This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

FluxCut: Shifting from Global Reconstruction to Local Geometry for Unsupervised Small Lesion Segmentation

Zhiwei Deng[1,2], Yonggang Shi[1,2,3] (✉)

[1] Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California (USC), Los Angeles, CA 90033, USA.
[2] Ming Hsieh Department of Electrical and Computer Engineering, Viterbi School of Engineering, University of Southern California (USC), Los Angeles, CA 90089, USA.
[3] Alfred E. Mann Department of Biomedical Engineering of Viterbi School of Engineering, University of Southern California (USC), Los Angeles, CA 90089, USA.
yonggang.shi@loni.usc.edu

Abstract. Detecting small lesions in brain Magnetic Resonance Imaging (MRI) is an important task for clinical early diagnosis, yet their subtle size and appearance restrict the feasibility of both manual annotation and automated segmentation. In addition, current reconstruction-based Unsupervised Anomaly Detection (UAD) methods try to model the global normal distribution, while suffering from the submerged local details and domain shifts between training and testing data. In this work, we propose *FluxCut*, a novel two-stage unsupervised framework that shifts from the global reconstruction mechanism to local geometric analysis. In the first stage, our network learns a general representation of small lesions that maximizes the flux measurement, the resulting flux map robustly enhances the small lesion locations. Then, we extract the localized patch features guided by this map and employ a learning-based graph-cut for fine-grained segmentation.This flux-guided segmentation paradigm encourages the network to focus on local discriminative features, preventing the subtle signatures of small lesions from being overwhelmed by background noise during global reconstruction-residual optimization. In our experiments, we compare with current state-of-the-art UAD methods in multiple 3D datasets to demonstrate that our method achieves significant improvement in all datasets.

Keywords: Unsupervised · Small Lesion · Anomaly Segmentation.

1 Introduction

Precise segmentation of small brain lesions is critical for early diagnosis and treatment for multiple diseases such as multiple sclerosis (MS) [7, 23], and white matter hyperintensities (WMH), which are highly associated with cognitive decline and cerebral vascular diseases [5]. Due to their inherently weak intensity and subtle boundaries, obtaining accurate annotations and applying supervised

learning to small lesion segmentation task is difficult [15]. In addition, supervised models can be easily biased by the extreme class imbalance between healthy tissues and sparse lesion voxels [25]. Therefore, unsupervised anomaly detection (UAD) has emerged as a promising alternative, leveraging an unbiased normative modeling paradigm.

Prevalent UAD approaches [3, 13, 17, 20, 24], typically models the health anatomy and identifies pathology as pixel-wise residuals between original scan and reconstructed '*normal*' scan. These methods are fundamentally fragile for small lesions due to the inherent assumption that anomalies will yield high residuals – a premise easily broken by small, low-contrast lesions. Small lesions usually possess limited fraction of voxels and present as slight local intensity variation compared to healthy tissue. Even with a well-trained reconstruction model, the residual of small lesion is still vulnerable to be overwhelmed by background noises arising from inevitable misalignment and global smoothing effects. In addition, reconstruction-based model particularly suffers from domain shifts. Since training is typically conducted on different cohorts, these models could generate systematic reconstruction artifacts on unseen domains, which will further obscure the subtle residual signals of small lesions. As discussed in [19], reconstruction-based models exhibit similar vulnerability to domain shifts. Consequently, simple histogram equalization followed by intensity thresholding can often outperform these learning-based UAD methods across multiple datasets. This finding supports our observation that subtle lesions cannot be reliably captured by global reconstruction-residual-based normative modeling. However, alternative non-reconstruction UAD approaches [16, 21] also remain sub-optimal for small lesions. Transformer-based models [21] rely on latent compression that inherently eliminates the high-frequency signatures of micro-lesions. Furthermore, Cycle-UNet [16] strictly require registered multi-modal data, as slight misalignment generates structural false positives that completely mask small pathologies.

To bridge the small lesion segmentation gap, in this paper, we shift the UAD paradigm from global reconstruction to local geometric modeling and propose a novel unsupervised framework requiring no additional training data, *FluxCut*. Following the classical R-CNN pipeline [9], we decompose the segmentation task into two stages: detection and segmentation. However, instead of relying on supervised bounding boxes and pixel-wise annotations, we leverage flux estimation for lesion localization and an unsupervised graph-cut formulation for mask generation. Overall, the contribution of this paper is three-fold: (1) We propose FluxCut, a novel UAD framework that fundamentally shifts the paradigm from global intensity reconstruction to local geometric modeling; (2) We explicitly formulate lesion detection through local flux estimation via a novel adaptively weighted flux objective that tailored for small lesions. (3) We integrate an unsupervised graph-cut module to segment precise lesion boundaries without any annotation. Comprehensive experiments show promising improvements in compared with previous UAD approaches.

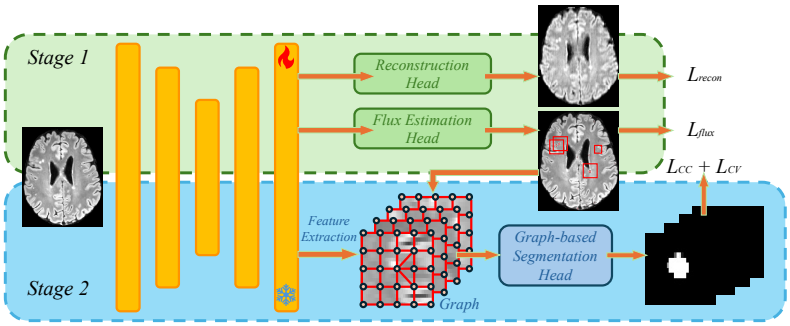

Fig. 1: **Network.** In Stage 1, a shared backbone drives the Reconstruction Head (RH) and Flux Estimation Head (FEH) to model boundary-aware geometric representations. In Stage 2, we freeze the backbone and feed the flux-guided lesion patches into the Graph-based Segmentation Head (GSH) to optimize a correlation clustering objective.

2 Method

As shown in Figure 1, FluxCut is a shared backbone network for feature extraction, augmented with three task-specific heads: a Reconstruction Head (RH), a Flux Estimation Head (FEH) and a Graph-based Segmentation Head (GSH). In stage 1, the shared backbone, the RH and the FEH are jointly trained to model the local geometric representations for every pixel by maximizing projected gradients along lesion boundaries. In stage 2, we crop local lesion patches using the flux map from stage 1 and pass them into the GSH to perform fine-grained segmentation. The segmentation network is optimized using a deep correlation clustering (CC) objective combined with a Chan-Vese regularization term to preserve both cohesive masks and accurate boundaries. Next we describe the two stages in detail.

2.1 Adaptively Weighted Flux Estimation

As discussed above, to robustly detect the small lesion with low intensity, we design a Flux estimation head (FEH) that captures the subtle boundary gradient changes. Inspired by [6], we formulate a generalized flux measurement tailored for irregular lesions. Specifically, FEH estimates voxel-wise flux responses from the original image domain, where abnormal local gradient variations around lesion boundaries are expected to produce distinctive flux patterns. The resulting flux map highlights potential lesion candidate regions for subsequent segmentation.

Formally, given a 3D image domain $\Omega \subset \mathbb{R}^3$ and an input volume $I : \Omega \to \mathbb{R}$, the FEH predicts K spatial scalar fields $\mathcal{R} = \{r_1, \ldots, r_K\}$. For each voxel $x \in \Omega$, $\mathcal{R}(x)$ explicitly estimates its distance to the lesion boundary along K uniformly sampled directions $\overrightarrow{d_i}$ on the unit sphere $\mathbb{S}^2$. Based on $\mathcal{R}(x)$, we tailor

a discretized flux measure as follows:

$$f(x, R(x)) = -\frac{1}{K}\sum\nolimits_i v(x + \overrightarrow{h_i}) \cdot \overrightarrow{d_i} \tag{1}$$

where $v(\cdot)$ is the image gradient function, $\overrightarrow{h_i} = r_i(x) \cdot \overrightarrow{d_i}$ and $r_i(x) \in R(x)$. Although estimating multiple radius on different directions successfully breaks the spatial symmetry to enable the model to fit irregular lesion boundaries, we observe that the simple average strategy on direction response remain vulnerable to unilateral edges. More specifically, the extreme gradient magnitudes invoked by sudden intensity changes, such as ventricle boundaries, can dominate the overall average responses, causing falsely high f response, which might overwhelm the flux response of small lesions with weak gradients.

To address this limitation, we propose an adaptive weighting scheme into Equation (1) to dynamically averaging the directional flux. Specifically, as presented in Equation (2) we define a directional asymmetry term, computed as the absolute difference between the flux responses of diametrically opposite sampling directions $\overrightarrow{d_i}$ and $\overrightarrow{d_j}$, where j is the index such that $\overrightarrow{d_i} = -\overrightarrow{d_j}$.

$$\Delta_i = |v(x + \overrightarrow{h_i}) \cdot \overrightarrow{d_i} - v(x + \overrightarrow{h_j}) \cdot \overrightarrow{d_j}| \tag{2}$$

With this asymmetry term, we weight the flux measure as

$$f_{AW}(x, R(x)) = -\frac{1}{K}\sum\nolimits_i exp(-\gamma\Delta_i) \cdot v(x + \overrightarrow{h_i}) \cdot \overrightarrow{d_i} \tag{3}$$

where γ is a scaling factor. This adaptive weighting scheme preserves high sensitivity to the bounded, cohesive gradients of even the smallest anomalies, while heavily penalizing unilateral edges.

Equipped with this adaptively weighted flux measure, we jointly train the shared backbone network alongside the Reconstruction Head (RH) and the Flux Estimation Head (FEH). Specifically, the FEH is trained to localize the accurate boundaries of irregular lesions that maximizes the average flux response over the whole image (as L_{flux} shown in Equation (4)), while the RH makes sure that the representations extracted by the shared backbone is semantically meaningful. This combination not only trains the network to isolate anomalies but also to generate good feature maps, which are crucial for the downstream segmentation task in next stage. Formally, the loss functions for the first stage training are computed as the combination of reconstruction loss $L_{recon} = MSE(\hat{I}, I)$ and flux loss, as follows:

$$\begin{aligned} L_{flux} &= -\frac{1}{N}\sum_{x\in\Omega} f_{AW}(x, R(x)) \\ L_{stage1} &= \lambda_1 L_{flux} + \lambda_2 L_{recon} \end{aligned} \tag{4}$$

where N is the total number of pixels and λ_1, λ_2 correspond to the coefficients of each objective.

2.2 Graph-based Lesion Segmentation

Automated Local Patch Extraction. We can obtain two fundamental components from the joint optimization in stage 1: the flux map that highlighted the lesion locations, and the feature maps from the shared backbone. Since our flux response is computed based on the lesion boundary gradients, which usually fluctuates due to the the inherent background noises and contrast variability of MRI scan, leading to non-uniform flux responses inside the lesions. Therefore, applying simple threshold as in [6], might lead to incomplete and fragmented segmentation results. For small lesions, such remaining sparse mask risks being filtered as noise during possible post-processing and causes missed detection. In addition, the loss of even a few constituent pixels severely compromises the morphological integrity for small lesions.

To prevent the subtle small lesion features from being submerged by massive background and maintain their precise boundary, we propose a local patch-based segmentation strategy. Specifically, we first straightforwardly apply a conservative percentile-based threshold to the flux map to make sure the lesion cores with high responses are identified. Based on these seed regions and simple connected component analysis as in [19], we then construct bounding boxes with sufficient spatial margin to ensure that the weakly responding peripheral boundaries are fully captured. In addition, overlapping candidate bounding boxes are merged according to their IoU overlap ratio. These bounding boxes are used as spatial anchors to perform an automated ROI-Align [9] operation to extract localized feature patches from the shared backbone.

Graph Construction. Within these extracted local ROIs, we reformulate lesion segmentation task as a graph partitioning problem. The cropped feature patch is modeled as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where nodes represent voxels and edges encode pairwise semantic affinities derived from deep features. Formally, given a local feature patch $\mathcal{F} \in \mathbb{R}^{C\times D\times H\times W}$ extracted from the frozen backbone, where C is the number of channels, and D, H, W are the spatial sizes. We flatten this spatial tensor into a set of node features $Z = \{z_1, z_2, \ldots, z_N\}$, where N is the total number of pixels in the patch, and $z_i \in \mathbb{R}^C$ denotes the deep semantic embedding of the i-th pixel. To capture the intrinsic semantic correlation, we define the affinity matrix W with the cosine similarity as follows:

$$W_{ij} = \frac{z_i \cdot z_j}{\|z_i\|_2 \|z_j\|_2} \tag{5}$$

Correlation Clustering. Our goal is to cluster the graph $\mathcal{G}$ into two disjoint sets: the lesion foreground and the background. To learn the best partition of graphs, we train the GSH to output the assignment matrix $S = GSH(\mathcal{F})$ using the extracted features. $S \in \mathbb{R}^{N\times 2}$ contains vectors representing the node's probability of belonging to a particular cluster. The segmentation head is optimized with the correlation clustering (CC) loss [1, 2] as follows:

$$L_{CC} = -Tr(WSS^T) \tag{6}$$

To prevent the clustering process from falling into trivial solutions, we also add a Chan-Vese level-set loss [14] to enforce the intensity homogeneity within each cluster based on the original image patch P as follows:

$$c_k = \frac{S_k^T P}{\mathbf{1}^T S_k + \epsilon}; \qquad L_{CV} = \sum_{k=1}^{2} S_k^T (P - c_k \mathbf{1})^2 \tag{7}$$

where c_k denotes the average image intensity of k-th cluster and $S_k \in \mathbb{R}^N$ is the pixel assignment of k-th cluster. So, the overall objectives of the second stage is expressed as

$$L_{stage2} = L_{CC} + \lambda_3 L_{CV} \tag{8}$$

where λ_3 is the coefficients of L_{CV}.

3 Experimental Results

Datasets. We use 3 public datasets containing small lesions to evaluate our model. MSLUB [15] provides 3D brain FLAIR scans from 30 patients with static MS. MICCAI-21 MS new lesion segmentation challenge dataset (MSSEG-2) [4] consists of 40 patients with longitudinal FLAIR scans of two time-points, it focuses on the detection and segmentation of new MS lesions. BrainMetShare [8] contains 105 subjects with brain metastases in T1GD modality.

For fair comparison with reconstruction-based baselines, we select 589 cognitive normal subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [12] dataset and use their FLAIR images as the 'healthy' cohort only for baseline training. Same pre-processing steps are applied on all datasets, including skull-stripping [10], N4-correction [26], and Z-score based normalization. In addition, MSSEG-2 dataset are further processed according to official pipeline.

Table 1: Performance on Different Datasets

	MSLUB				MSSEG-2				BrainMetShare			
Method	DSC	Sens.	Prec.	AP	DSC	Sens.	Prec.	AP	DSC	Sens.	Prec.	AP
Thresh	32.5	44.8	30.2	27.5	7.4	31.4	5.2	3.2	37.1	32.2	49.8	29.7
f-AnoGAN	26.8	29.5	28.7	22.7	19.5	43.6	20.7	17.4	20.4	18.6	34.8	12.5
DDPM	29.6	33.2	30.4	24.2	24.9	46.2	24.2	22.2	22.1	21.3	42.5	19.7
AE	23.3	27.4	24.1	18.1	21.6	50.6	22.3	20.5	-	-	-	-
DAE	30.4	32.6	32.8	25.8	28.6	48.3	29.5	25.5	-	-	-	-
FluxCut	**36.3**	**47.2**	**33.5**	**31.9**	**38.9**	**52.3**	**35.3**	**36.0**	**41.9**	**35.2**	**58.4**	**34.7**

Wilcoxon signed-rank tests are conducted, statistical significant results are underlined.

Implementation Details. We evaluated our model's performance against multiple established UAD baselines. These encompass a conventional intensity-based

thresholding approach [19], alongside two 2D slice-level generative models—the GAN-based f-AnoGAN [24] and the diffusion-based DDPM [20]. Furthermore, we compare against two 3D volume-level reconstruction methods, specifically the standard Autoencoder (AE) [3] and the Denoising AE (DAE) [13]. We do not include fully supervised methods, such as U-Net [22], nnU-Net [11], or recent MSSEG-2 pipelines [18], as main baselines because they rely on voxel-level lesion annotations or supervised downstream training. In contrast, FluxCut is designed for unsupervised anomaly detection without voxel-level lesion labels. We therefore focus on UAD baselines under the same supervision setting and regard supervised methods as references or upper bounds. Due to the lack of healthy T1GD data, we only trained the 3D baselines on the ADNI dataset and evaluated on MSSEG-2 and MSLUB dataset. However, all image slices in ADNI dataset and slices without lesion annotations in BrainMetShare dataset are used to train 2D baselines. Notably for MSSEG-2, baselines subtract error maps between time-points, while FluxCut utilizes subtracted flux maps for feature extraction. In addition, for all experiments with BrainMetShare dataset, a soft-gating strategy is applied by weighting the flux map with FLAIR intensities, effectively filtering out physiological vessels.

We utilized U-Net as our shared backbone and all task-specific heads are implemented as two convolutional layers followed by one fully connected layer. We performed a 5-fold cross-validation on all 3 evaluation datasets, and utilized 10% of the training data as the validation set to tune the hyper-parameters of networks. Finally, we empirically set the hyper-parameters for Stage 1 as $K = 64$, $\lambda_1 = 1$, $\lambda_2 = 0.2$, and $\gamma = 100$. For the Stage 2 training, the balancing weight is set to $\lambda_3 = 0.3$. For our FluxCut models, the input image patch size is set to $64 \times 64 \times 64$ voxels. Subsequently, the localized lesion patch is extracted via ROI-Align and spatially resampled to a fixed resolution of $20 \times 20 \times 20$ to preserve effective spatial representation. All experiments were conducted with Pytorch on 4 NVIDIA A6000 GPUs with a learning rate of 1e-4. We report four evaluation metrics in Table 1: Dice, sensitivity, precision, and voxel-wise AUPRC (area under the precision–recall curve, briefed as AP).

Performance on Small Lesion Segmentation. We can clearly observe from Table 1 that FluxCut outperforms all other comparing methods by a significant margin for all datasets. To further verify the statistical reliability of these improvements, we conduct Wilcoxon signed-rank tests on per-case results between FluxCut and the strongest baseline in each dataset. FluxCut achieves statistically significant improvements in Dice score ($p < 0.003$), Precision ($p < 0.009$), and AUPRC ($p < 0.0001$), suggesting that the observed gains are consistent at the case level. Beyond the improvement in Dice scores, our model achieves higher AUPRC scores. This indicates a superior precision-recall trade-off, demonstrating that FluxCut can effectively minimize false positives even as sensitivity increases. Such performance is particularly crucial for the robust detection of small lesions in highly imbalanced datasets. Furthermore, it is noteworthy that simple thresholding remains superior to other learning-based baselines for large

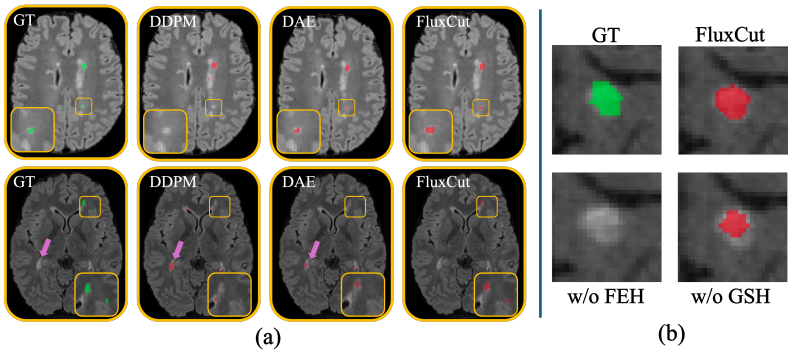

Fig. 2: **Qualitative comparison.** (a) presents the segmentation results of different models on MSSEG-2 dataset, showing FluxCut detects more small lesions while generating less false positives. (b) shows the ablation of each different modules with one image patch from BrainMetShare dataset

or high-contrast lesions, such as those in MSLUB and BrainMetShare. This observation supports the hypothesis in [19] that reconstruction-based models often attenuate distinctive pathological signals—treating them as noise to be smoothed—rather than modeling the actual lesion patterns. However, both traditional and reconstruction-based methods fail to detect the tiny, faint lesions in MSSEG-2. In contrast, our method fills this gap, showing a clear and significant improvement in cases where other models struggle. Visual comparisons in Figure 2 (a) demonstrate the advantages of FluxCut over the best-performing 2D and 3D baselines. While all models successfully identify prominent large-scale lesions,only FluxCut is able to detect the subtle small anomalies presented in the second example. Furthermore, our model generates fewer false positives (FP) than the other two counterparts as indicated by purple arrows. These visual observations validate that the quantitative improvements of FluxCut in Table 1 stem from accurate detection of small lesion and robust FP suppression.

Ablation Study. To investigate the contributions of each component in FluxCut, we performed ablation studies on MSSEG-2 dataset by removing each component of the network. The results are shown in Table 2, where we observe that removing the FEH causes a catastrophic DSC drop, highlighting it as the essential lesion locator. There are trade-offs between AWF and GSH indicating that AWF acts as a FP suppressor while GSH is a segmentation refiner as we developed in Section 2. Figure 2 (b) demonstrates that FEH is essential for lesion localization, while GSH contributes to accurate boundary delineation. Lastly, removing RH uniformly degrades performance, validating its role in enforcing structural homogeneity.

Furthermore, our ablation studies reveal that the number of estimated radii K and the output resolution of the ROI-Align operation (i.e., patch size) play pivotal roles in the final segmentation efficacy. Specifically, K dictates the geometric granularity of the flux estimation, whereas the ROI-Align pooling size

Table 2: Ablation Analysis of FluxCut. AWF means adaptively weighted flux.

Methods	DSC	Sens.	Prec.	AP
FluxCut	38.9	52.3	35.3	36.0
w/o AWF	32.4	53.7	28.2	31.7
w/o RH	36.8	51.2	34.0	34.3
w/o FEH	21.5	23.9	20.1	14.8
w/o GSH	35.3	49.8	37.3	34.8

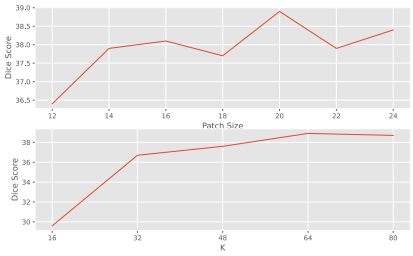

Fig. 3: Selection of patch size and K.

directly determines the topological connectivity of the constructed graph. As illustrated in Figure 3, we empirically established that a patch size of 20 and $K = 64$ yield the optimal balance between precise boundary fitting and stable graph topology, thereby maximizing segmentation performance.

4 Conclusion

In this paper, we propose FluxCut, a novel unsupervised framework for small lesion segmentation from 3D MRI data. Our method is designed to address existing challenges in small lesion in a reconstruction-regularized, flux-guided and local geometry-based paradigm, while reconstruction-based networks usually neglect the subtle local features of small lesions. In comparison to previous established baselines, we demonstrated that our method is able to achieve superior performance on zero-shot lesion delineation across highly heterogeneous multi-center datasets. For future work, we will apply it to broader clinical cohorts and modalities to examine its power for early-stage diagnosis.

Acknowledgments. This work was supported by the National Institutes of Health (NIH) under grants RF1AG077578, RF1AG064584, R01EB022744, RF1AG084072, RF-1AG072490, U19AG078109, P30AG066530 and S10OD032285.

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

References

1. Aflalo, A., Bagon, S., Kashti, T., Eldar, Y.: Deepcut: Unsupervised segmentation using graph neural networks clustering. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 32–41 (2023)
2. Bansal, N., Blum, A., Chawla, S.: Correlation clustering. Machine learning **56**(1), 89–113 (2004)
3. Baur, C., Denner, S., Wiestler, B., Navab, N., Albarqouni, S.: Autoencoders for unsupervised anomaly segmentation in brain mr images: a comparative study. Medical image analysis **69**, 101952 (2021)

4. 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 2021-24th international conference on medical image computing and computer assisted intervention. p. 126 (2021)
5. Debette, S., Markus, H.: The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. Bmj **341** (2010)
6. Deng, Z., Xu, S., Zhang, J., Zhang, J., Wang, D.J., Yan, L., Shi, Y.: Shape-aware 3d small vessel segmentation with local contrast guided attention. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 354–363. Springer (2023)
7. Filippi, M., Rocca, M.A., Ciccarelli, O., De Stefano, N., Evangelou, N., Kappos, L., Rovira, A., Sastre-Garriga, J., Tintorè, M., Frederiksen, J.L., et al.: Mri criteria for the diagnosis of multiple sclerosis: Magnims consensus guidelines. The Lancet Neurology **15**(3), 292–303 (2016)
8. Grøvik, E., Yi, D., Iv, M., Tong, E., Rubin, D., Zaharchuk, G.: Deep learning enables automatic detection and segmentation of brain metastases on multisequence mri. Journal of Magnetic Resonance Imaging **51**(1), 175–182 (2020)
9. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)
10. Hoopes, A., Mora, J.S., Dalca, A.V., Fischl, B., Hoffmann, M.: Synthstrip: skull-stripping for any brain image. NeuroImage **260**, 119474 (2022)
11. 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). https://doi.org/10.1038/s41592-020-01008-z
12. Jack Jr, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., L. Whitwell, J., Ward, C., et al.: The alzheimer's disease neuroimaging initiative (adni): Mri methods. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine **27**(4), 685–691 (2008)
13. Kascenas, A., Pugeault, N., O'Neil, A.Q.: Denoising autoencoders for unsupervised anomaly detection in brain mri. In: International Conference on Medical Imaging with Deep Learning. pp. 653–664. PMLR (2022)
14. Kim, Y., Kim, S., Kim, T., Kim, C.: Cnn-based semantic segmentation using level set loss. In: 2019 IEEE winter conference on applications of computer vision (WACV). pp. 1752–1760. IEEE (2019)
15. Kuijf, H.J., Biesbroek, J.M., De Bresser, J., Heinen, R., Andermatt, S., Bento, M., Berseth, M., Belyaev, M., Cardoso, M.J., Casamitjana, A., et al.: Standardized assessment of automatic segmentation of white matter hyperintensities and results of the wmh segmentation challenge. IEEE transactions on medical imaging **38**(11), 2556–2568 (2019)
16. Liang, Z., Anthony, H., Wagner, F., Kamnitsas, K.: Modality cycles with masked conditional diffusion for unsupervised anomaly segmentation in mri. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 168–181. Springer (2023)
17. Liang, Z., Guo, X., Noble, J.A., Kamnitsas, K.: Itermask 2: Iterative unsupervised anomaly segmentation via spatial and frequency masking for brain lesions in mri. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 339–348. Springer (2024)

18. Macar, U., Enamundram, N.K., 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). https://doi.org/10.48550/arXiv.2109.05409
19. Meissen, F., Kaissis, G., Rueckert, D.: Challenging current semi-supervised anomaly segmentation methods for brain mri. In: International MICCAI brainlesion workshop. pp. 63–74. Springer (2021)
20. Pinaya, W.H., Graham, M.S., Gray, R., Da Costa, P.F., Tudosiu, P.D., Wright, P., Mah, Y.H., MacKinnon, A.D., Teo, J.T., Jager, R., et al.: Fast unsupervised brain anomaly detection and segmentation with diffusion models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 705–714. Springer (2022)
21. Pinaya, W.H., Tudosiu, P.D., Gray, R., Rees, G., Nachev, P., Ourselin, S., Cardoso, M.J.: Unsupervised brain imaging 3d anomaly detection and segmentation with transformers. Medical Image Analysis **79**, 102475 (2022)
22. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer (2015). https://doi.org/10.1007/978-3-319-24574-4_28
23. Sastre-Garriga, J., Pareto, D., Battaglini, M., Rocca, M.A., Ciccarelli, O., Enzinger, C., Wuerfel, J., Sormani, M.P., Barkhof, F., Yousry, T.A., et al.: Magnims consensus recommendations on the use of brain and spinal cord atrophy measures in clinical practice. Nature Reviews Neurology **16**(3), 171–182 (2020)
24. Schlegl, T., Seeböck, P., Waldstein, S.M., Langs, G., Schmidt-Erfurth, U.: f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. Medical image analysis **54**, 30–44 (2019)
25. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: International Workshop on Deep Learning in Medical Image Analysis. pp. 240–248. Springer (2017)
26. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4itk: improved n3 bias correction. IEEE transactions on medical imaging **29**(6), 1310–1320 (2010)