2021

R2U3D: Recurrent Residual 3D U-Net for Lung Segmentation

Dhaval D. Kadia

MD Zahangir Alom

Ranga Burada

Tam Nguyen

Vijayan K. Asari

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 5, 2021, accepted June 9, 2021, date of publication June 16, 2021, date of current version June 28, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3089704

R^2U3D: Recurrent Residual 3D U-Net for Lung Segmentation

DHAVAL D. KADIA[1], (Member, IEEE), MD ZAHANGIR ALOM[2], (Member, IEEE), RANGA BURADA[3], TAM V. NGUYEN[1], (Senior Member, IEEE), AND VIJAYAN K. ASARI[4], (Senior Member, IEEE)

[1]Department of Computer Science, University of Dayton, Dayton, OH 45469, USA
[2]St. Jude Children's Research Hospital, Memphis, TN 38105, USA
[3]Microsoft, Redmond, WA 98052, USA
[4]Department of Electrical and Computer Engineering, University of Dayton, Dayton, OH 45469, USA

Corresponding author: Dhaval D. Kadia (kadiad1@udayton.edu)

This work was supported in part by the University of Dayton Open Access Fund and in part by the National Science Foundation (NSF) under Grant 2025234.

ABSTRACT 3D Lung segmentation is essential since it processes the volumetric information of the lungs, removes the unnecessary areas of the scan, and segments the actual area of the lungs in a 3D volume. Recently, the deep learning model, such as U-Net outperforms other network architectures for biomedical image segmentation. In this paper, we propose a novel model, namely, Recurrent Residual 3D U-Net (**R^2U3D**), for the 3D lung segmentation task. In particular, the proposed model integrates 3D convolution into the Recurrent Residual Neural Network based on U-Net. It helps learn spatial dependencies in 3D and increases the propagation of 3D volumetric information. The proposed R^2U3D network is trained on the publicly available dataset LUNA16 and it achieves state-of-the-art performance on both LUNA16 (testing set) and VESSEL12 dataset. In addition, we show that training the R^2U3D model with a smaller number of CT scans, i.e., 100 scans, without applying data augmentation achieves an outstanding result in terms of Soft Dice Similarity Coefficient (Soft-DSC) of 0.9920.

INDEX TERMS 3D Lung segmentation, R^2U3D, semantic segmentation, deep CNN, biomedical image analysis.

I. INTRODUCTION

Lung cancer is considered the second most common cancer type in both men and women [1]. The lung cancer patient is more likely to be successfully treated if it is found at an earlier stage, and before it has spread. Many patients having lung cancer report many kinds of delays in the diagnosis. The patients waited a median of 21 days before visiting a doctor and more than 22 days to complete the investigations. The median wait to start the treatment once the patients were seen at the cancer center was ten days. The total time from the development of the first symptoms to starting treatment was 138 days [2]. This affects the survival possibilities of the patients. Lung cancer screening using the low-dose CT is used to treat the patient to reduce lung cancer mortality.

Lung segmentation is important because it gives the volumetric information of the lungs. It is challenging because the lungs have irregular shapes, sizes, low contrasts, and complex boundaries [16]. Moreover, lung segmentation removes the unnecessary areas of the CT scan and segments the lungs' actual area, where attention is much essential. Lung segmentation prevents computer program to process irrelevant volumetric data that can produce false positives and leads to the erroneous diagnosis. Additionally, it can be considered as a necessary preprocessing for different lung disease analysis such as lung nodule detection or segmentation, pulmonary embolism (PE) diagnosis, Acute Respiratory Distress Syndrome (ARDS), and pneumothorax analysis [17], [19], [20].

Lung segmentation helps to save annotation time. Particularly for 3D segmentation applications, annotation is time-consuming, and such segmentation application can produce the segmentation that can be corrected with some additional efforts.

Traditional methods such as thresholding, edge tracking, region growing, contrast, and neighborhood homogeneity are applied for lung segmentation, but these methods do not give promising results when the CT scan of lungs is infected or has high attenuation patterns. They use edge detection filters

The associate editor coordinating the review of this manuscript and approving it for publication was Tao Zhou.

and other mathematical operations and algorithms. These methods have advantages; if the data are less diverse and domain knowledge is applied correctly, they give accurate results. Using the patch-based approach limits feature extraction by the number of patches, and that affects learning. The texture-based methods addressed such situations but gave poor results when some abnormalities were in the peripheral lung. Using traditional methods, the lung segmentation of a 3D CT scan can also be two-dimensional by applying 2D segmentation on each slice. The study shows that inter-slice smoothness is significantly smoother in 3D segmentation than 2D segmentation [15], [29].

In 1959, Arthur Samuel described machine learning as the ''field of study that gives computers the ability to learn without being explicitly programmed'' [30]. Deep learning (DL) is a subfield of machine learning, a field within artificial intelligence (AI). DL consists of a multilayer neural network that extracts features in more depth. The convolutional neural network (CNN) is one of the most powerful architectures in deep learning. CNN correlates nearby pixels of an image and produces different outputs using respective sets of weights. These sets of weights extract the features from an image. Repeating this process further gives us the features of an image. The initial stages of feature extraction give low-level features, and since the further stages extract the features from the previous or earlier stages, the later stages give high-level features. Low-level features help to correlate and understand small details, and high-level features represent the big picture or summarization of previous low-level features. A fully connected network is very bulky, whereas CNN has less trainable spatial feature extraction parameters. In recent years, deep convolutional neural networks are vital and outperforming state of the art in feature extraction, visual recognition, and object segmentation. The deep neural networks contain millions of parameters to solve complex problems, and hence, it is quite necessary to have the right data in the proper format and enough amount to train the parameters. With less data availability, it is necessary to discover the neural network architecture that can be trained using less amount of training data.

Traditional machine learning applications use techniques like support vector machines (SVMs) and random forests (RF). An issue with these approaches is that it requires collective efforts of field experts to approach useful features. Its optimization is time costly, and features are domain or problem-specific. Applicability of the same features among different domains is not always possible. Comparing to the traditional machine learning methods, deep learning has numerous advantages. Deep learning techniques learn useful features and do not require handcrafted features. Using transfer learning, the features learned from one dataset can be used to learn new features from different datasets. This gives importance to pre-trained deep learning systems trained on large datasets and sophisticated computational resources. They can be made available to the public to apply it to their applications. 3D convolutions are playing an essential role in spatial feature extraction in three-dimensions. Having fewer training data can be solved by training a 3D convolutional neural network on 3D patches of available data. This increases the training samples, and data augmentation helps further model generalization. Computer-aided diagnosis (CADx) requires sophisticated tools. It requires a deep neural network to learn complex features. Less training data will not let the deep neural network learn diverse features. So, instead of training a deep neural network from scratch, transfer learning can be applied, where the pre-trained model is trained on another dataset with enough diverse data. A significant transfer learning application is to use it for fine-tuning by freezing the initial layers of pre-trained convolutional neural networks and training the later layers. It works because high-level features differ among different datasets more than low-level features. It means better learned low-level feature extraction can be used to learn high-level features on a different dataset. During this process, the DL architecture remains the same; only the weights get updated. Another application of transfer learning is as a neural network weight initialization step. It helps the neural network to converge faster than other kinds of initialization approaches. Data augmentation generates new samples that increase the diversity in data points. Using such generated data for training reduces the probability of overfitting, and it overcomes the issue of the unbalanced dataset and helps generalize the neural network for testing dataset [14], [21].

Deep learning is also applicable for photoacoustic tomography artifact removal [24]. This paper uses a Fully Dense Unet (FD-Unet) for removing artifacts. DL can be used to design an annotation tool, and it is more helpful for multi-dimensional data like 3D CT scans or such time-series data. It can help medical professionals to estimate the initial annotations and make further corrections [25]. Additionally, the corrections can train the DL model to get better for the next use. D-UNet discusses the problems of computational resources for 3D CNN and demonstrates the combined neural network of both 2D and 3D CNN for chronic stroke lesion segmentation [31]. It uses four slices as both 2D and 3D context to apply them to its neural network and achieves better results while combining 3D CNN. AUNet proposed an attention-guided dense-upsampling network for an alternative to deconvolution commonly used for the upsampling [32]. It explained that the deconvolution was not as effective as bilinear upsampling for their application of breast mass segmentation in mammograms. The research [33] proposes a multidimensional region-based fully convolutional neural network and combines three views of 3D CT scan to give the possible shape of the detected nodule and its classification as malignant or benign. X-Net was developed to effectively extract features with fewer trainable parameters using depthwise separable convolution for brain stroke lesion segmentation [34]. It also designed Feature Similarity Module to extract a wide range of position-sensitive contextual information. Thus, having a vast data dimension and given computational resources, it is challenging to develop 3D CNN for achieving excellent performance. Furthermore, we propose our methods to overcome these problems and fulfill expectations.

The artificial intelligence algorithms can be called trained algorithms, and they are becoming more complex and sophisticated to solve complicated problems. This requires algorithmic regulation to systemically review them to prevent unexpected harm without constraining the innovation. It is essential to know how deep learning algorithms learn and reason from their learning. It is required to know the metrics of algorithmic responsibility and how it can be traced. Human responsibility plays a significant role in designing, improving, and maintaining the algorithm [26].

Deep neural networks have limitations to represent the learned knowledge to perform the assigned task explicitly. In such an environment, medical diagnostic tools need to be explainable, predictable, understandable, and transparent. This helps medical professionals, regulators, and patient's confidence and trust to understand how AI systems can be an integral part of routine diagnosis. Demonstrating the domain-specific features that help predict the output initially helps give an overview of a broader picture. Explainability is a necessary tool towards a trustworthy and ethical solution that is safe to use and has fairness in various aspects. It can be demonstrated using different approaches. Local interpretable methods give the reasoning for a single prediction, and the global methods give the abstract knowledge about the model, according to data [27]. The detailed analysis to understand the importance of a neuron is not limited by knowing the activation function's characteristics but necessitates its background learning process [28].

The physicians experience an increasing number of complex multi-dimensional visual readings, and this necessitates speed up clinical workflow with the help of deep learning and the technologies on top of that. While considering AI in medical imaging, we anticipate collaboratively using such technologies with physicians to decrease their burden, rather than replacing them.

The rest of this paper is organized as follows — Section II reviews the related work. Section III discusses details of the proposed framework. Section IV reports the experimental results, and Section V presents the evaluation of current research. Finally, Section VI concludes and paves the way to future work.

II. RELATED WORK

The current progress in the deep learning algorithms and available machine learning architectures provide neural networks to perform complex feature extraction [8]. Deep learning algorithms can be used to make computers analyze medical data accurately and generate multidimensional results. The pixel-wise classification of an image into logical related areas or volumes is called semantic segmentation. Different neural networks have their abilities in manipulating inputs and producing excellent results. Likewise, the deep learning technique U-Net outperforms the other network architectures for biomedical image segmentation. It is accurate and performs end-to-end semantic segmentation using an encoder and a decoder. U-Net is the popular approach for semantic medical image segmentation [7]. The first version of U-Net helped to crop and copy the feature map from the encoding unit to the decoding unit. It has significant advantages for segmentation tasks: first, the model allows the application of global location and context. Second, it gives good performance for the segmentation tasks with fewer training samples. It is using convolutional blocks and max-pooling in the encoder. The number of convolutional filters doubles at each level of the U-Net. The decoder uses de-convolution for up-sampling. While increasing the depth of a deep neural network, its accuracy may get saturated and degrades, and this degradation does not result from overfitting [23]. The residual learning makes the architecture less computationally complex, having the shortcut connection allowing the propagation of information without any degradation.

The deep Residual U-Net convolutional neural network uses a residual unit to extract discriminative features and overcomes the performance degradation by introducing a shortcut connection, which is an easier technique. Studies state that residual learning based neural network performs better than the sequential neural network [16]. This work applies data augmentation to generate synthetic data to enhance invariance property, which is shaped and illuminated. It applies online data augmentation that makes the number of augmented data equal to the number of total training data. The data augmentation includes flipping, shifting, rotation, and zooming. The analysis results state that shifting gives better improvement compared to rotation and flipping. This work applies post-processing to remove small areas of false-positives using connected components and applying thresholding. The publication has used a data dimension of 128 × 128 and achieved a DSC of 99.62 for 2D lung segmentation.

Multi-Scale Prediction Network gives predictions on multiple scales using single U-Net architecture. Using residual convolution blocks in a deep neural network solves gradient exploding and vanishing problems, providing the shortcut connection between input and output. The authors selected the lung CT scans from LUNA16 and NLST (National Lung Screening Trial) Dataset [18], having the criterion of selecting those CT scans that have interstitial lung disease and lung nodules attached on the lung wall [17].

The extension of the U-Net architecture using Recurrent Residual Convolutional Neural Networks called "R2U-Net" was evaluated in different fields of medical imaging [5]. The experimental results demonstrated better performance in 2D medical image segmentation. The residual units have an important role while training the deep architecture. The recurrent residual convolutional layers provide better feature representation for the segmentation tasks.

The recent 3D lung segmentation method – Extension of V-Net [6] is based on a modified version of the original V-Net [9]. It is using the max-pooling layer, in the beginning,

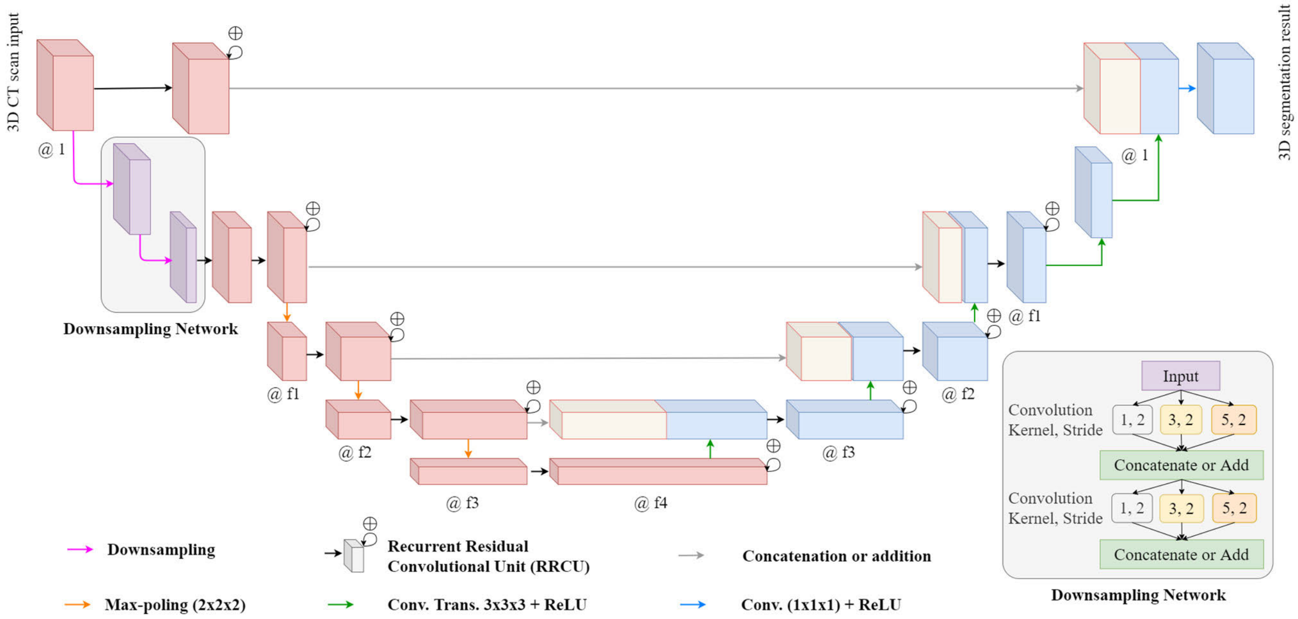


FIGURE 1. The overview of the proposed U-Net based R^{2}U3D architecture for lung segmentation.

to reduce the dimensions of the input scan from 256 × 512 × 512 to 64 × 128 × 128. This method randomly divide the available data of LUNA16 [3] into 700 scans and 188 scans for the validation of the neural network for 3D lung segmentation. The data augmentation techniques include spatial shifting and zooming along the depth axis.

III. PROPOSED FRAMEWORK

A. R^{2}U3D: RECURRENT RESIDUAL 3D U-NET

Inspired by the concept of volumetric image segmentation, we have developed Recurrent Residual 3D U-Net (R^{2}U3D) so that the proposed architecture can efficiently process volumetric data. The convolutional neural network learns by convolving over multi-dimensional data. The convolution layer represents the spatial features, and the higher the dimension, the better the spatial features will be. Hence, a 3D convolutional neural network extracts the features according to 3D local and, ultimately, over the entire 3D volume.

The Recurrent Neural Network learns the spatial dependencies over multiple steps, and the Residual Neural Network increases the propagation of 3D features. Considering these advantages, we have considered Recurrent and Residual Neural Networks based R^{2}U3D as a base architecture and improved it by applying different neural network module – Squeeze-and-Excitation Residual module, loss functions – Soft-DSC and Exponential Logarithmic Loss, optimizers – Adam, proper learning strategies, and appropriate hyperparameters. The analysis of deep neural networks becomes crucial while having high-dimensional data.

The proposed network architecture is illustrated in Fig. 1, consisting of a contracting path (left side) and an expansive path (right side). The left part is known as an encoder, and the right part is known as a decoder. The encoder consists of the down-sampling module, 3 × 3 × 3 convolutions, Recurrent Residual Convolutional Unit (RRCU) (shown in Fig. 2), and the max-pooling layer followed by 1 × 1 × 1 convolution. All of the convolutional units are followed by a Rectified Linear Unit (ReLU). Note that we down-sample the data in the beginning to avoid the hardware limitation. Instead of using the max-pooling layer, we are using Inception-like architecture for the down-sampling purpose. In particular, it has three convolutional layers with one filter with different kernel sizes. We either concatenate or add the output of each of them. This stretches the values (histogram) of the data and enhances the contrast. The decoder consists of 2 × 2 × 2 up-convolution, RRCU, and the concatenation of the feature map from the encoder followed by 1 × 1 × 1 convolution. The structure of **R^{2}U3D** in terms of number of filters is 1 1 1 f1 f2 f3 f4 f3 f2 f1 1 1 1. We have applied the dilation in the encoder and the recurrent convolution unit. The sigmoid activation function follows the final layer.

The Recurrent Residual Convolutional Unit (RRCU) is an important representative module of our proposed architecture. The Recurrent convolutional unit accumulates the features for different depths and gives better feature representation. It ensures low-level feature accumulation over the same levels of U-Net architecture.

B. R^{2}U3D VARIANTS

1) R^{2}U3D WITH DEFAULT PARAMETERS

Fig. 1 shows a typical R^{2}U3D architecture with the filters (f1, f2, f3, f4) = (40, 80, 160, 320). It has 20,306,691 parameters. The down-sampling network adds the outputs of convolutional layers. It uses Adam Optimizer with the

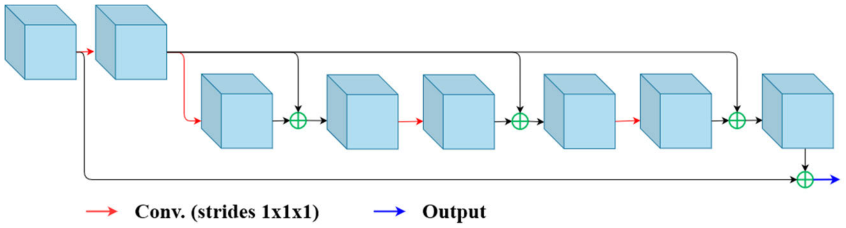


FIGURE 2. Recurrent Residual Convolutional Unit (RRCU) used in R^{2}U3D. RRCU with a depth of 3 units.

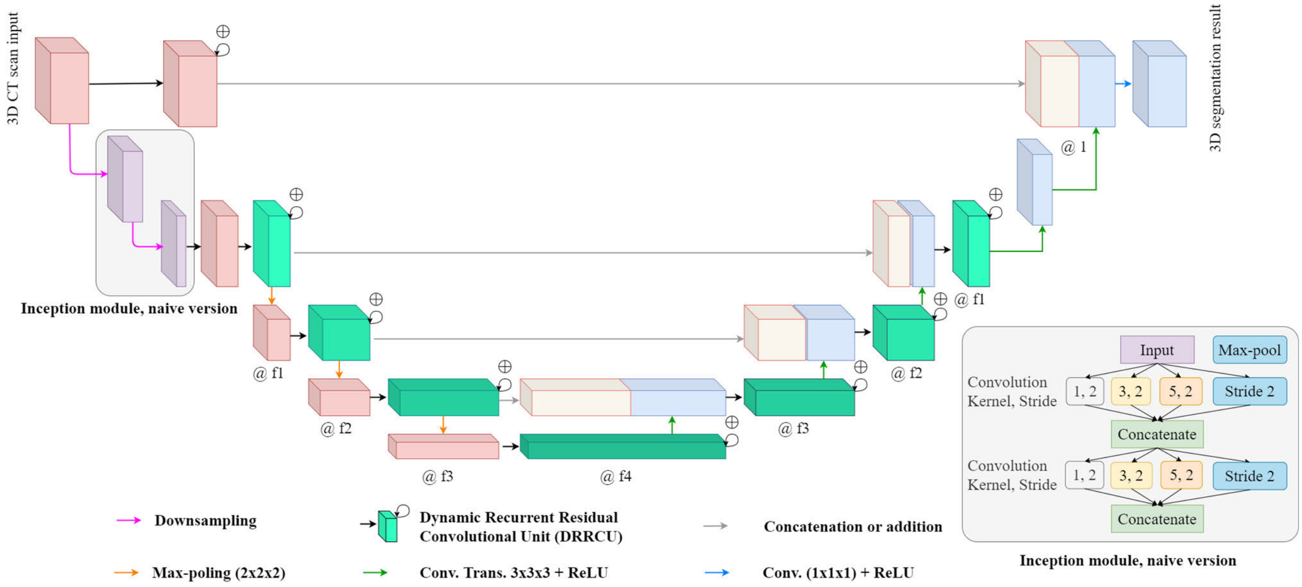

3D CT scan input
3D segmentation result
Inception module, naive version
@ f1
@ f2
@ f3
@ f4
@ 1
Input
Max-pool
Convolution Kernel, Stride
1, 2
3, 2
5, 2
Stride 2
Concatenate
Convolution Kernel, Stride
1, 2
3, 2
5, 2
Stride 2
Concatenate
Inception module, naive version
Downsampling
Max-poling (2x2x2)
Dynamic Recurrent Residual Convolutional Unit (DRRCU)
Conv. Trans. 3x3x3 + ReLU
Concatenation or addition
Conv. (1x1x1) + ReLU

FIGURE 3. R^{2}U3D architecture with Dynamic-Recurrent Residual Convolutional Unit (DRRCU).

learning rate of 0.001 and the loss function based on Dice Loss.

2) R^{2}U3D WITH DYNAMIC RECURRENT UNIT

The previous architecture has a less number of filters in each layer. It is necessary to increase the filters to make the deep neural network learn faster. To overcome this problem, we have introduced a Dynamic-Recurrent Residual Convolutional Unit (DRRCU). It primarily has a recurrent unit of different depth, with the Squeeze-and-Excitation Residual module in between. The purpose of applying different depth to the recurrent unit is to utilize the machine resources by eventually increasing the depth with approaching to the bottom layer of the architecture. That is less depth for the layers having a higher spatial dimension. This architecture has filters and depth of the recurrent unit $\{(f1, d1), (f2, d2), (f3, d3), (f4, d4))\} = \{(20, 1), (60, 2), (120, 3), (240, 4)\}$. It has 12,953,330 parameters. It uses Adam Optimizer with a learning rate of 0.001, and the loss function is the same as the previous architecture. We are using the Inception module, naïve version [12], with strides $2 \times 2 \times 2$ as a down-sampling network. It has three convolutional layers with one filter with different kernel sizes, along with a max-pooling layer. This architecture learns much faster than previous architectures. Fig. 3 shows the structure of R^{2}U3D with the dynamic recurrent unit.

Inspired by Squeeze-and-Excitation Networks [10], DRRCU is having the Recurrent Neural Network followed by the Squeeze-and-Excitation Residual module. Fig. 4 represents the DRRCU at different depths. This unit helps to accumulate the low-level features for higher depth and utilizing the available machine resources. As shown in Fig. 5, the Squeeze-and-Excitation Residual module proposes the channel interdependencies and nonlinear interactions among the channels. It uses the global information of each of the channels and feeds it to two Fully Connected Networks (FCN). The first layer has the activation function ReLU, and the second has Sigmoid to normalize the output values from zero to one. The output of FCN is then multiplied with the input and generates the scaled input. The input is then added to the scaled input, followed by ReLU activation.

IV. EXPERIMENTAL SETUP

We have implemented R^{2}U3D using Keras deep learning library, with TensorFlow [35] as backend and Nvidia GeForce RTX 2080 Ti having 12 GB graphics memory.

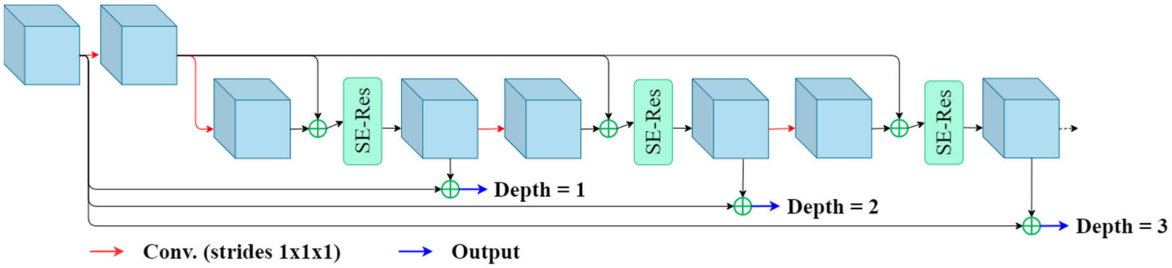


FIGURE 4. Dynamic-Recurrent Residual Convolutional Unit (DRRCU) with Squeeze-and-Excitation Residual module.

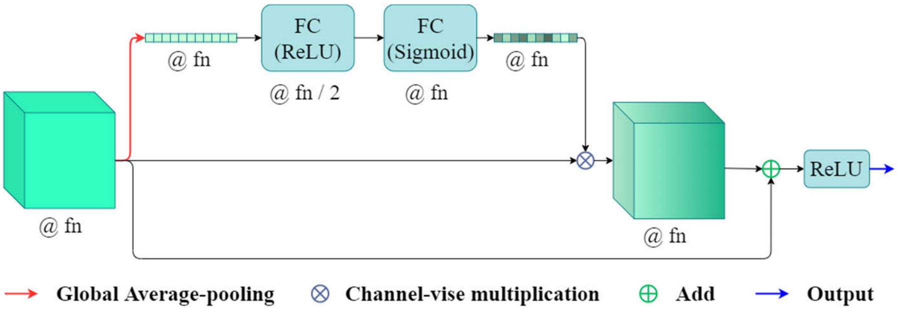


FIGURE 5. Squeeze-and-Excitation Residual module.

A. DATASET DETAILS

We used publicly available datasets – LUng Nodule Analysis 2016 (LUNA16) and VESsel SEgmentation in the Lung 2012 (VESSEL12) [3], [4]. LUNA16 consists of 888, and VESSEL12 consists of 20 three-dimensional lung CT scans, along with the segmented ground truth. We have considered 876 CT scans in our analysis, out of which 700 CT scans are for training, and 176 CT scans are for testing. Some of the ground-truth of LUNA16 scans are having holes inside lung areas, and most of them represent nodules. We have used CT scans from LUNA16 for both training and testing, and CT scans from VESSEL12 for testing.

B. DATA PREPARATION AND TRAINING SETTINGS

The spatial resolution of data is 256 × 512 × 512. The training and testing of 3D CT scans vary in the number of slices. Hence, to down-sample and up-sample 3D CT scans into 256 × 512 × 512 dimension, we repeat the slices if actual available slices are less than 256 and select equally over the available slices if they are more than 256. This is performed with the proper number of steps over the z-axis so that, the sampled data preserve the actual shape. We are normalizing each CT scan in the range from 0 to 1. We are not applying any data augmentation technique. Our training strategy is based on the random selection of training data from a certain part of the dataset. We are selecting five scans randomly from the set of first 100 scans, train them for five epochs, and repeat the procedure for 500 iterations. This strategy helps the model to overcome the problem of overfitting, particularly over the local set of training data. We used the learning rate of 0.001 and 0.0001 for 400 and 100 iterations, respectively. We kept batch-size one according to the available computational resources.

V. EVALUATION

A. PERFORMANCE METRICS AND LOSS FUNCTIONS

For the evaluation, we adopt the Dice Similarity Coefficient (*DSC*) as below.

$$DSC = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i + \sum_i^N g_i} \tag{1}$$

We consider *Soft-DSC* for the evaluation:

$$Soft - DSC = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{2}$$

where N is the number of voxels in each image, $p_i \in P$ is a voxel of predicted segmentation P, and $g_i \in G$ is a voxel of binary ground-truth G.

The Exponential Logarithmic Loss [11] is computed as below:

$$Loss_{ELL} = w_{DSC} Loss_{DSC} + w_{Cross} Loss_{Cross} \tag{3}$$

$$Loss_{DSC} = (-\ln(DSC))^{\gamma_{DSC}} \tag{4}$$

$$Loss_{Cross} = WCEL^{\gamma_{WCEL}} \tag{5}$$

The loss is calculated with the addition of *DSC* and Weighted Cross Entropy with Logits (*WCEL*) [36], with a ratio. *WCEL* is first applying the Logit function, which is the inverse of Sigmoid function, to the prediction, and then calculated the Weighted Cross Entropy. Logit is used to calculate the values before the Sigmoid Activation. Here, $w_{DSC} = 0.8$, $w_{Cross} = 0.2$ and $\gamma_{DSC} = \gamma_{WCEL} = 0.3$.

TABLE 1. Comparison of Soft-DSC of the results from proposed architecture with [6], for VESSEL 12 and LUNA16 (last 176 scans) as testing sets. The best results are marked as bold-faced.

	V-Net [10]	Extended V-Net [6]	**R^{2}U3D** (Default)	**R^{2}U3D** (Dynamic)
Training scans	700	700	100	100
Training iterations	8400	8400	15650	12500
Soft-DSC (VESSEL12)	0.972	0.987	0.9881	**0.9920**
Soft-DSC (LUNA16)	–	–	0.9831	0.9859

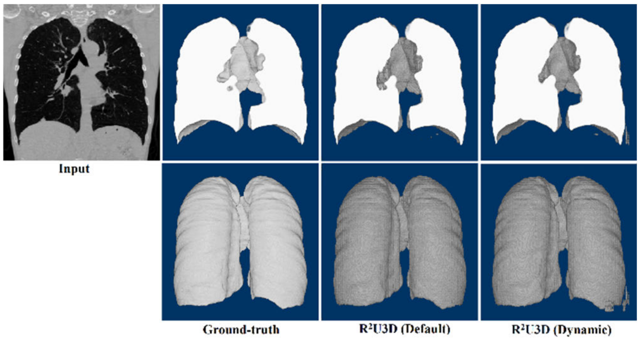


FIGURE 6. Visualization of 3D CT scan, ground-truth, and segmentation results using the proposed methods. The first row and second row show the dissected and entire lungs, respectively.

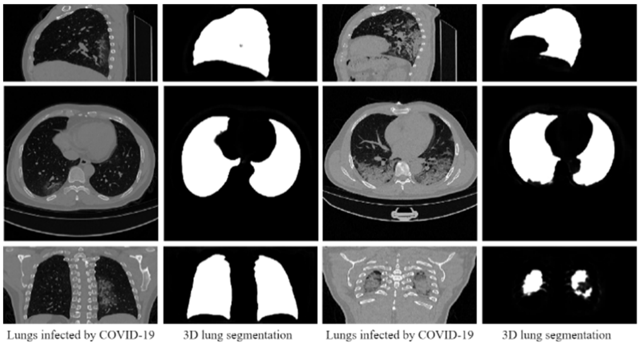


FIGURE 7. Visualization of lungs infected by COVID-19 and the corresponding 3D lung segmentation.

B. RESULTS AND DISCUSSIONS

The deep neural network that is the extension of V-Net (Extended V-Net) [6] is trained with 700 CT scans of LUNA16. Whereas, we have considered the first 100 CT scans from LUNA16 for the training set, tested our architecture with VESSEL12, and compared the results with Extended V-Net and V-Net as shown in Table 1. The **R^{2}U3D** (Default) and **R^{2}U3D** (Dynamic) provides the results of Soft-DSC as 0.9881 and 0.9920, respectively. Since the training data are enough, the testing data should be more than 20 number of 3D CT scans. Testing on less data does not guarantee the generalization and may overfit the model even if we observe a good testing accuracy. Therefore, we have tested our architecture with the last 176 CT scans of LUNA16. The **R^{2}U3D** (Default) and **R^{2}U3D** (Dynamic) give Soft-DSC of 0.9831 and 0.9859, respectively.

While testing all the remaining 776 scans of LUNA16, **R^{2}U3D** (Dynamic) gives the Soft-DSC 0.9828. While training one variant of **R^{2}U3D** (Default) with the rest of 700 LUNA16 CT scans in seven phases having 100 CT scans each and 700 CT scans in total, it shows an increment in the result for testing 176 CT scans of LUNA16. By training in batch of 100: 1 – 100, 100 – 200, 200 – 300, 300 – 400, 400 – 500, 500 – 600, by applying transfer learning, and testing on 700 – 876 (176 scans), the DSC is 0.9813, 0.9815, 0.982, 0.982, 0.9818, 0.9822. After that, training scans 1 – 700, further improves the DSC to 0.9827. Thus, the test results on 176 scans improve with more training data.

C. APPLICABILITY FOR COVID-19 DIAGNOSIS

The deep learning model is trained on LUNA16. While applying it on 3D CT scans having COVID-19 infection, the segmentation predictions are imperfect and fail at infected areas. 3D CT scans in Fig. 7 are selected from COVID-19 CT Lung and Infection Segmentation Dataset [22]. Lung infectious diseases are having diverse patterns of infections and lesions. It is necessary to train the deep neural network on infected lungs based data to get better results on such a testing dataset. This problem can be solved by training DNN on COVID-19 CT scans having lung masks or generating synthetic data and applying a generative adversarial network (GAN) to help segment infected lungs.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have proposed the **R^{2}U3D** network with its variants. We developed 3D deep neural network architecture of Dynamic-Recurrent Residual Convolutional Neural Network with a suitable down-sampling module and a Squeeze-and-Excitation Residual module, and trained with the Exponential Logarithmic Loss and Adam Optimizer. We trained our neural network on LUNA16 and tested it on both VESSEL12 and LUNA16 datasets. We have achieved better accuracies with less number of training data, and observed the improvement while training with the additional data.

Future work includes the application of medical imaging for diseases like lung cancer, chronic obstructive pulmonary disease (COPD), acute respiratory distress syndrome (ARDS), and pulmonary embolism (PE). These diseases damage the lungs and make imaging based tasks challenging. In addition, we aim to apply our methods for the segmentation of nodules from the segmented lungs, and classify them as malignant or benign. Furthermore, Coronavirus Disease 2019 (COVID-19) shows the regions having Ground-Glass Opacities (GGO) inside the lungs [13]. **R^{2}U3D** can segment the lungs and be applied further to segment the GGO region from the segmented lungs using an appropriate dataset. The lungs infected by COVID-19 show diverse types of infection patterns other than GGO. It is essential to segment the infected lungs properly, and we plan to train the proposed deep neural network on COVID-19 based datasets. Thus we plan to employ data augmentation techniques that have shown promising results for this modality and design novel methods for robust lung segmentation.

REFERENCES

[1] *American Cancer Society—Home*. Accessed: Nov. 4, 2020. [Online]. Available: https://www.cancer.org/
[2] P. M. Ellis and R. Vandermeer, "Delays in the diagnosis of lung cancer," *J. Thoracic Disease*, vol. 3, no. 3, p. 183, 2011.
[3] *LUNA16—Home*. Accessed: Nov. 4, 2020. [Online]. Available: https://luna16.grand-challenge.org/
[4] *VESSEL12—Home*. Accessed: Nov. 4, 2020. [Online]. Available: https://vessel12.grand-challenge.org/
[5] M. Zahangir Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*. [Online]. Available: http://arxiv.org/abs/1802.06955
[6] P. Sousa, A. Galdran, P. Costa, and A. Campilho, "Learning to segment the lung volume from CT scans based on semi-automatic ground-truth," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 1202–1206.
[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2015, pp. 234–241.
[8] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. Shamima Nasrin, B. C Van Esesn, A. A S. Awwal, and V. K. Asari, "The history began from AlexNet: A comprehensive survey on deep learning approaches," 2018, *arXiv:1803.01164*. [Online]. Available: http://arxiv.org/abs/1803.01164
[9] 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 Vis. (3DV)*, Oct. 2016, pp. 565–571.
[10] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[11] K. C. Wong, M. Moradi, H. Tang, and T. Syeda-Mahmood, "3D segmentation with exponential logarithmic loss for highly unbalanced object sizes," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2018, pp. 612–619.
[12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[13] S. Zhou, Y. Wang, T. Zhu, and L. Xia, "CT features of coronavirus disease 2019 (COVID-19) pneumonia in 62 patients in Wuhan, China," *Amer. J. Roentgenology*, vol. 214, no. 6, pp. 1287–1294, Jun. 2020.
[14] A. Ziabari, A. Shirinifard, M. R. Eicholtz, D. J. Solecki, and D. C. Rose, "A two-tier convolutional neural network for combined detection and segmentation in biological imagery," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2019, pp. 1–5.
[15] G. De Nunzio, E. Tommasi, A. Agrusti, R. Cataldo, I. De Mitri, M. Favetta, S. Maglio, A. Massafra, M. Quarta, M. Torsello, I. Zecca, R. Bellotti, S. Tangaro, P. Calvini, N. Camarlinghi, F. Falaschi, P. Cerello, and P. Oliva, "Automatic lung segmentation in CT images with accurate handling of the hilar region," *J. Digit. Imag.*, vol. 24, no. 1, pp. 11–27, Feb. 2011.
[16] A. Khanna, N. D. Londhe, S. Gupta, and A. Semwal, "A deep residual U-Net convolutional neural network for automated lung segmentation in computed tomography images," *Biocybern. Biomed. Eng.*, vol. 40, no. 3, pp. 1314–1327, Jul. 2020.
[17] Y. Gu, Y. Lai, P. Xie, J. Wei, and Y. Lu, "Multi-scale prediction network for lung segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 438–442.
[18] *NLST—The Cancer Data Access System*. Accessed: Nov. 4, 2020. [Online]. Available: https://cdas.cancer.gov/nlst/
[19] Y. Xin, G. Song, M. Cereda, S. Kadlecek, H. Hamedani, Y. Jiang, J. Rajaei, J. Clapp, H. Profka, N. Meeder, J. Wu, N. J. Tustison, J. C. Gee, and R. R. Rizi, "Semiautomatic segmentation of longitudinal computed tomography images in a rat model of lung injury by surfactant depletion," *J. Appl. Physiol.*, vol. 118, no. 3, pp. 377–385, Feb. 2015.
[20] S. Do, K. Salvaggio, S. Gupta, M. Kalra, N. U. Ali, and H. Pien, "Automated quantification of pneumothorax in CT," *Comput. Math. Methods Med.*, vol. 2012, pp. 1–7, Jan. 2012.
[21] B. Sahiner, A. Pezeshk, L. M. Hadjiiski, X. Wang, K. Drukker, K. H. Cha, R. M. Summers, and M. L. Giger, "Deep learning in medical imaging and radiation therapy," *Med. Phys.*, vol. 46, no. 1, pp. e1–e36, Jan. 2019.
[22] M. Jun, G. Cheng, W. Yixin, A. Xingle, G. Jiantao, Y. Ziqi, and H. Jian, "COVID-19 CT lung and infection segmentation dataset (version verson 1.0) [data set]," Univ. Sci. Technol., Nanjing, China, Tech. Rep., 2020, doi: 10.5281/zenodo.3757476.
[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[24] S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis, "Fully dense UNet for 2-D sparse photoacoustic tomography artifact removal," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 568–576, Feb. 2020.
[25] S. Park, L. C. Chu, E. K. Fishman, A. L. Yuille, B. Vogelstein, K. W. Kinzler, K. M. Horton, R. H. Hruban, E. S. Zinreich, D. Fadaei Fouladi, S. Shayesteh, J. Graves, and S. Kawamoto, "Annotated normal CT data of the abdomen for deep learning: Challenges and strategies for implementation," *Diagnostic Interventional Imag.*, vol. 101, no. 1, pp. 35–44, Jan. 2020.
[26] A. Tutt, "An FDA for algorithms," *Admin. L. Rev.*, vol. 69, p. 83, Jan. 2017.
[27] A. Singh, S. Sengupta, and V. Lakshminarayanan, "Explainable deep learning models in medical image analysis," 2020, *arXiv:2005.13799*. [Online]. Available: http://arxiv.org/abs/2005.13799
[28] R. Meyes, C. Waubert de Puiseau, A. Posada-Moreno, and T. Meisen, "Under the hood of neural networks: Characterizing learned representations by functional neuron populations and network ablations," 2020, *arXiv:2004.01254*. [Online]. Available: http://arxiv.org/abs/2004.01254
[29] M. Kim, J. Yun, Y. Cho, K. Shin, R. Jang, H.-J. Bae, and N. Kim, "Deep learning in medical imaging," *Neurospine*, vol. 16, no. 4, p. 657, 2019.
[30] M. Awad and R. Khanna, "Machine learning in action: Examples," in *Efficient Learning Machines*. Berkeley, CA, USA: Apress, 2015, pp. 209–240.
[31] Y. Zhou, W. Huang, P. Dong, Y. Xia, and S. Wang, "D-UNet: A dimension-fusion u shape network for chronic stroke lesion segmentation," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 3, pp. 940–950, May 2021.
[32] H. Sun, C. Li, B. Liu, Z. Liu, M. Wang, H. Zheng, D. Dagan Feng, and S. Wang, "AUNet: Attention-guided dense-upsampling networks for breast mass segmentation in whole mammograms," *Phys. Med. Biol.*, vol. 65, no. 5, Feb. 2020, Art. no. 055005.
[33] A. Masood, B. Sheng, P. Yang, P. Li, H. Li, J. Kim, and D. D. Feng, "Automated decision support system for lung cancer detection and classification via enhanced RFCN with multilayer fusion RPN," *IEEE Trans. Ind. Informat.*, vol. 16, no. 12, pp. 7791–7801, Dec. 2020.
[34] K. Qi, H. Yang, C. Li, Z. Liu, M. Wang, Q. Liu, and S. Wang, "X-Net: Brain stroke lesion segmentation based on depthwise separable convolution and long-range dependencies," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Cham, Switzerland: Springer, 2019, pp. 247–255.
[35] *TensorFlow*. Accessed: Nov. 4, 2020. [Online]. Available: https://www.tensorflow.org/
[36] *tf.nn.weighted_cross_entropy_with_logits-Home*. Accessed: Nov. 4, 2020. [Online]. Available: https://www.tensorflow.org/api_docs/python/tf/nn/weighted_cross_entropy_with_logits

DHAVAL D. KADIA (Member, IEEE) received the B.E. degree in computer science from The Maharaja Sayajirao University of Baroda, India, in 2017, and the M.S. degree in computer science from the University of Dayton, USA, in 2021.

He was a Research Intern with Defence Research and Development Organisation, India. He was a Teaching Assistant and Summer Fellow with the University of Dayton. He was a member with the Vision and Mixed Reality Laboratory and the Center of Excellence for Computational Intelligence and Machine Vision (Vision Lab), University of Dayton. He is currently working as a Staff Research Associate with the NCIRE—The Veterans Health Research Institute. His research interests include design and analysis of algorithms, image processing, computer vision, machine learning, deep learning, imaging, medical imaging, computer graphics, and mixed reality.

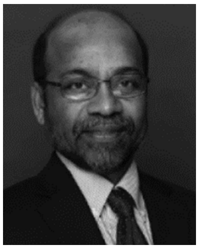
Mr. Kadia is a member of the Association for Computing Machinery (ACM). He received the Best Paper Award at the International Conference on Computing, Analytics, and Networks, in 2017, and the Graduate Student Summer Fellowship Award from the University of Dayton, in 2019.

MD ZAHANGIR ALOM (Member, IEEE) received the B.Sc. degree (Hons.) in computer science and engineering from the University of Rajshahi, Bangladesh, in 2008, the M.Eng. degree in computer engineering from Chonbuk National University, South Korea, in 2012, and the Ph.D. degree in electrical and computer engineering from the University of Dayton, OH, USA, in 2018.

He is currently working as a Bioinformatics Research Scientist at St. Jude Children's Research Hospital, USA. His research interests include machine learning, deep learning, biomedical imaging, medical informatics, computational pathology, computer vision, and big data analytics.

RANGA BURADA received the bachelor's degree in electrical and communications engineering from JNTU-HYD, India, in 2011, and the master's degree in electrical engineering from the University of Dayton, USA, in 2015, where he is currently pursuing the Ph.D. degree. He is currently working as an Image Quality Engineer at Microsoft. His research interests include developing camera image quality metrics and image processing algorithms.

TAM V. NGUYEN (Senior Member, IEEE) received the B.S. degree in computer science from the University of Science, Vietnam, in 2005, the M.Eng. degree from Chonnam National University, South Korea, in 2009, and the Ph.D. degree from the National University of Singapore (NUS), in 2013. He is currently an Assistant Professor and the Director of the Vision and Mixed Reality Laboratory, Department of Computer Science, University of Dayton. He has authored or coauthored more than 80 research articles with more than 1600 citations. His research interests include computer vision, machine learning, mixed reality, and multimedia analysis.



VIJAYAN K. ASARI (Senior Member, IEEE) is currently a Research Scholars Endowed Chair in wide area surveillance and a Professor with the Department of Electrical and Computer Engineering, University of Dayton, OH, USA, where he is also the Director of the Center of Excellence for Computational Intelligence and Machine Vision (Vision Lab). He has published, and coauthored with his graduate students and colleagues, more than 700 research articles, including more than 110 peer-reviewed journal articles, in the areas of image processing, computer vision, pattern recognition, machine learning, deep learning, and high performance digital system architecture design. He is a fellow of the SPIE. He has co-organized several IEEE and SPIE conferences and workshops. He has received many awards for his teaching, research, and technical leadership, including the Vision Award for Excellence, in August 2017, the Sigma Xi George B. Noland Award, in April 2016, and the Outstanding Engineers and Scientists Award for Technical Leadership, in April 2015.