Assessing and Enhancing the Robustness of Brain Tumor Segmentation using a Probabilistic Deep Learning Architecture

E Alwadee X Sun Y Qin FC Langbein

November 2023

Primary category: AI & Machine Learning — Diagnosis/Prediction
Secondary category: Analysis Methods — Segmentation
Keywords: Analysis/Processing, Cancer, Robustness

Synopsis

Motivated by the challenge of enhancing the robustness of deep neural network decisions against variable noise in MRI-based brain tumor segmentation, this study aims to evaluate the efficacy of probabilistic bottlenecks. Our approach simulates structured perturbations at increasing strength to assess their impact on segmentation performance utilizing the Wasserstein distance between per-sample Dice score distributions and the sensitivity with respect to the perturbation strength. Results show probabilistic bottlenecks significantly increase robustness to Gaussian noise, yet offer limited improvement towards Gaussian blur, with varying results for other perturbations, highlighting the perturbation-specific nature of network resilience.

Impact

This study provides a tool to assess and guard against various perturbations in deep learning. It specifically demonstrates that probabilistic bottlenecks boost robustness of performance with respect to certain noise types, but not all.

Abstract

1. Introduction

Advances in deep learning for medical image segmentation have greatly enhanced diagnosis, yet issues of confidence and robustness persist due to variabilities in MRI imaging [1]. Our study examines whether a probabilistic bottleneck embedded within our lightweight deep neural network LATUP-Net (see Figure 1) designed for brain tumor segmentation [2] can bolster the robustness of its decisions. A method to compare the robustness with respect to structured perturbations is introduced to achieve this. Model performance robustness under naturally-induced image variations, rather than adversarial learning, is a particular important, but a less studied factor in image-based and medical decision-making [3]. Our results are critical for brain tumor segmentation and may extend to other medical imaging analyses.

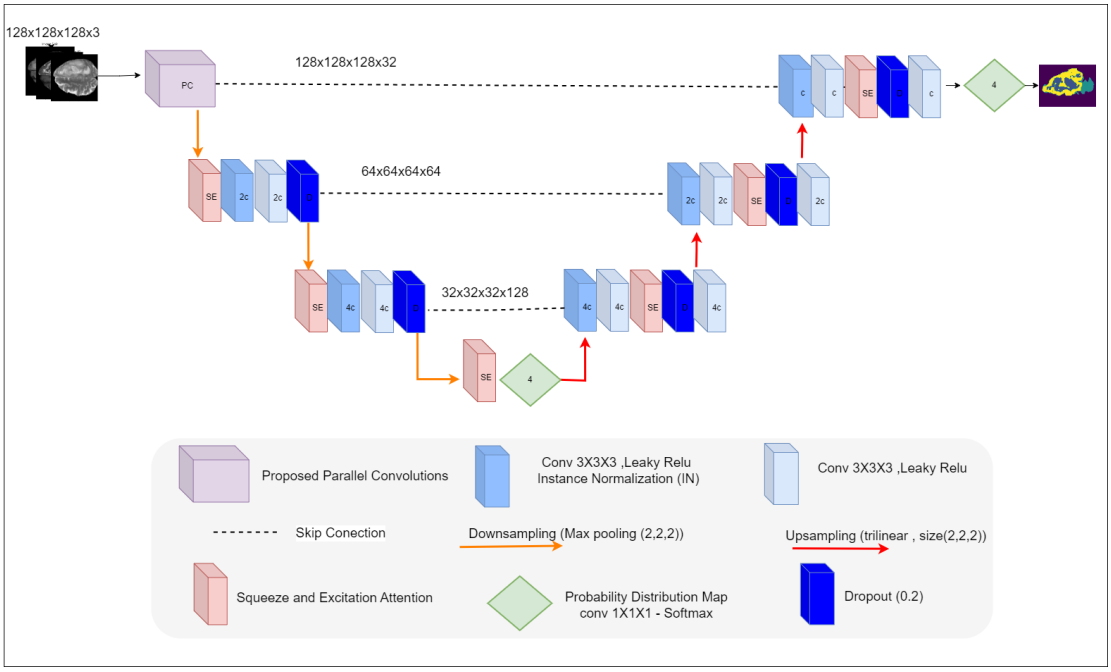

Figure 1: The LATUP-Net Model Architecture [2] Enhanced with Probabilistic Bottleneck [4].

2. Methods

To study robustness, we introduce structured perturbations describing different types of perturbation to MR images at increasing strength to the validation set for deep learning. Examples of perturbations are Gaussian noise at increasing variance, and Gaussian blur with increasing kernel size. Other perturbations, such as elastic deformation, scale, shift, and rotation, are also applied. Performance is measured by the per-sample Dice score distribution. The Wasserstein distance between consecutive distributions indicates performance deterioration as perturbation strength increases. Moreover, overall robustness is assessed using a box-spline fit to the performance distribution means and its sensitivity, i.e., the derivative with respect to perturbation strength.

This is employed to compare model robustness of our LATUP-Net architecture for brain cancer segmentation [2] with and without incorporating a probabilistic bottleneck, inspired by the successful application in variational autoencoders [4], using the BraTS 2020 dataset [5].

3. Results

Figures 2 to 4 show the results of our robustness analysis for Gaussian noise and blur on LATUP-Net. Figures 2 and 3 illustrate the model's Dice scores with and without the probabilistic bottleneck under varying levels of Gaussian noise. For both we observe a robustness plateau—a range of perturbation strengths where the model's performance remains relatively unaffected, followed by a performance drop. The model without probabilistic bottleneck (Figure 2) demonstrates an abrupt descent in Dice scores beyond the plateau, indicating a limited range of noise tolerance. Conversely, the incorporation of the probabilistic bottleneck (Figure 3) revealed a more substantial plateau with a lower slope of the drop, suggesting the model can sustain high accuracy across

a wider spectrum of noise before performance wanes.

The segmentation performance under Gaussian blur is captured in Figures 4 and 5. While a similar robustness plateau is observed, the probabilistic bottleneck's advantage is less pronounced (Figure 4), particularly at higher levels of blur where feature distortion is more severe. Despite this, the probabilistic model still outperforms the standard model at lower levels of blur (Figure 5), although the overall trend confirms that both models' performance inevitably declined once the noise exceeded a critical threshold.

Moreover, in both scenarios a jump in the Wasserstein distance indicates a change in behavior, which can be categorized by sections of constant or linear change at an approximately fixed slope (sensitivity).

Collectively, this underscores the nuanced influence of the probabilistic bottleneck on model robustness against Gaussian perturbations and reveals the limits of robustness. Similar behaviors are observed for other perturbations.

4. Discussion

Generally we observe a robustness plateau followed by one or two approximately linear sections. The plateau radius and slope (sensitivity) of the subsequent performance drop offer two critical parameters for evaluating model robustness, characterized by jumps in the Wasserstein distance. The plateau radius signifies the model's capacity to withstand noise without significant performance degradation. The slope characterizes the model's ability to resist noise influences beyond the plateau. Due to the approximate linear behavior seen empirically, the slope appears to be sufficient. Generally, behavior at higher perturbation strengths at low performance is not interesting.

Moreover, we found that robustness differs greatly depending on the perturbation type, as indicated by the reduced efficacy of the probabilistic bottleneck against Gaussian blur. This observation emphasizes the need for a differentiated approach to enhancing robustness, with specific requirements for training data and potentially augmentation, depending on specific perturbation characteristics.

5. Conclusion

Our research evaluates the robustness of a novel lightweight deep neural network (DNN) with probabilistic bottleneck for brain tumor segmentation on the BraTS2020 dataset. Introducing structured perturbations, we assess model performance using per-sample Dice score distributions, which demonstrate a robustness plateau radius and a subsequent linear performance drop. This yields two effective measures for robustness comparison. The study is constrained by not employing fully realistic MRI perturbations, relying instead on augmentation transforms. This limitation guides future work towards integrating more authentic perturbations to refine the DNN's reliability for clinical imaging applications.

References

1. Zou K, Yuan X, Shen X, Wang M, Fu H. TBraTS: Trusted Brain Tumor Segmentation. In: Medical Image Computing and Computer Assisted Intervention—MICCAI. Lecture Notes in Computer Science 2022;13438:503–513.

2. Alwadee E, Sun X, Qin Y, Langbein FC. LATUP-Net: A Lightweight 3D Attention U-Net with Parallel Convolutions for Brain Tumour Segmentation. Preprint; 2023. https://ca.

qyber.dev/bca/paper-patt-net/. Accessed November 7, 2023.

3. Drenkow N, Sani N, Shpitser I, Unberath M. A Systematic Review of Robustness in Deep Learning for Computer Vision: Mind the gap? arXiv; 2023;2112.00639

4. Myronenko A. 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. Lecture Notes in Computer Science 2019;11384:311–320.

5. Menze BH, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Transactions on Medical Imaging 2014;34(10):1993–2024.

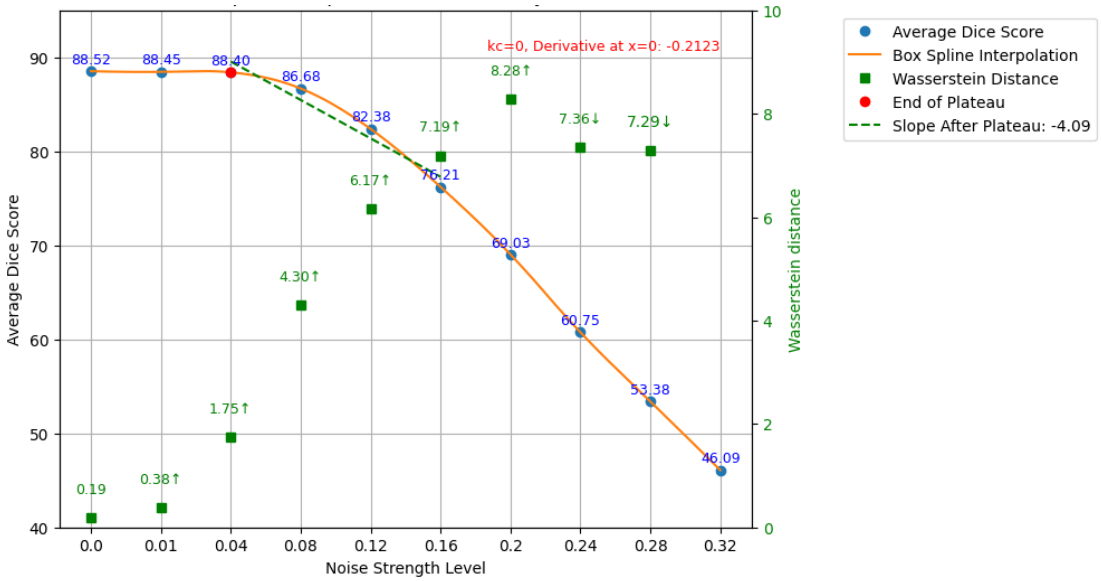

Figure 2: Comparative robustness of LATUP-Net, our proposed deep learning model, without a probabilistic bottleneck is shown through average per-sample Dice scores for the whole tumor segmentation class across various strengths of Gaussian noise. The green dots represent the Wasserstein distance between the current per-sample Dice score distribution and the next.

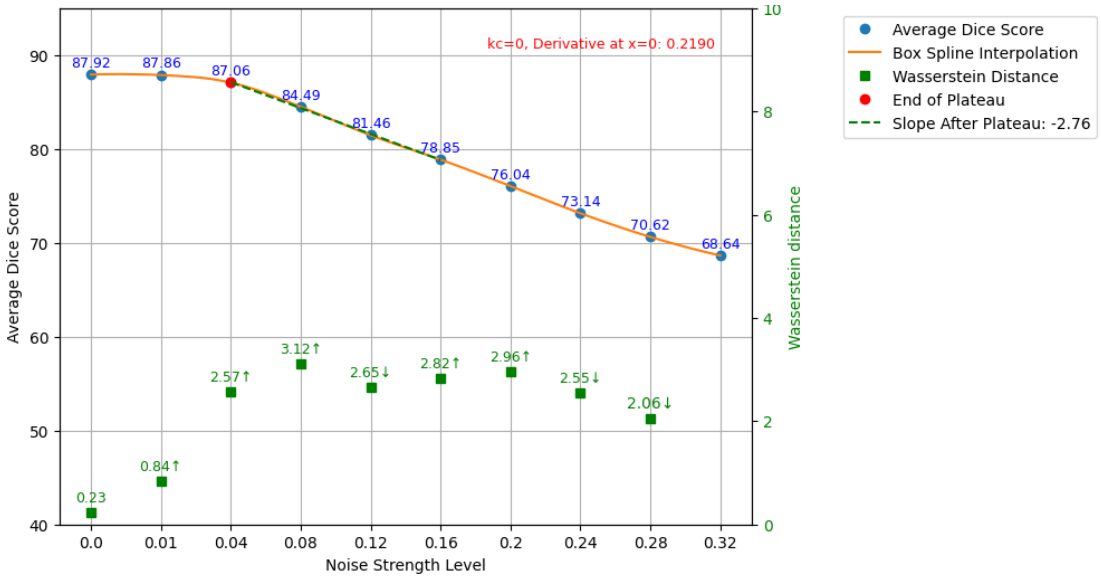

Figure 3: Comparative robustness of LATUP-Net, our proposed deep learning model, enhanced with probabilistic bottleneck is shown through average per-sample Dice scores for the whole tumor segmentation class across various strengths of Gaussian noise. The green dots represent the Wasserstein distance between the current per-sample Dice score distribution and the next.

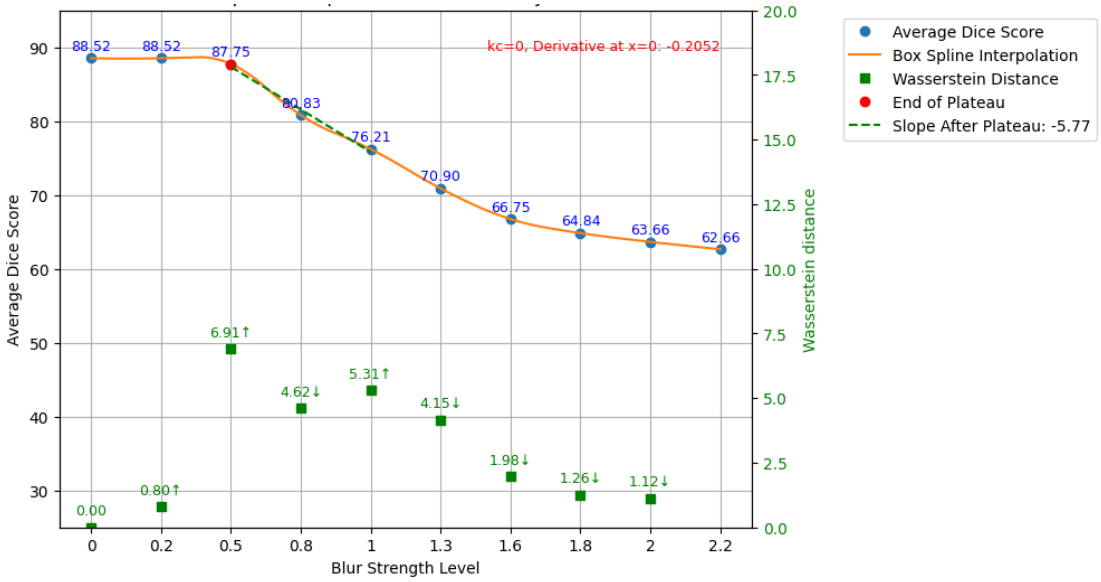

Figure 4: Comparative robustness of LATUP-Net, our proposed deep learning model, without a probabilistic bottleneck is shown through average per-sample Dice scores for the whole tumor segmentation class across various strengths of Gaussian blur. The green dots represent the Wasserstein distance between the current per-sample Dice score distribution and the next.

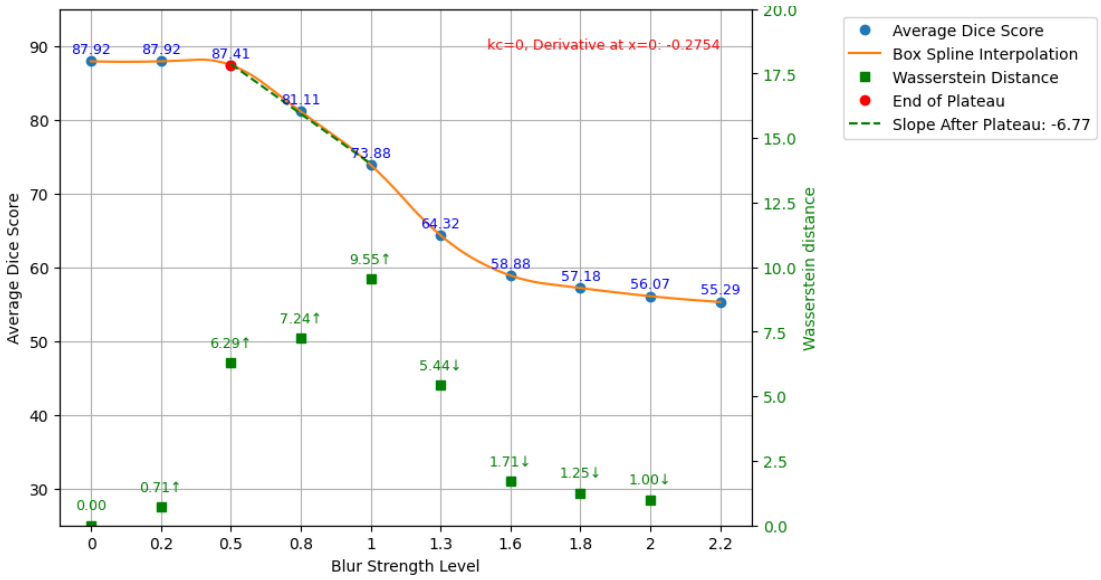

Figure 5: Comparative robustness of LATUP-Net, our proposed deep learning model, enhanced with probabilistic bottleneck is shown through average per-sample Dice scores for the whole tumor segmentation class across various strengths of Gaussian blur. The green dots represent the Wasserstein distance between the current per-sample Dice score distribution and the next.