New Multiple Sclerosis Lesion Segmentation via Calibrated Inter-patch Blending

Jin Ye[1,*], Son Duy Dao[1,*], Yicheng Wu[1], Yasmeen George[1], Thanh Nguyen-Duc[1], Daniel F. Schmidt[1], Hengcan Shi[2,†], Winston Chong[3,4], and Jianfei Cai[1]

[1] Department of Data Science & AI, Faculty of Information Technology, Monash University, Melbourne, VIC 3168, Australia
[2] College of Electrical and Information Engineering, Hunan University, Changsha, Hunan Province, 410082, China
[†] shihengcan@gmail.com
[3] Alfred Health Radiology, Alfred Health, Melbourne, VIC 3004, Australia
[4] School of Translational Medicine, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, VIC 3800, Australia

Abstract. Longitudinal monitoring of multiple sclerosis (MS) lesions provides crucial biomarkers for assessing disease progression and treatment efficacy. However, it remains challenging to detect and segment numerous MS lesion instances accurately. One key limitation lies in the common average blending of sliding-window predictions during inference, where unreliable patch-level outputs often lead to many false-positive results. To address this issue, we propose a ***Calibrated Inter-patch Blending*** *(CIB)* framework for new MS lesion segmentation, leveraging patch-level segmentation performance as blending weights. Specifically, our CIB model incorporates a multi-scale design with two additional prediction heads: one estimates the overall segmentation performance of the input patch, while the other predicts the performance of smaller grids within the patch. This dual-head architecture enables the model to capture both global and local contextual information, reducing overconfident lesion predictions. During inference, the predicted segmentation scores serve as calibration weights for adaptively blending patch predictions. Extensive experiments on the MSSEG-2 dataset demonstrate that our CIB model can significantly enhance both **new MS lesion detection** (*e.g.,* a 12.82% F1 gain) and **segmentation** (*e.g.,* a 4.01% Dice gain) across various backbones. Our code is available at https://github.com/Yejin0111/CIB.

Keywords: Multiple Sclerosis · Lesion Segmentation · Calibration

1 Introduction

Multiple sclerosis (MS) is a neurological disorder in which the immune system mistakenly attacks the protective myelin sheath surrounding nerve fibers. This

[*] Jin Ye and Son Duy Dao contributed equally to this work.

disruption interferes with signal transmission between the brain and the rest of the body, affecting millions of people worldwide [7]. Over time, irreversible harm or degradation of the nerve fibers will occur [20]. During the clinical treatment, accurately segmenting MS lesions is crucial for analyzing longitudinal activities, to assess disease progression and effectiveness of therapies [16]. However, identifying changes in MS lesions over multiple time points poses a considerable challenge for clinicians [1], because MS lesions are usually small, numerous, and can be misdiagnosed to other types of brain lesions (*e.g.,* ischemic vasculopathy [8]). Further, considering the high cost of manual analysis and the high requirement for clinical expertise, a precise segmentation model is highly desirable for correctly quantifying MS lesion instances in clinical applications.

Deep learning has significantly advanced the segmentation of MS lesions from brain MRI scans, by exploring diverse strategies for better performance [21,25,26]. For example, multiple MRI contrasts can be introduced for model training [12], alongside efforts to tackle class imbalance [28,19] and integrate anatomical priors [2]. Attention mechanisms [17] have been used to boost MS lesion segmentation. Furthermore, Krishnan et al. [10] developed a 3D, multi-arm U-Net specifically for the segmentation of T2 lesions, trained in a comprehensive multi-center clinical trial dataset for relapsing MS. Additionally, Zhang et al. [29] studied the multi-rater medical image segmentation for MS, noting the label quality on the algorithm's predictive accuracy. Recent attention has been shifted to the longitudinal MS lesions analysis [5,6], including the categorization of lesions as stable, newly formed, shrinking, or expanding [11]. Particularly, Coact-Seg [23] proposed to segment new MS lesions by utilizing both readily available single-time-point samples and heterogeneously annotated two-time-point data.

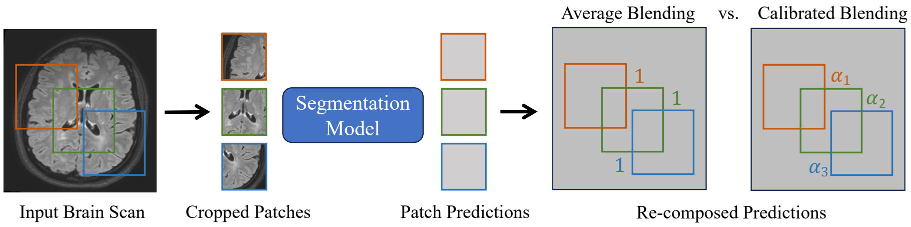

Fig. 1. Illustration of the sliding window blending technique. The common average blending heavily depends on the model segmentation quality and often leads to high false positives, while our proposed calibrated blending (right) weighs each patch prediction adaptively by its learnable calibration score.

We notice that most existing approaches receive 3D patch inputs and utilize a sliding window technique for inter-patch blending, as illustrated in Figure 1, which predicts MS lesions patch-by-patch and then averages the predictions for the final segmentation. However, this average blending approach heavily depends on the model segmentation quality and often leads to high false positives in MS segmentation due to the sparsity and size variation of MS lesions [23,3,19].

This motivates us to propose a Calibrated Inter-patch Blending (CIB) framework to mitigate the over-segmentation of MS lesions. Our key idea is to predict the segmentation performance of individual image patches, which are then leveraged to calibrate patch re-composition. Amazingly, such a simple idea significantly reduces false positives and improves the segmentation performance across both global metrics (Dice score) and instance-level evaluations (F1 score).

Overall, our contributions are three-fold:

- We propose a Calibrated Inter-patch Blending (CIB) framework for MS lesion segmentation, pointing out that re-composing patch-level predictions in a weighted way can significantly reduce false positives.
- Our CIB framework consists of two regularization heads that predict the segmentation performance at the patch level and the smaller grid level. The predicted segmentation performances are used as calibration scores for the final inter-patch blending.
- Extensive experiments demonstrate significant performance improvements in new MS lesion segmentation using our CIB framework across various backbones.

2 Method

2.1 Problem Definition

Preliminary Study. The training set is constructed as $(\boldsymbol{X},\ \boldsymbol{Y})$. Following [23], $\boldsymbol{X}$ is a set of two-time-point MRI scans [14], and $\boldsymbol{Y}$ is the new-lesion or full-lesion label. Figure 1 depicts the sliding window technique during inference, wherein the segmentation mask is predicted for each patch, and the patches are re-composed based on their locations. Since there are overlapping patches, the patch sum is usually averaged to derive the final segmentation P as

$$\boldsymbol{P} = \frac{\sum_{i=1}^{N} \alpha_i \times \boldsymbol{p}_i}{\sum_{i=1}^{N} \boldsymbol{A}_i} \tag{1}$$

where $\boldsymbol{p}_i$ is the predicted segmentation for each patch $x_i \in \boldsymbol{X}$, $\boldsymbol{A}_i$ is an all-one matrix, the sum $\sum_{i=1}^{N} \boldsymbol{A}_i$ counts the total number of overlaps for the patches, and N is the total number of patches. The weight α_i is introduced as a calibrated weight for each patch, with $\alpha_i = 1$ as “Average Blending”. Simply average strategy often produces a lot of false positives and achieves sub-optimal segmentation performance due to the sparsity and size variation of MS lesions [23,3,19]. Thus, we introduce the idea of **calibrated blending**. Table 1 gives a preliminary experiment, where we directly use the real segmentation performance (*i.e.*, Dice score between the ground truth and the predicted segmentation masks) of the patches as the blending weights α_i for the calibrated blending, denoted as “Calibrated Blending (Patch)”. The results show that this simple inference strategy

Table 1. Comparisons of different inter-patch blending techniques for new MS lesion segmentation on the MSSEG-2 dataset [3], with the CoactSeg baseline [23]. Note that, the blending is calibrated by the real segmentation performance.

Method	Dice(%)↑	Jaccard(%)↑	95HD(voxel)↓	ASD(voxel)↓	F1(%)↑
Average Blending	63.82	51.68	**30.35**	12.14	61.96
Calibrated Blending (Patch)	65.58	52.26	39.22	0.47	**79.40**
Calibrated Blending (Grid)	**69.52**	**55.75**	36.83	**0.43**	78.11

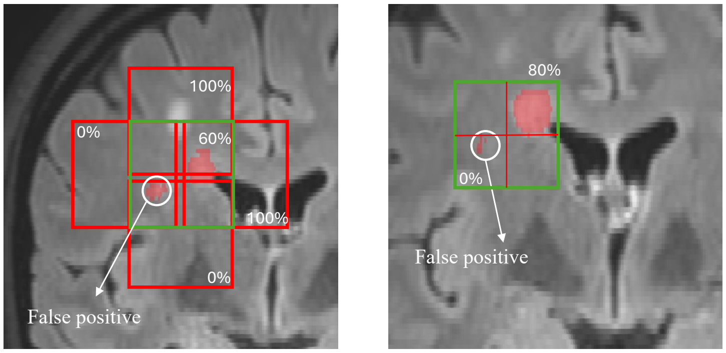

Fig. 2. Weighted blending effectively reduces false positives since it assigns different weights to each patch prediction. Patch weight (Left) can give a low weight to a false positive prediction based on the average score of its neighboring patches, while grid weight (Right) performs in a smaller region.

yields a notable improvement in both new lesion segmentation (a 1.76% gain in Dice) and detection (a 17.44% gain in F1).

We further extend the idea to a fine-grained level, *i.e.*, performing the blending at the grid (sub-patch) level as

$$\boldsymbol{P} = \frac{\sum_{i=1}^{N} \boldsymbol{p}_i}{\sum_{i=1}^{N} \boldsymbol{A}_i}; \ \boldsymbol{p}_i = \sum_{j=1}^{M} \alpha_j \times \boldsymbol{p}_{ij} \tag{2}$$

where $\boldsymbol{p}_{ij}$ is the predicted segmentation of the j^{th} grid in the i^{th} patch $\boldsymbol{p}_i$, and M is the number of grids (set to 8 for simplicity, forming a $2 \times 2 \times 2$ cube), and the weight α_j represents the grid weight. As shown in Table 1, the grid-calibrated blending, denoted as "Calibrated Blending (Grid)", exhibits a further 4% gain in Dice, compared to the "Calibrated Blending (Patch)".

Figure 2 illustrates how the application of patch or grid weights contributes to the reduction of false positives during the inference blending. The patch weighting scheme allows removing false positives based on its neighbor patch weights. As shown on the left of Figure 2, while the overall segmentation performance of the testing patch (green box) is the same for both true positive and false positive lesions, neighbor patches of false positives often yield low segmentation performance, while those of true positives exhibit significantly better performance.

Consequently, after inter-patch averaging, the region of false positives will have a low predicted value, leading to rejection during inference. Furthermore, as demonstrated on the right side of Figure 2, even with a high overall performance (*e.g.*, an 80% Dice score), false positives may still exist because their size is too small to influence the overall score. Here, the grid weighting scheme effectively isolates small lesions, assigning them a low score to facilitate their rejection during testing. Additionally, learning these contexts can further regularize the model training [27] and our experiments find that employing a simple Dice score as the context supervision has already achieved superior performance, see Table. 5.

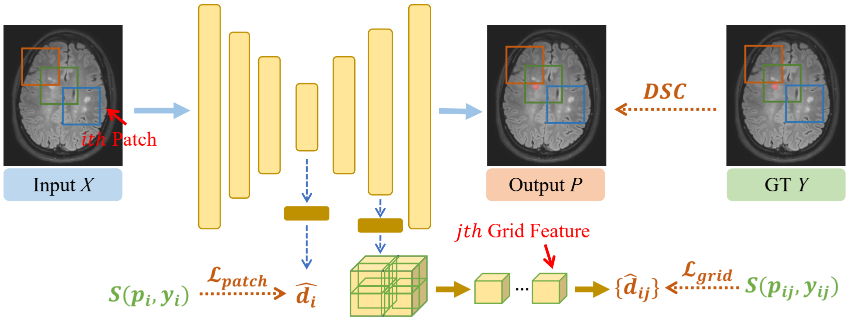

Fig. 3. Overview of our proposed CIB framework. We introduce two heads to predict the segmentation performance of the input patch and smaller grids, respectively.

2.2 Proposed Approach

Since ground truth during inference is unknown in practice or cannot be used for model tuning, here we describe our approach to predict the segmentation performance at the patch and grid levels, as shown in Figure 3. In particular, we set CoactSeg [23] as the baseline, which is trained on two types of datasets: full-lesion segmentation dataset (MS23-v1) and new-lesion segmentation dataset (MSSEG-2). The segmentation model follows a typical U-Net architecture, which includes an encoder and a decoder. The overall pipeline can be summarized by the following function:

$$\boldsymbol{p}_i = F_\theta(\boldsymbol{x}_i|c) \tag{3}$$

where c is a condition to control the segmentation setting, *e.g.*, full-lesion predictions for the single-time-point scan or new-lesion predictions for the given two-time-point data. F_θ denotes the baseline model [23].

We introduce two additional components: the patch head, which processes the highest feature from the encoder to estimate the segmentation performance for the corresponding patch, and the grid head, which utilizes intermediate features from the decoder to estimate the grid-level performance. The patch head consists of a linear layer followed by a sigmoid function. The grid head involves transforming grid features through a 3D convolutional layer, followed by average pooling and a linear layer with a sigmoid function.

We train the patch head and the grid head using the mean square error (MSE) loss, which calculates the difference between the predicted segmentation performance $\hat{d}$ and the ground truth d obtained by comparing the patch predictions $\boldsymbol{p}_i$ with the corresponding lesion mask $\boldsymbol{y}_i \in Y$. Specifically, the training losses for the patch head and grid head are:

$$d_i = S(\boldsymbol{p}_i, \boldsymbol{y}_i);\ \mathcal{L}_{patch} = MSE(\hat{d}_i, d_i) \tag{4}$$

$$d_{ij} = S(\boldsymbol{p}_{ij}, \boldsymbol{y}_{ij});\ \mathcal{L}_{grid} = \sum_{j=1}^{M} MSE(\hat{d}_{ij}, d_{ij}) \tag{5}$$

where S is a similarity function, set as a common Dice score. Finally, the total training loss is a weighted sum of a segmentation dice loss $DSC(\boldsymbol{p}_i, \boldsymbol{y}_i)$ (same as [23]) and our two regularization losses:

$$\mathcal{L}_{total} = DSC(\boldsymbol{p}_i, \boldsymbol{y}_i) + \lambda_1 \times \mathcal{L}_{patch} + \lambda_2 \times \mathcal{L}_{grid}. \tag{6}$$

3 Experiments and Results

Datasets. Following [23], our model is trained by two MRI datasets: MS23-v1 with full-lesion labels and MSSEG-2 with new-lesion labels. Therefore, there are 38 single-time-point [23] and 40 two-time-point [3] brain FLAIR scans, respectively. We adopt an identical dataset split as [23] and implement a weighted cropping strategy [28] to extract 3D brain patches sized $80 \times 80 \times 80$.

Implementation Details. We set the batch size as 8 and each batch contains 4 two-time-point samples and 4 single-time-point samples for the joint training. We utilize the Adam optimizer with a learning rate of 1e-2. Then, we train the baseline model with only the patch head for 10k iterations (*i.e.,* $\lambda_1 = 1$ & $\lambda_2 = 0$). Subsequently, both patch and grid heads are trained for additional 10k iterations (*i.e.,* $\lambda_1 = 1$ & $\lambda_2 = 1$). For fair comparisons, all experiments are conducted in the same environment. The computational complexity of the two prediction heads is 7.18 GFLOPs, and the number of parameters is only 7.35 K.

3.1 Performance of new MS Lesion Segmentation

The proposed CIB design is primarily evaluated on new-lesion segmentation with performance metrics following [23]. Here, F1 is used to assess the detection capability of MS lesion instances [4]. Table 2 shows that our method significantly enhances new-lesion segmentation performance, outperforming CoactSeg by 4% in Dice and 12.82% in F1. Furthermore, comparative analysis with individual human annotations (from the publicly available MSSEG-2 dataset [3]) validates the superior performance of our approach. Meanwhile, applying our CIB design to SNAC [14] and Neuropoly [13] also yields gains in Dice and F1 scores, indicating the general applicability of our method across diverse architectures.

Table 2. Comparisons of new MS lesion segmentation on MSSEG-2. The human experts' performance is shown based on their individually annotated results. Surpassed human metrics are denoted in grey.

Method	Dice(%)↑	Jaccard(%)↑	95HD(voxel)↓	ASD(voxel)↓	F1(%)↑
SNAC [14]	62.70	49.65	36.18	13.32	60.32
Neuropoly [13]	56.72	43.23	72.00	27.75	3.58
CoactSeg [23]	63.82	51.68	30.35	12.14	61.96
SNAC+CIB (Ours)	63.03	50.18	36.08	12.93	61.05
Neuropoly+CIB (Ours)	61.33	48.04	43.44	19.20	23.75
CoactSeg+CIB (Ours)	**67.83**	**54.22**	**21.84**	**9.58**	**74.78**
Human Expert #1	77.52	65.76	27.83	5.47	82.34
Human Expert #2	66.89	58.11	N/A	N/A	68.19
Human Expert #3	58.56	46.51	60.99	12.41	62.88
Human Expert #4	60.68	49.95	N/A	N/A	66.58

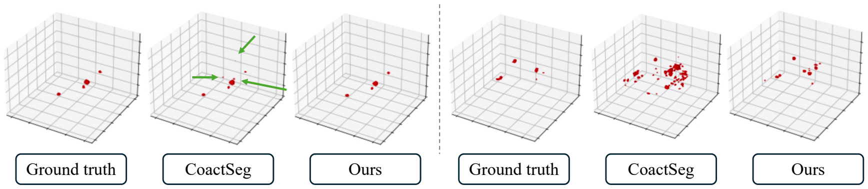

Fig. 4. Visual comparisons for new MS lesion segmentation between our proposed method and CoactSeg on MSSEG-2. Note that, most false-positive instances are successfully suppressed by our CIB framework.

Visualization Comparison. In Figure 4, we compare the new-lesion segmentation predictions of our approach with CoactSeg on the MSSEG-2 dataset. CoactSeg tends to generate false positive new lesions, a challenge effectively addressed by our method. On the left, our approach accurately identifies 4 new lesions, while CoactSeg generates several false positives (green arrows). The right side depicts a more challenging case, where CoactSeg produces numerous new lesions, whereas our approach significantly reduces false positives.

False Positives. We further report the False Discovery Rates (FDR, defined as FP/(FP+TP)) for CoactSeg and our model: 44.38% vs. 31.23%, demonstrating that our CIB design reduces FDR by 13%. Moreover, our model can be seamlessly integrated with the Tversky loss [19], specifically designed to address false positive and false negative issues. The combined model achieves a higher F1 score of 75.80%, yielding a 1.02% improvement in F1 and setting a new benchmark for new MS lesion segmentation on MSSEG-2.

3.2 Discussions

Ablation Study. Table 3 shows the ablation studies for each component of our proposed CIB in the new-lesion segmentation task on MSSEG-2, where “weight"

Table 3. Ablation studies for new MS lesion segmentation on MSSEG-2.

Weight	Patch head	Grid head	Dice(%)↑	Jaccard(%)↑	95HD(voxel)↓	ASD(voxel)↓	F1(%)↑
			63.82	51.68	30.35	12.14	61.96
	✓		63.81	50.99	29.71	11.87	68.07
	✓	✓	66.01	52.78	26.89	10.52	74.76
✓	✓		66.84	53.02	24.34	**9.17**	64.15
✓	✓	✓	**67.83**	**54.22**	**21.84**	9.58	**74.78**

refers to using the predicted performance to weigh each patch during inference. The findings reveal that: 1) Solely using the patch head for regularization improves performance across several metrics, and combining both the patch and grid heads as regularization further enhances performance across all metrics, indicating their contribution to segmentation quality. 2) Introducing the weighted calibration approach significantly improves segmentation performance across all metrics. 3) Utilizing the grid head for additional regularization at the grid level leads to a notable improvement in the F1 score.

Table 4. Performance of full MS lesion segmentation on MS23-v1 [23].

Method	Dice(%)↑	Jaccard(%)↑	95HD(voxel)↓	ASD(voxel)↓	F1(%)↑	FDR (%)-Medium↓	FDR (%)-Large↓
CoactSeg [23]	75.70	61.53	14.91	1.66	50.17	74.29	74.49
CoactSeg [23]+CIB	**76.22**	**61.76**	**7.87**	**0.88**	**52.20**	**72.97**	**62.48**

Full MS Lesion Segmentation. We further give the performance of full MS lesion segmentation on MS-23v1 [23] in Table 4. Implementing our CIB design with CoactSeg further improves performance by 0.52% in Dice and 2.03% in F1, highlighting the effectiveness of CIB for the full MS lesion segmentation task. Furthermore, we set 50 voxels as the threshold to separate MS lesions into medium and large categories following [15], and FDR can be reduced for both categories. Note that, tiny lesions (*i.e.*, < 3mm) are usually considered nonspecific and are filtered during inference to ensure fair comparisons [2,23].

Table 5. Comparisons by using different weighting strategies on MSSEG-2.

Method	Dice(%)↑	Jaccard(%)↑	95HD(voxel)↓	ASD(voxel)↓	F1(%)↑
Gaussian Filter [9]	64.08	50.81	30.77	12.29	66.16
CLS Weights	63.51	50.48	42.25	11.78	63.51
SEG Weights (Ours)	**67.83**	**54.22**	**21.84**	**9.58**	**74.78**

Weighting Strategies. Table 5 compares our proposed weighted method with the Gaussian filter method [9]. The results underscore the superiority of our approach. Unlike Gaussian filtering, which concentrates solely on the prediction at the center of the patch, our method acknowledges that lesions can appear at various locations. We also compare our calibration method using different types of context information. Here, we employ the classification performance (*i.e.,* is there a lesion or not) for the inter-patch blending, denoted as "CLS Weights". The results show that using the Dice score as the calibrated weights ("SEG Weights") is better than the classification-based one.

4 Conclusion

In this paper, we have presented a new sliding window blending strategy for multiple sclerosis lesion segmentation by introducing the segmentation performance as weights to calibrate the inter-patch blending. The proposed Calibrated Inter-patch Blending (CIB) framework is designed to predict the Dice scores for each input patch and smaller grids. During testing, the predicted Dice values are employed to adaptively weigh the patch predictions. Our comprehensive experiments have shown that our proposed CIB framework significantly enhances the performance of both all and new MS lesion segmentation tasks. Future work will include more statistical analysis [18,24,22].

Acknowledgments. This work was supported by the MRFF Grant, NCRI000074, "AI Precision Medicine for Multiple Sclerosis: Building Medical AI Capacity (AIMS)". We also appreciate the public repositories [23,14,13] and datasets [3].

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bai, L., Wang, D., Wang, H., Barnett, M., Cabezas, M., Cai, W., Calamante, F., Kyle, K., Liu, D., Ly, L., et al.: Improving multiple sclerosis lesion segmentation across clinical sites: A federated learning approach with noise-resilient training. Artificial Intelligence in Medicine **152**, 102872 (2024)
2. Basaran, B.D., Zhang, X., Matthews, P.M., Bai, W.: Seghed: Segmentation of heterogeneous data for multiple sclerosis lesions with anatomical constraints. arXiv preprint arXiv:2410.01766 (2024)
3. Commowick, O., Cervenansky, F., Cotton, F., Dojat, M.: Msseg-2 challenge proceedings: Multiple sclerosis new lesions segmentation challenge using a data management and processing infrastructure. In: MICCAI 2021. p. 126 (2021)
4. Commowick, O., Istace, A., Kain, M., Laurent, B., Leray, F., Simon, M., Pop, S.C., Girard, P., Ameli, R., Ferré, J.C., et al.: Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. Scientific Reports **8**(1), 13650 (2018)
5. Gessert, N., Bengs, M., Krüger, J., Opfer, R., Ostwaldt, A.C., Manogaran, P., Schippling, S., Schlaefer, A.: 4d deep learning for multiple-sclerosis lesion activity segmentation. In: MIDL 2020 (2020)
6. Gessert, N., Krüger, J., Opfer, R., Ostwaldt, A.C., Manogaran, P., Kitzler, H.H., Schippling, S., Schlaefer, A.: Multiple sclerosis lesion activity segmentation with attention-guided two-path cnns. Computerized Medical Imaging and Graphics **84**, 101772 (2020)

7. Gold, R., Kappos, L., Arnold, D.L., Bar-Or, A., Giovannoni, G., Selmaj, K., Tornatore, C., Sweetser, M.T., Yang, M., Sheikh, S.I., et al.: Placebo-controlled phase 3 study of oral bg-12 for relapsing multiple sclerosis. New England Journal of Medicine **367**(12), 1098–1107 (2012)
8. He, T., Zhao, W., Mao, Y., Wang, Y., Wang, L., Kuang, Q., Xu, J., Ji, Y., He, Y., Zhu, M., et al.: Ms or not ms: T2-weighted imaging (t2wi)-based radiomic findings distinguish ms from its mimics. Multiple Sclerosis and Related Disorders **61**, 103756 (2022)
9. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203—211 (2021)
10. Krishnan, A.P., Song, Z., Clayton, D., Jia, X., de Crespigny, A., Carano, R.A.: Multi-arm u-net with dense input and skip connectivity for t2 lesion segmentation in clinical trials of multiple sclerosis. Scientific Reports **13**(1), 4102 (2023)
11. Krüger, J., Opfer, R., Gessert, N., Ostwaldt, A.C., Manogaran, P., Kitzler, H.H., Schlaefer, A., Schippling, S.: Fully automated longitudinal segmentation of new or enlarged multiple sclerosis lesions using 3d convolutional neural networks. NeuroImage: Clinical **28**, 102445 (2020)
12. La Rosa, F., Abdulkadir, A., Fartaria, M.J., Rahmanzadeh, R., Lu, P.J., Galbusera, R., Barakovic, M., Thiran, J.P., Granziera, C., Cuadra, M.B.: Multiple sclerosis cortical and wm lesion segmentation at 3t mri: a deep learning method based on flair and mp2rage. NeuroImage: Clinical **27**, 102335 (2020)
13. Macar, U., Karthik, E.N., Gros, C., Lemay, A., Cohen-Adad, J.: Team neuropoly: Description of the pipelines for the miccai 2021 ms new lesions segmentation challenge. arXiv preprint arXiv:2109.05409 (2021)
14. Mariano, C., Yuling, L., Kain, K., Linda, L., Chenyu, W., Michael, B.: Estimating lesion activity through feature similarity: A dual path unet approach for the msseg2 miccai challenge. https://github.com/marianocabezas/msseg2
15. Nair, T., Precup, D., Arnold, D.L., Arbel, T.: Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. Medical Image Analysis **59**, 101557 (2020)
16. Péloquin, S., Schmierer, K., Leist, T.P., Oh, J., Murray, S., Lazure, P.: Challenges in multiple sclerosis care: results from an international mixed-methods study. Multiple Sclerosis and Related Disorders **50**, 102854 (2021)
17. Rondinella, A., Crispino, E., Guarnera, F., Giudice, O., Ortis, A., Russo, G., Di Lorenzo, C., Maimone, D., Pappalardo, F., Battiato, S.: Boosting multiple sclerosis lesion segmentation through attention mechanism. Computers in Biology and Medicine **161**, 107021 (2023)
18. Rondinella, A., Guarnera, F., Crispino, E., Russo, G., Di Lorenzo, C., Maimone, D., Pappalardo, F., Battiato, S.: Icpr 2024 competition on multiple sclerosis lesion segmentation—methods and results. In: ICPR 2024. pp. 1–16. Springer (2024)
19. Salehi, S.S.M., Erdogmus, D., Gholipour, A.: Tversky loss function for image segmentation using 3d fully convolutional deep networks. In: MLMI 2017. pp. 379–387. Springer (2017)
20. Sharmin, S., Bovis, F., Malpas, C., Horakova, D., Havrdova, E.K., Izquierdo, G., Eichau, S., Trojano, M., Prat, A., Girard, M., et al.: Confirmed disability progression as a marker of permanent disability in multiple sclerosis. European Journal of Neurology **29**(8), 2321–2334 (2022)
21. Tang, Z., Cabezas, M., Liu, D., Barnett, M., Cai, W., Wang, C.: Lg-net: lesion gate network for multiple sclerosis lesion inpainting. In: de Bruijne M. et al. (eds)

MICCAI 2021. pp. 660–669. Springer Cham (2021). https://doi.org/10.1007/978-3-030-87234-2_62

22. Wu, Y., Luo, X., Xu, Z., Guo, X., Ju, L., Ge, Z., Liao, W., Cai, J.: Diversified and personalized multi-rater medical image segmentation. In: CVPR 2024. pp. 11470–11479 (2024)
23. Wu, Y., Wu, Z., Shi, H., Picker, B., Chong, W., Cai, J.: Coactseg: Learning from heterogeneous data for new multiple sclerosis lesion segmentation. In: Celebi, M.E., et al. (eds) MICCAI 2023. vol. 14227, pp. 3–13. Springer Cham (2023). https://doi.org/10.1007/978-3-031-43993-3_1
24. Wu, Z., Wu, Y., Lin, G., Cai, J.: Reliability-adaptive consistency regularization for weakly-supervised point cloud segmentation. International Journal of Computer Vision **132**(6), 2276–2289 (2024)
25. Zeng, C., Gu, L., Liu, Z., Zhao, S.: Review of deep learning approaches for the segmentation of multiple sclerosis lesions on brain mri. Frontiers in Neuroinformatics **14**, 610967 (2020)
26. Zhan, G., Deng, J., Cabezas, M., Ouyang, W., Barnett, M., Wang, C.: Fed-cot: Co-teachers for federated semi-supervised ms lesion segmentation. In: Celebi, M.E., et al. (eds) MICCAI 2023. vol. 14393, pp. 357–366. Springer Cham (2023). https://doi.org/10.1007/978-3-031-47401-9_34
27. Zhang, H., Dana, K., Shi, J., Zhang, Z., Wang, X., Tyagi, A., Agrawal, A.: Context encoding for semantic segmentation. In: CVPR 2018. pp. 7151–7160 (2018)
28. Zhang, H., Nguyen, T.D., Zhang, J., Marcille, M., Spincemaille, P., Wang, Y., Gauthier, S.A., Sweeney, E.M.: Qsmrim-net: Imbalance-aware learning for identification of chronic active multiple sclerosis lesions on quantitative susceptibility maps. NeuroImage: Clinical **34**, 102979 (2022)
29. Zhang, L., Tanno, R., Xu, M., Huang, Y., Bronik, K., Jin, C., Jacob, J., Zheng, Y., Shao, L., Ciccarelli, O., et al.: Learning from multiple annotators for medical image segmentation. Pattern Recognition **138**, 109400 (2023)