sciendo

RIGA TECHNICAL
UNIVERSITY

Applying 3D U-Net Architecture to the Task of Multi-Organ Segmentation in Computed Tomography

Pavlo Radiuk[*]
Khmelnytskyi National University, Khmelnytskyi, Ukraine

Abstract – **The achievement of high-precision segmentation in medical image analysis has been an active direction of research over the past decade. Significant success in medical imaging tasks has been feasible due to the employment of deep learning methods, including convolutional neural networks (CNNs). Convolutional architectures have been mostly applied to homogeneous medical datasets with separate organs. Nevertheless, the segmentation of volumetric medical images of several organs remains an open question. In this paper, we investigate fully convolutional neural networks (FCNs) and propose a modified 3D U-Net architecture devoted to the processing of computed tomography (CT) volumetric images in the automatic semantic segmentation tasks. To benchmark the architecture, we utilised the differentiable Sørensen-Dice similarity coefficient (SDSC) as a validation metric and optimised it on the training data by minimising the loss function. Our hand-crafted architecture was trained and tested on the manually compiled dataset of CT scans. The improved 3D U-Net architecture achieved the average SDSC score of 84.8 % on testing subset among multiple abdominal organs. We also compared our architecture with recognised state-of-the-art results and demonstrated that 3D U-Net based architectures could achieve competitive performance and efficiency in the multi-organ segmentation task.**

Keywords – **Computed tomography volumetric images, fully convolutional neural networks, medical image analysis, multi-organ segmentation, Sørensen-Dice similarity coefficient.**

I. INTRODUCTION

Automated segmentation of medical images is a complex task due to substantial changes in the shape and size of anatomy among patients. Additionally, low contrast with surrounding tissues can complicate automated segmentation. Recent advances in this area have been mainly related to the use of deep learning methods [1], which allow for the configuration of effective models directly from the visualised data. Notably, the creation and practical implementation of convolutional neural networks (CNNs) [2] and their improvement to fully convolutional neural networks (FCNs) [3] have refined the semantic segmentation of both flat and volumetric images. In this paper, we investigate an application of three-dimensional FCNs (3D FCNs) that can provide multi-organ segmentation of volumetric medical images with high accuracy and computational efficiency.

A. Convolutional Neural Networks

The recent breakthroughs in computer visions relate primarily to the practical use of CNNs on graphics processing devices (GPUs). GPU acceleration has crucially sped up the matrix calculation, allowing for the training of multi-structural models for a shorter period on enormous datasets. CNNs have been commonly utilised in the segmentation and classification tasks owing to their efficient hierarchical feature representation of images based on a data-driven manner [4]. Attributes suitable for classification are extracted from images obtained only from the supervisory signal that determines the desired classification output. This technique, called "supervised learning", has enhanced biomedical and radiological visualisation [5], and significantly refined leading-edge approaches in medical image analysis [6].

Figure 1 illustrates an example of CNN architecture that generates prediction of the input image for a multi-level classification. Over the past few years, CNNs have shown exponential results in medical imaging tasks. Several examples of their prosperous implementation in radiology are the detection of pulmonary embolisms [7], the discovery of gastric cancer in endoscopic images [8], medical image registration [9], classification of human anatomy [10], nucleus segmentation [11], the reduction of false positives by unique morphological features for computer-aided polyp detection [12], and many others.

B. Fully Convolutional Neural Networks

Despite the essential advances of CNNs in computer vision tasks, they contain structural flaws that limit their use. One of the most remarkable drawbacks of these networks is the loss of spatial information of the input image when the convolutional features are transmitted into the final layer. However, spatial information is decisive for segmentation bulk images, and thus the application of raw CNNs to these tasks is insufficient and fruitless. In order to eliminate this limitation of CNNs, Shelhamer et al. [3] proposed FCN, in which the transposed convolutional layers replaced the final fully-connected layers.

[*] Corresponding author's e-mail: radiukpavlo@gmail.com

Transposed layers allow applying the trained sample set to low-resolution activation maps and restore the original spatial information of the input image. In FCN, the various levels of the network are bond by shortcut joints to store the features of an image that is "closer" to the original one. As presented in [13], such transformation allows CNN to retain multidimensionality and achieve a more detailed result in volumetric image segmentation. Fig. 2 depicts the schematic architecture of FCN for the segmentation of computer tomography (CT) image slices.
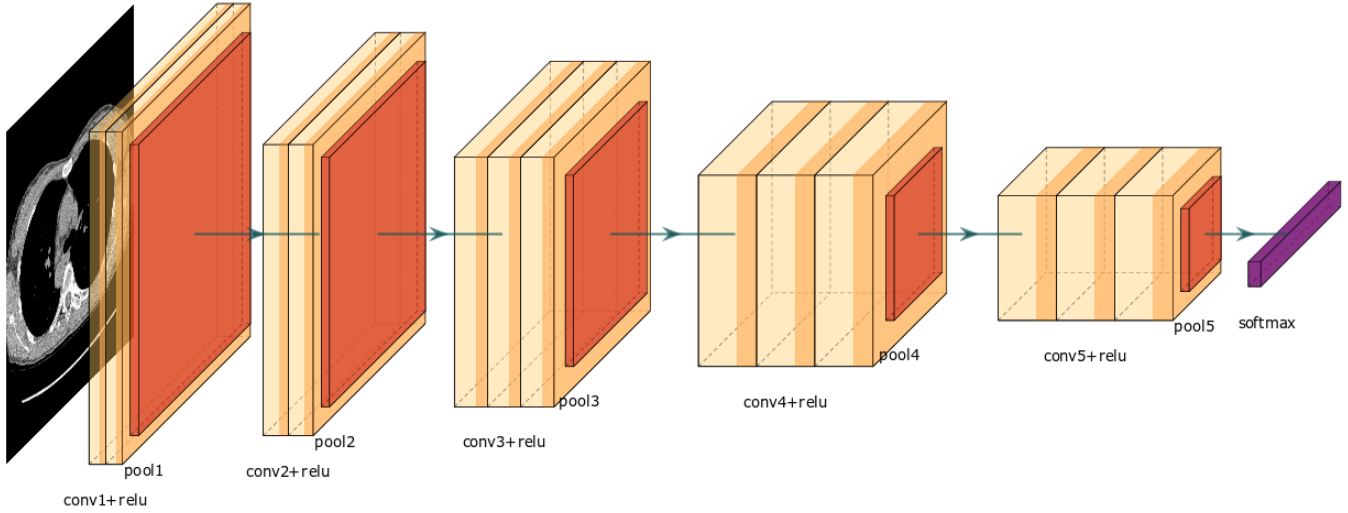

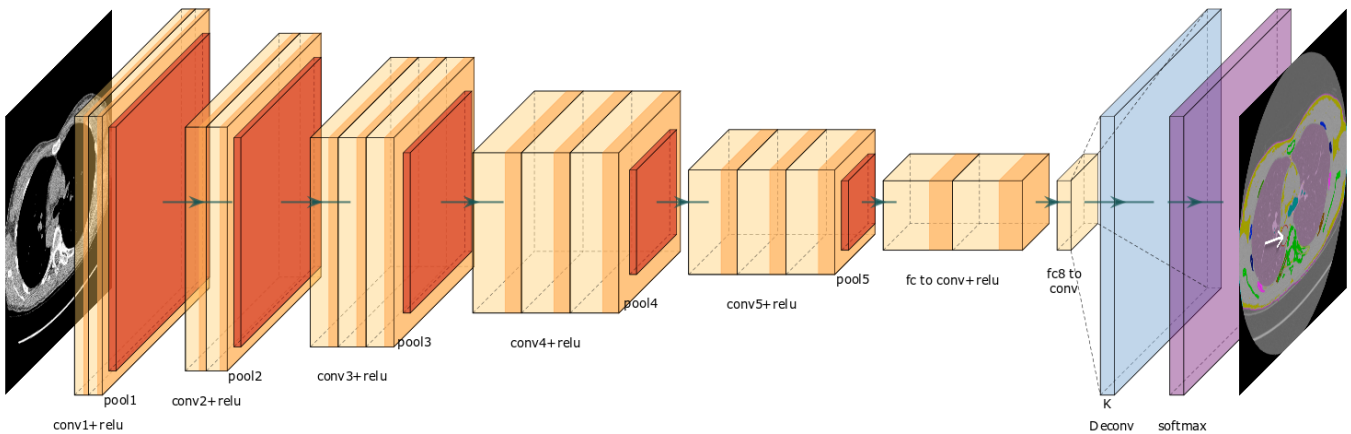Fig. 1. Volumetric scheme of CNN for anatomy segmentation in whole-body CT scans.

Fig. 2. Volumetric scheme of FCN for anatomy segmentation in whole-body CT scans.

II. RELATED WORKS

Semantic segmentation techniques of bulk medical images based on convolutional architectures are classified into two groups: 2D CNN based and 3D CNN based methods. 2D approaches usually work in a slice-by-slice manner. One of the most representative 2D CNN based techniques in medical segmentation is the U-Net architecture [14]. Because of the U-shaped block configuration, U-Net can perform image localisation by predicting the image pixel by pixel. In [15], authors suggested a setting of FCN that provides an end-to-end multi-class classification presenting an anatomical label as a voxel on a CT scan. Havaei *et al*. [16] proposed a two-pathway shallow network with different cascaded 2D architectures for high-grade glioblastomas segmentation in magnetic resonance imaging (MRI) of the human brain. Roth *et al*. [17] improved spatial aggregation by presenting holistically nested networks for pancreas segmentation in CT scans. In [18], authors extended a 2D CNN for volumetric segmentation of raw 3D tomographic images. The core of their network is a $3 \times 3 \times 3$ kernel with ReLU as an activation function.

Various enhancements to 2D CNN-based techniques have significantly improved image segmentation over traditional hand-generated feature-based methods. Nevertheless, 2D CNNs remain ineffective in volumetric image analysis. CNNs do not encode relative spatial long-range feature dependencies and therefore become invariant to massive transformations of the input data [19]. Moreover, plane convolutional layers drop

all the input data considering the pose and the orientation of the object and forward all the information to similar neurons.

To overcome the above-mentioned challenges, researchers proposed various 3D CNN based architectures [20]–[24]. Çiçek *et al.* [20] expanded the 2D U-Net into a three-dimensional version, called 3D U-Net. Their network comprises two stages: an analysis path, also known as an encoder, to allocate volumetric feature dependencies and a synthesis path (decoder) to generate a full-resolution segmentation. Both the analysis and synthesis paths include shortcut connections within convolutional layers of the same size. Milletari *et al.* [21] presented V-Net architecture that utilises residual connections and estimates the training accuracy based on the Dice coefficient function, aimed at addressing the situation of class imbalance. V-Net has demonstrated high performance in segmentation of bulk medical images. In [22], Zhu *et al.* proposed AnatomyNet – 3D U-Net with two extensions: new encoding design to allow auto-segmentation on whole-volume CT and 3D squeeze-and-excitation permanent blocks in encoding layers to improve feature representation.

Multi-modality input and multi-level classification were further implemented to obtain state-of-the-art results in multi-organ classification tasks. Multi-level CNNs provide the high accuracy of image classification and segmentation by using the algorithm that searches the surrogate model for hyperparameter configuration based on different strategy searches. For instance, Chen *et al.* [23] presented VoxResNet – the advancement of the 3D residual network, which stores multi-layer contextual information with multi-modality technique into the original network. In [24], authors took advantage of a cascade manner in 3D FCN and prove that two consecutive FCNNs can provide detail segmentation of abdominal organs and vessels.

III. THE PROBLEM STATEMENT

The main goal of the present research is the investigation of whether fully convolutional neural networks and their modifications can be successfully applied to multi-organ segmentation of bulk images. To achieve the goal, the following tasks have been set:

1) to select a fully convolutional network as an underlying architecture for the multi-class segmentation task;
2) to adjust the selected network for the addressed issue;
3) to prepare an appropriate dataset by compiling CT scans with divorce abdominal organs;
4) to explore the configured neural model by estimating it with an objective metric;
5) to compare the investigated architecture with state-of-the-art methods to prove its usefulness and efficiency in medical image segmentation tasks.

IV. APPLICATION

FCNs allow training models for pixel semantic segmentation in the end-to-end fashion [3]. According to the analysis of the literature provided above, the most promising successor of FCN is 3D U-Net architecture. In this section, we describe the 3D U-Net architecture settings and dataset improvements for further multi-organ segmentation of CT images.
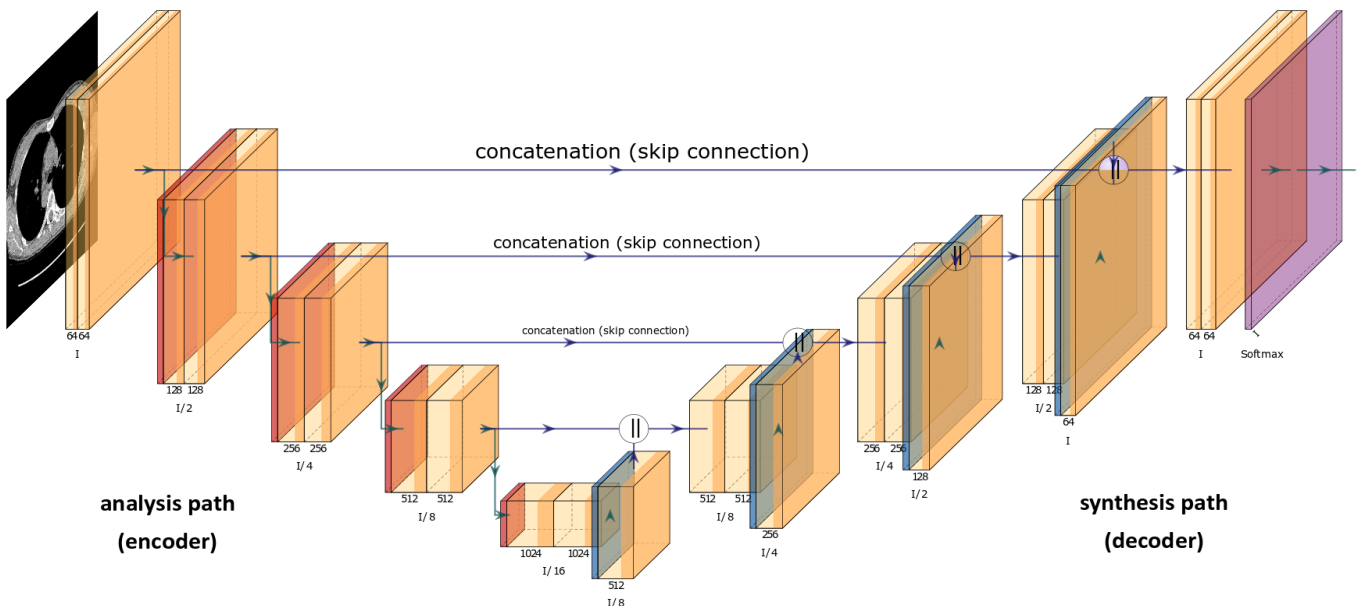

Fig. 3. Volumetric scheme of 3D U-Net, which applies an end-to-end architecture. The kernel of each convolutional layer has a size of $3 \times 3 \times 3$ and is followed by ReLU activations and $2 \times 2 \times 2$ Max Pooling layers.

A. 3D U-Net Configuration

As we have seen, FCNs can achieve promising results in segmentation and classification tasks in a data-driven manner. Let us consider a training set of input images and labels $T = (I_n, L_n)$, where I_n represents CT images of N, and L_n indicates the corresponding ground truth labels of N. Representation T allows the neural network to find a direct mapping from the original image to its segmentation version by

changing the value of numerous parameters. As in [20], our network comprises an encoder and a decoder. The encoder contains four resolution levels with two convolutional layers with $3 \times 3 \times 3$ kernels [25] at each level. ReLU activations [26] and Max Pooling of $2 \times 2 \times 2$ [27] follow convolutional layers with strides of two in each dimension. The decoder contains the transposed $3 \times 3 \times 3$ convolutional layers as the final layers, each of which employs ReLU activations. The final convolutional layer applies the voxel SoftMax activation function to calculate the three-dimensional probability activation map for each target organ as the output of our network. Fig. 3 illustrates the scheme of the proposed architecture.

To provide robust experiments with medical images, we choose the same input and output volume sizes as 3D U-Net and utilise convolution layers with $3 \times 3 \times 3$ kernels, which employ zero-padding throughout. We further apply randomly divided subvolumes selected from a few training CT scans to train the model. The scope of each subvolume is $64 \times 64 \times 64$, which fulfils feasible minibatch training on a single GPU. We choose batch size equal to 256 as it can lead to a better model approximation during the training [28]. Our architecture utilises shortcut connections in the symmetric analysis path as in the original 3D U-Net that leads to approximately 20 M trainable weights.

B. Benchmark Dataset & Data Augmentation

To estimate network training performance, we collected four hundred images with high-resolution of abdominal CT scans from the Cancer Imaging Archive [29]. Images from the dataset were related to patients with different types and stages of cancer. Fig. 4 shows representative instances of the CT scans from the crafted dataset.
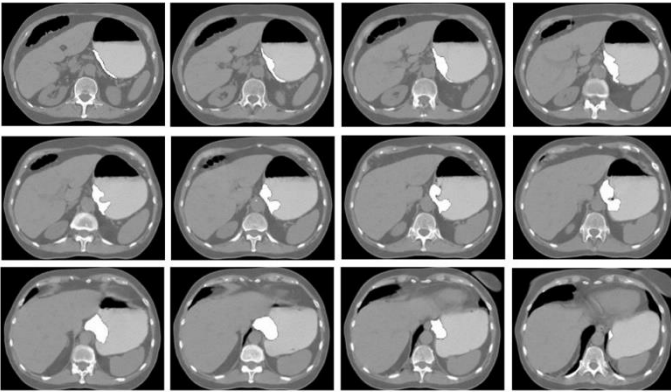

Fig. 4. The variety of grayscale images from the CT scans dataset [29].

Each CT volume has 450–1167 slices of 512×512 pixels. We chose voxel dimensions as [0.57–0.91, 0.58–0.91, 0.48–0.98] mm. All images delineate the arteries, gallbladder, liver, pancreas, spleen, stomach, and vein. Taken the advice from [21], we applied smooth *B*-spline distortions to the images and labelled training data. We randomly sampled the distortion maps from a smooth division with a maximum bias of 4 and an interlinear space of 24 voxels. In order to provide efficient calculation on a single GPU, we also reduced the extension of

all the original images by four times, getting axial image sizes of $128 \times 128 \times$ (number of slices) / 4.

Moreover, to generate decent distortions, we employed random rotations within $-20°$ and $+20°$, and translations of -20 to $+20$ voxels to every dimension at every iteration.

C. Minimisation of the Loss Function

Sørensen-Dice similarity coefficient (SDSC) is a metric that measures the amount of similarity between two binary regions [30]. For this reason, SDSC is commonly utilised to estimate the performance of different segmentation algorithms. The formula for classical SDSC is as follows:

$$SDSC = 2 \cdot \frac{|A \cap B|}{|A| + |B|}, \qquad (1)$$

where *A* represents a set of the ground truth, and *B* corresponds to the computed segmentation. Both sets are binary with values 0 or 1 at each of their voxels.

To extend the use of SDSC to volumetric images, Milletari *et al.* [21] proposed a differentiable version of SDSC, which we applied to train our 3D U-Net. We minimised the loss function in order to optimise the SDSC score on the training data as the following gradient

$$\frac{\partial L}{\partial p_j} = 2 \cdot \frac{g_j \left(\sum_{i=1}^{V} p_i^2 + \sum_{i=1}^{V} g_i^2 \right) - 2 p_j \left(\sum_{i=1}^{V} p_i g_i \right)}{\left(\sum_{i=1}^{V} p_i^2 + \sum_{i=1}^{V} g_i^2 \right)^2}, \qquad (2)$$

calculated according to the *j*-th voxel of the prediction; $p_i \in P$ is a predicted binary segmentation value and $g_i \in G$ is the ground truth binary value at each voxel *i* of a set of voxels *V* in each input image. To anticipate multiple classes for segmentation, our model computes total loss function as

$$L_{total} = \frac{1}{M} \sum_{j}^{M} w_j L_j, \qquad (3)$$

where *M* represents the number of both foreground and background classes and w_j is a weight indicator that can contribute to every label class *j*. As suggested in [31], we maintain $w_j = 0.5 \times 10^{-3}$ for all labels in our application.

V. EXPERIMENTS & RESULTS

A. Model Implementation

We trained and validated our model in Python v3.6, using the TensorFlow v.1.13 backend with Keras library as a wrapper [32]. The training of the model is based on automatic differentiation and eager execution technique. For the training process, we employed Adam optimisation method with an initial learning rate of 10^{-3}, the weight decay of $0.5 \cdot 10^{-3}$, and

momentum of 0.9. The model was trained for 950 epochs on a single NVIDIA GeForce GTX1080 GPU with 8 GB. The source code can be found via [33].

B. Model Estimation

We estimated the model by dividing the dataset into 360 training and 40 testing images. Fig. 5 depicts the training and testing curves of the loss function and the Dice coefficient.

Table I outlines the numerical results of multi-organ segmentation.

Our model achieved an average SDSC score of 0.855 ± 0.064 and 0.847 ± 0.065 by the end of the training and the testing, respectively. Fig. 6 illustrates the multi-organ segmentation of both axial and volumetric renderings.

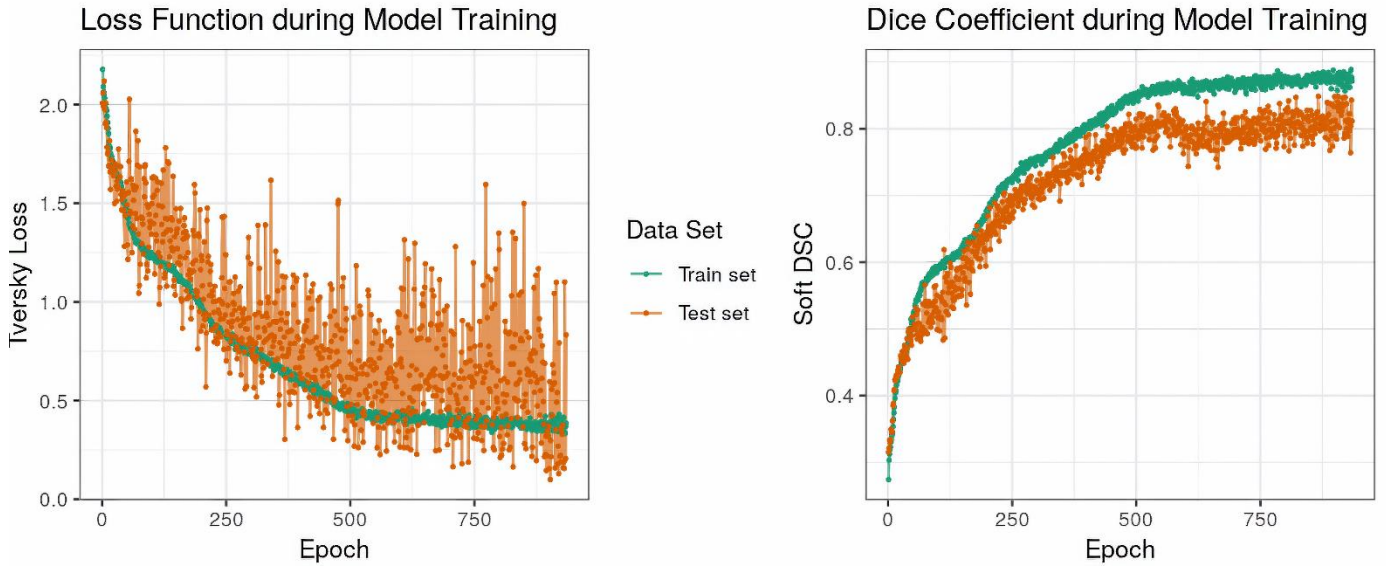Fig. 5. The training and testing accuracy of our 3D U-Net for multi-organ segmentation. Distribution of loss function among the 950 epochs is to the left. The curves of SDSC is to the right.

a) axial sample (ground truth)

b) axial sample (segmentation)

c) 3D rendering (ground truth)
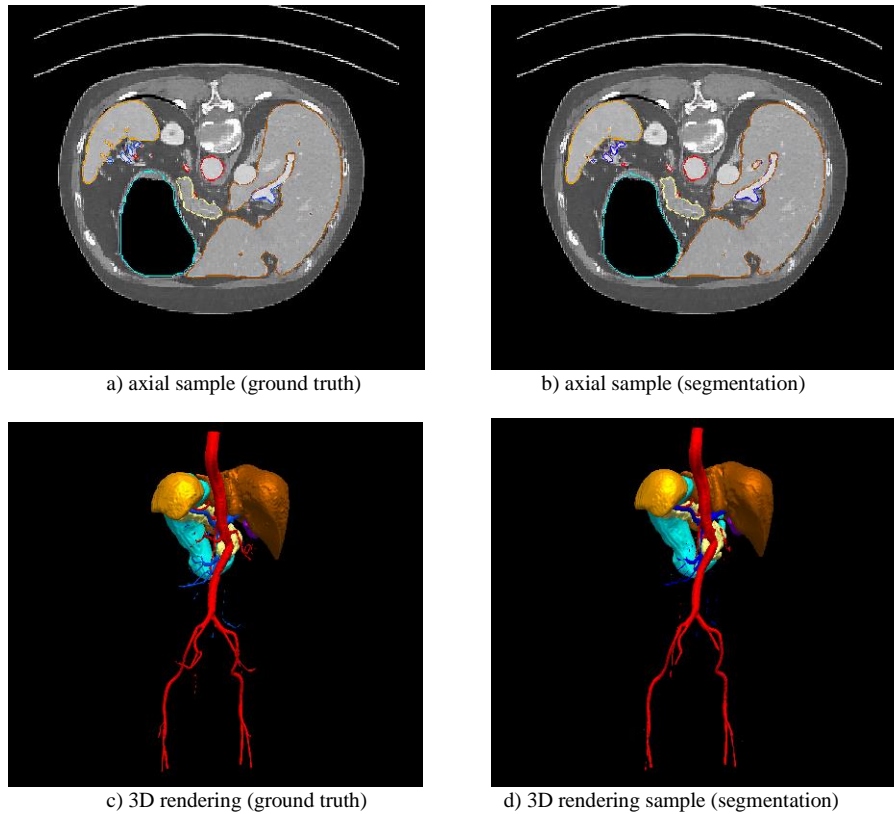
d) 3D rendering sample (segmentation)

Fig. 6. The axial (a, b) and 3D rendering (c, d) results of multi-organ segmentation of scanned CT images.

TABLE I

THE RESULTS OF THE 3D U-NET APPLICATION

SDSC	Training			Testing		
	Min.	Max.	Avg.	Min.	Max.	Avg.
artery	0.647	0.901	0.798	0.638	0.875	0.782
gallbladder	0.001	0.939	0.882	0.297	0.917	0.868
liver	0.883	0.944	0.932	0.891	0.946	0.924
pancreas	0.276	0.896	0.813	0.298	0.904	0.776
spleen	0.769	0.951	0.905	0.908	0.936	0.896
stomach	0.002	0.953	0.898	0.363	0.929	0.884
vein	0.274	0.867	0.745	0.373	0.841	0.736
Total Avg.	**0.397**	**0.918**	**0.856**	**0.631**	**0.913**	**0.848**

TABLE II

COMPARISON WITH STATE-OF-THE-ART METHODS

SDSC	Work	Approach	Avg. SDSC on the validation dataset	Training time, h
artery	[20]	3D U-Net	0.697	2.17
	[21]	2D FCN	0.701	2.32
	[22]	3D U-Net	0.797	4.14
	[24]	3D FCN	**0.801**	**0.97**
	Our approach	3D U-Net	0.782	1.65
gallbladder	[20]	3D U-Net	0.812	2.17
	[21]	2D FCN	0.835	2.32
	[22]	3D U-Net	**0.924**	4.14
	[24]	3D FCN	0.916	**0.97**
	Our approach	3D U-Net	0.868	1.65
liver	[20]	3D U-Net	0.823	2.17
	[21]	2D FCN	0.779	2.32
	[22]	3D U-Net	0.911	4.14
	[24]	3D FCN	0.903	**0.97**
	Our approach	3D U-Net	**0.924**	1.65
pancreas	[20]	3D U-Net	0.698	2.17
	[21]	2D FCN	0.716	2.32
	[22]	3D U-Net	**0.824**	4.14
	[24]	3D FCN	0.807	**0.97**
	Our approach	3D U-Net	0.776	1.65
spleen	[20]	3D U-Net	0.815	2.17
	[21]	2D FCN	0.819	2.32
	[22]	3D U-Net	0.875	4.14
	[24]	3D FCN	0.884	**0.97**
	Our approach	3D U-Net	**0.894**	1.65
stomach	[20]	3D U-Net	0.831	2.17
	[21]	2D FCN	0.827	2.32
	[22]	3D U-Net	0.868	4.14
	[24]	3D FCN	0.845	**0.97**
	Our approach	3D U-Net	**0.881**	1.65
vein	[20]	3D U-Net	0.634	2.17
	[21]	2D FCN	0.62	2.32
	[22]	3D U-Net	**0.872**	4.14
	[24]	3D FCN	0.856	**0.97**
	Our approach	3D U-Net	0.736	1.65

C. Direct Comparison to the State-of-the-Art

Here, we aim to compare our proposed architecture with state-of-the-art papers based on different approaches in multi-organ segmentation. We conducted all experiments on the same dataset compiled from the Cancer Imaging Archive [29] by dividing it into training and validation subsets. The characteristics of the work machine and software remained similar to the described above. The investigated methods were benchmarked on a validation subset by the average SDSC. Table II outlines the list of calculated indicators.

As it is seen in Table II, the investigated architecture outperforms the state-of-the-art methods for several separate organs, such as liver, spleen, and stomach achieving 92.4 %, 89.4 %, and 88.1 % of average SDSC, respectively. Our 3D U-Net architecture shows the efficiency on a single GPU and proves usefulness in research and practical use.

VI. Discussion

In this paper, we investigate a modified fixed-point 3D U-Net model for multi-organ segmentation. At the training phase, we applied a ground truth annotation to create a smaller input area. We applied Sørensen-Dice similarity coefficient as the primary evaluation criterion. The number of 950 epochs were performed at both the training and the testing stages to fit the model. The encoder-decoder structure with batch-normalisation serves as a baseline architecture. In the decoder, the final fully-connected layer was substituted with a transposed layer.

We accumulated a total of 400 images with high-resolution of abdominal CT scans to provide axial and multi-dimensional segmentation of seven human organs. The total dataset was split into training and testing subsets with 360 and 40 images, respectively. Moreover, we applied data augmentation to reduce the impact of overfitting.

The proposed approach of 3D image segmentation does not imply any restrictions on the form of segmented anatomy. This omission can lead to isolated areas at the edges of organs. For example, 3D rendering poorly details images of thin body parts with few voxels, such as arteries and veins (Fig. 6d). In addition, the authors admit that increasing numbers of training epochs might increase the precision variance. In this study, the number of epochs was limited in order to reduce the likelihood of the model overfitting. This issue is to be addressed in further research.

VII. Conclusion

Our hand-crafted architecture was trained and tested on the compiled dataset of CT scans. The investigated architecture achieved the average SDSC score of 84.8 % on testing subset among seven target abdominal organs, which were artery, gallbladder, liver, pancreas, spleen, stomach, and vein. We compared the proposed architecture with several state-of-the-art methods and showed that 3D U-Net based architectures could score high performance in the multi-organ segmentation tasks. Our study proved that a single GPU could achieve competitive results in 3D visualisation and could find practical use in medical image analysis.

In the future study, the authors aim to consider anatomical limitations of CT images to provide topologically correct segmentation results. We intend to utilise more GPU memory, which may allow processing the volume CT scans at a higher resolution. Furthermore, we will collect more CT images of the abdominal cavity for better case representation.

References

[1] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, vol. 19, no. 1, pp. 221–248, Jun. 2017. https://doi.org/10.1146/annurev-bioeng-071516-044442

[2] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, Apr. 2017. https://doi.org/10.1016/j.neucom.2016.12.038

[3] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, Apr. 2017. https://doi.org/10.1109/TPAMI.2016.2572683

[4] H. Suk, S. W. Lee, and D. Shen, "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *NeuroImage*, vol. 101, pp. 569–582, Nov. 2014. https://doi.org/10.1016/j.neuroimage.2014.06.077

[5] A. Hamidinekoo, E. Denton, A. Rampun, K. Honnor, and R. Zwiggelaar, "Deep learning in mammography and breast histology, an overview and future trends," *Medical Image Analysis*, vol. 47, pp. 45–67, Jul. 2018. https://doi.org/10.1016/j.media.2018.03.006

[6] G. Litjens *et al.*, "State-of-the-art deep learning in cardiovascular image analysis," *JACC Cardiovascular Imaging*, vol. 12, no. 8 Part 1, pp. 1549–1565, Aug. 2019. https://doi.org/10.1016/j.jcmg.2019.06.009

[7] J.-Z. Cheng *et al.*, "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans," *Scientific Reports*, vol. 6, no. 24454, Apr. 2016. https://doi.org/10.1038/srep24454

[8] T. Hirasawa *et al.*, "Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images," *Gastric Cancer*, vol. 21, no. 4, pp. 653–660, Jan. 2018. https://doi.org/10.1007/s10120-018-0793-2

[9] Y. Hu *et al.*, "Weakly-supervised convolutional neural networks for multimodal image registration," *Medical Image Analysis*, vol. 49, pp. 1–13, Oct. 2018. https://doi.org/10.1016/J.MEDIA.2018.07.002

[10] H. Takiyama *et al.*, "Automatic anatomical classification of esophagogastroduodenoscopy images using deep convolutional neural networks," *Scientific Reports*, vol. 8, no. 7497, pp. 1–8, May. 2018. https://doi.org/10.1038/s41598-018-25842-6

[11] X. Xie, Y. Li, M. Zhang, and L. Shen, "Robust segmentation of nucleus in histopathology images via mask R-CNN," *Springer*, pp. 428–436, Jan. 2019. https://doi.org/10.1007/978-3-030-11723-8_43

[12] Y. Ren, J. Ma, J. Xiong, Y. Chen, L. Lu, and J. Zhao, "Improved false positive reduction by novel morphological features for computer-aided polyp detection in CT colonography," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 1, pp. 324–333, Jan. 2019. https://doi.org/10.1109/JBHI.2018.2808199

[13] Q. Dou *et al.*, "3D deeply supervised network for automated segmentation of volumetric medical images," *Medical Image Analysis*, vol. 41, pp. 40–54, Oct. 2017. https://doi.org/10.1016/j.media.2017.05.001

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention, MICCAI 2015*. Lecture Notes in Computer Science, *Springer*, Champ, vol 935, pp. 234–241, Nov. 2015. https://doi.org/10.1007/978-3-319-24574-4_28

[15] X. Zhou, T. Ito, and R. Takayama, "Three-dimensional CT image segmentation by combining 2D fully convolutional network with 3D majority voting," in *Deep Learning and Data Labeling for Medical Applications*, *DLMIA 2016*. Lecture Notes in Computer Science, *Springer*, Cham, vol. 10008, pp. 111–120, Sep. 2016. https://doi.org/10.1007/978-3-319-46976-8_12

[16] M. Havaei *et al*., "Brain tumour segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18–31, Jan. 2017. https://doi.org/10.1016/j.media.2016.05.004

[17] H. R. Roth, L. Lu, N. Lay, A. P. Harrison, A. Farag, A. Sohn, and R. M. Summers, "Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localisation and segmentation," *Medical Image Analysis*, vol. 45, pp 94–107, Apr. 2018. https://doi.org/10.1016/j.media.2018.01.006

[18] E. Trivizakis *et al.*, "Extending 2-D convolutional neural networks to 3-D for advancing deep learning cancer classification with application to MRI liver tumor differentiation," *IEEE J. Biomed. Heal. Informatics*, vol. 23, no. 3, pp. 923–930, May 2019. https://doi.org/10.1109/JBHI.2018.2886276

[19] A. Sinha and J. Dolz, "Multi-scale guided attention for medical image segmentation," arXiv:1906.02849 [cs.CV], Jun. 2019.

[20] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-net: Learning dense volumetric segmentation from sparse annotation," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9901 LNCS, pp. 424–432, Oct. 2016. https://doi.org/10.1007/978-3-319-46723-8_49

[21] F. Milletari, N. Navab, and S. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*, pp. 565–571, Dec. 2016. https://doi.org/10.1109/3DV.2016.79

[22] W. Zhu *et al*., "AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy," *The International Journal of Medical Physics and Practice*, vol. 46, no. 2, pp. 576–589, Nov. 2018. http://dx.doi.org/10.1002/mp.13300

[23] H. Chen, Q. Dou, L. Yu, J. Qin, and P.-A. Heng, "VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images," *NeuroImage*, vol. 170, pp. 446–455, Apr. 2018. https://doi.org/10.1016/j.neuroimage.2017.04.041

[24] H. R. Roth *et al*., "An application of cascaded 3D fully convolutional networks for medical image segmentation," *Computerized Medical Imaging Graphics*, vol. 66, pp. 90–99, Jun. 2018. https://doi.org/10.1016/j.compmedimag.2018.03.001

[25] V. V. Romanuke, "An attempt of finding an appropriate number of convolutional layers in CNNs based on benchmarks of heterogeneous datasets," *Electrical, Control and Communication Engineering*, vol. 14, no. 1, pp. 51–57, Jul. 2018. https://doi.org/10.2478/ecce-2018-0006

[26] V. V. Romanuke, "Appropriate number and allocation of ReLUs in convolutional neural networks," *Research Bulletin of the National Technical University of Ukraine "Kyiv Polytechnic Institute"*, no. 1, pp. 69–78, 2017. https://doi.org/10.20535/1810-0546.2017.1.88156

[27] V. V. Romanuke, "Appropriate number of standard 2×2 Max Pooling layers and their allocation in convolutional neural networks for diverse and heterogeneous datasets," *Information Technology and Management Science*, vol. 20, no. 1, pp. 12–19, Jan. 2018. https://doi.org/10.1515/itms-2017-0002

[28] P. M. Radiuk, "Impact of training set batch size on the performance of convolutional neural networks for diverse datasets," *Information Technology and Management Science*, vol. 20, no. 1, pp. 20–24, Jan. 2017. https://doi.org/10.1515/itms-2017-0003

[29] The Cancer Imaging Archive, "TCIA Collections". [Online]. Available: https://www.cancerimagingarchive.net/#collections-list. [Accessed: Feb. 11, 2019].

[30] K. H. Zou, S. K. Warfield, A. Bharatha, C. M. C. Tempany M. R. Kaus, et al., "Statistical validation of image segmentation quality based on a spatial overlap index," *Academic Radiology*, vol. 11, no. 2, pp. 178–189, Feb. 2004. https://doi.org/10.1016/S1076-6332(03)00671-8

[31] Q. Huang, J. Sun, H. Ding, X. Wang, and G. Wang, "Robust liver vessel extraction using 3D U-Net with variant dice loss function," *Computers in Biology and Medicine*, vol. 101, pp. 153–162, Oct. 2018. https://doi.org/10.1016/j.compbiomed.2018.08.018

[32] M. Abadi *et al*., "TensorFlow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)*, pp. 265–283, Nov. 2016. [Online]. Available: https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi

[33] P. Radiuk, "Applying 3D U-Net architecture to the task of multi-organ segmentation in computed tomography," GitHub, Inc., Feb. 2020. [Online]. Available: https://github.com/soolstafir/3D-U-Net-in-CT [Accessed: Mar. 01, 2020].

Pavlo M. Radiuk was born in 1993 in Khmelnytskyi, Ukraine. He graduated from Khmelnytskyi National University in 2017 and received the Master's degree in Mathematical and Computer Modelling. In the same year, Pavlo Radiuk became a Doctoral postgraduate at Khmelnytskyi National University.

Mr Radiuk has been involved in teaching work at Khmelnytskyi National University. His primary study courses are system modelling, data mining, and machine learning. Mr Radiuk has published several scientific articles considering the application and optimisation of neural networks in medical image analysis. His current topics of interest concern statistical analysis, computer vision and numerical optimisation technique.

Address for correspondence: 11, Instytuts'ka str., Khmelnytskyi, 29016, Ukraine.

E-mail: radiukpavlo@gmail.com

ORCID iD: https://orcid.org/0000-0003-3609-112X