Instance-level explanations in multiple sclerosis lesion segmentation: a novel localized saliency map

Federico Spagnolo[1,2,3,4], Nataliia Molchanova[4,5], Roger Schaer[4], Mario Ocampo-Pineda[1,2,3], Meritxell Bach Cuadra[5,6], Lester Melie-Garcia[1,2,3], Cristina Granziera[1,2,3], Vincent Andrearczyk[4], and Adrien Depeursinge[4,7]

[1]Translational Imaging in Neurology (ThINK) Basel, Department of Medicine and Biomedical Engineering, University Hospital Basel and University of Basel, Basel, Switzerland, [2]Department of Neurology, University Hospital Basel, Basel, Switzerland, [3]Research Center for Clinical Neuroimmunology and Neuroscience Basel (RC2NB), University Hospital Basel and University of Basel, Basel, Switzerland, [4]MedGIFT, Institute of Informatics, School of Management, HES-SO Valais-Wallis University of Applied Sciences and Arts Western Switzerland, Sierre, Switzerland, [5]CIBM Center for Biomedical Imaging, Lausanne, Switzerland, [6]Radiology Department, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland, [7]Nuclear Medicine and Molecular Imaging Department, Lausanne University Hospital (CHUV) and University of Lausanne, Lausanne, Switzerland

Synopsis

Keywords: Other AI/ML, Machine Learning/Artificial Intelligence, Explainability, interpretability

Motivation: The use of AI in clinical routine is often jeopardized by its lack of transparency. Explainable methods would help both clinicians and developers to identify model bias and interpret the automatic outputs.

Goal(s): We propose an explainable method providing insights into the decision process of an MS lesion segmentation network.

Approach: We adapt SmoothGrad to perform instance-level explanations and apply it to a U-Net, whose inputs are FLAIR and MPRAGE from 10 MS patients.

Results: Our saliency maps provide local-level information on the network's decisions. Predictions of the U-Net rely predominantly on lesions' voxel intensities in FLAIR and the amount of perilesional volume.

Impact: These results cast some light on the decision mechanisms of deep learning networks performing semantic segmentation. The acquired new knowledge can be an important step to facilitate AI integration into clinical practice.

Abstract
(https://cds.ismrm.org/protected/24MProceedings/PDFfiles/2231_RrAZecQBe.html)

Keywords