

CONFLUNET: IMPROVING CONFLUENT LESION IDENTIFICATION IN MULTIPLE SCLEROSIS WITH INSTANCE SEGMENTATION

Maxence Wynen^{†1,2}, Maxime Istasse¹, Pedro M. Gordaliza^{3,4}, Anna Stölting², Pietro Maggi^{2,5},
Benoît Macq^{*1}, Meritxell Bach Cuadra^{*3,4}

¹ICTEAM, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

²Louvain Neuroinflammation Imaging Lab (NIL), Université Catholique de Louvain, Brussels, Belgium

³CIBM Center for Biomedical Imaging, Switzerland

⁴Department of Radiology, Lausanne University Hospital, Lausanne, Switzerland

⁵Cliniques Universitaires Saint-Luc, Université Catholique de Louvain, Brussels, Belgium

ABSTRACT

In multiple sclerosis (MS), white matter lesion (WML) *instance* masks are very relevant to enhance diagnosis and disease monitoring. Yet, all existing automated WML segmentation methods aim at improving a *semantic* segmentation model, and postprocessing it to group voxels together into lesion instances. A large majority of studies use connected components (CC) analysis for this final step. In this paper, we show both theoretically and empirically that CC is sub-optimal for WML instance segmentation due to the presence of confluent lesions (CLs), i.e. lesions whose segmentation encompasses two or more individual lesions. We address this issue by proposing ConFLUNet - the first end-to-end instance segmentation model designed to detect and segment WML instances in MS. We evaluate ConFLUNet against two baseline methods, and show that it improves lesion detection metrics while maintaining similar segmentation performance. The results shown in this paper pave the way for more in-depth analysis of instance segmentation applied in the context of MS. Source code is available on Github.

Index Terms— Instance segmentation, Multiple Sclerosis, White matter lesion segmentation, Confluent lesions

1. INTRODUCTION

Multiple sclerosis (MS) is a chronic autoimmune disease, and represents the most common cause of non-traumatic neurological disability in young adults [1]. An important characteristic of MS is the presence of central nervous system demyelinated white matter lesions (WML), which are visible on conventional MRI scans. The total count and cumulative volume of these lesions play a crucial role for both MS diagnosis [2] and prognosis [3], and both can be derived from a lesion instance mask. Lesion *instance* masks (LIMs, Fig. 1b-1d) differ from lesion *semantic* masks (LSMs, Fig. 1a) because,

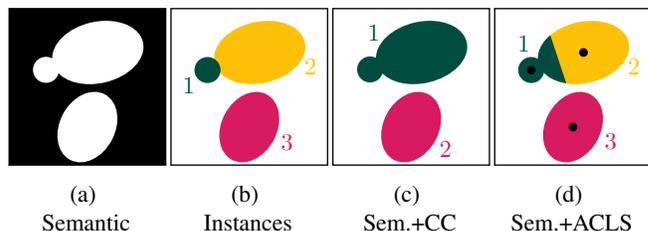


Fig. 1: Scheme of semantic (Sem), instance masks, and instantiation outputs (CC, connected components; ACLS, Automated Confluent lesion Splitting).

in addition to a class label (lesion/non-lesion), each voxel in the LIM is also associated with a unique lesion identifier (id).

Because manually annotating any kind of 3D lesion mask is tedious and time consuming, many methods have been proposed to automate this process [4, 5]. However, these methods' outputs often necessitate manual correction [6, 7], especially when confluent lesions are involved. Confluent lesions (CLs) are lesions whose semantic segmentation encompasses two or more pathologically independent lesions, which we refer to as confluent lesion units (CLUs). As an illustration, the instance 1 in Fig 1c conceptualizes a CL encompassing two CLUs, displayed as instances 1 and 2 in Fig 1b. CLs occur either due to pathological reasons, when two lesions merge into one, or when the instance segmentation is unable to discriminate their enclosed CLUs. In that last case, the presence of CLs is closely related to most methods' choice to use connected components (CC) on the predicted LSM to produce the final LIM. We argue that the presence of CLs in MS patients renders the CC step inappropriate for WML instance segmentation, as their nature inherently disallows identification through this method (Fig 1c). Given that CLUs make up 10 to 20% of all lesions (refer to [7] and section 3.2), any WML segmentation method unable of detecting them will inevitably remain sub-optimal.

[†] Corresponding author. ^{*} The last two authors contributed equally.

Moreover, we further identified a dissonance between 1) the will of most methods [4, 8] and challenges [9, 10] to evaluate *detection* performance, and 2) the fact that most methods actually aim at improving these detection metrics by producing better *semantic* segmentation methods. Interestingly, this is occurring alongside the continued use of the sub-optimal CC method. Though efforts have been made to replace CC as a method to produce LIMs from LSMs [7, 11], surprisingly no previous research has explored the application of *end-to-end* instance segmentation models, to our knowledge. We aim to fill this gap by proposing ConfLUNet, a simple model adapted from a 3D U-Net architecture [12] and inspired by a state-of-the-art panoptic segmentation model for 2D image analysis [13]. We demonstrate that by adapting a state-of-the-art WML semantic segmentation model, we achieve better detection metrics while preserving a segmentation performance that is on par with or superior to the baseline methods.

2. RELATED WORK

While there is no published end-to-end method for WML instance segmentation, the prevailing WML semantic segmentation methods often evaluate their performance using both segmentation and detection metrics [4]. These methods mostly follow the same two-step approach, first computing a LSM [4, 8], then applying a post-processing technique to produce a LIM from the predicted WML mask, a technique we refer to as *instantiation*. As stated before, the go-to instantiation method is the connected-components (CC) analysis, wherein adjacent voxels are grouped together according to a connectivity parameter. However, this method falls short in detecting CLUs within lesions (Fig. 1c).

Another instantiation method, introduced in [7], uses an automatic confluent lesion splitting method termed ACLS. ACLS is divided in a lesion center detection step [11] and a voxel clustering step. The center detection step uses the eigenvalues of the hessian matrix calculated on the lesion probability mask, identifying center voxels based on negative eigenvalues in all three directions. This relies on the hypothesis that predicted lesion voxel probabilities align with MS lesion pathology, suggesting initial vein damage leads to outward inflammation with less impact on the lesion periphery. Identified center voxels are then clustered in a lesion center using CC. ACLS’s second step clusters remaining predicted lesion voxels by assigning them to their closest lesion center, producing a LIM. While promising and leveraging MS pathology insights, this method overlooks lesion size in the clustering algorithm, potentially leading to suboptimal results. This is illustrated in Figure 1d, where voxels belonging to larger lesions with a distant center have a higher chance of being erroneously assigned to closer centers. Finally, a recent study evaluating three WML segmentation tools with CC and ACLS found that ACLS prioritized sensitivity over specificity and worsened segmentation quality [14].

3. MATERIALS AND METHODS

ConfLUNet consists of an adapted 3D U-Net followed by an instantiation layer. The model uses FLAIR volumes as input, and has three outputs: a semantic segmentation, a center heatmap, where the value at each voxel represents its probability to be a lesion center, and an offset map with x , y and z components, corresponding for each voxel to an offset vector pointing to the center of the most probable lesion it belongs to. The three components are then combined during instantiation to produce a final LIM.

Model architecture: ConfLUNet adopts a 3-leveled 3D U-Net backbone, identical to [15]. Inspired by Panoptic DeepLab [13], the final convolutional layer is changed to output six channels, split among three task-specific heads, instead of the two employed in binary semantic segmentation. As input, the model takes 3D patches of $96 \times 96 \times 96$ voxels, randomly sampled during training and using a sliding window algorithm with Gaussian weights for inference [16].

Semantic segmentation head: The first two channels are for semantic segmentation. As in [17], a weighted combination of the focal loss [18] and the dice loss [19] was used as segmentation loss: $\mathcal{L}_{seg} = \mathcal{L}_{focal} + 0.5 * \mathcal{L}_{dice}$.

Center prediction head: To retrieve the ground truth (GT) center heatmap, each lesion instance is represented by its center of mass, computed by averaging the coordinates of all voxel positions within that instance. During training, GT instance centers are encoded using a 3D Gaussian with a standard deviation of $\pm 2mm$ pixels. The Mean Squared Error (MSE) loss is used for the center heatmaps prediction, corresponding to the third output channel [13].

Offsets prediction head: For every voxel (x, y, z) categorized as lesion in the GT, ConfLUNet predicts the offsets $\mathcal{O}(x, y, z)$ to its respective center of mass in the \vec{x} , \vec{y} and \vec{z} directions [13], corresponding to the last three output channels. This offsets map is learned through the minimization of a L_1 loss, which is only computed at GT lesion voxels.

Training Loss: The final training loss is a weighted sum between the semantic segmentation loss \mathcal{L}_{seg} , the center heatmap loss \mathcal{L}_{center} and the offsets loss $\mathcal{L}_{offsets}$:

$$\mathcal{L}_{total} = \mathcal{L}_{seg} + \alpha \mathcal{L}_{center} + \beta \mathcal{L}_{offsets} \quad (1)$$

where α and β are tunable parameters.

Lesion Instantiation: First, a LSM is obtained by applying a threshold of 0.5 to the softmaxed segmentation output. For the instantiation, the same principles as in [13] are applied: first obtain the predicted object centers $\hat{P} = \{\mathcal{C}_p : (x_p, y_p, z_p)\}$ by max pooling the predicted center heatmap and filtering out the values that have changed before and after the pooling. Next, every voxel in the LSM is assigned to the center lying closest to the voxel’s predicted offset vector $\mathcal{O}(x, y, z)$. Specifically, $\hat{c}(x, y, z)$, the predicted instance id

for a voxel at position (x, y, z) , is found by computing:

$$\hat{c}(x, y, z) = \underset{k}{\operatorname{argmin}} \|\mathcal{C}_k - ((x, y, z) + \mathcal{O}(x, y, z))\|^2 \quad (2)$$

3.1. Evaluation Metrics

We evaluate ConfLUNet by computing both semantic and instance segmentation metrics, relying on the *Metrics Reloaded*'s recommendations [20]. All metrics are first computed patient-wise, then averaged across patients in the partition set.

Matching strategy: To pair prediction instances to GT instances, we calculate the Intersection over Union (IoU) of every predicted lesion with respect to all GT lesions, selecting the GT instance for which the IoU is the highest above a predetermined threshold. We set this threshold to 0.1 instead of the standard 0.5 used in 2D imaging [13] to better align with the reality of 3D image processing and the instances size [20]. Consequently, we label unpaired predicted lesions as false positives, and unpaired GT lesions as false negatives.

Confluent lesion units identification: We further identify a subset of CLU $\mathcal{I}_{clu} \subset \mathcal{I}$ defined by:

$$\mathcal{I}_{clu} = \{i \mid 0 < \operatorname{IoU}(i, j) < 1; \forall i \in \mathcal{I}, j \in \mathcal{I}_{cc}\} \quad (3)$$

where \mathcal{I} is the set of all GT instances, and \mathcal{I}_{cc} is the set of instances obtained through CC post-processing on the GT binary lesion mask.

Semantic segmentation, or voxel-wise **metrics**, include the Dice Score (DSC), DSC_{TP} , the dice score computed only on correctly true positive (TP) instances and $\operatorname{DSC}_{TP,CLU}$, the dice score computed only on TP CLUs.

Detection metrics, computed based on the matching strategy described earlier, include Precision, Recall, F1 and absolute difference in count (DiC) between the number of predicted and GT instances. We also introduce the confluent lesion unit recall ($\operatorname{Recall}_{CLU}$), designed to measure how often CLUs have correctly been detected:

$$\operatorname{Recall}_{CLU} = \frac{|\operatorname{TP}_{CLU}|}{|\mathcal{I}_{clu}|} \quad (4)$$

where TP_{CLU} is the number of correctly predicted CLUs.

Instance segmentation-specific metrics: Finally, we also compute Panoptic Quality (PQ). The PQ measure combines Segmentation Quality (SQ) and Recognition Quality (RQ). Specifically, PQ is computed as follows:

$$\operatorname{SQ} = \frac{\sum_{(i,j) \in \mathcal{M}} \operatorname{IoU}(i, j)}{|\operatorname{TP}|}; \quad \operatorname{RQ} = \frac{|\operatorname{TP}|}{|\operatorname{TP}| + \frac{1}{2}|\operatorname{FP}| + \frac{1}{2}|\operatorname{FN}|} \quad (5)$$

$$\operatorname{PQ} = \operatorname{SQ} \times \operatorname{RQ} \quad (6)$$

where $\mathcal{M} \subset \mathbb{R}^2$ is the set of matched instance pairs such that $|\mathcal{M}| = |\operatorname{TP}|$, FP is the set of false positive instances and FN the set of false negative instances.

3.2. Dataset

Our dataset includes 63 MS patients, aged 22-66 years. Brain images of participants were acquired using a 3T Signa Premier General Electrics MRI scanner at Saint-Luc University Hospital in Brussels, Belgium. Reference lesion instance segmentation was performed by two experts (A.S & P.M.) upon consensus, using 3D-FLAIR images. After exclusion of cortical and infratentorial lesions, 1104 lesions were identified, of which 221 were considered as CLUs according to (3).

3.3. Experimental setting

The dataset was split into 47 subjects for training, 13 for validation, and 13 for dedicated testing, with randomized partitions to ensure comparable lesion count and volume distributions. ConfLUNet was compared to a semantic 3D U-Net baseline on the test set. This baseline generates a semantic WML mask, further instantiated via CC and ACLS to produce instance lesion masks termed *UNet+CC* and *UNet+ACLS*. For all methods, connected components or instances whose size were below $3mm$ in any axes or below $14mm^3$ in volume were removed, according to the clinical definition of a MS lesion [21]. This setup allows to minimize variations from semantic segmentation quality.

3.4. Experiments

In all experiments, all parameters were fixed except for α and β , respectively the weights of center and offsets losses, and the learning rate ($\{1e^{-4}, 1e^{-5}\}$). To give roughly the same importance to every loss, a grid search was conducted with values of $\alpha \in \{500, 1000, 1200, 1500, 2000\}$ and values of $\beta \in \{0.1, 0.3, 0.5\}$ totaling to 30 experiments. All models were trained for 300 epochs. Finally, all validation metrics were computed for the 10 models achieving the highest DSC on the validation set, and the best-performing model ($\alpha = 2000, \beta = 0.5, lr=1e^{-4}$) was compared with the baselines on the test set. A comparison of this model with the baselines on the validation and test set is shown on Table 1.

4. RESULTS

Table 1 provides a summary of ConfLUNet and the baseline methods' performance, assessed on the validation and test sets. This section only considers the results on the test set. Regarding detection metrics first, ConfLUNet outperforms *UNet+CC* and *UNet+ACLS* in Precision, Recall and F1 measures, and has mixed results concerning DiC. Regarding segmentation metrics, ConfLUNet achieves higher DSC and $\operatorname{DSC}_{TP,CLU}$, but has mixed results when it comes to DSC_{TP} , performing better than *UNet+ACLS* but worse than *UNet+CC*. Finally, ConfLUNet also surpasses the baseline methods when it comes to PQ. Figure 2 shows illustrative examples of the LIM produced by each method.

	Method	Precision (%)	Recall (%)	F1 (%)	Recall _{CLU} (%)	DiC \searrow	DSC (%)	DSC _{TP} (%)	DSC _{TP, CLU} (%)	PQ (%)
Validation	3D U-Net + CC	76.2 ± 19.5	66.0 ± 16.0	70.7 ± 18.1	44.2 ± 12.4	2.8 ± 5.1	70.3 ± 12.0	77.0 ± 4.9	65.2 ± 13.1	45.1 ± 12.7
	3D U-Net + ACLS	67.7 ± 16.2	78.8 ± 13.3	72.8 ± 13.7	82.9 ± 23.6	12.6 ± 21.1	66.4 ± 10.6	64.8 ± 6.9	68.0 ± 12.0	35.9 ± 7.9
	ConFLUNet (ours)	66.8 ± 14.3	89.5 ± 10.8	76.5 ± 13.3	85.4 ± 18.6	8.2 ± 10.8	71.1 ± 6.7	76.6 ± 5.9	74.3 ± 9.5	47.7 ± 6.9
Test	3D U-Net + CC	61.5 ± 18.3	63.0 ± 20.6	62.2 ± 16.9	19.4 ± 20.8	4.8 ± 4.1	58.1 ± 15.7	73.6 ± 11.3	31.6 ± 34.1	36.4 ± 11.2
	3D U-Net + ACLS	56.6 ± 16.8	78.5 ± 19.9	65.7 ± 20.4	54.6 ± 50.3	11.5 ± 13.0	55.0 ± 16.8	62.3 ± 9.3	36.2 ± 36.8	30.2 ± 9.2
	ConFLUNet (ours)	63.1 ± 18.2	82.1 ± 22.7	71.4 ± 21.7	70.0 ± 21.2	6.6 ± 8.1	59.5 ± 14.3	71.8 ± 9.5	53.3 ± 25.3	40.0 ± 10.9

Table 1: Comparison between ConFLUNet and 3D U-Net semantic segmentation baseline model + Connected Components (CC) or Automated Confluent Lesion Splitting (ACLS) as post-processing. Results are on the Validation and Test sets. Values averaged over all patients in the partition set along with the standard deviation. PQ: Panoptic Quality, DSC: Dice Score, DSC_{TP}: Dice Score at True Positives, DSC_{TP, CLU}: Dice Score at true positive confluent lesions, DiC: Absolute Different in Count.

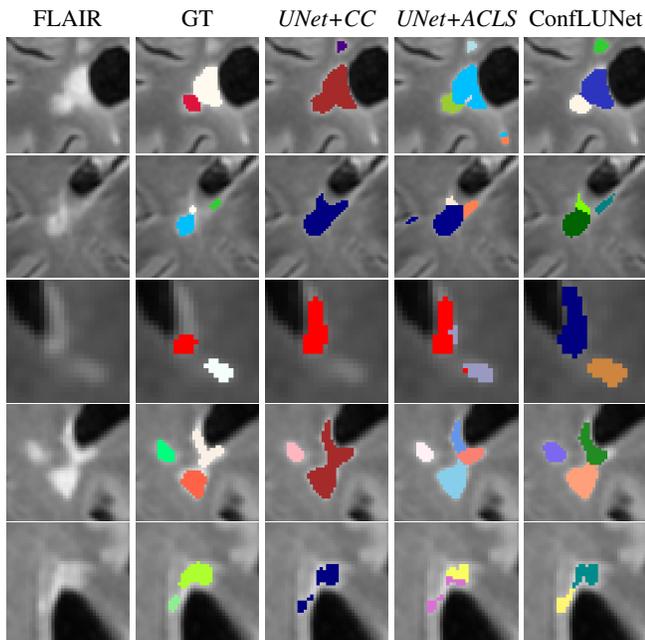


Fig. 2: Examples of lesion instance masks produced by *UNet+CC*, *UNet+ACLS* and *ConFLUNet*, drawn from the test dataset. GT: Ground Truth; CC: connected components; ACLS: Automated Confluent Lesion Splitting.

ConFLUNet clearly outperforms previous instantiation strategies by improving detection metrics and maintaining high segmentation performance. It surpasses *UNet+CC* by tripling CLUs’ detection rate and nearly doubling their segmentation metric, empirically proving this method is sub-optimal to detect CLUs (cf sec. 1 & 2; Figs. 1c and 2). The difference in DiC can be attributed to *UNet+CC*’s tendency to predict a lower number of lesions, missing most CLUs. Regarding DSC_{TP}, *UNet+CC*’s slight superiority is potentially due to the fact it detected fewer lesions than ConFLUNet (116 v. 145). A paired t-test revealed no significant difference in that metric for the two methods ($p \approx 0.3$). Furthermore, ConFLUNet surpasses *UNet+ACLS* in all metrics, and achieves

1.5 times its score on the segmentation of CLUs.

Among the four subjects for which models performed the worst in the test set, three had noisy images due to motion artifacts, and one had only one large tumefactive lesion. This, along with the small sizes of the different datasets, contributes to the high variability across patients, and the poor generalization. Additional analysis showed that most discrepancies between ConFLUNet and experts involved small lesions: undetected GT lesions had lower volumes than detected ones ($p < .05$), and FP lesions had lower volumes than TP lesions ($p < .0001$). Additionally, out of the 94 FP lesions, 28 exhibited gliosis, frequently including one or multiple small vascular lesions, 22 were correctly outlined as cortical or infratentorial lesions, and 31 were re-evaluated as TP.

5. DISCUSSION AND CONCLUSION

The results presented in this paper provide empirical evidence supporting the theoretical argument that both the CC and ACLS instantiation methods are suboptimal for accurately detecting or segmenting CLUs. CC struggles in detecting CLUs, whereas ACLS, despite improving their detection rate, lacks precision and faces challenge with their segmentation. Contrary to the baseline methods, ConFLUNet learns both to detect lesion centers *and* how to correctly assign voxels to each center. However, additional refinement through advanced hyperparameter tuning and thorough validation on larger in- and out-of-domain datasets, including annotations from multiple experts, is needed. Currently, the major hurdle for extensive exploration of end-to-end WML instance segmentation is the scarcity of publicly available datasets with instance segmentation masks. Challenges like those suggested in [9, 10] could significantly enhance the field by promoting better model generalization across diverse datasets, similar to the progress seen in semantic segmentation [22]. This improvement would facilitate the integration of these methods into clinical practice. Still, ConFLUNet’s superiority across nearly all metrics underscores the potential of end-to-end instance segmentation approaches to address the challenge of CLs in WML instance segmentation.

6. COMPLIANCE WITH ETHICAL STANDARDS

All procedures performed in this study were in accordance with the ethical standards of the Belgian national research committee (n° B4032020000104).

7. ACKNOWLEDGMENTS

M.W. is funded by TRAIL and the Walloon region. M.B.C. and P.M.G. acknowledge the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), École polytechnique fédérale de Lausanne (EPFL), University of Geneva (UNIGE) and Geneva University Hospitals (HUG).

8. REFERENCES

- [1] Daniel S. Reich, Claudia F. Lucchinetti, and Peter A. Calabresi, “Multiple Sclerosis,” *The New England Journal of Medicine*, vol. 378, no. 2, pp. 169–180, Jan. 2018.
- [2] Alan J. Thompson et al., “Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria,” *The Lancet. Neurology*, vol. 17, no. 2, pp. 162–173, 2018.
- [3] on behalf of the MAGNIMS study group, “MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis—establishing disease prognosis and monitoring patients,” *Nature Reviews Neurology*, vol. 11, no. 10, pp. 597–606, Oct. 2015.
- [4] Amrita Kaur et al., “State-of-the-Art Segmentation Techniques and Future Directions for Multiple Sclerosis Brain Lesions,” *Archives of Computational Methods in Engineering*, vol. 28, no. 3, pp. 951–977, May 2021.
- [5] Chenyi Zeng et al., “Review of Deep Learning Approaches for the Segmentation of Multiple Sclerosis Lesions on Brain MRI,” *Frontiers in Neuroinformatics*, vol. 14, pp. 610967, Nov. 2020.
- [6] Germán Barquero et al., “RimNet: A deep 3D multi-modal MRI architecture for paramagnetic rim lesion assessment in multiple sclerosis,” *NeuroImage: Clinical*, vol. 28, pp. 102412, Jan. 2020.
- [7] Carolyn Lou et al., “Fully Automated Detection of Paramagnetic Rims in Multiple Sclerosis Lesions on 3T Susceptibility-Based MR Imaging,” *NeuroImage: Clinical*, p. 102796, Aug. 2021.
- [8] Chenyi Zeng, Lin Gu, Zhenzhong Liu, and Shen Zhao, “Review of Deep Learning Approaches for the Segmentation of Multiple Sclerosis Lesions on Brain MRI,” *Frontiers in Neuroinformatics*, vol. 14, pp. 610967, Nov. 2020.
- [9] Martin Styner et al., “3D Segmentation in the Clinic: A Grand Challenge II: MS lesion segmentation,” *The MIDAS Journal*, Nov. 2008.
- [10] Aaron Carass et al., “Longitudinal multiple sclerosis lesion segmentation data resource,” *Data in Brief*, vol. 12, pp. 346–350, 2017.
- [11] J. D. Dworkin et al., “An Automated Statistical Technique for Counting Distinct Multiple Sclerosis Lesions,” *American Journal of Neuroradiology*, vol. 39, no. 4, pp. 626–633, Apr. 2018, Publisher: American Journal of Neuroradiology Section: ADULT BRAIN.
- [12] Olaf Ronneberger et al., “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *arXiv:1505.04597 [cs]*, May 2015, arXiv: 1505.04597.
- [13] Bowen Cheng, “Panoptic-DeepLab: A Simple, Strong, and Fast Baseline for Bottom-Up Panoptic Segmentation,” Mar. 2020, arXiv:1911.10194 [cs].
- [14] Maxence Wynen, Pedro M. Gordaliza, Anna Stölting, Pietro Maggi, and Meritxell Bach Cuadra, “Lesion Instance Segmentation in Multiple Sclerosis: Assessing the Efficacy of Statistical Lesion Splitting,” 2024, ISMRM.
- [15] Francesco La Rosa et al., “Multiple sclerosis cortical and WM lesion segmentation at 3T MRI: a deep learning method based on FLAIR and MP2RAGE,” *NeuroImage: Clinical*, vol. 27, pp. 102335, Jan. 2020.
- [16] “MONAI: Medical Open Network for AI,” .
- [17] Andrey Malinin et al., “Shifts 2.0: Extending The Dataset of Real Distributional Shifts,” Sept. 2022, arXiv:2206.15407 [cs, stat].
- [18] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár, “Focal Loss for Dense Object Detection,” Feb. 2018, arXiv:1708.02002 [cs].
- [19] Carole H. Sudre et al., “Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations,” vol. 10553, pp. 240–248. 2017, arXiv:1707.03237 [cs].
- [20] Lena Maier-Hein et al., “Metrics reloaded: Recommendations for image analysis validation,” June 2023, arXiv:2206.01653 [cs].
- [21] S. Grahl et al., “Evidence for a white matter lesion size threshold to support the diagnosis of relapsing remitting multiple sclerosis,” *Multiple Sclerosis and Related Disorders*, vol. 29, pp. 124–129, Apr. 2019.
- [22] Francesco La Rosa et al., “A deep learning-based pipeline for longitudinal white matter lesion segmentation using diverse FLAIR images,” 2023.