UDC 004.932.2; 004.93'11

A COMPARATIVE STUDY OF THREE DEEP LEARNING-BASED METHODS FOR SEGMENTATION OF LUNG LESIONS ASSOCIATED WITH COVID-19 IN 3D CT IMAGES

V.A. Kovalev
*Head of Biomedical Image Analysis department,
UIIP NASB, PhD*

V.A. Liauchuk
Senior Researcher, UIIP NASB, PhD

S.V. Trukhan
*PhD Student, UIIP NASB,
Norwegian University of Life Sciences (NMBU)*

E.V. Snezhko
*Leading Researcher,
UIIP NASB, PhD*

*United Institute of Informatics Problems of the National Academy of Sciences of Belarus (UIIP NASB), Belarus.
Norwegian University of Life Sciences (NMBU), Norway.
E-mail: vassili.kovalev@gmail.com.*

V. A. Kovalev

Dr. Vassili Kovalev has been working in the field of biomedical image analysis for about 30 years. He has extensive experience in 3D MRI image analysis, 3D texture, quantification of normal brain ageing, and related neuroscience problems. Currently, his research interests are focused on the computer-assisted diagnosis of lung diseases and cancer diagnosis based on the whole slide imaging using Deep Learning methods. Administratively, he is a Head of the Biomedical Image Analysis Department of UIIP NASB.

V. A. Liauchuk

Graduated from the Faculty of Physics of Belarusian State University in 2012. Acquired PhD in Computer Sciences in 2019 at UIIP NASB. Current position – senior researcher at the Biomedical Image Analysis department. Research interests include image analysis, deep learning, computer-aided diagnosis.

S. V. Trukhan

Graduated from the Faculty of Applied Mathematics and Computer Science, Belarusian State University in 2014. Acquired Master Degree in Applied Mathematics and Computer Science at Belarussian State University in 2015. Started a co-tutorship Ph. D. program at NMBU and UIIP NASB in 2017. Research interests include infrared, histological, and radiological image analysis using deep learning, image registration.

E. V. Snezhko

Graduated the Faculty of Applied Mathematics and Informatics, Belarusian State University in 2003. Acquired PhD in Computer Sciences in 2007 at UIIP NASB. Leading Researcher at the Mathematics Cybernetics Department of UIIP NASB. Research interests include satellite and aerial image processing and analysis, Deep Learning, medical image analysis, including CT, NMR, X-ray and Ultrasound.

Abstract. In this paper, we present three different methods for segmentation of lung lesions associated with COVID-19 virus disease in 3D CT images. All the methods are based on using state-of-the art deep learning techniques. The proposed methods were tested in the framework of COVID-19 Lung CT Lesion Segmentation Challenge 2020 (COVID-19-20), a MICCAI-associated event which was held on the «Grand Challenge» platform. The best achieved result corresponded to the 6th rank among 98 challenge finalists from all over the world. The results of this study demonstrate the importance of development of robust image analysis solutions applicable to the image data acquired in different hospitals.

Keywords: COVID-19, Computed Tomography, Lesions, Segmentation.

Introduction.

Machine learning and in particular Deep Learning (DL) techniques become more and more popular in the biomedical image analysis domain during the last several years [1, 2]. The Deep Learning technologies gave the community well-grounded promises to become an effective tool in biomedical image analysis and computer-assisted diagnosis of diseases. The reasons for such a success are the ability of Convolutional Neural Networks (CNNs) to identify the very complex dependencies in the image data as well as their colossal capacity as measured by the amount of convolution filters in all the layers. These key characteristics for both scientists and engineers allow them to choose DL-based methods without hesitation in case the data size and their complexity are very large. It should be noted, however, that despite the high efficiency of these approaches the success of DL methods strongly depends on the availability of large training datasets accompanied by appropriate professional annotations. Nevertheless, the use of such techniques as transfer learning and image augmentation allows to use deep CNNs even with the relatively limited image datasets.

The purpose of this paper is to present the image analysis methods suggested by the authors during their participation in COVID-19 Lung CT Lesion Segmentation Challenge 2020 (COVID-19-20) which was held on the «Grand Challenge» platform [3]. The main task of the COVID-19-20 challenge was to automatically detect and segment lung lesions associated with COVID-19 in 3D Computed Tomography (CT) images.

Materials and metrics.

All the data used with this study were provided by the challenge organizers and included unenhanced CT images of patients with positive RT-PCR for SARS-CoV-2 and ground truth annotations of COVID-19 lesions in the lungs [4, 5]. The data were split by the organizers into three sets: Training set (199 CTs), Validation set (50 CTs) and Test set (46 CTs).

The ground-truth segmentation of COVID-19 lesions was performed in a semi-automatic way. At first, the lesions were segmented automatically using an AI-based solution provided by NVIDIA [6]. Then the segmentation results were manually corrected by a board of certified radiologists with use of «itk-SNAP» software. It should be noted that such kind of ground-truth labeling performed in a semi-automated manner can be biased towards the image segmentation methods which share some common characteristics (data preprocessing details, neural network architecture fingerprints) with the algorithms used at the automatic segmentation stage.

Figure 1 shows an example of the original CT image slices along with the ground-truth segmentation overlay.

The ground-truth lesion masks from the Training set were available for the challenge participants and could be used for training the algorithm and preliminary evaluation of the results. The masks for Validation and Test sets were not available to the participants and were used for evaluation of the results by the challenge organizers. Users could submit their results for the Validation set multiple times. After a successful submission, the evaluation results were returned to the user as the values of the calculated evaluation metrics. The key metric used for comparison of the results was the average Dice score. During the Test phase, the participants could submit only once. The submitted results were preliminary ranked using the average Dice score values, and the final ranking was performed with use of statistical analysis methods described in [7].
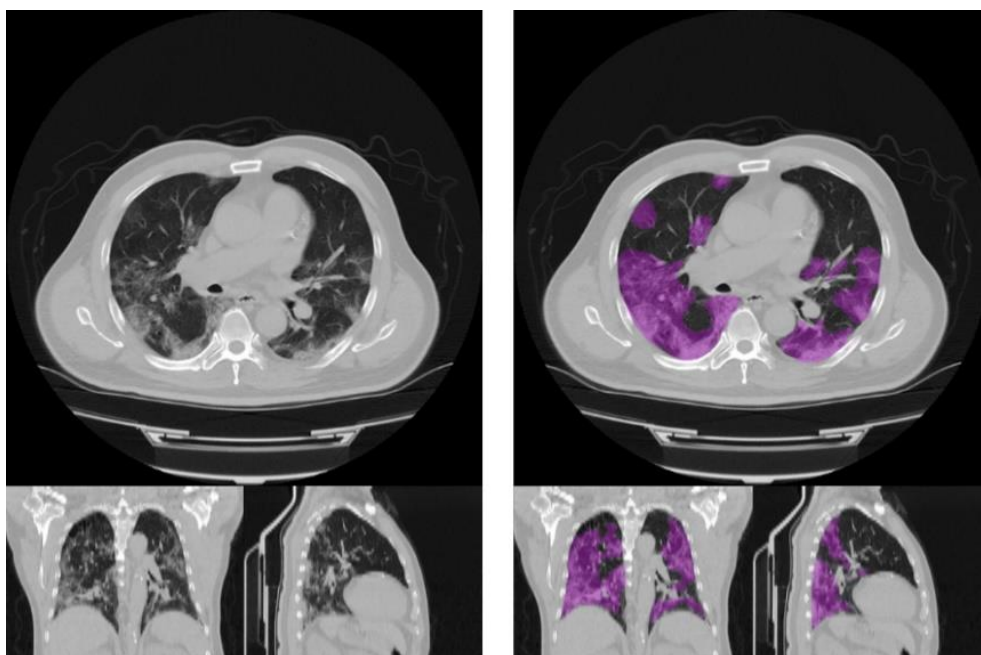
Figure 1. Examples of the original CT image slices (left) and the corresponding ground-truth segmentation colored overlay (right)

Methods.

All the methods proposed by the authors were based on utilizing the deep convolutional neural networks (CNNs) operated in a sliding-window manner. The methods differ in data preprocessing, sampling and the CNN models employed.

The first and the second methods are based on a modified version of the MONAI baseline framework provided by the challenge organizers [8]. MONAI is a freely available, community-supported, PyTorch-based framework for deep learning in healthcare imaging. It provides domain-optimized foundational capabilities for developing healthcare imaging training workflows in a native PyTorch paradigm.

The MONAI baseline framework allows to run both training and inference on the CT image data provided by the challenge organizers. The default training pipeline involves image preprocessing including resizing to a fixed voxel size and intensity scaling as well as on-the-fly image data augmentation including random affine transforms, flips and noising. The default CNN model used is 3D UNet [9]. Default loss function is a sum of binary cross-entropy and Dice loss.

In total, three image segmentation methods were proposed and tested by the authors.

Method-1: VGG16-based 2D-to-3D ImageNet-pretrained CNN.

This method is based on the MONAI baseline framework with certain changes. During the training, the CT images are interpolated to 1.25×1.25×5 mm voxel size. Then, the image samples of 192×192×32 voxels are extracted and inputted into the CNN. In contrast to the default framework, the cross-entropy term in the combined loss function is multiplied by 10 to potentially balance the impact of the two terms. Also, the image data augmentation is enriched by the increase of probabilities of the existing transforms (affine, noising) as well as by adding contrast adjustments, Gaussian smoothing and Gaussian sharpening.

The key difference was in using a custom CNN model (Figure 2). The CNN model used performs slice-wise 2D convolutions at the downsampling stage and then performs both 2D and 3D convolutions at the upsampling stage. The downsampling part of the CNN reproduces the conventional VGG16 model that allows the use of ImageNet-trained weights for initialization. Each downsampling block has the same number of convolutional layers as the corresponding block of the original VGG16 network. The convolution layers with 3×3 kernels are replaced with the corresponding layers with 3×3×1 kernel, numbers of filters in each layer being preserved. All the MaxPooling and Upsampling layers had 2×2×1

kernel. At the upsampling stage, UNet-like skip connections were used. Thus, the utilized CNN model resembles the 3D Anisotropic Hybrid Network described in [10].
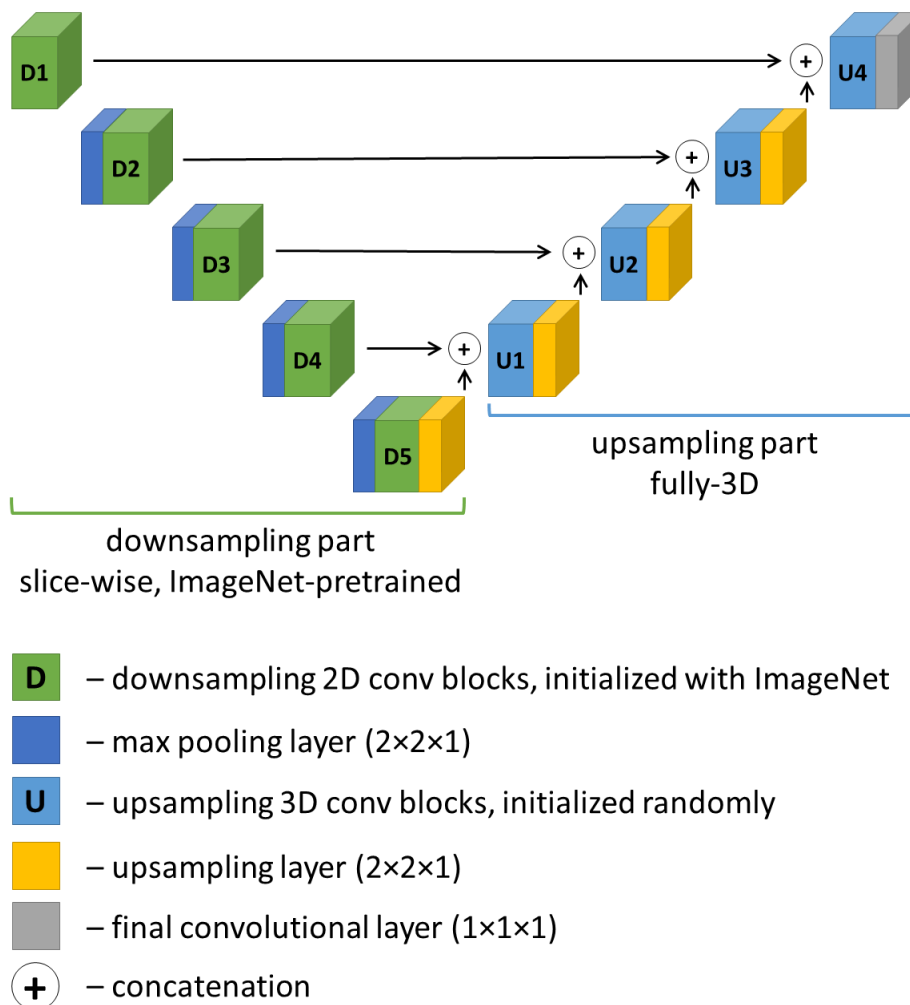

Figure 2. Scheme of the proposed 2D-to-3D ImageNet-pretrained CNN

Method-2. U-Net for 3D anisotropy data.

The method is based on 3D UNet architecture with minor modifications. Input images have a shape of 192×192×16 voxels and a spacing of 1.25×1.25×5 mm. We halve the number of output features of each convolutional layer in order to fit the memory constraints. As suggested in [9], to address the problem of anisotropic voxel size, convolutions and poolings at the finest-resolution levels are kept 2D until the voxel dimensions are approximately isotropic. At all the remaining resolution levels, these operations are applied along all three dimensions. The MONAI baseline framework was used in training and inference.

Method-3. Sliding samples and MobileNetV2-based approach.

The CNN used here consists of encoder and decoder parts. The encoder has MobileNetV2 architecture without fully-connected parts, while the decoder consists of upsampling layers with softmax at the end. A number of encoder and decoder layers are connected in a UNet manner (Figure 3).

The CNN takes as an input 224×224×3 samples of original CT images and produces a 224×224×2 mask (one-hot encoding). During training dataset preparation, we perform sampling with a sliding window, step=60 along X, Y/ step=1 along Z of the original CT images/masks. Size of each sample is 264×264×3 (264×264×1). Initial sampling size of 264=224+40 was chosen to discard border effects during augmentation: for each image/mask pair we apply non-rigid random deformation to these pairs and further crop the deformed samples with border=20 in XY plane. Thus, we get an image/mask pair and a number of their deformed versions of size 224×224×3 (224×224×1). The loss function used was

dice_loss; optimizer Adam, initial lr=0.001 with decreasing factor 0.2, patience period = 2. No pretrained weights were used during the training phase.
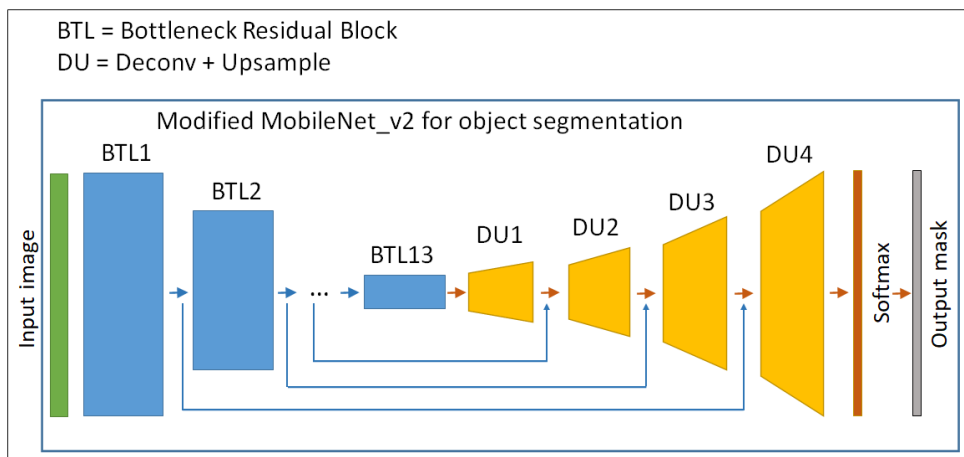

Figure 3. Architecture of the MobileNetV2-based CNN for lesions segmentation

During the inference phase we extract 224×224×3 samples of CT images using sliding window technique, then perform inference of all samples to get the predicted masks (one-hot encoded). Then we restore the 3D mask of lesions for the original CT image using coordinates of the samples stored in advance.

Results

The three methods were examined during both Validation and Test phases of the challenge. The basic metric used for evaluating the results was average Dice score. At the Test phase, some additional metrics were also employed for ranking the results. A total of 225 registered participants submitted their results at the Validation phase and 98 of them also submitted to the Test phase. The COVID-19 lesion segmentation methods described above demonstrated moderate results during the Validation stage. The method-1 and method-2 fell in the top-100 and method-3 was ranked within the top-150 best results. During the Test stage, method-1 got into the top-10 best results with average Dice score exceeding 0.64. Methods 2 and 3 found themselves in the top-50 and top-75 leaderboard positions respectively. The detailed information about the achieved results is listed in Table 1.

Table 1. Results achieved with the proposed methods

Result	Method-1	Method-2	Method-3
Validation phase average Dice score	0.7036	**0.7205**	0.6475
Validation phase rank	86	**68**	138
Test phase average Dice score	**0.6461**	0.6094	0.5608
Test phase non-final rank	**8**	44	73
Test phase rank	**6**	48	78

Figure 4 shows two examples of the automatic segmentation of COVID-19 lesions obtained with use of method-1. The first example (on the left) corresponds to a relatively high agreement between automatic and ground-truth segmentation with a Dice score of 0.88. The second example (on the right) corresponds to a poor agreement with 0.39 Dice score.
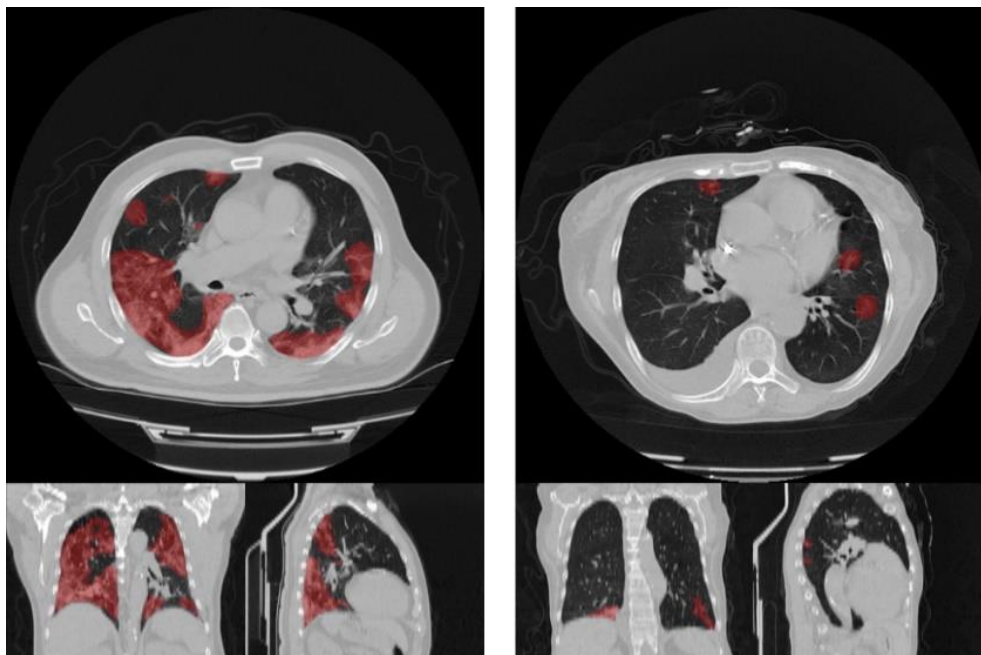
Figure 4. Examples of the automatic segmentation of COVID-19 lesions obtained with use of method-1 with Dice score of 0.88 (left) and 0.39 (right)

Conclusion

Results obtained with this study allow to draw the following conclusions:

1. Deep Learning-based methods can be used for automatic segmentation of COVID-19 lesions in chest CT images with average Dice score not less than 0.64.

2. ImageNet-trained weights can be used for initialization of CNN models for 3D medical image segmentation in order to increase the robustness of the final models.

Obviously low Dice score values obtained during the competition even of those participants at the top of the Leaderboard (the 1st place reached 0.6659 ± 0.2390), as well as significant decrease of average Dice values for Test set in comparison to the Validation set for all the participants, allow to assume the following possible reasons:

1. The Training set consisted of only 199 CT series, including obtained from patients of various age, gender, and stages of the disease, which causes wide variations of lung abnormalities manifestation. Such an amount is likely not enough for generalization using even advanced deep learning techniques.

2. Bias of the Ground Truth manual annotation performed by specialists-radiologists as a result of correction of preliminary automatic annotation by NVidia's CNN during Training/Validation/Test sets preparation.

The issues mentioned above provide support for the high dependency of the results on quality and quantity of the original annotated medical data when using supervised deep learning techniques, especially for such sensitive areas as computer-aided medical diagnostics and therapy monitoring.

The developed lesion segmentation techniques are planned to be used for segmentation of tuberculosis lesions in CT images for NIAID TB Portals.

References

[1] A survey on deep learning in medical image analysis / G. Litjens [et al.] // Medical Image Analysis. – 2017. – Vol. 42. – P. 60-88.

[2] Ker J. Deep Learning Applications in Medical Image Analysis / J. Ker, L. Wang, J. Rao, T. Lim // IEEE Access. – 2018. – Vol. 6. – P. 9375-9389.

[3] COVID-19 Lung CT Lesion Segmentation Challenge - 2020 (COVID-19-20) [Electronic resource]. – Mode of access: https://covid-segmentation.grand-challenge.org/ – Date of access: 25.02.2021.

[4] An P., Xu S., Harmon S., Turkbey E., Sanford T., Amalou A., Kassin M., Varble N., Blain M., Anderson V., Patella F., Carrafiello G., Turkbey B., Wood B. (2020). CT Images in Covid-19 [Data set]. The Cancer Imaging Archive. – Mode of access: https://doi.org/10.7937/tcia.2020.gqry-nc81 – Date of access: 25.02.2021.

[5] The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository / K. Clark [et al.] // Journal of Digital Imaging. – 2013. – Vol. 26(6). – P. 1045-1057.

[6] A pre-trained model for volumetric (3D) segmentation of COVID-19 affected region from CT image) [Electronic resource]. – Mode of access: http://ngc.nvidia.com/catalog/models/nvidia:clara_train_covid19_ct_lesion_seg – Date of access: 25.02.2021

[7] Methods and open-source toolkit for analyzing and visualizing challenge results / M. Wiesenfarth [et al.] // Scientific Reports. – 2021. – Vol. 2369. https://doi.org/10.1038/s41598-021-82017-6.

[8] MONAI Tutorials [Electronic resource]. – Mode of access: https://github.com/Project-MONAI/tutorials – Date of access: 25.02.2021.

[9] U-Net: deep learning for cell counting, detection, and morphometry / T. Falk [et al.] // Nature methods. – 2019. – Vol. 16. – P. 67-70.

[10] 3D anisotropic hybrid network: Transferring convolutional features from 2D images to 3D anisotropic volumes / S. Liu [et al.] // In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham. – 2018. – P. 851-858. https://arxiv.org/pdf/1711.08580.pdf

[11] COVID-19 CT Lesion Segmentation [Electronic resource]. – Mode of access: https://github.com/skliff13/CompetitionsParticipation/tree/master/Covid19Segmentation2020 – Date of access: 25.02.2021.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ТРЕХ МЕТОДОВ СЕГМЕНТАЦИИ НОВООБРАЗОВАНИЙ В ЛЕГКИХ, АССОЦИИРОВАННЫХ С COVID-19, НА ТРЕХМЕРНЫХ КТ-ИЗОБРАЖЕНИЯХ

В.А. КОВАЛЕВ
Заведующий лабораторией анализа биомедицинских изображений, ОИПИ НАН Беларуси, кандидат технических наук

В.А. ЛЕВЧУК
Старший научный сотрудник, ОИПИ НАН Беларуси, кандидат технических наук

С.В. ТРУХАН
Младший научный сотрудник ОИПИ НАН Беларуси, Норвежский университет наук о жизни

Э.В. СНЕЖКО
Ведущий научный сотрудник, ОИПИ НАН Беларуси, кандидат технических наук

Объединенный институт проблем информатики Национальной академии наук Беларуси (ОИПИ НАН Беларуси), Беларусь
Норвежский универсистет наук о жизни (NMBU), Норвегия
E-mail: vassili.kovalev@gmail.com

Аннотация. В данной работе представлены три различных метода сегментации новообразований в легких, ассоциированных с поражением вирусом COVID-19, на трехмерных КТ-изображениях. Все методы основаны на использовании современных методов глубокого обучения. Предложенные методы были протестированы в рамках международного соревнования «COVID-19 Lung CT Lesion Segmentation Challenge 2020 (COVID-19-20)» – ассоциированного с MICCAI мероприятия, которое проводилось на платформе «Grand Challenge». Наилучший достигнутый результат соответствует 6-му месту среди 98 финалистов со всего мира. Результаты данного исследования демонстрируют важность разработки надежных решений для анализа изображений, применимых к медицинским изображениям, полученным в различных учреждениях.

Ключевые слова: COVID-19, компьютерная томография, новообразования, сегментация.