This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3024277, IEEE Access

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.Doi Number

Automatic Segmentation using a Hybrid Dense Network Integrated with an 3D-Atrous Spatial Pyramid Pooling Module for Computed Tomography (CT) Imaging

Abdul Qayyum[1,6], Iftikhar Ahmad[2], Wajid Mumtaz[3],Madini O. Alassafi[4], Rayed AlGhamdi[5] and Moona Mazher[6]

[1]Department of Information and Engineering, ImViA Laboratory, University of Burgundy, Dijon, 21000, France
[3]Department of Electrical Engineering, National University of Science and Technology, Pakistan
[2,4,5]Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[6]Centre for Intelligent Signal and Imaging Research, Department of Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, Perak, Malaysia

Corresponding author: Abdul Qayyum (e-mail: engr.qayyum@ gmail.com).

This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant No. No. DF-148-611-1441. The authors, therefore, gratefully acknowledge DSR for technical and financial support.

ABSTRACT Computed tomography (CT) with a contrast-enhanced imaging technique is extensively proposed for the assessment and segmentation of multiple organs, especially organs at risk. It is an important factor involved in the decision making in clinical applications. Automatic segmentation and extraction of abdominal organs, such as thoracic organs at risk, from CT images are challenging tasks due to the low contrast of pixel values surrounding other organs. Various deep learning models based on 2D and 3D convolutional neural networks have been proposed for the segmentation of medical images because of their automatic feature extraction capability based on large labeled datasets. In this paper, we proposed a 3D-atrous spatial pyramid pooling (ASPP) module integrated with a proposed 3D DensNet encoder–decoder network for volumetric segmentation to segment abdominal organs from CT. The proposed network used a 3D-ASPP block to capture spatial information in multiscale input feature maps from the decoder side. We also proposed a 3D-ASPP block with a 3D DensNet network for automatically processing 3D medical volumetric images. The proposed hybrid network was named 3D-ASPPDN for volumetric segmentation via CT medical imaging. We tested our proposed approach on a public dataset, Thoracic Organs at Risk (SegTHOR) 2019.The proposed solution showed excellent performance in comparison with other existing state-of-the-art DL methods. The proposed method achieved Dice scores 97.89% on the SegTHOR dataset. Results presented that 3D-ASPPDN exhibited enhanced performance in volumetric biomedical segmentation. The proposed model could be used for volumetric segmentation in clinical applications to diagnose problems in multi class organs.

INDEX TERMS 3D volumetric segmentation, 3D deep learning models, 3D-atrous spatial pyramid pooling (ASPP), SegTHOR

I. INTRODUCTION

The contrast-enhanced Computed Tomography (CT) has been used effectively for thoracic diseases and has great importance in medical field. The radiologist spent lot of the time for identification and localization of organs in CT images. The automatic segmentation and localization of organs from CT images could be helpful to diagnose the thoracic organs at risk in CT images. The CT images contain multiple organs have three-dimensional(3D) structure and manually segmented of each slice from 3D volume is time consuming process. Automatic segmentation based on deep learning models could be

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3024277, IEEE Access

helpful and can produce effective results for segmentation and detection of organs in 3D CT images. Lot of works have been done for automatic segmentation of CT images in biomedical fields. Mostly the segmentation in medical fields has been done based on individual slice of each patient such as 2D U-net model. The U-net model consisted of encoder and decoder part with skip connection and produce better results in medical image segmentation.

In 3D deep learning models, the segmentation of all slices of the CT images done at once. However, the first issue in 3D segmentation deep learning models is computationally complexity. The second issue is due to large number of weights parameters can make the model to overfitting due limited training data. The second could be resolve by using small 3D convolutional layers or decreasing the number of channels per convolutional layer. This idea might not be work well especially for small training dataset. Semantic segmentation produces an effective and rich representation of tumors and achieves accurate extraction of contact surface area [1], irregularity [2], and morphological traits.

Automatic segmentation is generally used in addition to semiautomatic and manual segmentation. Manual segmentation requires expertise to deal with tasks. It is subjective, requires considerable time for processing, and is poorly reproducible. It entirely depends on human-made handcrafted features; hence, it is impractical for real applications [3]. Similarly, semiautomatic segmentation initially depends on human involvement that could cause mistakes and errors. On the contrary, automatic segmentation could be accurate, produce minimal errors, and help surgeon's segment. However, structural variability, noise existence, the complexity of 3D spatial multiclass features, large-scale spatial variability, partial volume effects, and the similarity of nearby organs to make automatic segmentation a difficult task [3].

Convolutional neural networks (CNNs) have recently been used for classification and other important applications, such as object detection and volumetric image segmentation. They achieve state-of-the-art performance compared with traditional machine learning models. CNN models can automatically extract hierarchical features from raw images without depending on handcrafted features. The deep layers in CNN models capture global information in a broad way due to large receptive fields [4], while shallow layers grasp only local information. The automatic extraction of structural information and the delineation of organs from images [5] are highly needed to perform visual augmentation [6], interventions [7], and computer-assisted diagnosis [8]. The interventional and diagnostic imaging consisting of 3D images and volumetric segmentations can consider the entire volume content at once, which is the requirement for the automatic segmentation of biomedical images. The depth of 3D CNNs is limited compared with 2D networks due to many constraints, such as the memory consumption of the graphics processing unit (GPU), high computational cost, and the requirement for a large number of annotated datasets. These 3D CNN models are also not flexible and efficient for 3D image sequences due to the large number of parameters and the limited number of hardware resources.

Many approaches based on patch-wise image classification, postprocessing steps, and different techniques, such as Markov random fields [9], voting strategies [10], and level sets [11], are used in combination with CNN models to perform volumetric segmentation. These techniques globally fail to produce accurate and efficient volumetric segmentation. Patch-wise approaches could not perform well due to high computation cost and produce redundant information that makes the algorithm runtime high. Therefore, efficient computational schemes are required to tackle 3D CNN segmentation problems.

Fully convolutional networks (FCNs) [12] have been widely used for dense segmentation. They change fully connected layers with upsampling layers to retain the spatial structure. The upsampling layers use downsample feature maps that can restore input images into the original resolution. FCN models automatically produce a pixel-to-pixel-based segmentation map for medical image segmentation [12]. For the dense semantic segmentation task, the dense label provides accurate solutions for small lesions or objects and produces improved solutions with reduced classification error rate. This approach, however, has two issues. First, the receptive field could be large by using a convolutional layer with pooling or striding. Second, the resolution of input images is downsampled, and small objects with different scales and shapes in multiclass scenarios are difficult to classify [13].

Thus, the application of deconvolution or bilinear interpolation by using an upsampling layer at the decoder part of the network could not assure an accurate segmentation map. Various state-of-the-art networks, such as DeepLab with an atrous spatial pyramid pooling (ASPP) module [14], U-Net [15], FCN [12], dense FCN, and residual dense FCN, have been proposed to handle this issue. DeepLabv2 [16] introduces ASPP for semantic segmentation.

The ASPP network handles objects at multiple scales [17] to capture multiscale information on the basis of parallel atrous-based CNN layers by using multiple atrous rates. It achieves efficient and robust performance in dense semantic labeling and detects small objects in various scales and shapes.

ASPP consists of different numbers of atrous CNN layers and combines these layers in a parallel branch. Each layer uses different multiple kernel functions to capture feature maps with specific field of view. Simple 3D CNN models might not perform well, and we need a hybrid solution to tackle 3D segmentation problems and incorporate some valuable information at the dataset level and in model architecture. In this study, we proposed a 3D-ASPP module

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3024277, IEEE Access

in a 3D DensNet network (3D-DN) model for the automatic segmentation of CT images. ASPP block could be integrated within any CNN model. The ASPP module was embedded with the proposed 3D-DN as a bottom layer to extract contextual information with multiple resolutions. We extended 2D ASPP into a 3D-ASPP block and integrated it with the proposed 3D-DN for volumetric segmentation. In accordance with the literature review and our knowledge, the 3D-ASPP module was proposed for the first time. A deep learning (DL)-based model, which processed volumetric slices for volumetric segmentation on the basis of abdominal CT volumes, was proposed. We also proposed volumetric convolutions that processed 3D input slices as a volume. The main contributions of this work are follows:

- First, we proposed 3D-DN for 3D segmentation. 3D-DN involved an encoder–decoder structure. Various 3D DensNet blocks were introduced into the encoder–decoder part of the network.
- Second, we introduced a 3D-ASPP module into 3D-DN at the bottom of the encoder–decoder module. The 3D-ASPP block captured the spatial information at multiscale and produced improved segmentation. ASPP extracted contextual information in a multiscale form and recovered spatial information with different fields of view to determine sharp object boundaries from encoder to decoder.

Our approach for 3D volumetric segmentation provided an automatic solution by indicating the complete volume of a patient at once for accurate and robust segmentation. In our proposed method, no postprocessing steps were required for further processing and evaluation. We tested our model on abdominal CT datasets, Thoracic Organs at Risk (SegTHOR) 2019 [18]. The proposed technique was generalized for the SegTHOR 2019 dataset, and its performance was compared with that of existing state-of-the-art segmentation models.

II. RELATED WORK

DL models have been used for natural language processing, image analysis [19], image classification, and image segmentation. These deep neural network (DNN) models have been successfully used in medical imaging challenges [20]. Meanwhile, CNN-based models have shown best performance in the medical imaging domain by using either patch-based or multiscale pixel-based segmentation approaches that could increase the segmentation results.

For example, Zhang et al. introduced CNN-based brain tissue segmentation based on multimodal magnetic resonance imaging (MRI) [21]. Pereira et al. introduced a CNN model for the segmentation of brain tumors in multimodal MRI [22] and claimed satisfactory results for complete tumors. Lee et al. proposed a DL-based CNN model for brain segmentation features [23]. Li et al. proposed a 2D CNN model using CT slices for segmentation and compared its performance with that of traditional machine learning approaches, such as AdaBoost [24], random forests [25], and support vector machine [26]. This study determined that CNNs had limitation in tumor segmentation due to unclear borders and uneven density. However, existing CNN models have been used for segmentation on the basis of 2D slices that are extracted from 3D volumes. These models could be a choice on the basis of using low computational resources and memory consumptions. We required an automatic 3D volumetric segmentation solution based on DL models that could consider the axial direction of 3D volumes to assist medical doctors and would be beneficial in health applications.

In early studies, Shakeri et al. introduced a 2D CNN incorporated with a 3D conditional random field algorithm as postprocessing for tumor detection from brain slices [27] to achieve volumetric consistency. Cihik et al. used first 3D CNN-based U-Net for the segmentation of sparsely sequential volumetric images and then a 3D model on a large scale [28]. Dolz et al. proposed a 3D CNN based on an FCN model for the segmentation of brain MRI images [29]. They introduced small kernels in DNNs to reduce computational complexity and memory cost in 3D CNN models. Andermatt et al. [30] introduced a 3D recurrent neural network for the segmentation of gray and white matters in a brain MRI dataset.

Bui et al. introduced a 3D dense CNN for the segmentation of a brain volumetric dataset [31]. A densely connected 3D model captured multiscale contextual information and achieved fast convergence with an effective discrimination capability [32]. Oktay et al. introduced a 3D attention U-Net model for medical image segmentation. This novel technique captured target structures of different shapes and sizes [33] from a medical imaging dataset. Nevertheless, 3D CNNs still encounter bottlenecks due to hardware limitation, complex datasets, and memory constraints. In the literature, most deep 3D CNN models use a brain MRI dataset. On the contrary, we covered some recent advancements in abdominal datasets by using 3D CNN models.

Lu et al. presented a 3D CNN model with graph cut technique for the segmentation of liver tumors but tested it on only one dataset to judge the generalized behavior of the model [34]. Few authors [35-38] used the deep learning models for segmentation of other CT organs such as Liver and tumor and they applied FCNs models based on encoding and decoding techniques. The most of DL models that have been used in liver or lesion segmentation are based on 2D slices that are extracted from 3D image volumes. These models do not fully consider the spatial information from 3D volumes.

We tested our proposed model with the existing publicly available SegTHOR 2019 dataset. The main challenge in this dataset is the shape and position of each organ at each slice vary greatly and contrast in CT images is very low. This is a multiclass segmentation problem and the dataset contains 4

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3024277, IEEE Access

organs as risk such as heart, aorta, trachea, esophagus. Recently, the deep learning-based models using SegTHOR 2019 dataset has been used published for segmentation problem [40]. Min Feng et al. [41] presented Dense V-Net deep learning model and achieved optimal performance with some postprocessing steps. They used patch processing approach using Dense V-Net to avoid extra computational burden. Vladimir Kondratenko et al.[42] introduced 2d T-Net approach using organ at risk dataset. Tao He at al. [43] proposed 3D U-Net with multitasking techniques and also applied some auxiliary task for generalization the proposed using SegTHOR 2019 dataset. Sulaiman Vesal at al.[44] presented 2D dilated residual U-net model and produced better results. Miaofei Han at al.[45] presented V-Net model for organ at risk dataset. Pan Chen et al.[46] used cascading approach based on 3D U-Net two models that could produce more computationally complexity during training their proposed models. S. Kim et al. [47] introduced ensemble approach to fuse three-axis information for segmentation of SegTHOR dataset. Li Zhang et al.[48] proposed two-level deep learning model and achieved better performance on organ at risk dataset. The detail of each DL-based method can be found in recently published works [40-48] and is shown in Table. 1 in result section.

III. MATERIALS AND METHODS

A. DATASETS

2) SEGTHOR 2019 DATASET

The SegTHOR 2019 [18] dataset contains 40 cases in training and 20 cases in testing. The images or slices vary from 150 to 284 with a plane input image size of 512×512. The spatial resolution also varies from 0.90 mm to 1.37 mm with a slice thickness between 2 and 3.7 mm. In this experiment, we used 32 CT cases (the number of patients) for training, 8 for validation, and 20 for testing. The problem was a multiclass segmentation one, and each patient in the dataset consisted of four classes (aorta, esophagus, trachea, and heart). All CT scans comprised a high variation in z direction with anisotropic dimension. The dataset is publicly available on all grand challenge websites. The test sample based on SegTHOR 2019 is shown in Fig. 1.

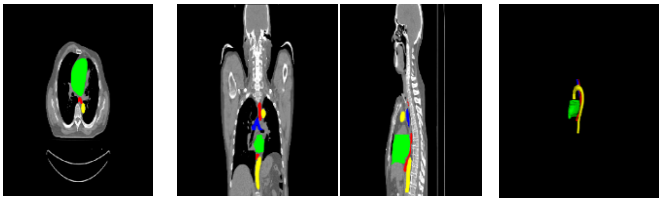

FIGURE 1. Axial, coronal, and sagittal views for the SegTHOR 2019 dataset. The heart (green color), esophagus (red), trachea (blue), and aorta (yellow) are presented.

B. PROPOSED METHOD

Fig. 4 shows the proposed model using a 3D-ASPP module. Each block with detailed explanation is provided in the following subsection.

1) 3D-ASPP LAYER

3D-ASPP comprised various atrous convolutional layers with multiple rates of convolution kernels. It used blocks of different atrous convolutional layers with a spatial pyramid pooling layer. It captured spatial information from input feature maps in multiple scales and shapes for accurate semantic segmentation and classification. However, ASPP provided contextual information by using multiple atrous layers blocks. Atrous convolution produced promising results by capturing the resolution of features with different rates by using deep CNNs. It adjusted the receptive field in such a way to acquire multiscale information from the set of features. Atrous convolution has been applied to input image (x) with some kernel filters w and produce output y, as shown as follows [16]:

$$y[i] = \sum_{k} x(i + r.k)z[k], \tag{1}$$

where r represents the atrous factor that controls the stride for input image. Atrous convolution inserted some zeros, e.g., $(r-1)$ values, among kernel filters and convolved the input x with those filters. The receptive field of filters could be modified using different rate r values.

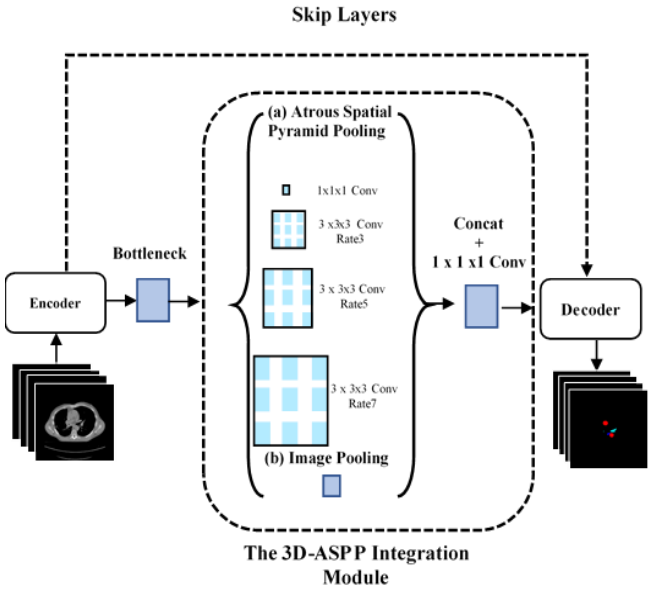

FIGURE 2. Proposed model using 3D encoder–decoder integrated with a 3D-ASPP module.

The proposed 3D-DN was integrated into the proposed 3D-ASPP network to improve the performance of volumetric segmentation. ASPP used a combination of four parallel atrous convolutions by using different atrous rates and one global average pooling layer. The ASPP module consisted of

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3024277, IEEE Access

a parallel 3×3 3D convolutional layer with atrous rates of 3, 5, and 7 and one 1×1 3D convolution. One average pooling layer was used with the same input feature maps as image level features. The feature maps from all branches were bilinearly upsampled. After the concatenation of all layers, 1×1×1 convolution was used. The 3D-ASPP module was applied on the feature maps from the proposed 3D decoder part. After the transition layer, the output feature maps were fed into the 3D decoder part of the network, as shown in Fig. 2.

2) PROPOSED MODEL

Our model was based on a hybrid approach using 3D-ASPP with 3D-DN, as shown in Fig. 3. We incorporated the 3D-ASPP module at the bottoms of the 3D encoder and decoder densNet network to boost the contextual information of the spatial features of the input volume. The decoder part is a regular 3D-DN based on 3D proposed dense blocks, except the bottom layer. The proposed model consisted of an encoder and decoder path with some skip connection and bottom block. Appropriate features were extracted using a 3D dense based convolutional block at each level in the decoder, and their resolution was reduced using appropriate stride with 3D stride convolutional layer. The decoder path was divided into different stages with varying resolutions. Each stage in the decoder path consisted of one to three desne based convolutional layers having volumetric kernel of 3×3×3.The input data resolution was halved by passing every stage by using a convolutional layer with 2×2×2 voxel kernels applied with stride 2. The number of features was doubled for each stage in the decoder block of the proposed network. After the convolutional layer, an activation function (PReLU) was applied throughout the network. The dense block was used at each stage of the decoder and encoder path. The densNet function consisted of convolutional and PReLU with batch normalization (BN) layers with repeated blocks. The dense block consisted of a combination of 3D convolutional layer, ReLu, and BN layer. The input for every stage in the dense block passed to the combination of convolution and nonlinearities (ReLu+BN). The information obtained from the last layer was used for the next stage of the dense block.

The size of input signal was reduced through downsampling, and the receptive field was increased in the successive network layer. The proposed 3D-ASPP module was integrated at the bottom of the decoder and encoder path of the proposed 3D-DN.

The module used the feature maps from the last bottom layers from the decoder side and fed feature maps at the de-convolutional layer from the encoder side of the network. In the encoder path, after each stage, the deconvolutional layer increased the input dimension size in the block of convolutional layers. The network extracted and expanded the spatial feature size from a low resolution to gather essential information for 3D volumetric segmentation. The feature maps from the dense block in the decoder side were concatenated with a downsampled densNet block in the encoder side. The 1×1×1 convolutional layer used to compute two feature maps produced output with the same input volume size. The softmax layer produced a segmentation map for foreground and background voxel-wise regions. The proposed model is shown in Fig. 3.

3) LOSS FUNCTION

The Dice coefficient (DC) proposed in [10] was used to compute the loss function. The loss function between predicted and GT segmentation for 3D was optimized using (2). The loss L is shown in (2) and directly evaluates the similarity of two samples.

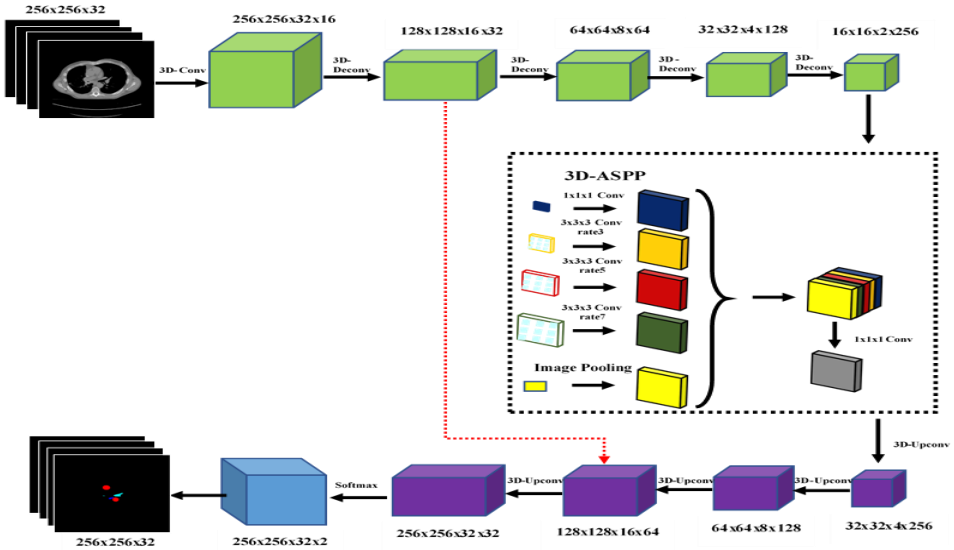

FIGURE 3. Proposed 3D-DN based on an encoder–decoder integrated with a 3D-ASPP module.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3024277, IEEE Access

$$D(p,g)=\frac{2\sum_{i=1}^{N}s_i g_g}{\sum_{i=1}^{N}s_i^2+\sum_{i=1}^{N}g_i^2}, \qquad (2)$$

where s_i and g_i denote the predicted segmentation map and manually provided segmentation map, respectively; and N is the total number of voxels.

4) IMPLEMENTATION DETAIL

The proposed model was built using PyTorch library. All models were trained from scratch. The Adam optimizers used the defined learning rate of 0.0008. The number of epochs was set to 10000 for SegTHOR 2019. The batch size was set to 2. The training of the proposed model was conducted using an NVIDIA GTX 1080 GPU having 12GM memory.

IV. RESULTS

Eighty percent of the dataset was used for training, and the remaining 20% was used for the prediction on the SegTHOR 2019 dataset. The input slices for all cases were downsampled to 256×256 in plane resolution to simplify the computation. In the SegTHOR 2019 dataset, the input resolution size of each volume was set to 256×256. Slices of lesion with five additional slices empty from the start and end of each volume slice were selected. The qualitative and quantitative results are demonstrated in the next section.

A. VISUALIZATION RESULTS BASED ON THE PROPOSED METHOD

Fig. 4 demonstrates the visualization segmentation results of the proposed model. Two different cases have been described in this manuscript. The axial, coronal, sagittal, and 3D views for GT and predicted segmentation visualization are shown in Fig. 4.

Any interpolation in tumor slices to avoid information loss from input volume was not needed. Lesion or masks with a large size were segmented, and some lesions with small organs were hardly segmented using our proposed model.

Some other organs, such as esophagus, trachea, and aorta, have a small size, and a low contrast exists surrounding the heart. Our proposed model could still segment such small organs. In Fig. 4, the first row represents the GT, and the second row represents the predicted values based on the proposed model. Fig. 5 shows the segmentation map for another subject based on the SegTHOR 2019 dataset.

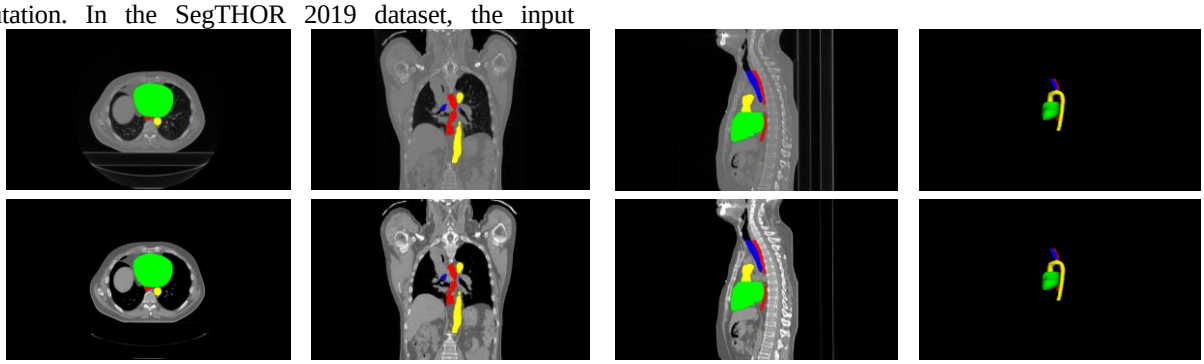

FIGURE 4. The first row shows the GT of axial, coronal, sagittal, and 3D views of the SegTHOR dataset. The second row shows the predicted values of axial, sagittal, coronal, and 3D views by using the proposed model. The heart (green color), esophagus (red), trachea (blue), and aorta (yellow) are presented.

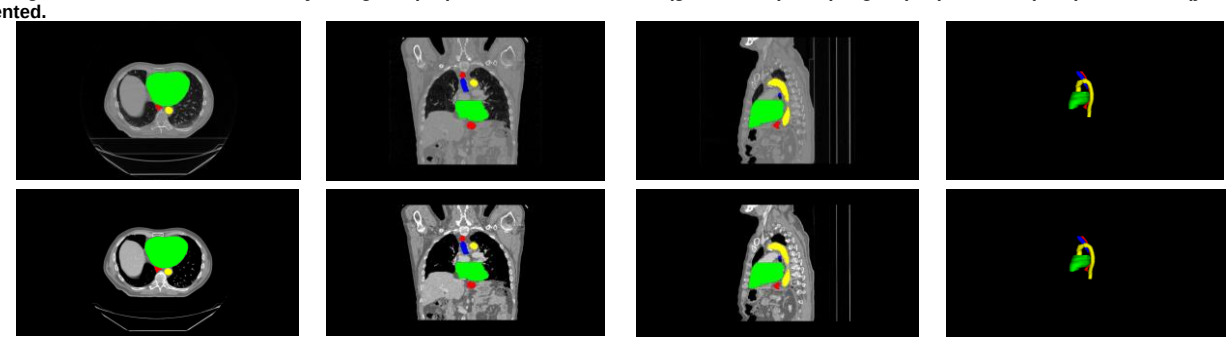

FIGURE 5. The first row represents the manual masks for axial, sagittal, coronal, and 3D views of the SegTHOR dataset. The second row shows the predicted values of axial, coronal, sagittal, and 3D views by using the proposed model. The heart (green color), esophagus (red), trachea (blue), and aorta (yellow) are presented.

B. PERFORMANCE METRICS

Performance metrics [47] were used for the test and evaluation of our proposed model on the datasets. Specifically, the following evaluation metrics were adopted

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3024277, IEEE Access

to test the proposed model. The results were compared with those of existing DL models.

1) SENSITIVITY

Sensitivity was used to compute the positive portion of voxels by using GT and predicted segmentation masks.

$$\text{Specificity} = TPR = TP/(TP+FN) \quad (3)$$

2) SPECIFICITY

Specificity, also called true negative rate (TNR),was used to compute the performance on the basis of GT and predicted segmentation masks.

$$\text{Specificity} = TNR = TN/(TN+FP) \quad (4)$$

3) JACCARD COEFFICIENT (JC)

JC is defined as

$$J(A,B) = |A \cap B| / |A \cup B|, \quad (5)$$

where A is the GT, and B denotes the predicted volume.

4) DICE SIMILARITY COEFFICIENTS (DCS)

DC is commonly used for the validation of medical volume segmentations. It is also called overlap index. It is used to measure the overlap between GT and predicted segmentation masks. For GT and predicted masks, DC is defined as

$$Dice(A,B) = 2|A \cap B| / |A \cup B|. \quad (6)$$

Two types of DCs could be measured. The first is called global Dice that is applied on complete volume for all cases and second is the local dice that was used for individual test sample.

5) VOLUME OVERLAP ERROR (VOE)

VOE is defined as

$$VOE(A,B) = 1 - |A \cap B| / |A \cup B|. \quad (7)$$

6) RELATIVE VOLUME DIFFERENCE (RVD)

RVD is expressed as follows:

$$RVD(A,B) = \left(|B| - |A|\right) / |A|. \quad (8)$$

7) SURFACE DISTANCE METRICS

Surface distance was used to compute the surface between GT and predicted masks. Let $S(A)$ denote the set of surface voxels of A.

$$d(v, S(A)) = \min_{S_A \in S(A)} \|v - S_A\| \quad (9)$$

$$D(A,B) = \frac{1}{|S(A)| + |S(B)|} \left(\sum_{S_A \in S(A)} d(S_A, (S(B)) + \sum_{S_B \in S(B)} d(S_B, (S(A))) \right) \quad (10)$$

8) HAUSDORFF DISTANCE (HD)

Symmetric HD was used to compute the (symmetric) HD between the binary objects in the two segmentation masks. It is defined as the maximum surface distance (MSD) between the objects.

$$MSD(A,B) = \max\left\{ \max_{S_A \in S(A)} d(S_A, S(B)), \max_{S_B \in S(B)} d(S_B, S(A)) \right\} \quad (11)$$

C. COMPARISON WITH STATE-OF-THE-ART METHODS

The segmentation models provided an automatic segmentation map and could be evaluated using different performance metrics.

Various performance metrics, such as accuracy, DCs, JC, MSD, RVD, and average symmetric surface distance (ASSD), were used to compute the performance of the proposed and existing state-of-the art DL models. High values of Dice and three other metrics (JC, sensitivity, and specificity) indicate improved segmentation performance.

The proposed 3D-DN with a 3D-ASPP block was evaluated and compared with existing state-of-the-art methods for SegTHOR 2019 dataset. Our proposed model produced optimal results compared with existing methods. It was avalidated to have robust performance by using the recently published SegTHOR 2019 dataset. We reimplemented 3D simple-VNet and 3D attention U-Net models and compared their results with those of our proposed model and existing state-of-the-art results. Our proposed model outperformed existing segmentation models on SegTHOR 2019. The DCs for 20 test patients based on SegTHOR 2019 were reported. These Dice scores were computed using the proposed, attention U-Net, and simple-VNet models. As shown in Fig. 5, the proposed model produced excellent results compared with existing models. The Dice values by using the proposed and existing models based on 20 test samples from the SegTHOR 2019 dataset are presented in Fig. 5. The proposed model clearly showed good performance compared with existing models. The correlation coefficients based on GT and predicted values by using the proposed and existing 3D models are shown in Fig. 6. These correlation coefficients validated that our proposed model obtained better performance compared with existing models. The results confirmed that the proposed model successfully segmented on the SegTHOR 2019

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3024277, IEEE Access

dataset and could be used for the evaluation of 3D volumetric biomedical images. A comparison of the proposed model with state-of-the-art models by using various performance metrics on the basis of the SegTHOR 2019 dataset is shown in Table 1. The simple 3D-VNet and the 3D-VNet with an attention module were reimplemented, and the results were compared with those of the proposed model. The proposed model produced better DCs and other performance metrics compared with existing 3D models.

Table 1: Performance analysis based on the SegTHOR 2019 dataset by using the proposed model and reimplemented (VNet and Attention U-Net) architectures with other state-of-the-art models.

DL Models	DC(%)	JC	VOE(%)	RVD(%)	ASSD	MSD	Sens	Spec
Proposed	**97.89**	0.5422	0.2920	0.1612	1.0114	1.3462	97.67	96.29
Simple-VNet [51]	93.61	90.98	3.044	3.678	2.198	1.253	91.90	94.40
Attention U-Net [34]	94.35	91.89	2.931	2.071	2.011	2.234	92.38	95.23
Feng et al. [41]	92.34	0.3345	-	-	-	-	-	-
Vladimir et al. [42]	87.46	0.4123	-	-	-	-	-	-
He et al. [43]	91.35	0.3419	-	-	-	-	-	-
Sulaiman et al. [44]	97.49	0.2500	-	-	-	-	-	-
Miaofei et al. [45]	93.45	0.2467	-	-	-	-	-	-
Kim et al. [46]	90.12	0.8823	-	-	-	-	-	-
Chen et al. [47]	91.34	0.4023	-	-	-	-	-	-
Zhang et al. [48]	88.56	0.6324	-	-	-	-	-	-

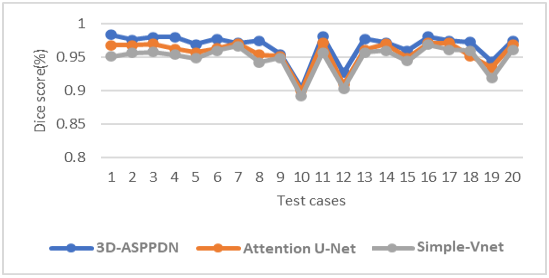

FIGURE 6. DCs based on the SegTHOR 2019 dataset for 20 patients.

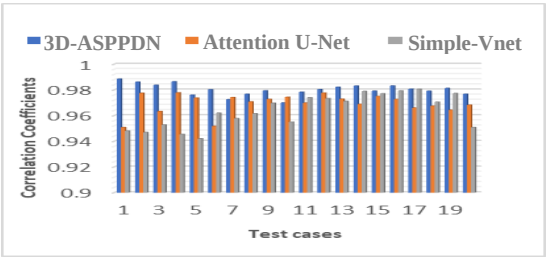

FIGURE 7. Correlation coefficients between prediction and GT for 20 cases. Correlation coefficients by using the proposed and existing 3D models based on the SegTHOR 2019 dataset.

D. DISCUSSION

The considerable improvement in computer hardware and data availability in 3D medical imaging has enabled 3D medical segmentation by utilizing spatial information. Comprehensive information can be produced in any direction on the basis of volumetric images rather than be viewed in a single direction in 2D approaches. A deep model could usually be used to extract highly informative features from complicated organs via segmentation algorithms for volumetric images. The main challenge is the training of these deep networks for 3D models. Most DL architectures are based on FCN, and hybrid-based FCN models are proposed for tumor and liver segmentation tasks. The FCN used a fixed receptive field and could fail in the segmentation of objects with varying sizes. The fixed-size receptive field issue in FCN could be resolved by increasing the field of view.

A sliding window based on complete images by using uniform patches can be used in the FCN model to handle the problem of fixed-size receptive field. The proposed model based on ASPP extracted multiscale information from input images and provided an efficient solution for fixed-size receptive field networks. DNNs encounter some issues during the training of these networks. The first issue is the

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3024277, IEEE Access

overfitting, which would occur during the training of the proposed model due to the less samples of the datasets in comparison with the weights of 3D DL models [30].

A small training dataset could produce an overfitting problem, which could be minimized using the data augmentation and dropout layer. The dropout layer was used in the proposed model to handle the overfitting issue. The considerable training time would be the second issue if we have limited hardware resources. Recently, the computation time for training could be minimized using convolution based on stride [19] that would provide the same effect as that with a pooling layer while achieving faster convergence. BN was also used for rapid convergence of the deep models. Pooling and downsampling techniques could reduce the performance and might loss beneficial information. Gradient vanishing is the third issue, which could happen during the deep network training. We used a desne with carried feature maps from previous layers to handle the vanishing gradient problem during the backpropagation while training the proposed model.

Target organs present a heterogeneous appearance that depends on shape, location, and size from patient to patient [42] and could induce a great challenge in pixel-based image segmentation.

The limited contrast and ambiguous boundary that arise with target organs and surrounding tissues refer to the fourth issue and are usually caused by the attenuation coefficient in CT [23]. Superpixel information and different weight-based techniques with different weighting used in class imbalance could be used to handle such an issue [12].

The aforementioned issues in 3D DL models might lead to decreased performance when handling 3D volumetric datasets because of less data samples and a large number of parameters compared with input 3D data samples and the low variance among voxels with neighboring voxels. The proposed model used multiscale contextual feature information by utilizing an ASPP module and handled the heterogeneous appearance and varying sizes, shapes, and locations of target organs and neighboring tissues.

The main objective of this study is to enhance the accuracy and quality of segmentation based on volumetric 3D CT images. A model based on 3D-ASPP was proposed to improve the segmentation accuracy by capturing multiscale features. It could estimate the pixels at the decoder side of the network with an improved capability to reconstruct small organs with different sizes, shapes, and irregular structures from abdominal CT medical images. The visual and quantitative experimental result reported that the 3D-ASPP-integrated denseNet model achieved better performance in 3D medical segmentation compared with existing segmentation models.

Automatic segmentation based on abdominal CT scans could provide enhanced solutions based on morphometric features. However, the computational complexity would also increase if the number of feature maps and the filter size are large in the encoder–decoder part of the architecture. In consideration of the limited computational resources and GPU memory, we need to design an effective and deep 3D model for improved segmentation performance. The proposed model produced excellent results by using limited memory and computational resources.

V. CONCLUSION

The proposed model in this paper was applied on two different datasets for the segmentation of liver and tumor from CT images and other abdominal CT scans. The proposed approach utilized 3D-ASPP and could acquire multiscale features in a 3D way with 3D-DN based on encoder–decoder to extract discriminative features. Consequently, the proposed model achieved accurate and detailed segmentation results on the basis of CT images. It also achieved excellent performance metrics compared with existing 3D models in the biomedical segmentation field. It could be used for volumetric segmentation in clinical applications to diagnose problems from medical images without the intervention of medical doctors and might be helpful in the medical community for the classification and segmentation of medical images.

In the future, we will explore new models based on the attention mechanism or incorporate different parameters into the proposed model to generalize and train well for another medical dataset in the biomedical domain.

ACKNOWLEDGMENT

This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant No. DF-148-611-1441. The authors, therefore, gratefully acknowledge DSR for technical and financial support.

REFERENCES

[1] Leslie, Scott, Inderbir S. Gill, Andre Luis de Castro Abreu, Syed Rahmanuddin, Karanvir S. Gill, Mike Nguyen, Andre K. Berger et al. "Renal tumor contact surface area: a novel parameter for predicting complexity and outcomes of partial nephrectomy." European urology 66, no. 5 (2014): 884-893.

[2] Yap, Felix Y., Darryl H. Hwang, Steven Y. Cen, Bino A. Varghese, Bhushan Desai, Brian D. Quinn, Megha Nayyar Gupta et al. "Quantitative contour analysis as an image-based discriminator between benign and malignant renal tumors." Urology 114 (2018): 121-127.

[3] Li, Wen, Fucang Jia, and Qingmao Hu. "Automatic segmentation of liver tumor in CT images with deep convolutional neural networks." Journal of Computer and Communications 3, no. 11 (2015): 146.

[4] Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." In European conference on computer vision, pp. 818-833. springer, Cham, 2014.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3024277, IEEE Access

[5] Bernard, Olivier, Johan G. Bosch, Brecht Heyde, Martino Alessandrini, Daniel Barbosa, Sorina Camarasu-Pop, Frederic Cervenansky et al. "Standardized evaluation system for left ventricular segmentation algorithms in 3D echocardiography." IEEE transactions on medical imaging 35, no. 4 (2015): 967-977.

[6] Moradi, Mehdi, Purang Abolmaesumi, D. Robert Siemens, Eric E. Sauerbrei, Alexander H. Boag, and Parvin Mousavi. "Augmenting detection of prostate cancer in transrectal ultrasound images using SVM and RF time series." IEEE Transactions on Biomedical Engineering 56, no. 9 (2008): 2214-2224.

[7] Zettinig, Oliver, Amit Shah, Christoph Hennersperger, Matthias Eiber, Christine Kroll, Hubert Kübler, Tobias Maurer et al. "Multimodal image-guided prostate fusion biopsy based on automatic deformable registration." International journal of computer assisted radiology and surgery 10, no. 12 (2015): 1997-2007.

[8] Porter, Christopher R., and E. David Crawford. "Combining artificial neural networks and transrectal ultrasound in the diagnosis of prostate cancer." ONCOLOGY-WILLISTON PARK THEN HUNTINGTON THE MELVILLE NEW YORK-17, no. 10 (2003): 1395-1418.

[9] Kamnitsas, Konstantinos, Christian Ledig, Virginia FJ Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, Daniel Rueckert, and Ben Glocker. "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation." Medical image analysis 36 (2017): 61-78.

[10] Milletari, Fausto, Seyed-Ahmad Ahmadi, Christine Kroll, Annika Plate, Verena Rozanski, Juliana Maiostre, Johannes Levin et al. "Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound." Computer Vision and Image Understanding 164 (2017): 92-102.

[11] Cha, Kenny H., Lubomir Hadjiiski, Ravi K. Samala, Heang-Ping Chan, Elaine M. Caoili, and Richard H. Cohan. "Urinary bladder segmentation in CT urography using deep-learning convolutional neural network and level sets." Medical physics43, no. 4 (2016): 1882-1896.

[12] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431-3440. 2015.

[13] Zhao, Hengshuang, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. "Pyramid scene parsing network." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881-2890. 2017.

[14] Chen, Liang-Chieh, George Papandreou, Florian Schroff, and Hartwig Adam. "Rethinking atrous convolution for semantic image segmentation." arXiv preprint arXiv:1706.05587 (2017).

[15] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Springer, Cham, 2015.

[16] Chen, Liang-Chieh, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs." IEEE transactions on pattern analysis and machine intelligence 40, no. 4 (2017): 834-848.

[17] Peng, Chao, Xiangyu Zhang, Gang Yu, Guiming Luo, and Jian Sun. "Large Kernel Matters--Improve Semantic Segmentation by Global Convolutional Network." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4353-4361. 2017.

[18] Lambert, Z., C. Petitjean, B. Dubray, and S. Ruan. "SegTHOR: Segmentation of Thoracic Organs at Risk in CT images." arXiv preprint arXiv:1912.05950 (2019).

[19] Liu, Weibo, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E. Alsaadi. "A survey of deep neural network architectures and their applications." Neurocomputing 234 (2017): 11-26.

[20] Litjens, Geert, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I. Sánchez. "A survey on deep learning in medical image analysis." Medical image analysis 42 (2017): 60-88.

[21] Zhang, Wenlu, Rongjian Li, Houtao Deng, Li Wang, Weili Lin, Shuiwang Ji, and Dinggang Shen. "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation." NeuroImage 108 (2015): 214-224.

[22] Pereira, Sérgio, Adriano Pinto, Victor Alves, and Carlos A. Silva. "Brain tumor segmentation using convolutional neural networks in MRI images." IEEE transactions on medical imaging 35, no. 5 (2016): 1240-1251.

[23] Lee, Noah, Andrew F. Laine, and Arno Klein. "Towards a deep learning approach to brain parcellation." In 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 321-324. IEEE, 2011.

[24] Collins, Michael, Robert E. Schapire, and Yoram Singer. "Logistic regression, AdaBoost and Bregman distances." Machine Learning 48, no. 1-3 (2002): 253-285.

[25] Breiman, Leo. "Random forests." Machine learning 45, no. 1 (2001): 5-32.

[26] Furey, Terrence S., Nello Cristianini, Nigel Duffy, David W. Bednarski, Michel Schummer, and David Haussler. "Support vector machine classification and validation of cancer tissue samples using microarray expression data." Bioinformatics 16, no. 10 (2000): 906-914.

[27] Shakeri, Mahsa, Stavros Tsogkas, Enzo Ferrante, Sarah Lippe, Samuel Kadoury, Nikos Paragios, and Iasonas Kokkinos. "Sub-cortical brain structure segmentation using F-CNN's." In 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 269-272. IEEE, 2016.

[28] Çiçek, Özgün, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. "3D U-Net: learning dense volumetric segmentation from sparse annotation." In International conference on medical image computing and computer-assisted intervention, pp. 424-432. Springer, Cham, 2016.

[29] Dolz, Jose, Christian Desrosiers, and Ismail Ben Ayed. "3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study." NeuroImage 170 (2018): 456-470.

[30] Andermatt, Simon, Simon Pezold, and Philippe Cattin. "Multi-dimensional gated recurrent units for the segmentation of biomedical 3D-data." In Deep Learning and Data Labeling for Medical Applications, pp. 142-151. Springer, Cham, 2016.

[31] Bui, Toan Duc, Jitae Shin, and Taesup Moon. "3D densely convolutional networks for volumetric

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3024277, IEEE Access

segmentation." arXiv preprint arXiv:1709.03199 (2017).

[32] Dou, Qi, Hao Chen, Yueming Jin, Lequan Yu, Jing Qin, and Pheng-Ann Heng. "3D deeply supervised network for automatic liver segmentation from CT volumes." In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 149-157. Springer, Cham, 2016.

[33] Oktay, Ozan, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori et al. "Attention U-Net: learning where to look for the pancreas." arXiv preprint arXiv:1804.03999 (2018).

[34] Lu, Fang, Fa Wu, Peijun Hu, Zhiyi Peng, and Dexing Kong. "Automatic 3D liver location and segmentation via convolutional neural network and graph cut." International journal of computer assisted radiology and surgery 12, no. 2 (2017): 171-182.

[35] Christ, Patrick Ferdinand, Florian Ettlinger, Felix Grün, Mohamed Ezzeldin A. Elshaera, Jana Lipkova, Sebastian Schlecht, Freba Ahmaddy et al. "Automatic liver and tumor segmentation of CT and MRI volumes using cascaded fully convolutional neural networks." arXiv preprint arXiv:1702.05970 (2017).

[36] Kaluva, Krishna Chaitanya, Mahendra Khened, Avinash Kori, and Ganapathy Krishnamurthi. "2D-Densely Connected Convolution Neural Networks for automatic Liver and Tumor Segmentation." arXiv preprint arXiv:1802.02182 (2018).

[37] Bi, Lei, Jinman Kim, Ashnil Kumar, and Dagan Feng. "Automatic liver lesion detection using cascaded deep residual networks." arXiv preprint arXiv:1704.02703 (2017).

[38] Li, Xiaomeng, Hao Chen, Xiaojuan Qi, Qi Dou, Chi-Wing Fu, and Pheng-Ann Heng. "H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes." IEEE transactions on medical imaging 37, no. 12 (2018): 2663-2674.

[39] Feng, M., Huang, W., Wang, Y., & Xie, Y. (2019). Multi-organ Segmentation using Simplified Dense V-net with Post-processing. In SegTHOR@ ISBI.

[40] Kondratenko, V., Denisenko, D., Pimkin, A., & Belyaev, M. SEGMENTATION OF THORACIC ORGANS AT RISK IN CT IMAGES USING LOCALIZATION AND ORGAN-SPECIFIC CNN.

[41] He, T., Guo, J., Wang, J., Xu, X., & Yi, Z. (2019). Multi-task Learning for the Segmentation of Thoracic Organs at Risk in CT images. In SegTHOR@ ISBI.

[42] Vesal, S., Ravikumar, N., & Maier, A. (2019). A 2D dilated residual U-Net for multi-organ segmentation in thoracic CT. arXiv preprint arXiv:1905.07710.

[43] Han, M., Yao, G., Zhang, W., Mu, G., Zhan, Y., Zhou, X., & Gao, Y. (2019). Segmentation of CT Thoracic Organs by Multi-resolution VB-nets. In SegTHOR@ ISBI.

[44] Kim, S., Jang, Y., Han, K., Shim, H., & Chang, H. J. (2019, January). A cascaded two-step approach for segmentation of thoracic organs. In CEUR Workshop Proceedings (Vol. 2349). CEUR-WS.

[45] Chen, P., Xu, C., Li, X., Ma, Y., & Sun, F. (2019). Two-stage Network for OAR segmentation. In SegTHOR@ ISBI.

[46] Zhang, L., Wang, L., Huang, Y., & Chen, H. (2019). Segmentation of Thoracic Organs at Risk in CT Images Combining Coarse and Fine Network. In SegTHOR@ ISBI.

[47] Lalande, A., Garreau, M., & Frouin, F. (2015). Evaluation of cardiac structure segmentation in cine magnetic resonance imaging. Multi-Modality Cardiac Imaging: Processing and Analysis, 169-215.

[48] Milletari, Fausto, Nassir Navab, and Seyed-Ahmad Ahmadi. "V-net: Fully convolutional neural networks for volumetric medical image segmentation." In 2016 Fourth International Conference on 3D Vision (3DV), pp. 565-571. IEEE, 2016.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/.