ASTRA: Atomic Surface Transformations for Radiotherapy Quality Assurance

Amith Kamath[1], Robert Poel[1,2], Jonas Willmann[3,4], Ekin Ermiş[2], Nicolaus Andratschke[3], and Mauricio Reyes[1]

***Abstract*— Treatment for glioblastoma, an aggressive brain tumour usually relies on radiotherapy. This involves planning how to achieve the desired radiation dose distribution, which is known as treatment planning. Treatment planning is impacted by human errors, inter-expert variability in segmenting (or outlining) the tumor target and organs-at-risk, and differences in segmentation protocols. Erroneous segmentations translate to erroneous dose distributions, and hence sub-optimal clinical outcomes. Reviewing segmentations is time-intensive, significantly reduces the efficiency of radiation oncology teams, and hence restricts timely radiotherapy interventions to limit tumor growth. Moreover, to date, radiation oncologists review and correct segmentations without information on how potential corrections might affect radiation dose distributions, leading to an ineffective and suboptimal segmentation correction workflow. In this paper, we introduce an automated deep-learning based method: atomic surface transformations for radiotherapy quality assurance (ASTRA), that predicts the potential impact of local segmentation variations on radiotherapy dose predictions, thereby serving as an effective dose-aware sensitivity map of segmentation variations. On a dataset of 100 glioblastoma patients, we show how the proposed approach enables assessment and visualization of areas of organs-at-risk being most susceptible to dose changes, providing clinicians with a dose-informed mechanism to review and correct segmentations for radiation therapy planning. These initial results suggest strong potential for employing such methods within a broader automated quality assurance system in the radiotherapy planning workflow. Code to reproduce this is available at `https://github.com/amithjkamath/astra` Clinical Relevance: ASTRA shows promise in indicating what regions of the OARs are more likely to impact the distribution of radiation dose.**

***Index Terms*— Radiotherapy, Treatment Planning, Deep Learning, U-Net, Automated Quality Assurance.**

I. Introduction

Approximately 45% of all malignant primary brain tumors are accounted for by aggressive tumors such as glioblastoma [1]. Treatment consists of surgery, adjuvant radiotherapy (RT), and concomitant and adjuvant chemotherapy [2]. RT planning aims to conform the dose to the target volume (i.e., tumor or resection cavity, with adjacent areas of potential microscopic spread) while sparing organs-at-risk (OAR), thereby ensuring optimal tumor control and limiting normal tissue toxicity [3]. Accurate segmentation of the tumor target volumes and OARs is critical to achieving this objective. Radiation oncologists or dosimetrists draw contours around OAR and tumor target volumes, either manually or semi-automatically. This is time-consuming and can take up to seven hours per patient in the head and neck anatomy [4]. In a multi-institutional delineation study among radiation oncologists, incorrect target volume segmentation has been reported to have caused 25% of non-compliant treatment plans [5]. Tumour target volume and OAR segmentation are amongst the most error-prone and time-consuming steps in the RT planning process. This has led to efforts to create segmentation standards and develop RT Quality Assurance (RTQA) systems [6].

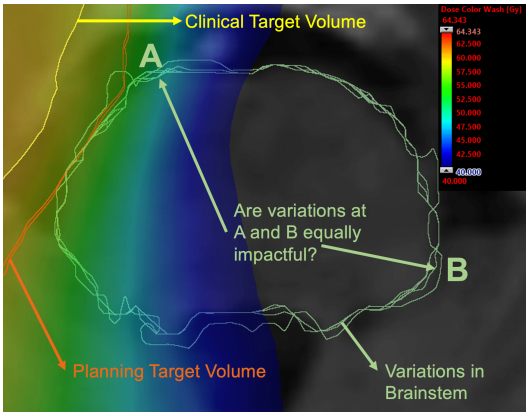

Fig. 1. Visualizing variations in segmentation: are variations at locations A and B equally impactful from a RT perspective? To date, radiation oncologists review and correct segmentations without information on how potential corrections might affect radiation dose distributions, leading to an ineffective and suboptimal segmentation correction workflow.

Motivation for visualizing impact of local variations: Fig. 1 shows multiple cyan lines representing different brainstem segmentations performed by several expert radiation oncologists. The clinical target volume is shown in yellow, and the planning target volume in orange. The underlying heat map indicates the dose distribution (color wash) for a treatment plan in units of Gray (red: high, blue: low dose) and provides the dose context. If OAR volumes are drawn larger than their actual extent, it leads to overestimating the OAR dose and potentially under-dosing the tumor target volume to spare the OAR from excess dose. This negatively

[1]ARTORG Center for Biomedical Engineering Research, University of Bern, Switzerland {amith.kamath, mauricio.reyes}@unibe.ch

[2]Department of Radiation Oncology, University Hospital Bern, University of Bern, Switzerland {robert.poel, ekin.ermis}@insel.ch

[3]Department of Radiation Oncology, University Hospital Zurich, University of Zurich, Switzerland {jonas.willmann, nicolaus.andratschke}@usz.ch

[4]Center for Proton Therapy, Paul Scherrer Institute, ETH Domain, Villigen, Switzerland

impacts tumor control. Conversely, missing the actual extent of the OAR (under-segmentation) would result in an under-estimation of the true dose. This leads to excess toxicity to the OAR.

Treatment dose plan computation is currently done by medical physics experts and is a separate step in the RT process. Treatment delays occur if this dose is not protocol compliant and necessitates re-computation and reviewing any potentially incorrect segmentation. This time between image acquisition and RT planning completion is reported to be 9.63 days on average [7]. It is unknown beforehand if local corrections made at a specific region of the segmentation lead to better or worse dosimetric outcomes. As re-planning and dosimetric assessment are both time-consuming, they demand faster workflows. To address this, deep learning-based dose prediction models were proposed in a head & neck cancer challenge [8]. Recently, in [9], we presented the first such model for glioblastoma. In this paper, we introduce an automated deep-learning-based method called 'atomic surface transformations' for radiotherapy quality assurance (ASTRA) that predicts the potential impact of local segmentation variations on RT dose predictions in the form of a sensitivity map for faster review of dosimetric effects of local segmentation changes. We believe this can help prioritize review efforts on the most critical areas.

Hypothesis and Contributions: Our hypothesis is that a deep learning dose prediction model offering near-instant dosimetry can aid radiation oncologists to focus their review efforts on locations where segmentation variations are dosimetrically most critical.

The proposed approach identifies locations of the OARs that most contribute to variations in dose impact, presenting radiation oncologists with an assessment that highlights the relative significance of local segmentation accuracies.

We demonstrate this using a novel method called ASTRA, which performs voxel-wise local transformations in a high-throughput manner across the OAR's surface, to then compute the mean absolute difference between the dose predicted with and without transformations to estimate the dosimetric impact of local segmentation changes. Our contributions in this paper are threefold:

- First, we introduce and show representative visualizations with ASTRA - modeled as spheres added onto the surface of the OAR segmentation, to evaluate how these local variations impact radiation dose computations.
- Then, we analyze the sensitivity of dose predictions to atomic surface transformations (with over 2000 inference predictions made on 10 test subjects each) and compute correlations with the smallest distance-to-target volume and local dose gradient. We observe stronger correlation between dose changes and distance-to-target (on larger OARs), and a weaker correlation with local dose gradient magnitude (across most OARs).
- Finally, we show how the dose changes are impacted by the size of the transformation - using three different sizes of spheres to simulate segmentation changes of varying magnitudes. We show that the sensitivity map is robust to changes in this parameter.

II. Materials and Methods

Data: Our data set included imaging and segmentation data from 100 patients diagnosed with glioblastoma. This included CT imaging data, and binary segmentation masks of 13 OARs as well as the Planning Target Volume. Each of these subjects also had a reference dose plan, calculated using a standardized clinical protocol with Eclipse (Varian Medical Systems Inc., Palo Alto, USA). This reference was a double arc co-planar volumetric modulated arc therapy (VMAT) plan with 6 mega volt flattening filter free beams, optimized (Varian photon optimizer version 15.6.05) to deliver 30 times 2 Gray while maximally sparing the OARs. Sixty randomly chosen subjects formed the training set, 15 were used as validation (five samples excluded due to missing segmentations) and the rest of the 20 were used as the test set.

Model: We used a two-level C3D U-Net [10] as the dose prediction network. The model input was a normalized CT volume and binary segmentation masks for each of the 13 OARs and target volume, and predicted a continuous-valued dose volume (up-scaled from [0, 1] to 0 to 70 Gray) of the same dimension as the input. The loss was computed as a weighted sum of L1 losses between outputs of the first and second U-Nets versus the reference dose. All volumes were resampled to 128^3 voxels, due to GPU memory constraints. As indicated in [9], this model had a mean absolute error (MAE) of 0.906 Gray, making it suitable for this work.

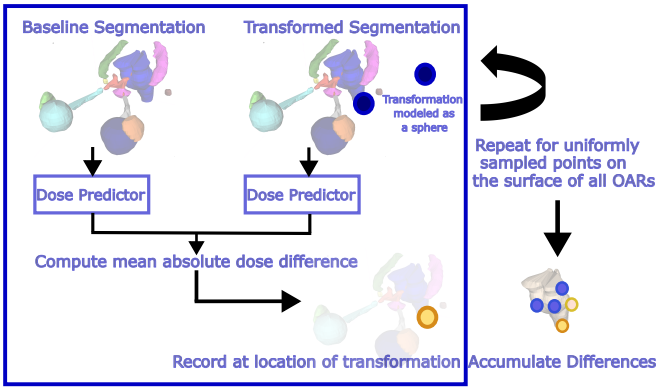

Fig. 2. How are sensitivity maps constructed from 'atomic surface transformations'? We use a Deep Learning based dose prediction inference model sampled uniformly at each perturbed point on the surface of each OAR.

What are Atomic Surface Transformations: Fig. 2 describes the process of generating 'atomic surface transformations' to estimate the relative impact of local variations in segmentation to overall dose predictions. From the baseline segmentation an initial dose prediction is generated. Then, a sphere with a radius of 3 (in voxel units, equivalent to approximately 7.5 mm in the axial plane) is added to a single point on the surface of a single OAR, yielding a perturbed segmentation, on which a new dose prediction is generated. The mean absolute difference between these dose predictions

TABLE I

SENSITIVITY TO SEGMENTATION CHANGES PER OAR MEASURED AS MEAN ABSOLUTE DOSE DIFFERENCE (IN GRAY) IN THE BRAIN, AVERAGED OVER ALL THE TRANSFORMATION POINTS ON ITS SURFACE.

ID	Brainstem	Eye L	Eye R	Hipp. L	Hipp. R
(a)	0.0069	0.0115	0.0092	0.1041	0.0819
(b)	0.0163	0.0485	0.0593	0.3737	0.2710
(c)	0.0097	0.0419	0.0286	0.3085	0.2626
(d)	0.0191	0.0613	0.0514	0.3673	0.3151

is computed and recorded at the center of the corresponding sphere. This is repeated over a uniformly sampled set of points on the OAR's surface, yielding a sensitivity map, as shown in Fig. 3.

Sensitivity experiments: Beyond constructing the sensitivity maps using 'atomic surface transformations', we run two experiments to evaluate the ability of sensitivity maps to describe dose-sensitive areas of segmentation changes. We first look at the correlation between the mean average dose difference due to an atomic surface transformation with the minimum distance of the transformation location to the tumor target volume. We expect that points closer to the tumor target volume will be more highly impacted by dose differences, indicating their relative importance to the segmentation quality. The second experiment is to compute the correlation of dose differences against the local magnitude of the dose gradient. The hypothesis here is that steeper dose gradients indicate regions where dose estimates have to be sensitive to segmentation changes, leading to a positive correlation.

Transformation size experiments: Next, we investigate what happens when the size of the atomic surface transformations change - and how, if at all, would it impact dose predictions. The size in physical units of each voxel is generally in the range of 2.7 mm in the axial plane and 2 mm in the transverse direction. We experiment with transformations of radii $3, 5, 7$ (in voxel units) on the brainstem of two test set subjects.

III. RESULTS

Visualizing Atomic Surface Transformations: Fig. 3 includes four examples of visualizing the relative importance of local regions on the surface of segmentation for its impact on dose prediction. The tumor target volume (represented in red) is displayed as an isosurface. The sensitivity map is overlaid on the surfaces of five OARs: brainstem, eye (left and right) and hippocampus (left and right). The colors indicate the mean absolute dose difference in the brain region due to a transformation at that location, and are measured in Gray. This is done using the MATLAB® Medical Imaging Toolbox. Table. I indicates the average mean absolute dose difference per OAR for the same four cases in Fig. 3.

Fig. 3 (a) shows a relatively small tumor target volume far from the entire OAR (about 3 cm). The most impactful points on the OARs are indeed the ones that are closer to the tumor target, while the magnitude of these changes are small (maximum of 0.168). Fig. 3 (b) shows a larger tumor target, with a wider distribution of dose differences across the surfaces of the OARs. The maximum mean absolute dose difference is also higher at 0.609 on the left hippocampus. The yellow arrow indicates a region on the surface of the brainstem that is away from the tumor target, yet has more impact on dose than points above it closer to the tumor target.

Fig. 3 (c) has a tumor target with a complex shape, leading to higher overall mean absolute dose differences. The yellow arrow indicates a region at the edge of the left hippocampus to be the most impactful, as it likely lies on critical beam angles. Fig. 3 (d) shows how the map changes when the tumor is closer to the OARs. Note the yellow arrow indicating the inferior parts of the eye more sensitive than superior, which differs from other cases. Another interesting point is both the hippocampi have a higher average mean absolute dose difference (see Table. I) even though the tumor target volume being close to only one of them. These observations are non-trivial and demonstrate the utility of this approach.

Sensitivity analysis: We demonstrate these results on ten test set subjects, featuring more than 2000 transformation points (more points on larger OARs like Brainstem and Eye) across all 13 OARs per subject. Table. II shows the correlation between the smallest distance to tumor target volume from the point of transformation and the mean average dose difference caused due to atomic surface transformations. Note that for larger OARs, like brainstem and eyes, the correlation is more strongly negative, indicating that points that are closer to the tumor target volume are more impacted by the transformations from a dose difference perspective. The correlation was found to be lower for smaller OARs like the cochlea or the lacrimal glands because the extent of these OARs does not vary much with respect to the distance from the tumor target.

A similar analysis was done to compute the correlation between local dose gradient magnitude versus the mean average dose difference. Results show a lower correlation than correlations obtained for distance-to-targets. However, as expected, there is still a positive overall correlation, confirming that dose estimates are sensitive to segmentation changes in locations where dose gradients are higher in magnitude.

Transformation size analysis: Fig. 4 shows the impact of changing the radius of the atomic surface transformation in the range $3, 5$ and 7. The dose differences correlate well by location (0.949 between radius 3 and 5, and 0.939 between 5 and 7), indicating the stability of finding important regions across this parameter.

Discussion: In this paper, we demonstrated with experiments on a test set of ten subjects with more than 20000 transformations at various points on the surfaces of OARs (each being an inference run of our deep learning-based dose prediction model) that the proposed 'atomic surface transformation' method can identify locations on segmentation boundaries that most impact dose computation. This would not have been practically feasible to compute with pre-deep learning methods. A limitation of this setup is that separate

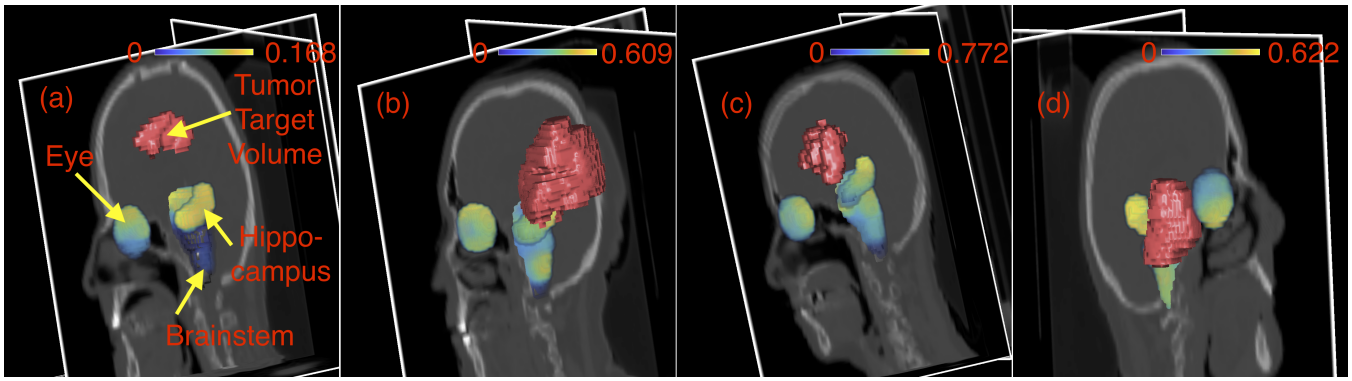

Fig. 3. Visualizing atomic surface transformations: demonstrated using selected OARs. Tumor target volume is shown in red; brighter yellow regions overlaid on OAR segmentations are most impactful on dose predictions, while darker blue regions describe the lowest impact. (a) to (d) demonstrate increasingly interesting situations: simple; large tumor; complex tumor shape; and tumor close to hippocampus and brainstem.

TABLE II

CORRELATION BETWEEN DOSE DIFFERENCE AND MINIMUM DISTANCE TO TUMOR TARGET (XCORR - DIST), AND LOCAL GRADIENT OF DOSE (XCORR - GRAD). AVERAGE SIZE OF THE OAR (IN VOXEL UNITS) INCLUDED FOR REFERENCE.

OAR	Size	XCorr (dist)	XCorr (grad)
Brainstem	2133.14	-0.39	0.22
Chiasm	48.71	-0.20	0.05
Cochlea L	8.14	-0.17	0.18
Cochlea R	8.42	-0.33	0.18
Eye L	637.42	-0.76	0.06
Eye R	628.14	-0.63	0.01
Hippocampus L	195.00	-0.35	0.41
Hippocampus R	206.85	-0.14	0.50
Lacrimal Gland L	78.57	0.58	0.39
Lacrimal Gland R	80.28	0.01	-0.22
Optic Nerve L	55.28	-0.03	0.05
Optic Nerve R	63.42	0.29	-0.15
Pituitary	61.00	-0.26	0.26

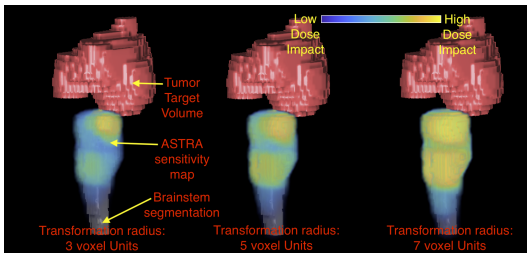

Fig. 4. When radius of atomic surface transformation is varied, the dose impact scales appropriately: results on representative brainstem with more than 900 atomic surface transformations. Radius measured in voxel units.

models need to be trained for every tumor location, delivery machine, and planning software. We deem this to be minor in practice, though, but also a target of potential future research.

Through this work, we hope to assist the clinical workflow of reviewing segmentations by drawing the attention of radiation oncologists to specific local regions of highest sensitivity to the RT plan. We believe clinicians would benefit from more detailed feedback in the form of sensitivity maps rather than a single numeric score. Next, we hope to improve the interpretation of the visualization and link it with dose constraints, such that locations, where transformations cause changes beyond the maximum dose limits for each OAR, could be highlighted. In the longer term, we plan to build on this further to create an automated dosimetry-aware segmentation quality assurance system.

IV. ACKNOWLEDGMENTS

We are funded by Swiss Cancer Research (KFS-5127-08-2020). We report no financial relationship or conflicts of interest.

REFERENCES

[1] J. R.McFaline-Figueroa and E. Q.Lee, "Brain tumors," *The American Journal of Medicine*, vol. 131, no. 8, pp. 874–882, 2018.

[2] R.Stupp, W. P.Mason, M. J.Van Den Bent, M.Weller, B.Fisher, M. J.Taphoorn, K.Belanger, A. A.Brandes, C.Marosi, U.Bogdahn, et al., "Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma," *New England Journal of Medicine*, vol. 352, no. 10, pp. 987–996, 2005.

[3] C.Scaringi, L.Agolli, and G.Minniti, "Technical advances in radiation therapy for brain tumors," *Anticancer research*, vol. 38, no. 11, pp. 6041–6045, 2018.

[4] I. J.Das, V.Moskvin, and P. A.Johnstone, "Analysis of treatment planning time among systems and planners for intensity-modulated radiation therapy," *Journal of the American College of Radiology*, vol. 6, no. 7, pp. 514–517, 2009.

[5] L. J.Peters, B.O'Sullivan, J.Giralt, T. J.Fitzgerald, A.Trotti, J.Bernier, J.Bourhis, K.Yuen, R.Fisher, and D.Rischin, "Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02.02," *Journal of Clinical Oncology*, vol. 28, no. 18, pp. 2996–3001, 2010.

[6] M.Niyazi, M.Brada, A. J.Chalmers, S. E.Combs, S. C.Erridge, A.Fiorentino, A. L.Grosu, F. J.Lagerwaard, G.Minniti, R.-O.Mirimanoff, et al., "ESTRO-ACROP guideline "target delineation of glioblastomas"," *Radiotherapy and Oncology*, vol. 118, no. 1, pp. 35–42, 2016.

[7] C.Guo, P.Huang, Y.Li, and J.Dai, "Accurate method for evaluating the duration of the entire radiotherapy process," *Journal of Applied Clinical Medical Physics*, vol. 21, no. 9, pp. 252–258, 2020.

[8] A.Babier, B.Zhang, R.Mahmood, K. L.Moore, T. G.Purdie, A. L.McNiven, and T. C.Chan, "Openkbp: The open-access knowledge-based planning grand challenge and dataset," *Medical Physics*, vol. 48, no. 9, pp. 5549–5561, 2021.

[9] A.Kamath, R.Poel, J.Willmann, N.Andratschke, and M.Reyes, "How sensitive are deep learning based radiotherapy dose prediction models to variability in organs at risk segmentation?," in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2023, pp. 1–4.

[10] S.Liu, J.Zhang, T.Li, H.Yan, and J.Liu, "Technical note: A cascade 3d u-net for dose prediction in radiotherapy," *Medical Physics, https://doi.org/10.1002/mp.15034*, 2021.