Eurographics Workshop on Visual Computing for Biology and Medicine (2021)
K. Nieselt and T. Schultz (Editors)

Visual Assessment of Growth Prediction in Brain Structures after Pediatric Radiotherapy (Supplementary Material)

C. Magg[1], L. Toussaint[2], L. P. Muren[2], D. J. Indelicato[3], R. G. Raidou[1]

[1]TU Wien, Austria, [2]Danish Centre for Particle Therapy, Aarhus, Denmark, [3]University of Florida, Jacksonville, USA

1. Dataset

The available dataset consists of 20 folders, corresponding to 20 pediatric patients with previously treated brain tumors. The data include CT pre-treatment scans (referred to as t_0, in the paper) and 1–3 additional pre- and post-treatment MRI data (referred to as $t_1 - t_3$). All the data are registered to the coordinate system of each patient. For the pre-treatment CT scans, segmentation masks of over 100 structures are available. In total, the dataset contains 298 different segmentation labels. However, the label names are not consistent and, therefore, the same structure can occur with multiple labels. This is the reason for the low number of occurrences of label names in Figure 1. Over 190 of the label names are used less than 5 times, whereas only 47 labels occur over 18 times.

For this project, the following brain structures were chosen, based on the work by Toussaint et al. "*Radiation doses to brain substructures associated with cognition in radiotherapy of pediatric brain tumors*", *Acta Oncol. 58*, 10 (2019):

- Brain (Brain)
- Cerebel (Cerebell POST YL)
- Cingulum (Cingulum left and right)
- Corpus callosum (Corpus callosum)
- CTV (CTV, CTV1, CTV2)
- Fornix (Fornix)

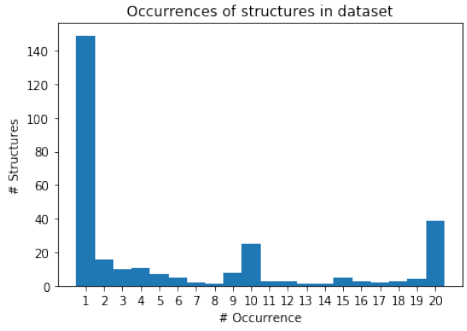

Figure 1: *Overview of the occurrence of structure labels in the provided dataset.*

- GTV (GTV)
- Hypothalamus (Hypothalamus)
- Papez (PapezCircle)
- PTV (PTV1, PTV2)
- Scalp (Scalp)
- TemporalLobe (TemporalLobeLt and TemporalLobeRt)
- Thalamus (Thalamus ant L and R, Thalamus left and right)

2. Segmentation Prediction: Hyperparameter Search

The segmentation prediction is conducted through an active contour model approach. The active contour model is initialized with the dilated version of the snake of the pre-treatment segmented structures to predict the post-treatment segmentations. A hyperparameter search is conducted to obtain adequate parametrizations of the algorithm (dilation and active contour model). An overview of all the possible hyperparameters and their values, as investigated in Section 4 "Segmentation Prediction" of the paper, is given in Table 1. Three parameters—kernel size of dilation k, smoothness of the snake shape β, and maximal number of iterations used to optimize the snake n—influence the result the most, as discussed in Section 4 "Segmentation Prediction" of the paper. The final parameter values for each structure are derived by a majority vote over the data of all patients and are listed in Table 2.

3. Prediction Accuracy: Support Vector Regression (SVR)

To quantify the prediction accuracy, we follow the approach proposed by Kohlberger et al. in *"Evaluating segmentation error without ground truth"*, Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2012, pp. 528-–536. The segmentation of the pre-treatment data (t_0) are predicted using the active contour model and together with the ground truth information, they build the training data for the SVR. At test time, the accuracy of post-treatment segmentations ($t_1 - t_3$) is predicted. The following notation is used for describing the SVR variables:

- I represents the image data, I_i the pixel value at position i.
- S stands for the segmentation mask indices, δS for the segmentation contour indices and $\bar{S}$ for the remaining pixel indices.
- G declares the pixels belonging to the ground truth segmentation mask, δG the corresponding contour indices.

Table 1: *Overview of the parameters involved in the segmentation prediction process. The values marked with * are determined for each structure, by means of a sparse hyperparameter search. For the remaining fixed parameters, the default settings were used. The best structure-specific parameter values are shown in Table 2.*

Parameter	*Description*	*Value*
k	*kernel size of structuring element for dilation*	*structure-specific**
α	*length shape parameter*	0.1
β	*smoothness shape parameter*	*structure-specific**
γ	*explicit stepping parameter*	0.1
w_{line}	*attraction to brightness*	0
w_{edge}	*attraction to edge*	1
c	*convergence criteria*	0.1
m	*max pixel distance to move per iteration*	1
n	*maximal number of iteration steps*	*structure-specific**
B	*boundary conditions for the contour*	*periodic*

- The weights are a Cauchy distribution function $w(I_1,I_2) = \frac{1}{1+\beta(\frac{I_1-I_2}{M})^2}$ with I_1, I_2 being two image intensities, $\beta = 10^4$ and M being the maximum L1 norm.
- The vertex v weight is defined as $w(v) = \frac{1}{D_v}\sum_{i:(v,i)\in E} w(I_v,I_i)$ with the degree of the vertex D_v.
- The out-weight is defined as:
$$w_+(I_i,I_j) = \begin{cases} w(I_1,I_2) & if\ I_1 > I_2 \\ 1 & otherwise \end{cases}$$
- The in-weight is defined as:
$$w_-(I_i,I_j) = \begin{cases} w(I_1,I_2) & if\ I_1 < I_2 \\ 1 & otherwise \end{cases}$$

The 35 independent variables of the SVR are a combination of 2 unweighted geometry features, 4 weighted geometry features, 7 intensity features, 10 gradient features, and 12 selected ratios of those. A complete list is given in Table 3. The 5 dependent variables are error metrics (Table 4). The best dependent variable combination was tested for each brain structure label (21 in total). The results for single- and multi-output are shown in Tables 5 and 6. The tables show the error metrics and C value used to achieve the best SVR score for a particular brain structure. The last column provides the SVR score calculated with the best error metrics over all brain structures used as reference metric. The best value for the SVR score is 1.0 and the score can be negative for arbitrary bad models. For the single-output SVR, the overall best metric is the Jaccard distance with 10 counts (i.e., the results are good for 10/21 labels). For the multi-output SVR, the best combination is the Jaccard distance and Dice Coefficient with 14 counts (i.e., the results are good for 14/21 labels). Table 7 holds the averaged values of the Jaccard distance for each brain structure for the test set of the SVR evaluation. The actual Jaccard distance value is calculated by comparing the ground truth with the segmentation mask generated by the segmentation pipeline. The generated Jaccard distance is the prediction of the SVR based on the independent features. The results are also shown in Figure 2. Although segmentations of larger structures are more accurately predicted, the growth behavior of all structures is learned correctly by the SVR.

Table 2: *Overview of the values for structure-specific parameters (kernel size of dilation k, smoothness of the snake shape β, and maximal number of iterations used to optimize the snake n) involved in the segmentation prediction process. These parameter values were derived by majority vote in a patient-specific evaluation.*

Structure	k	β	n
Brain	4	0.05	4
Cerebell POST YL	4	0.05	4
Cingulum left	10	0.01	10
Cingulum right	10	0.02	10
Corpus callosum	4	0.05	4
CTV	10	0.05	10
CTV1	10	0.05	10
CTV2	10	0.2	10
Fornix	4	0.05	4
GTV	10	0.2	10
Hypothalamus	10	0.05	10
PapezCircle	10	0.05	10
PTV1	10	0.05	10
PTV2	10	0.05	10
Scalp	4	0.05	4
TemporalLobeLt	10	0.05	10
TemporalLobeRt	4	0.05	4
Thalamus ant L	10	0.05	10
Thalamus ant R	10	0.05	10
Thalamus left	10	0.2	10
Thalamus right	10	0.2	10

Table 3: *Overview of all 35 features used as independent variables for the SVR. A combination of unweighted (2) and weighted (4) geometry, intensity (7) and gradients (10) features and ratios of those (12) are used.*

Feature	Description
Unweighted geometry features	
Volume	$\|S\|$
Surface area	$\|\delta S\|$
Weighted geometry features	
Weighted volume	$\sum_{v\in S} w(v)$
Weighted cut	$\sum_{i,j:i\in S,j\in\bar{S}} w(I_i,I_j)$
Low-hi weighted cut	$\sum_{i,j:i\in S,j\in\bar{S}} w_+(I_i,I_j)$
Hi-low weighted cut	$\sum_{i,j:i\in S,j\in\bar{S}} w_-(I_i,I_j)$
Intensity features	
Mean intensity	$\mu_I = \frac{1}{\|S\|}\sum_{v\in S} I_v$
Median intensity	$median(I_v : v \in S)$
Sum of intensities	$\sum_{v\in S}^{I_v}$
Minimum intensity	$min_{v\in S} I_v$
Maximum intensity	$max_{v\in S} I_v$
Interquartile distance of intensities	$(Q_3 - Q_1)/2$
Standard deviation of intensities	$\frac{1}{\|S\|-1}\sum_{v\in S}(\|\|I_v\|\|_1 - \mu_I)$
Gradient features	
Sum of L1 norm	$\sum_{v\in S} \|\|\nabla I_v\|\|_1$
Sum of L2 norm	$\sum_{v\in S} \|\|\nabla I_v\|\|_2$
Mean of L1 norm	$\frac{1}{\|S\|}\sum_{v\in S} \|\|\nabla I_v\|\|_1$
Mean of L2 norm	$\frac{1}{\|S\|}\sum_{v\in S} \|\|\nabla I_v\|\|_2$
Median of L1 norm	$median(\|\|\nabla I_v\|\|_1 : v \in S)$
Minimum of L1 norm	$min_{v\in S}\|\|\nabla I_v\|\|_1$
Maximum of L1 norm	$max_{v\in S}\|\|\nabla I_v\|\|_1$
Interquartile distance of L1 norm	$(Q_3 - Q_1)/2$
Standard deviation of L1 norm	$\frac{1}{\|S\|-1}\sum_{v\in S}(\|\|I_v\|\|_1 - \mu_G)$
Standard deviation of L2 norm	$\frac{1}{\|S\|-1}\sum_{v\in S}(\|\|I_v\|\|_2 - \mu_G)$
Ratio features	
Weighted cut divided by volume	
Unweighted cut divided by volume	
Low-hi weighted cut divided by weighted volume	
Low-hi weighted cut divided by unweighted volume	
Hi-low weighted cut divided by weighted volume	
Hi-low weighted cut divided by unweighted volume	
Weighted cut divided by unweighted cut	
Low-hi weighted cut divided by weighted cut volume	
Low-hi weighted cut divided by unweighted cut volume	
Hi-low weighted cut divided by weighted volume	
Hi-low weighted cut divided by unweighted volume	
Sum of L2 norm divided by sum of L1 norm - “blur index”	

Table 4: *Overview of the five error metrics used as independent variables in the SVR.*

Feature	Description
Jaccard distance (E_J)	$\frac{\lvert S \cap G \rvert}{\lvert S \cup G \rvert}$
Dice coefficient (E_D)	$\frac{2 \cdot \lvert S \cap G \rvert}{\lvert S \rvert + \lvert G \rvert}$
Hausdorff distance (E_H)	$max\{sup_{x \in \delta S}\ inf_{y \in \delta G}\ d(x,y) sup_{x \in \delta G}\ inf_{y \in \delta S}\ d(x,y)\}$
Modified Hausdorff distance (E_{MH})	$mean\{sup_{x \in \delta S}\ inf_{y \in \delta G}\ d(x,y), sup_{x \in \delta G}\ inf_{y \in \delta S}\ d(x,y)\}$
Average surface error (E_S)	$\frac{1}{2}\{\frac{1}{\lvert \delta S \rvert} \sum_{x \in \delta S} min_{y \in \delta G}\ d(x,y) + \frac{1}{\lvert \delta G \rvert} \sum_{x \in \delta G} min_{y \in \delta S}\ d(x,y)$

Table 5: *Best single-output SVR results (**SVR score**) per structure, based on employed error metrics (**best output**) and C value (**best C**), and final used Jaccard distance, as reference (**best ref**). The best value for the **SVR score** and the **best ref** is 1.0, and both scores can be negative for arbitrary bad models.*

Structure	best output	SVR score	best C	best ref
Brain	E_{MH}	0.711	85	0.025
Cerebell POSTYL	E_J	0.72	99	0.72
Cingulum left	E_S	-0.098	9	-17.325
Cingulum right	E_S	-0.019	1	-5.652
Corpus callosum	E_J	0.53	5	0.53
CTV1	E_J	0.748	13	0.748
CTV2	E_D	0.63	14	0.556
CTV	E_J	0.742	13	0.742
Fornix	E_Jt	0.848	99	0.848
GTV	E_D	0.66	6	0.592
Hypothalamus	E_D	0.504	1	0.368
Papez Circle	E_D	0.8	99	0.779
PTV1	E_J	0.898	3	0.898
PTV2	E_J	0.865	21	0.865
Scalp	E_J	0.89	64	0.89
TemporalLobe Lt	E_J	0.925	74	0.925
TemporalLobe Rt	E_J	0.72	89	0.72
Thalamus ant L	E_H	-0.657	1	-18.287
Thalamus ant R	E_H	-0.383	99	-47.152
Thalamus left	E_D	0.862	1	0.783
Thalamus right	E_D	0.836	1	0.817

Table 6: *Best multi-output SVR results (**SVR score**) per structure, based on employed error metrics (**best output**) and C value (**best C**), and final used [Dice coefficient, Jaccard distance], as reference (**best ref**). The best value for the **SVR score** and the **best ref** is 1.0, and both scores can be negative for arbitrary bad models.*

Structure	best combi	SVR score	best C	best ref
Brain	E_{MH}, E_S	0.353	7	-0.321
Cerebell POSTYL	E_J, EM	0.707	99	0.658
Cingulum left	E_H, E_S	-0.509	97	-10.203
Cingulum right	E_H, E_S	-0.982	59	-4.587
Corpus callosum	E_D, E_J	0.522	5	0.522
CTV1	E_D, E_J	0.739	7	0.739
CTV2	E_D, E_J	0.582	91	0.582
CTV	E_D, E_J	0.714	13	0.714
Fornix	E_D, E_J	0.842	33	0.842
GTV	E_D, E_J	0.617	96	0.617
Hypothalamus	E_D, E_J	0.433	2	0.433
Papez Circle	E_D, E_J	0.789	99	0.789
PTV1	E_D, E_J	0.876	3	0.876
PTV2	E_D, E_J	0.861	5	0.861
Scalp	E_D, E_J	0.857	45	0.857
Temporal Lobe Lt	E_D, E_J	0.906	57	0.906
Temporal Lobe Rt	E_J, E_{MH}	0.649	89	0.5
Thalamus ant L	E_H, E_{MH}	-1.162	1	-12.523
Thalamus ant R	E_H, E_S	-1.016	99	-29.226
Thalamus left	E_D, E_J	0.818	16	0.818
Thalamus right	E_D, E_J	0.826	1	0.826

Table 7: *Averaged Jaccard distance for each brain structure for the predicted segmentation mask with the active contour model (ACM) and the SVR prediction.*

Structure	ACM	SVR
Brain	0.931	0.887
Cerebell POSTYL	0.755	0.737
Cingulum left	0.049	0.116
Cingulum right	0.046	0.083
Corpus callosum	0.565	0.554
CTV1	0.381	0.374
CTV2	0.253	0.25
CTV	0.444	0.408
Fornix	0.242	0.249
GTV	0.253	0.251
Hypothalamus	0.107	0.107
Papez Circle	0.238	0.248
PTV1	0.491	0.487
PTV2	0.384	0.375
Scalp	0.656	0.643
Temporal Lobe Lt	0.644	0.633
Temporal Lobe Rt	0.833	0.825
Thalamus ant L	0.034	0.075
Thalamus ant R	0.042	0.106
Thalamus left	0.277	0.252
Thalamus right	0.289	0.267

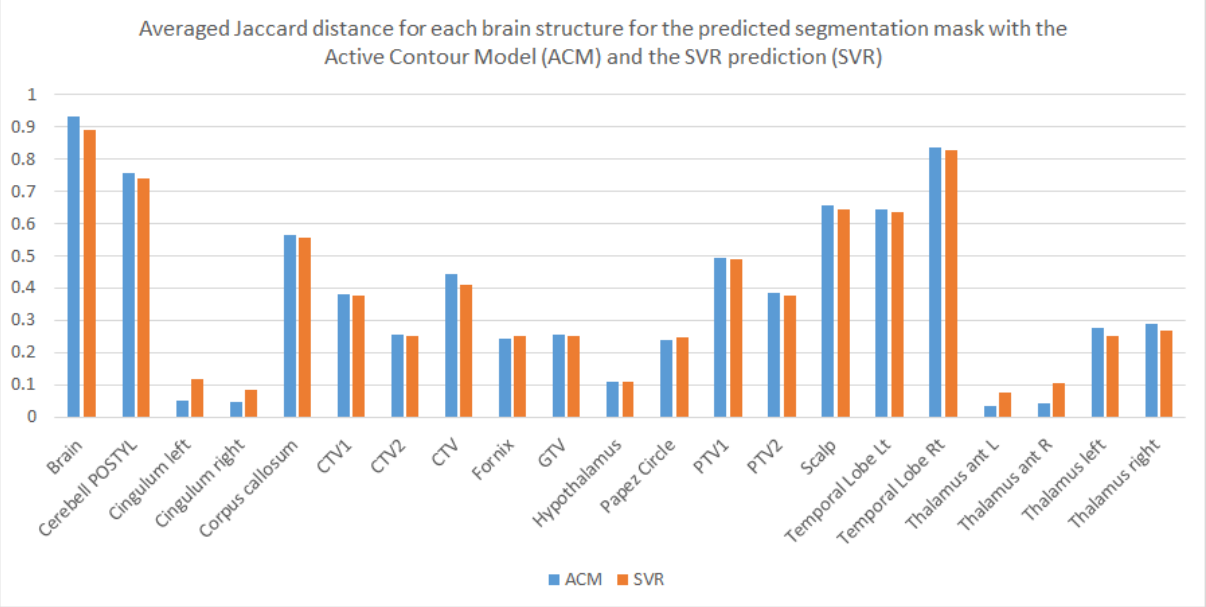

Figure 2: *Averaged Jaccard distance for each brain structure for the predicted segmentation mask with the active contour model (ACM, in blue) and the SVR prediction (in orange).*