

# UNCERTAINTY-DRIVEN SANITY CHECK: APPLICATION TO POSTOPERATIVE BRAIN TUMOR CAVITY SEGMENTATION

Jungo A.<sup>1</sup>, Meier R.<sup>2</sup>, Ermis E.<sup>3</sup>, Herrmann E.<sup>3</sup>, Reyes M.<sup>1</sup>

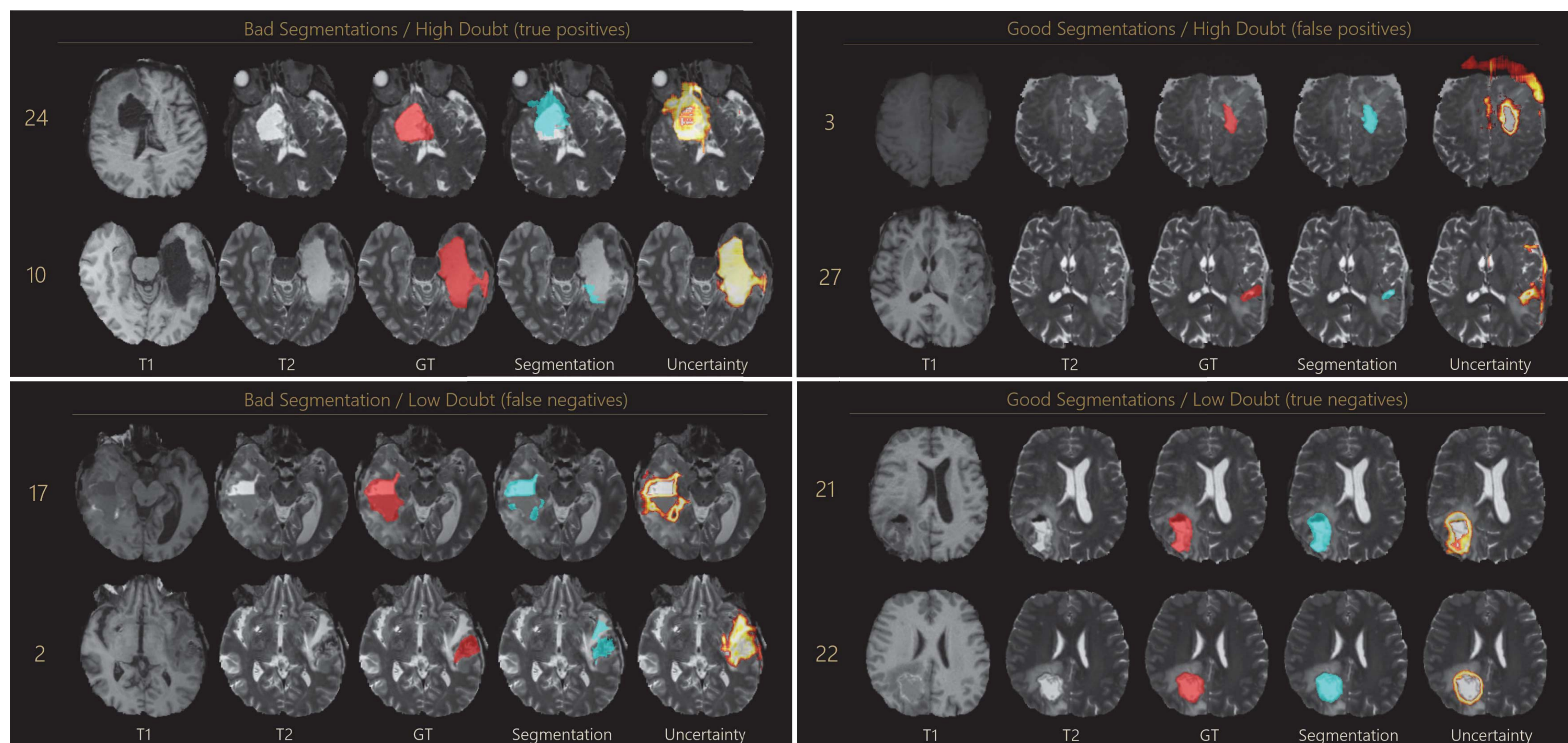
<sup>1</sup>Institute for Surgical Technology and Biomechanics, University of Bern, Switzerland

<sup>2</sup>Support Center for Advanced Neuroimaging, Institute for Diagnostic and Interventional Neuroradiology, Inselspital, Bern University Hospital, University of Bern, Switzerland

<sup>3</sup>University Clinic for Radio-oncology, Inselspital, Bern University Hospital, University of Bern, Switzerland

## Conclusion

The proposed method produces accurate segmentations of postoperative brain tumor cavities and identifies cases necessitating user monitoring by using the model's uncertainty. Additionally, the results point out the importance of the spatial prior introduced in the *doubt score* to improve the reliability of the score as indicator of challenging cases. Overall, the results suggest that a transfer of the uncertainty-driven sanity check to other segmentation tasks is possible.



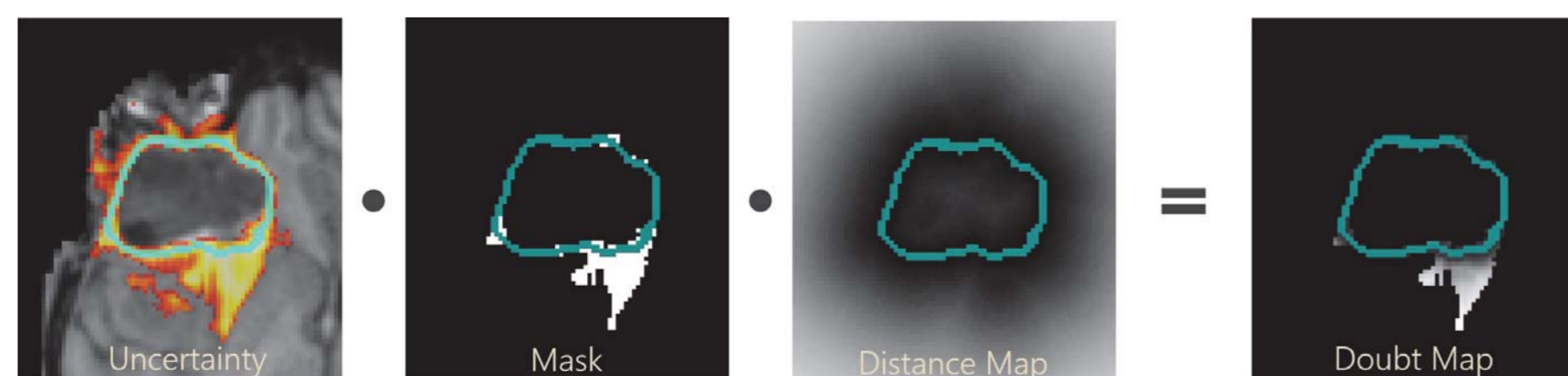
## Introduction

Uncertainty estimates can increase the understanding and foster acceptance of deep learning technologies by providing additional information about their predictions. This additional information enables building systems with a self-assessment capability for cases necessitating human monitoring. We propose an approach based on a fully-convolutional neural network for postoperative brain tumor cavity multi-sequence image segmentation that employs model's uncertainty to automatically identify challenging cases for expert review.

## Method

We use a fully-convolutional DenseNet [1] architecture to perform the segmentation of the tumor resection cavity. The architecture processes the three-dimensional brain volumes as three separated two-dimensional plane-wise predictions.

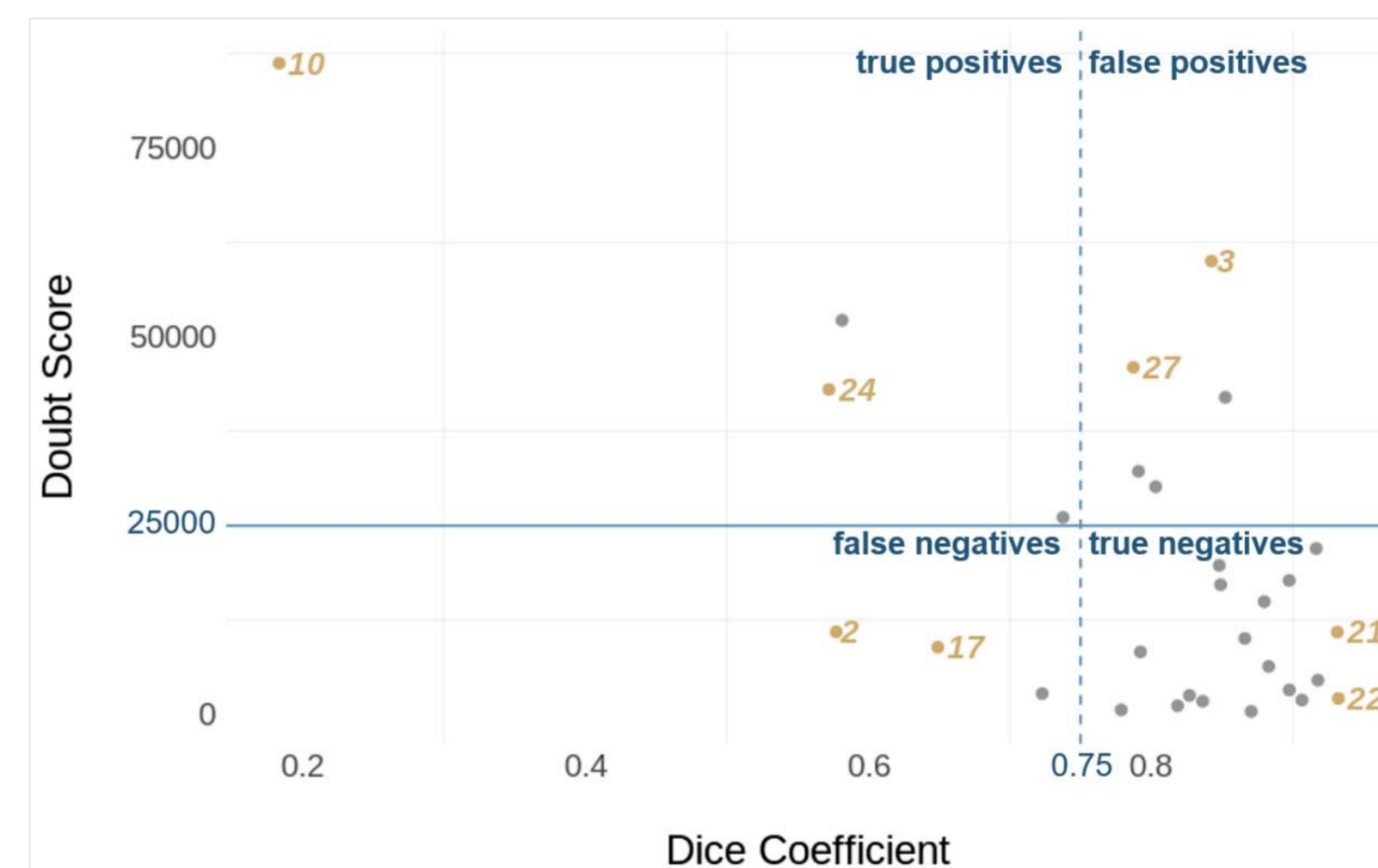
We compute the model's uncertainty by applying Monte Carlo dropout [2] sampling. As a result, for each segmentation we obtain a corresponding uncertainty map indicating the regions of high and low confidence. We aggregate the overall uncertainty with the help of a spatial prior, penalizing uncertainties being distant from the segmentation result, and define the resulting scalar value as *doubt score*. High *doubt score* values indicate a high likelihood that a user correction is needed. We evaluated the segmentation as well as the error detection performance of our approach on 30 standard multi-sequence (T1, T1c, T2, FLAIR) postoperative brain tumor cavity magnetic resonance images with six-fold cross-validation.



Building blocks of the introduced *doubt score*: Uncertainty estimates from Monte Carlo dropout [2], mask with most uncertain regions, distance map to segmentation boundary. The *doubt score* is the sum of the doubt map values.

## Results

Regarding segmentation performance, we achieved an average Dice coefficient of  $0.792 \pm 0.154$  with a median of 0.839. The average Hausdorff distance is  $16.24 \pm 9.07$ mm with a median of 14.74mm. By applying uncertainty-driven sanity check, we were able to detect the worst segmentation result (Dice coefficient: 0.184) and three out of the four cases with Dice coefficients below 0.6. The doubt score also pointed us to cases with Dice coefficients greater than 0.7 containing abnormalities in the images.



Relation between *doubt score* and Dice coefficient. The horizontal line is the chosen sanity check threshold. With the vertical line, the lines form 4 regions of outcomes.

## References

- [1] Simon Jégou et al. The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on, pages 1175–1183. IEEE, 2017.
- [2] Yarin Gal and Zoubin Ghahramani. Bayesian Convolutional Neural Networks with Bernoulli Approximate Variational Inference. In 4th International Conference on Learning Representations (ICLR) workshop track, 2016.

## Acknowledgments

This work was supported by the Swiss National Foundation by grant number 169607.

Correspondence to [alain.jungo@istb.unibe.ch](mailto:alain.jungo@istb.unibe.ch)