EasyChair Preprint

№ 2103

# Clinical evaluation of deep learning methods for brain tumor contouring

Alexandra Dalechina, Egor Krivov, Mikhail Belyaev, Valery Kostjuchenko, Boris Shirokikh, Aleksey Shevtsov, Andrey Golanov, Mikhail Galkin, Amayak Durgaryan and Ivan Osinov

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

December 5, 2019

# Clinical evaluation of deep learning methods for brain tumor contouring

Alexandra Dalechina†1, Egor Krivov[2], Mikhail Belyaev[2,3], Valery Kostjuchenko[1], Boris Shirokikh[3], Aleksey Shevtsov[2], Andrey Golanov[4], Mikhail Galkin[4], Amayak Durgaryan[4], Ivan Osinov[1],

[1]*Moscow Gamma Knife Center, 1-st Tverskoy – Yamskoy lane,13/5, 141205, Moscow, RUSSIA.*

[2]*Kharkevich Institute for Information Transmission Problems, Bolshoy Karetny per.,19/1, 127051, Moscow, RUSSIA.*

[3]*Skolkovo Institute of Science and Technology, Bolshoy Boulevard 30, bld. 1, 121205 Moscow, RUSSIA.*

[4]*Burdenko neurosurgery institute, 4th Tverskaya-Yamskaya St., build. 16, 141205, Moscow, RUSSIA.*

## Introduction

The efficacy and safety of radiosurgical treatment are highly dependent on tumor delineation. But there is no consensus of target contouring. As large inter - and intra- observer variability in target delineation exists [1], it could be challenging to evaluate treatment outcomes. Tumor segmentation is a time-consuming process especially in case of multiple lesions. Deep learning methods may significantly reduce time of tumor segmentation and address the challenge with tumor contour standardization. This study aimed to evaluate the quality of the automatically generated contours by a deep convolutional neural network (CNN) and the time reduction using these contours within the radiosurgery treatment planning.

## Materials & Methods

Ten patients who underwent Gamma Knife radiosurgery in the period from November to December 2018 at Moscow Gamma Knife Center were selected from routine clinical practice. The dataset comprised four cases of meningioma, two cases of vestibular schwannoma and four cases of multiple brain metastases. Four experts with experience in segmenting tumors (ranged from 3 to 15 years) were involved in this study. Delineation was performed in Leksell Gamma Plan (version 11.1.0, Elekta AB) by User 1 and User 2 and in iPlan (version 4.5, BrainLab) by User 3 and User 4. For automatic brain tumor segmentation, we used 3D-Unet architecture with residual connections, trained with custom loss function and sampling procedure [2], optimized for metastases segmentation. We used deep_pipe library for computational experiments [3]. A quarter of the training dataset cases was delineated by User 1.

We compared the following four types of contours. (C1), manual contours: the users delineated tumors using standard tools available in the corresponding planning system. (C2), CNN-initialized contours: the users adjusted automatically generated contours in their planning system. (C3), leave-one-out average (LOO-average) contours: to estimate "gold standard" delineation of a tumor for a user, we averaged manual contours of the same tumor created by the other three users. (C4) LOO-CNN average contours: the same as (3), but CNN-initialized contours were averaged instead of manual ones. To quantify differences in contouring, we compared Dice coefficient (DC) between individual contours (C1, C2) and LOO average contours (C3, C4). Comparing C1 to C3 allows us to measure current inter-rater variability for a specific user whereas comparing C2 to C3 we estimate the effect of CNN-initialization on the same user. At the same time, comparing C2 to C4 allows us to (indirectly) measure the level of additional standardization provided by DL. To investigate the differences in Dice scores and time reduction we performed the Sign test and the Wilcoxon test respectively. P-values smaller than 0.05 were assumed to be statistically significant.

## Results

Qualitative and time-saving results are presented in Tables 1-2. We observed better agreement between contours created by the expert and the reference one when the starting point of the segmentation process was automatically generated contours, even if the reference contour was generated completely manually (except for user 1).

**Table 1:** *Quality evaluation in contouring.*

| | Median Dice Coefficient (DC) | | | P values | |
|---|---|---|---|---|---|
| | C1 vs C3 (DC1) | C2 vs C3 (DC2) | C2 vs C4 (DC3) | DC1 vs DC2 | DC1 vs DC3 |
| User 1 | 0.938 | 0.947 | 0.969 | 2.85E-01 | 7.00E-06 |
| User 2 | 0.930 | 0.941 | 0.968 | 7.01E-03 | 7.00E-06 |
| User 3 | 0.915 | 0.920 | 0.934 | 2.29E-03 | 2.26E-03 |
| User 4 | 0.918 | 0.935 | 0.968 | 1.40E-02 | 3.55E-02 |

C1, ..., C4 corresponds to contours types defined above. C1 vs C3 means that we calculated DC between contours C1 and C3. Median DC scores were calculated using all contours of the corresponding user. DC1 vs DC2 means that we evaluated the hypothesis that median difference between DC1 and DC2 is equal to zero.

**Table 2:** *The time reduction in tumor delineation ($p < 0.05$, $r > 0.6$).*

| User | Median manual time (mm:ss) | Range (mm:ss) | Median time reduction (mm:ss) | Range (mm:ss) |
|---|---|---|---|---|
| User 1 | 13:15 | 07:00 - 35:06 | 06:54 | 00:40-17:06 |
| User 2 | 05:30 | 02:17 - 15:20 | 02:16 | 00:48-08:20 |
| User 3 | 12:00 | 03:00 - 44:00 | 09:00 | 01:00 -26:00 |
| User 4 | 06:30 | 03:00 - 23:30 | 05:27 | 03:00-17:35 |

The automatic contours were generated within five seconds on modern GPU. The time required to import the deep learning contours to the treatment planning systems was less than one minute. The total median time needed to delineate a tumor manually was 9.15 min. (ranged from 3.15 min. to 29.18 min).

## Discussion & Conclusions

Results from performed evaluation show improvements in contouring quality. Deep learning methods can be used to reduce the variability in delineation of targets providing speed up in contouring.

## References

[1] Sandström H., et al. Multi-institutional study of the variability in target delineation for six targets commonly treated with radiosurgery// Acta Oncol. 2018. P. 1515-1520.

[2] Krivov E. et al. Tumor Delineation For Brain Radiosurgery by a ConvNet and Non-Uniform Patch Generation // 3rd Int. W. on Patch-based Techniques in Medical Imaging. 2018. 122-129.

[3] https://github.com/neuro-ml/deep_pipe

†Corresponding Author: adalechina@nsi.ru