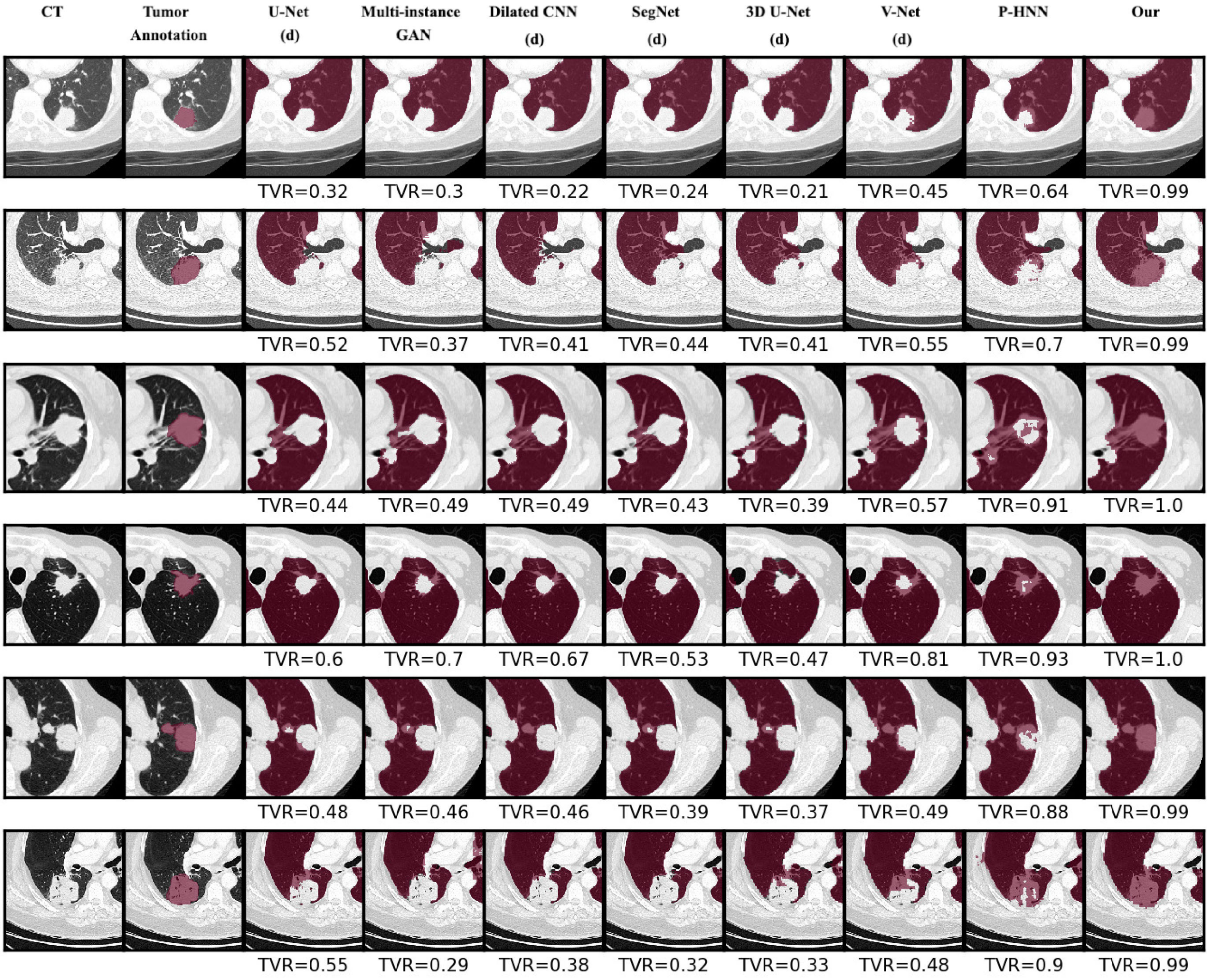


Fig. 9. Figure illustrating cases from NSCLC dataset. The tumor voxel recall (TVR) is shown below each case. Only the central slice of the tumor is shown here. Our method is showing results of 3D CNN trained with 10–50 mm tumor masks and after Flood-fill refinement.

split to obtain the performance of segmentation algorithms for normal cases. Here the ground truth segmentation mask is available and therefore dice coefficient is taken as the performance metric [40]. The formulation of dice coefficient is defined below:

$$Dice\ coefficient = \frac{2TP}{2TP + FP + FN}, \tag{5}$$

where, *TP* is True positive, *FP* is False positive and *FN* is False negative. We follow the convention used in LOLA11 challenge to evaluate the lung overlap. Due to variability in defining the lung border, voxels in a 2 mm slack border are not considered during the evaluation [32]. The dice coefficient results for left and right lung separately as well as for the entire lung volume are shown in Table II. As can be seen from Table II that the proposed model performs on-par with other segmentation models such as Dilated CNN, SegNet etc.

For some use cases the segmentation masks needs to be very crisp and should provide good estimate of the volume surface. For such use-cases the symmetric average surface distance

TABLE II

DICE COMPARISON OF LUNG VOLUME SEGMENTATION ON LUNA TEST SET. (MEAN ± STD)

Model	Left lung	Right lung	Total lung
U-Net (d) [4]	0.9973±0.0038	0.9968±0.0047	0.9977±0.003
V-Net (d) [6]	0.9973±0.0024	0.997±0.003	0.9973±0.0023
3D U-Net (d) [5]	0.9966±0.0031	0.9958±0.0043	0.9964±0.0031
SegNet (d) [37]	0.9943±0.0061	0.9876±0.0178	0.993±0.0093
Dilated CNN (d) [24]	0.99±0.0355	0.9891±0.0281	0.9904±0.031
U-Net (ce) [4]	0.9944±0.0106	0.9931±0.0105	0.9941±0.0086
V-Net (ce) [6]	0.9973±0.0025	0.9966±0.0039	0.9971±0.0027
3D U-Net (ce) [5]	0.9959±0.0059	0.9946±0.0093	0.9955±0.0057
SegNet (ce) [37]	0.9968±0.005	0.9967±0.0043	0.9971±0.0038
Dilated CNN (ce) [24]	0.9975±0.0035	0.9948±0.0103	0.9966±0.0052
P-HNN (Luna) [26]	-	-	0.9769±0.0116
P-HNN (Public) [26]	-	-	0.992±0.0113
Multi-instance GAN [39]	-	-	0.9947±0.0055
Our (w/o Flood,10-20)	0.9972±0.0025	0.9965±0.0027	0.997±0.0023
Our (w Flood,10-20)	0.9973±0.0025	0.9971±0.0024	0.9973±0.0022
Our (w/o Flood,10-30)	0.9973±0.0025	0.9968±0.0027	0.9971±0.0023
Our (w Flood,10-30)	0.9974±0.0025	0.997±0.0026	0.9973±0.0023
Our (w/o Flood,10-50)	0.9973±0.0025	0.9968±0.0027	0.9972±0.0023
Our (w Flood,10-50)	0.9975±0.0024	0.9974±0.0024	0.9975±0.0022

(d) is dice loss and and (ce) is cross-entropy loss.

TABLE III

AVERAGE SURFACE DISTANCE COMPARISON OF LUNG VOLUME SEGMENTATION ON LUNA TEST SET ($in\ mm$). (MEAN ± STD)

	Distance Percentile	90%			95%			All		
	Lung	Left	Right	Total	Left	Right	Total	Left	Right	Total
Dice	U-Net [4]	0.1±0.19	0.2±0.75	0.08±0.11	0.2±0.45	0.38±1.17	0.14±0.19	0.64±1.08	1.0±1.78	0.43±0.54
	3D U-Net [5]	0.44±0.21	0.43±0.08	0.42±0.08	0.53±0.39	0.51±0.16	0.5±0.13	0.74±0.72	0.68±0.39	0.68±0.39
	V-Net [6]	0.37±0.07	0.37±0.07	0.37±0.06	0.43±0.1	0.43±0.08	0.43±0.07	0.57±0.23	0.56±0.14	0.54±0.13
	SegNet [37]	0.12±0.3	0.12±0.33	0.08±0.05	0.22±0.53	**0.2**±0.51	0.15±0.16	0.58±0.98	0.53±0.78	0.41±0.46
	Dilated CNN [24]	0.76±2.52	0.63±2.01	0.61±1.85	1.07±2.93	0.93±2.34	0.88±2.21	1.84±3.39	1.88±2.72	1.69±2.65
Cross Entropy	U-Net [4]	0.49±1.44	0.84±2.35	0.38±1.12	0.97±2.01	1.46±3.28	0.76±1.55	2.1±2.84	2.72±4.28	1.63±2.11
	3D U-Net [5]	0.57±1.44	0.5±0.51	0.51±0.56	0.68±1.67	0.58±0.69	0.6±0.8	0.97±1.95	0.79±0.91	0.85±1.09
	V-Net [6]	0.49±0.54	0.46±0.09	0.45±0.1	0.65±1.45	0.53±0.1	0.54±0.17	1.01±2.48	0.74±0.24	0.81±0.69
	SegNet [37]	**0.1**±0.14	**0.1**±0.17	**0.09**±0.11	**0.18**±0.28	0.22±0.41	**0.14**±0.17	**0.43**±0.57	0.64±0.9	**0.36**±0.34
	Dilated CNN [24]	0.43±2.41	0.16±0.65	0.23±1.18	0.66±2.77	0.27±0.87	0.38±1.54	1.35±3.17	0.68±1.28	0.94±1.97
	P-HNN (Luna) [26]	-	-	1.06±0.32	-	-	1.27±0.45	-	-	1.67±0.73
	P-HNN (Public) [26]	-	-	0.69±1.75	-	-	1.04±2.13	-	-	2.05±2.61
	Multi-instance GAN [39]	-	-	0.25±0.34	-	-	0.48±0.75	-	-	1.17±1.21
	Our (w/o Flood, 10-20)	0.38±0.07	0.39±0.07	0.38±0.07	0.45±0.1	0.45±0.08	0.44±0.07	0.6±0.26	0.58±0.13	0.57±0.12
	Our (w Flood, 10-20)	0.2±0.05	0.21±0.05	0.21±0.05	0.27±0.09	0.27±0.06	0.27±0.05	0.44±0.3	0.41±0.12	0.4±0.13
	Our (w/o Flood, 10-30)	0.38±0.08	0.39±0.07	0.38±0.07	0.45±0.1	0.45±0.08	0.44±0.07	0.6±0.23	0.57±0.13	0.57±0.12
	Our (w Flood, 10-30)	0.21±0.06	0.21±0.05	0.21±0.05	0.27±0.1	0.27±0.06	0.27±0.06	0.44±0.27	**0.4**±0.12	0.4±0.14
	Our (w/o Flood, 10-50)	0.42±0.07	0.44±0.07	0.43±0.06	0.49±0.09	0.51±0.08	0.49±0.07	0.63±0.22	0.64±0.14	0.62±0.12
	Our (w Flood, 10-50)	0.22±0.07	0.22±0.05	0.21±0.05	0.29±0.11	0.28±0.06	0.27±0.06	0.46±0.29	0.41±0.12	0.41±0.14

(*ASD*) metric [40] is used which is defined in Equation (6):

$$ASD(S,F) = (\frac{\sum_{d_i \in e_S} \min_{d_j \in e_F} \text{dist}\,(d_i,d_j)}{2 * \sum_{d_i \in e_S} 1} + \frac{\sum_{d_j \in e_F} \min_{d_i \in e_S} \text{dist}\,(d_j,d_i)}{2 * \sum_{d_j \in e_F} 1}, \tag{6}$$

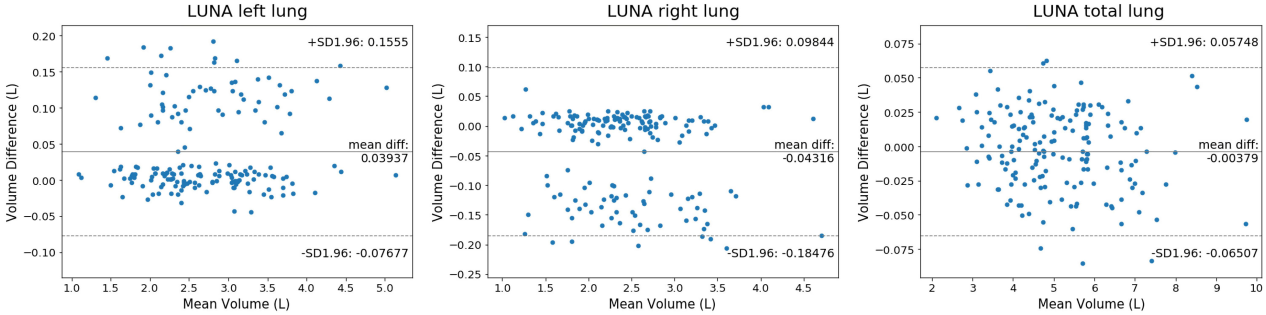
where S is ground truth segmentation, F predicted segmentation, V volume size, e_S surface of ground-truth segmentation, e_F surface of predicted segmentation, d_i vertices on the surface e_S, d_j vertices on the surface e_F, and $\text{dist}(d_i,d_j)$ euclidean distance between vertices d_i and d_j. The results of surface distance performance is shown in Table III. 2D scan based method sometimes suffer from holes in their output which results into higher surface distance for those voxels. Hence, to be fair while calculating the average distance we considered various percentile of surface voxel distances. In Table III, ASD is calculated in three ways, first considering all surface voxels, second considering lowest 95% surface distance voxels and third with lowest 90% surface distance voxels. Ground truth mask is available for left, right, and total lung volume for Luna CT scans and therefore we plot the agreement between the predicted lung volume and the ground truth volume using a Bland-altman plot as shown in Fig. 10. For comparison purpose we also plot the Bland-Altman plot between output of Stage 1 V-Net and ground truth volume. The plot shows the lung volume (in Liter) estimated by proposed algorithm in comparison to ground truth volume.

C. Time of Execution Results

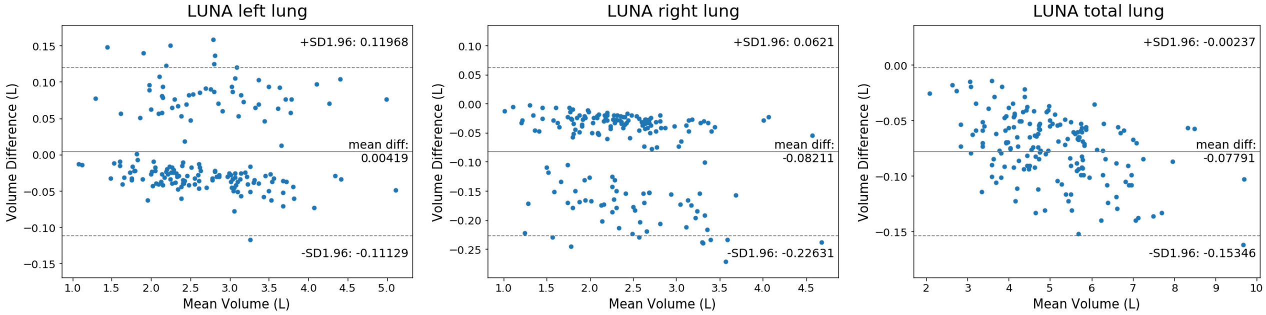
We also calculate the mean execution time for all the models on the LUNA test split. The results for time of execution is shown in Table IV. It can be observed that the proposed model is significantly faster in comparison to many state-of-the-art models. The slowest of them all is P-HNN model which comprises four VGG-16 networks making it significantly slower. The fastest one is the V-Net model. Both the volume based segmentation model have lower execution time because of performing segmentation on a lower resolution volume. However, like other CNN models in comparison both of them suffer in pathological cases having tumors as shown in Fig. 9. Also because of lack of details at the surface they have higher symmetric average surface distance, refer Table III. The refinement performed on the proposed model's output lowers the average surface distance and adds additional 1.2 seconds on average in total execution time.

D. LOLA11 Results for Other Pathologies

Finally, we also test the performance of the proposed method in segmenting other pathological cases such as emphysema etc. For this purpose we evaluate the algorithm on the LObe and Lung Analysis 2011 (LOLA11) dataset. LOLA11 dataset comprises 55 lung CT cases with a number of pathologies such as scoliosis, emphysema, pleural effusion etc. The ground truth segmentation mask for this dataset is not publicly available and the evaluation is obtained from the challenge website [32]. The performance of the proposed method and other models in comparison is shown in Table V. Evaluation in this case is done for left, right lung individually and entire lung volume using the Intersection Over Union (IoU) metric [40]. Multi-instance GAN and P-HNN model provide entire lung segmentation mask and hence we could not obtain their performance on LOLA11 dataset where submission requires left and right lung masks separately. Results in Table V demonstrates that the proposed algorithm performs satisfactorily in presence of other pathologies such as scoliosis, emphysema etc. Fig. 11 shows few qualitative results of the proposed method on pathological cases from LOLA11 dataset.



Agreement between output of proposed method(with Flood, 10-50mm) and Ground truth in LUNA dataset



Agreement between output of Stage 1 V-Net and Ground truth in LUNA dataset

Fig. 10. Bland Altman plot showing the agreement between the segmentation mask obtained by our method and the ground truth mask for left, right and total lung volume in LUNA dataset. For comparison, Bland Altman plot of Stage 1 V-Net output and Ground truth mask is also shown.

TABLE IV
MEAN EXECUTION TIME COMPARISON ON LUNA TEST SET. (MEAN ± STD)

Model	Time (in sec)
U-Net [4]	9.66±5.46
V-Net [6]	**1.43**±0.56
3D U-Net [5]	1.87±0.63
SegNet [37]	19.8±10.51
Dilated CNN [24]	10.32±5.57
P-HNN [26]	20.91±11.38
Multi-instance GAN [39]	5.02±2.63
Our (Without Flood)	1.47±0.57
Our (With Flood)	2.69±1.21

TABLE V
IOU COMPARISON OF LUNG VOLUME SEGMENTATION ON LOLA11 DATASET. (MEAN ± STD)

Model	Left lung	Right lung	Total lung
U-Net (ce) [4]	0.9644±0.1341	0.9675±0.1345	0.9660
V-Net (ce) [6]	0.965±0.1098	0.9678±0.1305	0.9664
3D U-Net (ce) [5]	0.9616±0.1037	0.9641±0.1308	0.9629
SegNet (ce) [37]	0.9641±0.1240	0.9670±0.1319	0.9656
Dilated CNN (ce) [24]	**0.9684**±0.0957	0.9665±0.1339	**0.9674**
U-Net (d) [4]	0.9574±0.1327	0.9659±0.1319	0.9616
V-Net (d) [6]	0.9665±0.104	0.9666±0.1331	0.9665
3D U-Net (d) [5]	0.9606±0.1308	0.965±0.1317	0.9628
SegNet (d) [37]	0.965±0.1328	0.9593±0.1394	0.9621
Dilated CNN (d) [24]	0.9632±0.0959	0.9579±0.1355	0.9605
Our (w/o Flood,10-20)	0.9631±0.1326	0.9685±0.133	0.9658
Our (w Flood,10-20)	0.9636±0.1328	0.9691±0.133	0.9663
Our (w/o Flood,10-30)	0.9631±0.1323	0.9687±0.1329	0.9659
Our (w Flood,10-30)	0.9635±0.1325	**0.9691**±0.1329	0.9663
Our (w/o Flood,10-50)	0.9647±0.1038	0.9679±0.1329	0.9663
Our (w Flood,10-50)	0.9654±0.1026	0.9684±0.1329	0.9669

V. DISCUSSION

A. Tumor Recall Improvement Due to Structure Correction

It can be observed from Table I that in cases of large juxtapleural nodules/tumors the recall of voxels drops drastically for all the other models. However, due to the presence of structure correction done by the 3D CNN, the proposed model is able to recover such regions and significantly outperforms other models. P-HNN model performs the second best among the models in comparison with a mean large tumor recall of 0.756 when the publicly available weights are used and obtains a mean recall of 0.723 when trained using only LUNA dataset. The best configuration of proposed method (with Flood, trained with 10–50 mm resized corruption masks) gets a mean recall of 0.869 which is better than the second best model (P-HNN) by around 15%, see Table I. A Wilcoxon signed rank test indicated that the tumor recall obtained by the proposed method is statistically significantly higher than the recall obtained by the second best P-HNN model (public weights) with $Z = -3.062945$, $p = 0.0011 (< 0.05)$.

Without any structure correction the output is same as that of up-sampled Stage 1 V-Net (with dice loss) where the mean recall of large juxtapleural tumor voxels is around 0.57. Again, the mean recall obtained from proposed method is statistically significantly higher with $Z = -8.977491$, $p = -1.11022\mathrm{e} - 16 (< 0.05)$. This shows that in absence of pathological training cases, shape correction is necessary to avoid missing important juxtapleural tumors/nodules. Except for P-HNN all other 2D slice based models have significantly lower

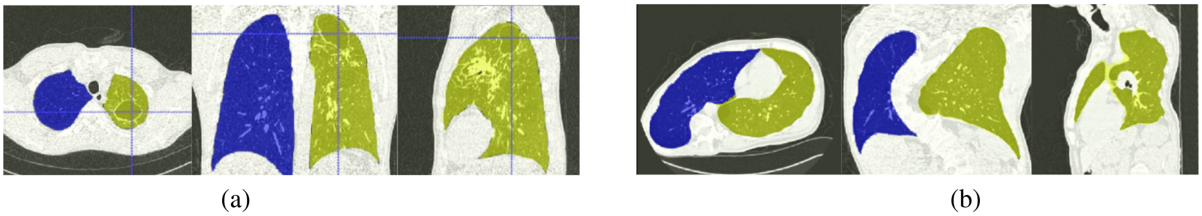

(a) (b)

Fig. 11. Figure illustrating the pathological cases from LOLA11 dataset. In (a) Case 42 is shown having emphysema. The proposed algorithm is able to capture the focal regions of emphysema. In (b) Case 5 is shown where the input CT has severe abnormality in shape due to scoliosis. The proposed method does not hallucinate region outside of the lung volume despite the presence of severe abnormality.

recall at around 0.5. This also demonstrates that P-HNN architecture having progressive multi-path scheme is better suited for this purpose. Surprisingly, we also observe that U-Net trained with dice loss has higher tumor voxel recall than a GAN based network proposed in [39].

We also observe that while training the structure correction CNN, the nodules should be resized to a range such that it covers all the tumors under consideration. A model trained only on smaller nodules is not able to capture larger tumors as can been seen in Table I. The mean recall of large juxtapleural tumors for model trained with 10–20 mm tumors is 0.751 while for the model trained with 10–50 mm tumors is 0.869.

B. Enhancement Due to Flood-Filling

In Table III it can be seen that the parallel flood-fill operation improves the symmetric average surface distance over only up-sampled segmentation mask. However, the 2D CNN model such as SegNet have lower surface distance which is due to the fact that it operates on full resolution CT slices and is able to preserve the very fine edge details [27]. Dilated CNN and U-Net have higher surface distance because they produce patchy output in many cases resulting into surface voxels far from the ground truth surface [37]. The proposed model has higher average surface distance compared to SegNet but is still below 0.5 mm which is typically the x-y resolution in CT scans. In use-cases where the surface distance is not a critical requirement, the flood-filling operation can be avoided. However, from the time-of-execution results it can be seen that the cost of performing flood-filling is around an additional second. The flood-fill operation also improves the Tumor voxel recall slightly. Our approach demonstrates that using a 3D segmentation model can give lower execution time. However, to use a 3D model one needs to use lower resolution to be able to train it on a single GPU. Also, the loss of high resolution details lost due to down-sampling can be recaptured by performing proposed flood-fill operation instead of performing costly convolution operation at full resolution.

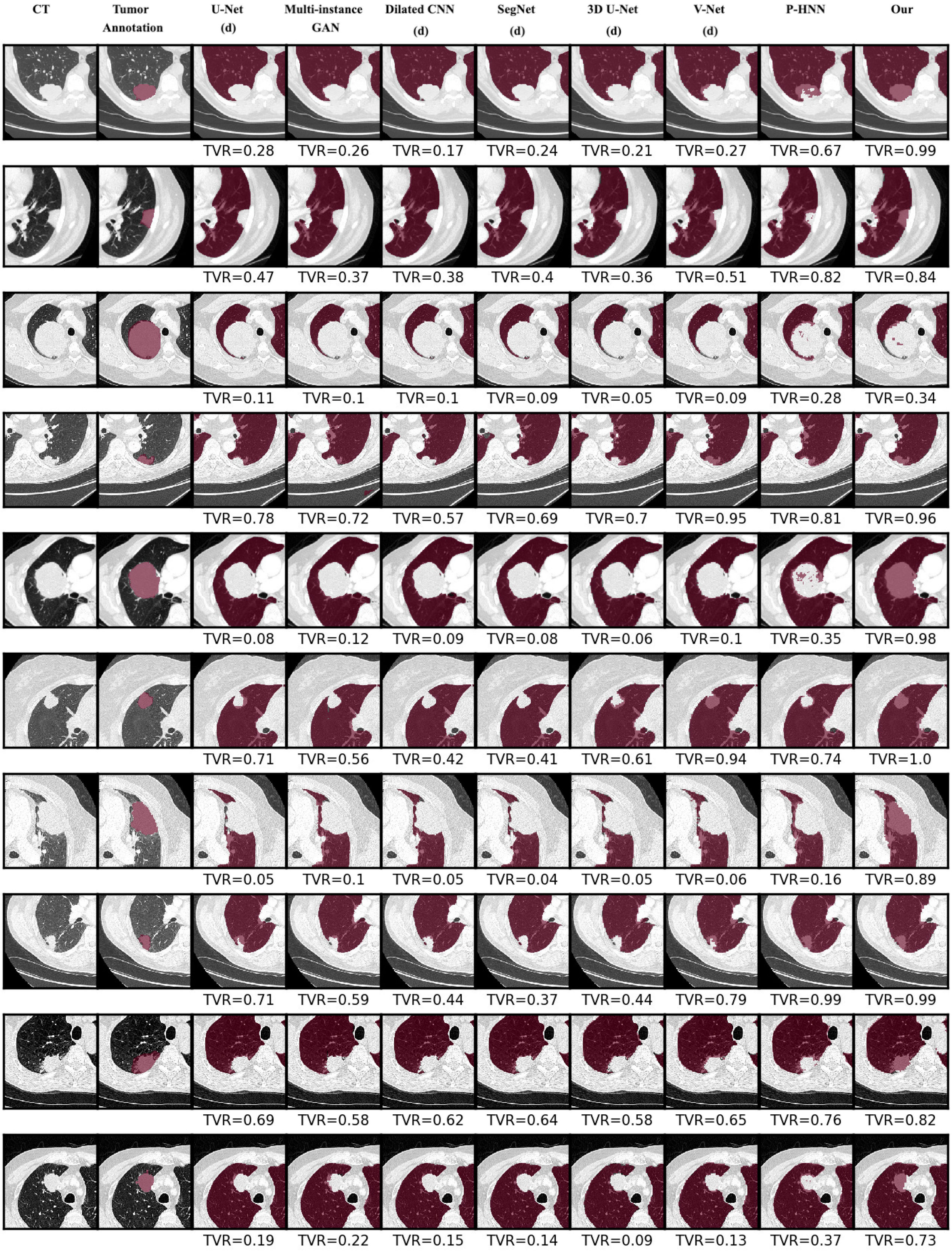
C. Limitation

The proposed algorithm has major limitation in segmenting lung volumes with pleural effusions where the entire or majority of the lung region is filled with fluid. An example of such case from LOLA11 dataset is shown in Fig. 12. This is because the output of the V-Net model in such cases is devoid of any lung region and the structure correction CNN is not able to correct such cases. It should be noted that this limitation is common in all the CNN models compared in absence of pathological training volumes with pleural effusion. For such cases labelled training data would be needed. Also, nodules are not commonly seen in the regions filled with pleural effusions so this kind of case will not affect the performance of nodule detection.

Fig. 12. Figure illustrating case 45 from LOLA11 dataset which has severe pleural effusion. The left lung is entirely filled with fluid and the stage 1 V-Net fails to segment it.

VI. CONCLUSION

We propose a novel approach for lung volume segmentation from CT scans. The benefits of our design choices is demonstrated by experiments on a number of use-cases. The model performs segmentation on a downsampled 3D volume which alleviates the problem of large GPU memory required for training 3D CNN and reduces the inference time. The proposed structure correction method helps in incorporating shape characteristics into the segmentation model. The benefits of such structure correction is demonstrated by a substantial increase in the recall of tumor voxels which are missed by current CNN based segmentation models. Also to compensate for the loss of fine edge details a parallel flood-fill operation is proposed which recovers the fine surface details without incurring significant cost. The proposed method provides a fast and robust alternative to current lung segmentation methods when the average surface distance is not a critical requirement. One future extension of this work could be to obtain a large collection of un-labelled pathological lung CT volumes and to obtain the segmentation masks for them using the proposed algorithm. Using these generated segmentation masks the V-Net in Stage 1 could be trained. This will eliminate the need for structural correction in Stage 2 and can further simplify the model.


CT
Tumor Annotation
U-Net (d)
Multi-instance GAN
Dilated CNN (d)
SegNet (d)
3D U-Net (d)
V-Net (d)
P-HNN
Our
TVR=0.28
TVR=0.26
TVR=0.17
TVR=0.24
TVR=0.21
TVR=0.27
TVR=0.67
TVR=0.99
TVR=0.47
TVR=0.37
TVR=0.38
TVR=0.4
TVR=0.36
TVR=0.51
TVR=0.82
TVR=0.84
TVR=0.11
TVR=0.1
TVR=0.1
TVR=0.09
TVR=0.05
TVR=0.09
TVR=0.28
TVR=0.34
TVR=0.78
TVR=0.72
TVR=0.57
TVR=0.69
TVR=0.7
TVR=0.95
TVR=0.81
TVR=0.96
TVR=0.08
TVR=0.12
TVR=0.09
TVR=0.08
TVR=0.06
TVR=0.1
TVR=0.35
TVR=0.98
TVR=0.71
TVR=0.56
TVR=0.42
TVR=0.41
TVR=0.61
TVR=0.94
TVR=0.74
TVR=1.0
TVR=0.05
TVR=0.1
TVR=0.05
TVR=0.04
TVR=0.05
TVR=0.06
TVR=0.16
TVR=0.89
TVR=0.71
TVR=0.59
TVR=0.44
TVR=0.37
TVR=0.44
TVR=0.79
TVR=0.99
TVR=0.99
TVR=0.69
TVR=0.58
TVR=0.62
TVR=0.64
TVR=0.58
TVR=0.65
TVR=0.76
TVR=0.82
TVR=0.19
TVR=0.22
TVR=0.15
TVR=0.14
TVR=0.09
TVR=0.13
TVR=0.37
TVR=0.73

Disclaimer: The concepts, information and software presented in this paper are based on research and are not commercially available. Its future availability cannot be guaranteed.

REFERENCES

[1] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, "Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky noisy-or network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3484–3495, Nov. 2019.

[2] W. Zhu, C. Liu, W. Fan, and X. Xie, "Deeplung: Deep 3dDdual path nets for automated pulmonary nodule detection and classification," in *Proc. IEEE Winter Conf. Appl. Comput. Vision*, 2018, pp. 673–681.

[3] W. Zhu, Y. S. Vang, Y. Huang, and X. Xie, "DeepEM: Deep 3D convnets with em for weakly supervised pulmonary nodule detection," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2018, pp. 812–820.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2015, pp. 234–241.

[5] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2016, pp. 424–432.

[6] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vision*, 2016, pp. 565–571.

[7] W. Zhu *et al.*, "Anatomynet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy," *Med. Phys.*, vol. 46, no. 2, pp. 576–589, 2019.

[8] O. Oktay *et al.*, "Anatomically constrained neural networks (ACNNs): Application to cardiac image enhancement and segmentation," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 384–395, Feb. 2017.

[9] A. Mansoor *et al.*, "A generic approach to pathological lung segmentation," *IEEE Trans. Med. Imag.*, vol. 33, no. 12, pp. 2293–2310, Jul. 2014.

[10] G. Chen *et al.*, "Automatic pathological lung segmentation in low-dose CT image using eigenspace sparse shape composition," *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1736–1749, Jul. 2019.

[11] C. Shi, Y. Cheng, J. Wang, Y. Wang, K. Mori, and S. Tamura, "Low-rank and sparse decomposition based shape model and probabilistic atlas for automatic pathological organ segmentation," *Med. Image Anal.*, vol. 38, pp. 30–49, 2017.

[12] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. N. Metaxas, and X. S. Zhou, "Towards robust and effective shape modeling: Sparse shape composition," *Med. Image Anal.*, vol. 16, no. 1, pp. 265–277, 2012.

[13] G. Li, X. Chen, F. Shi, W. Zhu, J. Tian, and D. Xiang, "Automatic liver segmentation based on shape constraints and deformable graph cut in CT images," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5315–5329, Dec. 2015.

[14] S. Sun, C. Bauer, and R. Beichel, "Automated 3-D segmentation of lungs with lung cancer in CT data using a novel robust active shape model approach," *IEEE Trans. Med. Imag.*, vol. 31, no. 2, pp. 449–460, Feb. 2012.

[15] P. P. Rebouças Filho, P. C. Cortez, A. C. da Silva Barros, V. H. C. Albuquerque, and J. M. RS Tavares, "Novel and powerful 3D adaptive crisp active contour method applied in the segmentation of CT lung images," *Med. Image Anal.*, vol. 35, pp. 503–516, 2017.

[16] K. Nakagomi, A. Shimizu, H. Kobatake, M. Yakami, K. Fujimoto, and K. Togashi, "Multi-shape graph cuts with neighbor prior constraints and its application to lung segmentation from a chest CT volume," *Med. Image Anal.*, vol. 17, no. 1, pp. 62–77, 2013.

[17] L. Zhang, E. A. Hoffman, and J. M. Reinhardt, "Atlas-driven lung lobe segmentation in volumetric X-ray CT images," *IEEE Trans. Med. Imag.*, vol. 25, no. 1, pp. 1–16, Jan. 2006.

[18] E. M. van Rikxoort, B. de Hoop, M. A. Viergever, M. Prokop, and B. van Ginneken, "Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection," *Med. Phys.*, vol. 36, no. 7, pp. 2934–2947, 2009.

[19] A. Mansoor *et al.*, "Segmentation and image analysis of abnormal lungs at CT: Current approaches, challenges, and future trends," *RadioGraphics*, vol. 35, no. 4, pp. 1056–1076, 2015.

[20] J. E Roos *et al.*, "Computer-aided detection (CAD) of lung nodules in CT scans: Radiologist performance and reading time with incremental CAD assistance," *Eur. Radiol.*, vol. 20, no. 3, pp. 549–557, 2010.

[21] G. Singadkar, A. Mahajan, M. Thakur, and S. Talbar, "Automatic lung segmentation for the inclusion of juxtapleural nodules and pulmonary vessels using curvature based border correction," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 2016, 2018, Art. no. 2962047.

[22] S. Zhou, Y. Cheng, and S. Tamura, "Automated lung segmentation and smoothing techniques for inclusion of juxtapleural nodules and pulmonary vessels on chest CT images," *Biomed. Signal Process. Control*, vol. 13, pp. 62–70, 2014.

[23] J. Wang and H. Guo, "Automatic approach for lung segmentation with juxta-pleural nodules from thoracic CT based on contour tracing and correction," *Comput. Math. Methods Medicine*, vol. 2016, 2016, Art. no. 2962047.

[24] M. Anthimopoulos, S. Christodoulidis, L. Ebner, T. Geiser, A. Christe, and S. Mougiakakou, "Semantic segmentation of pathological lung tissue with dilated fully convolutional networks," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 714–722, Mar. 2019.

[25] A.-Al-Zubaer Imran, A. Hatamizadeh, S. P. Ananth, X. Ding, N. Tajbakhsh, and D. Terzopoulos, "Fast and automatic segmentation of pulmonary lobes from chest CT using a progressive dense V-network," *Comput. Methods Biomech. Biomed. Eng.: Imag. Visualization*, pp. 1–10, 2019.

[26] A. P. Harrison, Z. Xu, K. George, L. Lu, R. M. Summers, and D. J. Mollura, "Progressive and multi-path holistically nested neural networks for pathological lung segmentation from CT images," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2017, pp. 621–629.

[27] S. E. Gerard, J. Herrmann, D. W. Kaczka, G. Musch, A. Fernandez-Bustamante, and J. M. Reinhardt, "Multi-resolution convolutional neural networks for fully automated segmentation of acutely injured lungs in multiple species," *Med. Image Anal.*, vol. 60, 2019, Art. no. 101592.

[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[29] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3D medical image segmentation: A review," *Med. Image Anal.*, vol. 13, no. 4, pp. 543–563, 2009.

[30] A. A. A. Setio *et al.*, "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The luna16 challenge," *Med. Image Anal.*, vol. 42, pp. 1–13, 2017.

[31] S. Bakr *et al.*, "A radiogenomic dataset of non-small cell lung cancer," *Scientific Data*, vol. 5, 2018, Art. no. 180202.

[32] Lola11 challenge, 2011. [Online]. Available: https://lola11.grand-challenge.org/details/, Accessed: Sep. 1, 2011.

[33] S. G. Armato III *et al.*, "The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, no. 2, pp. 915–931, 2011.

[34] M. F. McNitt-Gray *et al.*, "The lung image database consortium (LIDC) data collection process for nodule detection and annotation," *Academic Radiol.*, vol. 14, no. 12, pp. 1464–1474, 2007.

[35] M. C. Hancock and J. F. Magnan, "Lung nodule malignancy classification using only radiologist-quantified image features as inputs to statistical learning algorithms: Probing the lung image database consortium dataset with two statistical learning methods," *J. Med. Imag.*, vol. 3, no. 4, 2016, Art. no. 044504.

[36] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, Y. Bengio and Y. LeCun, Eds., San Diego, CA, USA, May 2015.

[37] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[38] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Advances Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[39] T. Zhao, D. Gao, J. Wang, and Z. Tin, "Lung segmentation in CT images using a fully convolutional neural network with multi-instance and conditional adversary loss," in *Proc. IEEE 15th Int. Symp. Biomed. Imag.*, 2018, pp. 505–509.

[40] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Med. Imag.*, vol. 15, no. 1, p. 29, 2015.