AUTOMATIC SEGMENTATION OF THE BRAIN TUMORS BY RESULTS OF RADIATION DIAGNOSTICS USING A MODIFIED NEURAL NETWORK UNET

[1]OSMAKOV ILYA A., [2]ULYANOV PAVEL, [3]ZUBITSKIY PAVEL S., [4]KOSYRKOVA ALEXANDRA V., [5]BELSKII DENIS B., [6]LETUNOVSKAYA MARINA V.

[1,5,6] "Burtsev Laboratory" LLC
[1,2,3,5] "NeuroSoft Diagnostic" LLC
[1,4,5,6]NMIC of Neurosurgery named after academician N.N. Burdenko
4 – "NeuroHive" LLC
E-mail: ilya.osmakov@gmail.com

Abstract - In this paper described using Convolution Neural Network UNET with specification of boundaries. The modelhastrained for segmentation and detection tasksof glial brain tumors. Mixed datasetconsists from BraTS'2019 and own data has used for training.The article also highlights the data pre-processing and formation of data.The following metrics of the quality of neural networks were obtained: Sørensen–Dice coefficient= 0.708,Sensitivity = 0.937; Specificity = 0.903. Final model realization will be presented in medicine diagnostic software by authors and "NeuroSoft Diagnostic" LLC.

Keywords - Convolutional Neural Networks, UNET, Border Refinement, Dicom Images, Image Segmentation, Glial Tumors, Oncology, Radiation Diagnostics, Brats Dataset, Machine Learning, UNET Network, Image Recognition, Computer Vision, Data Pre-Processing.

I. INTRODUCTION

In our time notice stable growing of cancer incidence rate of central nervous system[1]. One of the main methods in the combined treatment of most malignant brain neoplasms is surgical removal. However, the definition of the boundaries of glial tumors is a nontrivial task by reasons ofits characteristics of growing along myelinated nerve fibers and vessels deep into the healthy white matter of the brain [2, 3]. All this led to incompleteremoval of the tumors and high incidence of postoperative recurrence.

The need for demarcation of glial tumors at the planning and during the operation stages is due to the infiltrative nature of tumor growth near to functionally significant areas of the brain. Glial tumors, among which the most malignant is glioblastoma multiforme, doesn't form capsules. All thisled to the need to the tumor demarcationnot only intraoperative but in DICOM images[4,5].

Currently, in Russia no one use tools for automatic segmentation of pathologies outside of scientificpurposes. Nevertheless, over the past decade, more and more research has been observed in the application of machine learning methods and computer vision to various types of medical data for diagnosing diseases [6, 7, 8, 9]. That gives false impression of the existing good conditions for the healthcare big data. However, almost all of these works were done by using open datasets or using a small amount of data extracted independently. This situation may be due to the fact that in medicine and healthcare in general exists the big administrative barrier to the bigmedical data.The reason for this may be the lack of a sufficient number of prepared and labeled objects for training. According to the authors, this situation has arisen due to the fact that thehealthcareinnovations, and in particular digitalization, are implementing with a delay of 5–20 years due to various regulatory agencies (for example, the FDA or the RosZdravNadzor).Moreover, high cost of markup specialists (medical experts) who could process the data retrospectively.

Because of all these reasons, the automation of this task is relevant to improving the quality of radiation diagnostics. The article proposed a method for automatic segmentation of DICOM images with glial tumors of patients diagnosed with an MR scanner.

II. MATERIAL AND METHODS

In solvingpattern recognition problems, at the moment, the highest efficiency is demonstrating by neural networks of various architectures [10]. In this paper, a modified UNET neural network has beenrealizedby using the Python programming language and its frameworks Keras and TensorFlow. The training was conducted on a PC with the Ubuntu operating system, with an Intel Core i7 central processor, 32 GB of RAM and two GeForce GTX 1080 Ti video cards. The training took about 12 hours.

The main approaches to solving the problem:

1. For solutions, a modified convolutional neural network was used from linkhttps://github.com/selimsef/dsb2018_topcoders, which is distinguished by the feature that the boundary contrast between the main masks of the segmented object is used.
2. As a loss-function were used combinedcrossentropy [11]and soft

dicecoefficient [12]. That allowed to minimize the effect of pixel imbalance. In learning process used RMSProp optimizer [13].

3. BraTS'2019 dataset looks like as multimodal scans are available as NIfTI files (.nii.gz) and describe a) native (T1) and b) post-contrast T1-weighted (T1Gd), c) T2-weighted (T2), and d) T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes, and were acquired with different clinical protocols and various scanners from multiple (n=19) institutions. All the imaging datasets have been segmented manually, by one to four raters, following the same annotation protocol, and their annotations were approved by experienced neuro-radiologists. Annotations comprise the GD-enhancing tumor (ET — label 4), the peritumoral edema (ED — label 2), and the necrotic and non-enhancing tumor core (NCR/NET — label 1), as described in the BraTS reference paper, published in IEEE Transactions for Medical Imaging (also see Fig.1). The provided data are distributed after their pre-processing, i.e. co-registered to the same anatomical template, interpolated to the same resolution (1 mm^3) and skull-stripped [14, 15].
4. Our own data was gained retrospective from clinic and did not have single structure.
5. The resulting model was saved as a set of K different weights.

Data Preprocessing

For consistency of objects in dataset, the T1 mode data in the BraTS Dataset is deleted, since it was incomplete in its own dataset. In turn, the data obtained from the clinic resulted in the resolution of BraTS Dataset objects. Then, the resulting dataset was normilized to the maximum intensity value, and then separate slices were taken from each mode in the axial plane and merged into one PNG file, where each mode were assigned a color channel: FLAIR mode - green, T1CE mode - red, T2 mode - Blue. The visualization of the data preprocessing is shown in Figure 1. In this case, the labels in the file with the segmentation were combined into one.

The result was 40,174 independent objects. Thus, it became possible to pose the problem of binary classification for the detection of glial tumors in MRI images.

III. QUALITY METRICS AND RESULTS

There were 4,001 independent objects from individual slices in the test sample. This allows each colored file to be treated as a separate patient, as the neural network considers them separately. Thus, the following performance metrics of the resulting model were obtained:

Sørensen–Dice coefficient

The dice coefficient is calculated by the formula (1) and its parameters visualized on figure 2 and table 1.

Mean value	Standard Deviation	Maximum	Minimum	1st Quantile	2nd Quantile	3rd Quantile
0.8153	0.2931	1	0	0.7523	1	1

Table 1. Math metrics of Dice Coefficient across all test sample

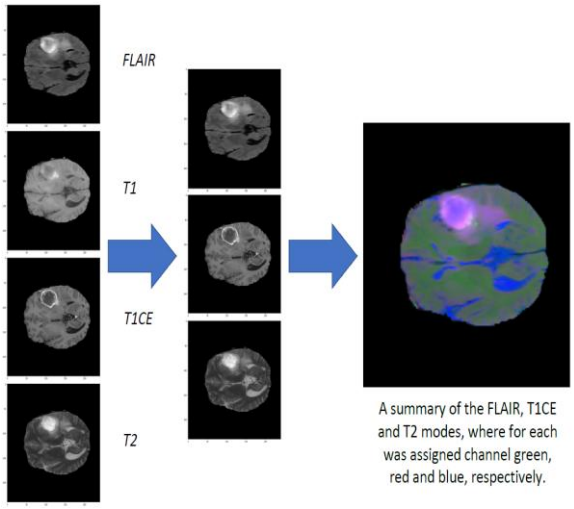


Figure1. Visualization of data preprocessing.

$$Dice\ Coef. = \frac{2 * Right\ Prediction}{Prediction\ + Ground\ Truth} (1)$$

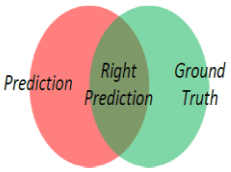


Figure 2. Visualization Dice Coefficient parameters

To study the number typeI and type II errors,presented in the table 2, where the true-positivepredictionsare those objects where the model correctly predicted at least one pixel belonging to the tumor (identifying the confidence level of this metricwill be presented as a separate research); true-negativepredictions, where the model correctly predicted the absence of a tumor, while the value of the Dice coefficient for them was equal to 1; false-negativepredictions, where the model incorrectly predicted the absence of a tumor; false-positivepredictions, where the model predicted a tumor, but it was not there.

		Ground Truth		
		True	False	Numberof Predictions
Model Predictions	True	1484	148	1632
	False	167	2202	2359
	Number of Ground Truth	1651	2350	4001

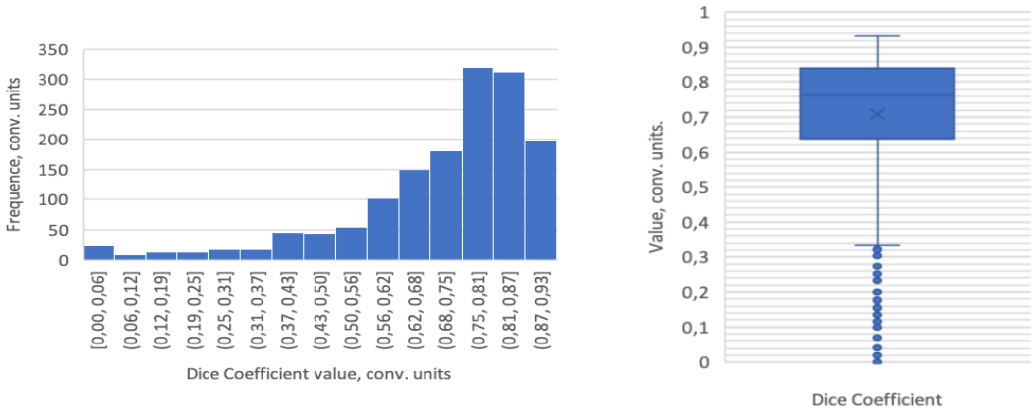
Table2. Number of type I and type II errors

From these results, presented in table 2, we can conclude about the little imbalance of classes. Visualization of this data for the across sample would

be not representative. Thus, it was decided to study the quality metrics of true-positive predictions separately. The distribution of the measured Dice coefficient present on histogram in the figure 3.a and the figure of the box-plot 3.b and at the table 3.

Mean value	Standard Deviation	Maximum	Minimum	1st Quantile	2nd Quantile	3rd Quantile
0.7143	0.1723	0.932	0.0168	0.6413	0.7649	0.8417

Table 3. Math metrics of Dice Coefficient across all test sample



a) Histogram of Dice Coefficient b) Box-plot of Dice Coefficient

Figure3. The visualization of measuring the Dice coefficient

Sensitivity and Specificity

Sensitivity and Specificity are statistical indicators of the effectiveness of a diagnostic test for identifying cases (sick) and non-cases (non-sick) and is calculated by using formulas (2) ,(3) and data from table 2, where TP Rate is the number of true-positive predictions, FN Rate - the number of false-negative predictions, TN Rate - the number of true-negative predictions, FP Rate - the number of false-positive predictions. In our research,theyare equal to 92.95% and 90.93%, respectively.

$$Sensitivity = \frac{TP\ Rate}{TP\ Rate + FN\ Rate} \quad (2)$$

$$Specifity = \frac{TN\ Rate}{TN\ Rate + FP\ Rate} \quad (3)$$

The following is an example ofthe test image segmentation that was used to obtain the quality metrics of the neural network. Figure 4 shows one of the test objects. In Figure 5 shows the segmentation produced by the radiologist. In Figure 6 shows the segmentation predicted by our model with the measured Dice coefficient.

IV. DISCUSSION OF THE RESULTS

By histogram 3.a of the distribution of the Dice coefficient it is possible to evaluate the work of the model in the case of a correct prediction, where you can see a strong shift in the sample relative to level 0.5 conv. units, which is also seen in Figure 3.b. By box-plot graph you can see the inter-quantile range, which has a width of 0.2. conv. units. At the same time, there is a small number of objects with values above 0.92 conv. Units. It is also should noting that the model has never generated a segmentation that absolutely matched with the BraTS dataset segmentation (i.e., the Dice coefficient is 1), which was made by the radiologist, except cases of true-negative predictions. However, there are studies comparing segmentation obtained from BraTS Dataset and other radiologists [16], from which it can be concluded that there is a confusionofreading MRI image data between radiologists.

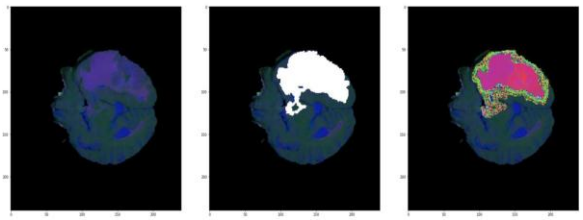


Figure4. RGB mode Figure5. Ground truth mask Figure6. Predicted mask with Dice Coef. = 0.87

The measured sensitivity and specificity metrics interpret the results presented in table 2 more clearly. They may indicate a slightly higher sensitivity than specificity, but this difference is relatively small and equal to 2.02%, which may indicate a conditional “balance” of the diagnostic test based on this neural network.

Figures 4-6 show the quality of the predicted mask, where the Dice coefficient value is 0.87.

V. CONCLUSIONS

Neural networks have shown their effectiveness in solving the problem of segmentation and pathology detection in DICOM images. Even with all the approximations that were made in solving this problem, it was possible to obtain a model with relatively high sensitivity and specificity metrics, which also creates image markup.

However, the fact that there is no absolute matching of the predicted and pseudo-true area may indicate the need for a more maintance assessment of the obtained results when planning the removal of pathology. Undoubtedly, with the expansion of the test train dataset, the improvement of the neural network structure and the elimination of admitted approximations, the quality of the predictions can be increased to a much higher level, which will be reflected in the next works of the co-authors. The final version of the neural network will be presented in the products of NeuroSoft Diagnostic LLC.

Special thankTsoy AlinaиLomanova Daria. This work was supported by the RFBR grant «19-29-01154».

REFERENCE

[1] De Robles P., Fiest K.M., Frolkis A.D., et al. The worldwide incidence and prevalence of primary brain tumors: a systematic review and meta-analysis // Neuro-Oncology. – 2015. – Vol. 17(6). – P. 776-783. doi:10.1093/neuonc/nou283

[2] Claes A., Idema A.J., Wesseling P. Diffuse glioma growth: a guerilla war // Acta Neuropathol. – 2007. – Vol. 114. – P. 443-458. doi:10.1007/s00401-007-0293-7;

[3] Sutter M., Eggspuehler A., Grob D., et al. The validity of multimodal intraoperative monitoring (MIOM) in surgery of 109 spine and spinal cord tumors // Eur Spine J. – 2007. – Vol. 16, Suppl. 2. – P. 197-208.

[4] Zhukov VY, Goryaynov SA, Ogurtsova AA, Ageev IS, Protskiy SV, Pronin IN, Tonoyan AS, Kobyakov GL, Nenashev EA, Smirnov AS, Batalov AI, Potapov AA.

[5] Diffusion tensor imaging tractography and intraoperative neurophysiological monitoring in surgery of intracranial tumors located near the pyramidal tract]. ZhVoprNeirokhirIm N N Burdenko. 2016;80(1):5-18. Russian.

[6] Osmakov I.A., Savelieva T.A., Loschenov V.B., Goryajnov S.A., Potapov A.A. Cluster analysis of the results of intraoperative optical spectroscopic diagnostics In brain glioma neurosurgery. Biomedical Photonics. 2018;7(4):23-34. (In Russ.) https://doi.org/10.24931/2413-9432-2018-7-4-23-34

[7] Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Arxiv.Org. 2015. URL: arxiv.org/abs/1505.04597/.],

[8] Svensson C. M. et al. Automated detection of circulating tumor cells with naive Bayesian classifiers //Cytometry Part A. – 2014. – T. 85. – №. 6. – C. 501-511.

[9] Manak M. S. et al. Live-cell phenotypic-biomarker microfluidic assay for the risk stratification of cancer patients via machine learning //Nature biomedical engineering. – 2018. – T. 2. – №. 10. – C. 761.

[10] De Boer P. T. et al. A tutorial on the cross-entropy method //Annals of operations research. – 2005. – T. 134. – №. 1. – C. 19-67

[11] Bishop C. M. et al. Neural networks for pattern recognition. – Oxford university press, 1995.

[12] Jimenez S., Becerra C., Gelbukh A. Soft cardinality: A parameterized similarity function for text comparison //Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation. – Association for Computational Linguistics, 2012. – C. 449-453.

[13] Tieleman T., Hinton G. Lecture 6.5-rmsprop, coursera: Neural networks for machine learning //University of Toronto, Technical Report. – 2012

[14] Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K and others "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694

[15] Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117

[16] Menze B. H. et al. The multimodal brain tumor image segmentation benchmark (BRATS) //IEEE transactions on medical imaging. – 2014. – T. 34. – №. 10. – C. 1993-2024.

★★★