



**HAL**  
open science

## Uncertainty-based Quality Control for Subcortical Structures Segmentation in T1-weighted Brain MRI

Benjamin Lambert, Florence Forbes, Senan Doyle, Alan Tucholka, Michel Dojat

► **To cite this version:**

Benjamin Lambert, Florence Forbes, Senan Doyle, Alan Tucholka, Michel Dojat. Uncertainty-based Quality Control for Subcortical Structures Segmentation in T1-weighted Brain MRI. ISMRM 2023 - International Society for Magnetic Resonance in Medicine, Jun 2023, Toronto, Canada. pp.1-4. hal-04126201

**HAL Id: hal-04126201**

**<https://hal.science/hal-04126201>**

Submitted on 13 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Uncertainty-based Quality Control for Subcortical Structures Segmentation in T1-weighted Brain MRI

Benjamin Lambert<sup>1,3</sup>, Florence Forbes<sup>2</sup>, Senan Doyle<sup>1</sup>, Alan Tucholka<sup>1</sup>,  
Michel Dojat<sup>3</sup>

<sup>1</sup>*Pixyl, Research and Development Laboratory, 38000 Grenoble, France,*

<sup>2</sup>*Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France,*

<sup>3</sup>*Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, 38000 Grenoble, France*

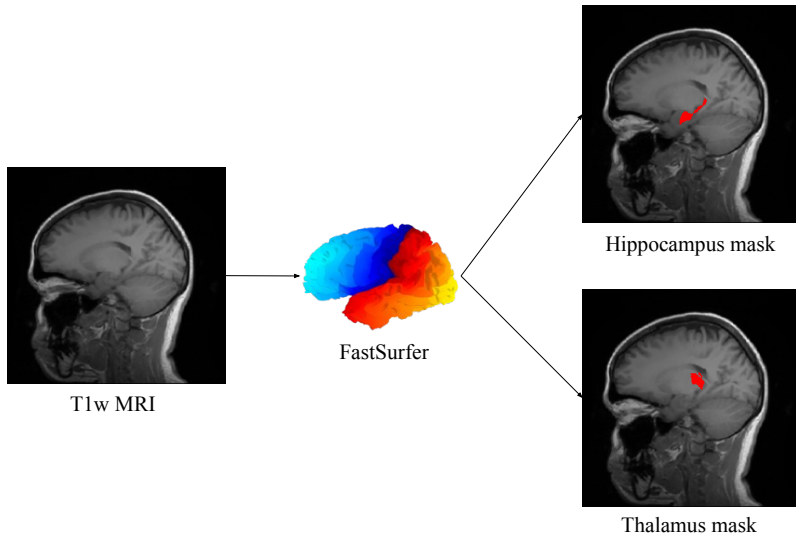
**Introduction :** Deep Learning (DL) models are presently the gold standard for medical image segmentation. However, their performance may drastically drop in the presence of characteristics in test images not present in the training set. The automatic detection of these Out-Of-Distribution (OOD) inputs is the key to prevent the silent failure of DL models, especially when the visual inspection of the input is not systematically carried out [1]. For MRI segmentation, a wide range of covariables can perturbate a DL model : noise, artifacts or MR sequence parameters. Deterministic Uncertainty Methods (DUM) [2] are novel and promising techniques for OOD detection. They propose to analyze the intermediate activations of a trained segmentation DL model to detect OOD inputs. In a previous study, we demonstrated that DUM achieved high OOD detection performance on a task of Multiple Sclerosis lesions segmentation in T2-weighted FLAIR MRI [3]. To evaluate the generalization capability of this technique, we propose to evaluate DUM in the context of automatic subcortical structures segmentation. We focus our results on the hippocampus and thalamus structures segmentation from T1-weighted MR brain scans of healthy subjects.

**Methods :** For this work, we used the IXI dataset<sup>1</sup> composed of 581 healthy T1-weighted brain MRI. For each scan, we used FastSurfer in order to obtain a parcellation of the brain composed of 95 structural classes [4], which we used to isolate 2 subcortical regions : the hippocampus and thalamus (Figure 1). We then split the dataset into a training set (381 subjects) and a testing set (200 subjects). As these test images shared the same origin as the training images, we referred to them as in-distribution (ID) test images. For each ID test image, 7 synthetic OOD images were obtained by using the TorchiO's library [5] transformations : Downsample, Bias, Ghost, Motion, Gaussian Noise, Scale and Spikes (Figure 2). These OOD images are representative of common MRI artifacts that can be encountered in clinical practice. On the training set, we trained 2 Attention U-Nets 3D [6] to respectively segment the hippocampus and thalamus. We then explored two ways of detecting OOD inputs from the trained models : Maximum Softmax Probability (M1) and DUM (M2).

**M1. Maximum Softmax Probability (MSP)** [7] proposes to analyze the output probabilities of the segmentation model to detect OOD. Allegedly, ID images should yield to greater probabilities than OOD images, which should allow for their detection. To implement this method, we retrieved the MSP for each voxel, corresponding to the probability of the predicted class. We then computed the mean MSP across the MRI volume to get a single conformity score for each image.

---

1. <https://brain-development.org/ixi-dataset>

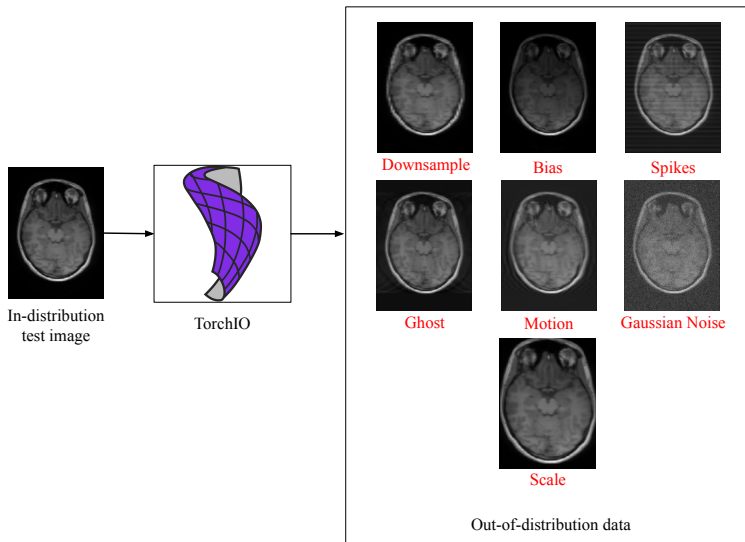


**Figure 1** – Subcortical region masks generation using FastSurfer.

**M2. Deterministic Uncertainty Method (DUM)** aims at detecting OOD images based on the intermediate activations of a trained segmentation model. We implement DUM as follows [8] : for a given query image, we first gathered the activations  $F$  of the penultimate layer of the trained segmentation model.  $F$  corresponds to a 4-dimensional array of shape  $NHWD$ ,  $HWD$  being the dimension of the MRI and  $N$  being the number of features, which is 32 in our DL model. We then computed the Singular Value Decomposition (SVD) of  $F$ , and defined its spectral signature as the vector of singular values. The final conformity score corresponds to the minimum euclidean distance between the spectral signature of the test image, and the signatures of all training images. The higher the distance, the more abnormal the input.

Our evaluation protocol was two-fold. First, we evaluated the segmentation quality of the subcortical structures using the Dice score (DSC) between the predictions and the FastSurfer parcellation. Second, we evaluated the ability to distinguish between ID and OOD images using both MSP and DUM. We defined OOD detection as a binary classification problem, where ID test images and OOD images were respectively negative and positive samples. We then computed the AUROC score for each type of MRI artifact by comparing the conformity scores obtained for ID and OOD data.

**Results** : Segmentation and OOD detection performances are presented in Table 1. DSC scores decrease on OOD data, yet this drop is only catastrophic for two types of artifacts : Spikes and Gaussian Noise. For these types, only the DUM approach allows a robust identification of the OOD cases, as indicated by the AUROC score of 1.0. Overall, DUM surpasses MSP on 6 out of 7 OOD datasets and on 5 out of 7 for the hippocampus and thalamus segmentation task



**Figure 2** – Generation of Out-of-distribution images using the Torchio library. From a clean in-distribution test image, 7 different types of artifacted image are generated.

respectively.

**Discussion** : Overall, DUM surpasses MSP on the task of OOD detection, for both the hippocampus and thalamus models. In cases where the segmentation quality drastically decreases, i.e. Spikes and Gaussian Noise artifacts, DUM allows a perfect identification of non-conform images, which would prevent poor-decision making based on the DL model predictions. Additionally, such Quality Control is easy to implement and computationally efficient as it only requires a trained segmentation model.

**Conclusion** : We compared the performance of two Out-of-Distribution detection frameworks equipping DL segmentation models. Experiments on a set of 7 synthetic MRI artifacts and on 2 subcortical structures segmentation tasks show the superiority of the DUM approach. This confirms the robustness and the versatility of DUM for the identification of non-conform MRIs.

|                 | Hippocampus model |             |             | Thalamus model |       |             |
|-----------------|-------------------|-------------|-------------|----------------|-------|-------------|
|                 | DSC               | AUROC       |             | DSC            | AUROC |             |
|                 |                   | MSP         | DUM         |                | MSP   | DUM         |
| In-distribution | 0.90              | NA          | NA          | 0.93           | NA    | NA          |
| Downsample      | 0.87              | 0.72        | <b>0.78</b> | 0.92           | 0.65  | <b>0.86</b> |
| Bias            | 0.89              | 0.67        | <b>0.71</b> | 0.92           | 0.68  | <b>0.91</b> |
| Spikes          | 0.04              | 0.05        | <b>1.00</b> | 0.21           | 0.33  | <b>1.00</b> |
| Ghost           | 0.87              | 0.79        | <b>0.81</b> | 0.87           | 0.82  | <b>0.90</b> |
| Motion          | 0.86              | <b>0.79</b> | <b>0.79</b> | 0.88           | 0.80  | <b>0.87</b> |
| Gaussian Noise  | 0.00              | 0.01        | <b>1.00</b> | 0.49           | 0.85  | <b>1.00</b> |
| Scale           | 0.83              | 0.55        | <b>0.98</b> | 0.85           | 0.56  | <b>1.00</b> |

**Table 1** – Segmentation quality (DSC) and Out-of-Distribution (OOD) detection performance (AUROC) for hippocampus and thalamus models, on in and out-of-distribution datasets. Top performing OOD detection method is highlighted in bold. MSP : Max Softmax Probability. DUM : Deterministic Uncertainty Method. NA : non-applicable.

## Références :

- [1] Jingkang, Y. Zhou, K. Li, Y. and Liu, Z., 2021. Generalized out-of-distribution detection : A survey. arXiv preprint arXiv :2110.11334.
- [2] Postels, J. Segù, M. Sun, T. Sieber, LD. Van Gool, L. Yu, F. and Tombari F., 2022. Proceedings of the 39th International Conference on Machine Learning, PMLR 162 :17870-909.
- [3] Lambert B., Forbes F., Doyle S., Tucholka A. and Dojat, M., 2022. Improving Uncertainty-based out-of-distribution detection for medical image segmentation, arxiv preprint (2022).
- [4] Henschel L., Conjeti S., Estrada S., Diers K., Fischl B. and Reuter M., 2020. Fastsurfer-a fast and accurate deep learning based neuroimaging pipeline. NeuroImage, 219, p.117012.
- [5] Pérez-García F., Sparks R. and Ourselin, S. TorchIO : a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Computer Methods and Programs in Biomedicine 208 (2021) : 106236.
- [6] Ozan O. et al. Attention u-net : Learning where to look for the pancreas. arXiv preprint arXiv :1804.03999 (2018).
- [7] Hendrycks D. and Gimpel K., 2016. A baseline for detecting misclassified and out-of-distribution examples in neural networks. arXiv preprint arXiv :1610.02136.
- [8] Karimi D. and Gholipour A., 2020. Improving calibration and out-of-distribution detection in medical image segmentation with convolutional neural networks. arXiv preprint arXiv :2004.06569.