Robust Multiple Sclerosis Lesion Inpainting with Edge Prior

Huahong Zhang[1], Rohit Bakshi[2], Francesca Bagnato[3], and Ipek Oguz[1(✉)]

[1] Vanderbilt University, Nashville, TN 37235, USA
{huahong.zhang,ipek.oguz}@vanderbilt.edu
[2] Brigham and Women's Hospital, Boston, MA 02115, USA
[3] Vanderbilt University Medical Center and Nashville VA Medical Center, Nashville, TN 37212, USA

Abstract. Inpainting lesions is an important preprocessing task for algorithms analyzing brain MRIs of multiple sclerosis (MS) patients, such as tissue segmentation and cortical surface reconstruction. We propose a new deep learning approach for this task. Unlike existing inpainting approaches which ignore the lesion areas of the input image, we leverage the edge information around the lesions as a prior to help the inpainting process. Thus, the input of this network includes the T1-w image, lesion mask and the edge map computed from the T1-w image, and the output is the lesion-free image. The introduction of the edge prior is based on our observation that the edge detection results of the MRI scans will usually contain the contour of white matter (WM) and grey matter (GM), even though some undesired edges appear near the lesions. Instead of losing all the information around the neighborhood of lesions, our approach preserves the local tissue shape (brain/WM/GM) with the guidance of the input edges. The qualitative results show that our pipeline inpaints the lesion areas in a realistic and shape-consistent way. Our quantitative evaluation shows that our approach outperforms the existing state-of-the-art inpainting methods in both image-based metrics and in FreeSurfer segmentation accuracy. Furthermore, our approach demonstrates robustness to inaccurate lesion mask inputs. This is important for practical usability, because it allows for a generous over-segmentation of lesions instead of requiring precise boundaries, while still yielding accurate results.

Keywords: Multiple sclerosis · Deep learning · Inpainting

1 Introduction

Multiple Sclerosis (MS) is a common autoimmune disease of the central nervous system characterized by focal demyelinating lesions visible on brain MRI. In the long term, the clinical progression of the disease is closely linked to brain atrophy, which makes the estimation of brain atrophy an important tool for longitudinal monitoring [11]. In order to measure the GM and WM atrophy, algorithms such

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-59861-7_13) contains supplementary material, which is available to authorized users.

M. Liu et al. (Eds.): MLMI 2020, LNCS 12436, pp. 120–129, 2020.
https://doi.org/10.1007/978-3-030-59861-7_13

as brain tissue classification and cortical surface reconstruction/thickness measurement are necessary. However, the presence of MS lesions is problematic for most automated algorithms and can lead to biased morphological measurements. To mitigate this problem, many lesion inpainting methods have been proposed [2–4,6,9,12–14,17,18]. These methods, whether conventional or deep-learning-based, consider the areas of lesions as "missing" during inpainting. Discarding the original lesion voxels is a common practice in MS lesion inpainting, and is consistent with the inpainting techniques in computer vision. Nevertheless, much structural information is lost when the lesion voxels are discarded. As a consequence, the algorithms are highly likely to produce inaccurate predictions if the lesion masks cross the tissue boundaries.

In this work, we re-consider the difference between the MS lesion inpainting and the general-purpose inpainting task in computer vision. We hypothesize that, specifically for MS lesion refilling (and potentially for other medical image inpainting tasks), the areas that require inpainting are not "missing" but rather contain information useful for inpainting and, thus, should not be discarded. Even though the lesions have different intensities than normal-appearing brain tissue, the structural information around lesions can and should be preserved. Therefore, we introduce the edge information from the input images as a prior.

Contributions. The main contributions of this paper are summarized below:

- Instead of considering the lesion areas as missing, we use the edge information extracted from the input image as a prior to guide the inpainting. This is more appropriate for MS inpainting as it preserves the shape and boundary of normal-appearing tissue around the lesions in the inpainted images. Different from [10] which infers the edges with an extra network, we believe the input image itself already contains the information for constraining the inpainting.
- Our method makes no assumptions about the characteristics of the lesions, which makes it suitable for both white matter lesions and gray matter lesions. This could also allow for the generalization to different modalities.
- We demonstrate that the proposed method can handle highly inaccurate lesion masks, which makes it robust to errors in lesion delineations. This reduces the dependence on highly accurate lesion segmentation, which is an elusive goal even for experts. Also, while incorrect lesion masks can cause artifacts for some methods such as FSL [2] and SLF [17], our method does not create such artifacts.
- The introduction of edge information as a prior to inpainting algorithms is architecture-agnostic. It can be readily used in conjunction with other deep learning algorithms, which may further improve performance.

2 Methods

2.1 Datasets and Pre-processing

We use an in-house dataset of 15 healthy controls and 40 patients with relapsing-remitting MS (RRMS) or secondary progressive MS (SPMS). At Brigham and

Women's Hospital, 3T T1-w MPRAGE and FLAIR images were acquired for each subject, at 1 mm^3 isotropic resolution. FLAIR images were co-registered to T1-w space [1]. All were skull-stripped with BET [15], followed by N4ITK [16].

In this paper, we refer to the healthy controls as original healthy control images (OHC) and the generated images as simulated lesion images (SL, see Sect. 3), in contrast to real lesion images (RL) from MS patients.

We use the original healthy control images (without any lesions, real or simulated) in the training phase. This forces the network to only learn the appearance of healthy tissue. In the test phase, we use the simulated lesion images (see Sect. 3) for both the qualitative and quantitative experiments; we also present qualitative results from the real lesion images.

2.2 Edge Detection

Due to the lack of ground truth pairs for training, a simple translation between the lesion and lesion-free images, e.g. using pix2pix [7], cannot be performed. Instead, the current inpainting methods consider the lesion areas as "missing" and train the network to infer the masked regions based on their surrounding areas. This idea is straightforward but causes the network to, by construction, ignore the lesion areas in the test phase, which means all the structure or shape information within the lesion masks is lost. Instead, our goal is to preserve such information in the test phase, while making no assumption on lesion appearance (e.g., a specific intensity profile relative to surrounding tissue) during training.

We use the classical Canny edge detector to extract the edge information from the lesion image. In Fig. 1, A-1 is a RL image, and A-2 is the corresponding edge map ($\sigma = 0.8$ for Gaussian filter, first step of edge detection). The boundaries between different brain tissues remain visible even when the lesions are present.

In the training phase, we use only OHC images. We provide random ROIs as the lesion mask, and the network has to learn to reconstruct the original T1w image. However, this means all the edges from the input will be preserved in the network output for the training set, whereas for the testing set, it is desirable that the edges from the lesion are ignored while the edges from tissue (e.g., WM/GM boundary) will be preserved. To alleviate this discrepancy, and to teach the model to ignore certain edges while preserving others, we use augmentations of the input edges during the training phase. This is illustrated in Fig. 1, where B-1 is a slice from OHC image, and B-2 is the corresponding edge map. By adding Gaussian noise to B-1 and detecting the edges from it, we get B-3, which contains some random edges compared to (B-2). We hypothesize that, with these augmentations, the network will learn to deal with the edges caused by the lesions.

2.3 Network Structure and Loss Functions

We adopted the network architecture described in [10] for MS inpainting. This network follows the encoder-decoder structure as shown in Fig. 2. Similar to the 2.5D methods in [19], the training input is the cropped 128×128 2D slices from

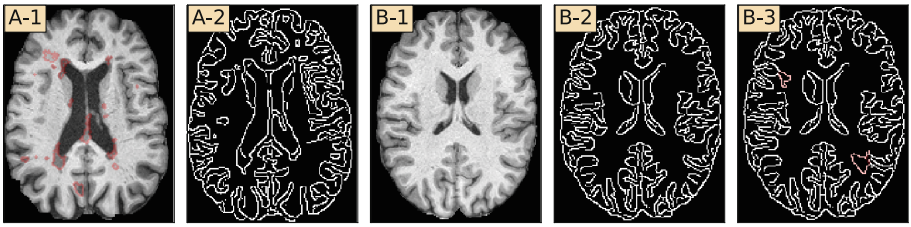

Fig. 1. Edge maps. (A-1) T1-w image with real lesions (RL) outlined in red; (A-2) Edges from A-1. (B-1) T1-w image of OHC; (B-2) Edges from B-1; (B-3) Edges after adding Gaussian noise to B-1. Orange pixels in (B-3) mark the newly generated edges. (Color figure online)

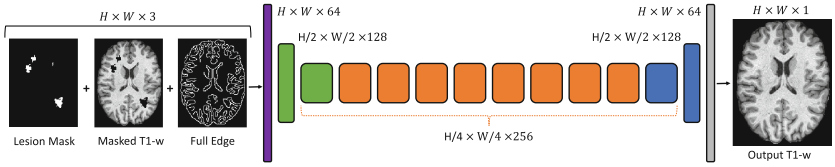

Fig. 2. Overview of the network architecture. The input is the concatenation of the binary lesion mask, the masked T1-w and edge detection of T1-w after adding random noise. The output is the inpainted T1-w image. Purple: 7×7 Convolution-SpectralNorm-InstanceNorm-ReLU; Green: 4×4 Convolution-SpectralNorm-InstanceNorm-ReLU layer, $stride = 2$ for down-sampling; Orange: residual blocks; Blue: similar to green, up-sampling; Gray: scaled $tanh$. (Color figure online)

all the 3 orientations and the test output is the consensus of the prediction from axial, coronal and sagittal views.

Our loss function is defined as $\mathcal{L}_{total} = \mathcal{L}_{rec} + 250\mathcal{L}_{style} + 0.1\mathcal{L}_{perceptual}$, where $\mathcal{L}_{rec}$ is the reconstruction loss, i.e., the pixel-wise L1 distance between the ground truth and the output image. $\mathcal{L}_{perceptual}$ and $\mathcal{L}_{style}$ are the perceptual loss and content loss introduced by [8], and they capture the difference of high-level features instead of raw images. We use perceptual loss and content loss to help the network generate realistic inpainting results with natural texture.

3 Experiments

Ground truth. Unlike lesion segmentation, which usually has gold standard lesion delineations, there is no gold standard for pairs of T1-w images with lesions and the corresponding lesion-free images. We thus use a lesion simulator on OHCs to evaluate in-painting methods: 1) the OHC image serves as 'ground truth' for what the corresponding inpainting result should look like, and 2) the tissue classification of the OHC image serves as 'ground truth' for the tissue classification of the corresponding inpainted images.

With our in-house dataset, we split 4/5 of the OHC subjects into the training set and 1/5 into the test set. We then obtained 11 SL-OHC pairs for the test set by simulating multiple SL images for each OHC image.

Using a publicly available lesion simulator[1], several lesion images with different lesion size, location and volume were generated for each healthy control. The lesion load followed a log-normal distribution $Log - N(log(15), \frac{log(3)}{3})$.

Implementation Details. For training, we use an Adam optimizer with a momentum of 0.5 and the initial learning rate (lr) of 0.0002. We use the initial learning rate for 100 epochs and let it decay to 0 within another 100 epochs. For validation, we choose the model with best F1 score (see below) for each fold. During training, we perform random lesion mask dilation for data augmentation.

Evaluation Metrics. The performance is evaluated from two aspects:

- *Compare the two images in the image domain (i.e., synthesis error).*
  - We calculate the mean squared error (MSE) to measure the distance between the inpainted images and the original images. For a more intuitive presentation, we report PSNR, which is based on MSE.
  - We compare the edge detection result between the output image and the ground truth images. To take both precision and recall into consideration, the F_1 score is calculated from the binary edge maps.
- *Compare the brain segmentation results (i.e., volume error and tissue overlap).* The brain segmentation results will be affected by MS lesions, however, the lesion filling methods are expected to reduce this effect. We use FreeSurfer [5] for segmentation.
  - Absolute volume difference (AVD) for WM and GM is computed as AVD $= \frac{|V_o - V_{gt}|}{V_{gt}} \times 100\%$ where V_o and V_{gt} are the volume of a given class in the segmentation of the inpainting output and the ground truth, respectively.
  - The F1 score and Jaccard similarity coefficient (IoU) over all classes are also reported to provide an overall evaluation of the segmentations.

With the exception of PSNR, all the metrics are reported on the lesion neighborhoods to focus on the changes caused by lesions, since there are practically no differences away from the lesion ROIs. The neighborhoods are defined by the dilation of lesion masks (3D dilation using octahedron kernel, $k = 5$).

Compared Methods. We compared our approach with three existing inpainting algorithms on T1-w images: the lesion filling tool included in FSL library [2], the SLF toolbox [17], and the recently proposed non-local partial convolutions (NLPC) inpainting method which is deep-learning-based [18]. Since NLPC is not publicly available, to make a fair comparison, we kept everything the same as our methods for training and evaluating NLPC except the network architecture.

[1] https://github.com/CSIM-Toolkits/LesionSimulatorExtension.

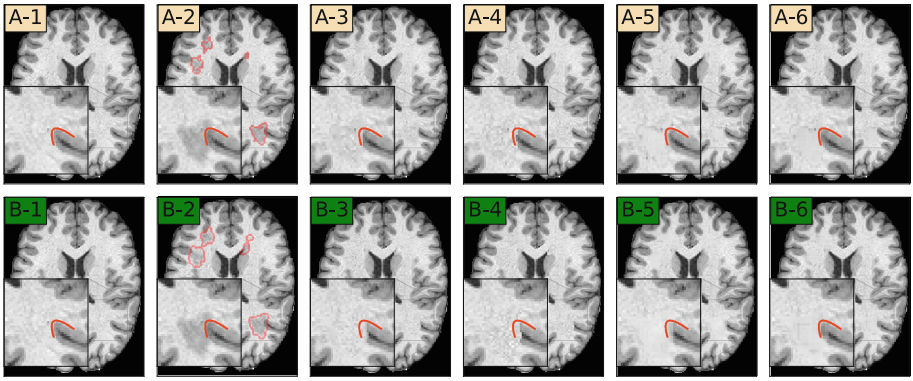

Fig. 3. Qualitative evaluation (simulated dataset). Group (A) uses "ground truth" lesion mask inputs. Group (B) uses dilated ($k = 3$) lesion mask inputs. (1) T1-w image of a healthy control; (2) T1-w image with lesions (lesion boundaries are marked in red); (3–6) Inpainted results using FSL, SLF, NLPC and our method, respectively. (Color figure online)

4 Results

Qualitative Analysis. The qualitative evaluation results are shown in Fig. 3.

In group (A), an OHC image (A1) and the corresponding lesion simulation results (A2) are shown. The inpainting results by FSL (A3), SLF (A4), NLPC (A5) and our proposed method (A6) are shown. Panel insets show a zoomed-in view near the lesion in temporal lobe. We note that SLF and FSL failed to preserve the brain structure of the sulcus by erroneously inpainting GM tissue. While for the deep-learning-based approaches, NLPC and the proposed method provide reasonable inpainting results. However, compared to NLPC, our method produced a more similar GM sulcus to the ground truth T1-w image.

Next, we explore the effect of inaccurate lesion delineations by dilating the input lesion masks. Group (B) contains the results filled with the dilated (k = 3) lesion masks. Similar results with dilation kernels k = 2 and k = 4 are presented in the appendix. In group (B), we can observe that SLF and FSL performance deteriorates because of the dilated masks. Additionally, because NLPC discards voxels within the lesion mask, it lacks adequate information and cannot preserve GM structure either. Our method, however, still presents results similar to ground truth and preserves the boundary between WM and GM.

In (A-5) and (A-6), we observe discontinuities between the inpainted regions and the surrounding tissue. This is due to the fusing of non-lesion regions of the input and the lesion regions of the output. Even though the lesion masks from the simulation are considered to be the 'ground truth', voxels outside the lesion mask can still contain abnormal intensities. This discontinuity disappears in group (B) since most abnormal voxels are included in the inpainted regions.

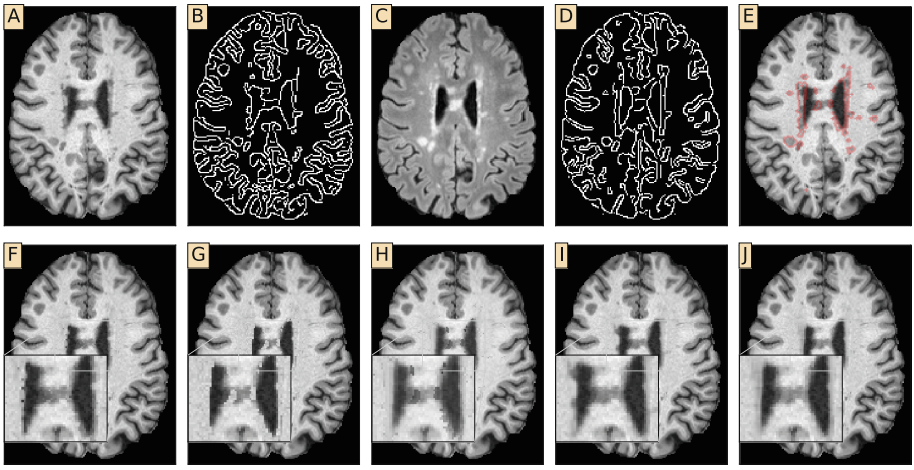

Fig. 4. Qualitative evaluation (real lesions). (A) T1-w image with real lesion; (B) Edge detection from A; (C) FLAIR image corresponding to (A); (D) Edge detection from C; (E) T1-w with segmented lesions (red pixels mark the boundary of lesions); (F–J) Inpainted results of FSL, SLF, NLPC, ours with T1-w edge and ours with FLAIR edge. (F–H) contain many small holes/artifacts, whereas our method avoids this problem. We note that the lesion between the lateral ventricles is missed by the lesion segmentation algorithm, and thus is not inpainted by any of the methods.

For some real lesion (RL) images, we observed that the lesions in the T1-w image can mislead the inpainting model by affecting the edges generated by the Canny edge detector. However, this can be solved by using the edge detection results from some other modalities (FLAIR, T2, etc) or their combinations. Qualitative results illustrating this extension are shown in Fig. 4. In this example, the lesion segmentation is obtained using [19]. This figure also contains examples of the artifacts generated by other inpainting methods. Our method provides the smoothest and most accurate boundaries after inpainting.

Quantitative Analysis. In Fig. 5, we show boxplots of the evaluation metrics described above. To demonstrate the effect of inaccurate lesion delineations, the x-axes are the dilation kernel sizes. We also report the same metrics for the simulated lesion images (without any inpainting) as a baseline.

In the image domain, our method has the best PSNR with all k values; its PSNR also decreases much slower than the other methods with increasing k-value. Even when $k = 4$, our method showed better PSNR than the simulated images. For the F1 score of the edge detection results, our method again decreases with the smallest slope among the methods. It falls under the lines of simulated images when $k = 4$, and we believe this is due to the variability of edge detection. The superior performance of our method on the other metrics supports this.

More importantly, for the metrics of segmentation results using FreeSurfer, our methods performed much better than the comparing methods. For the AVD

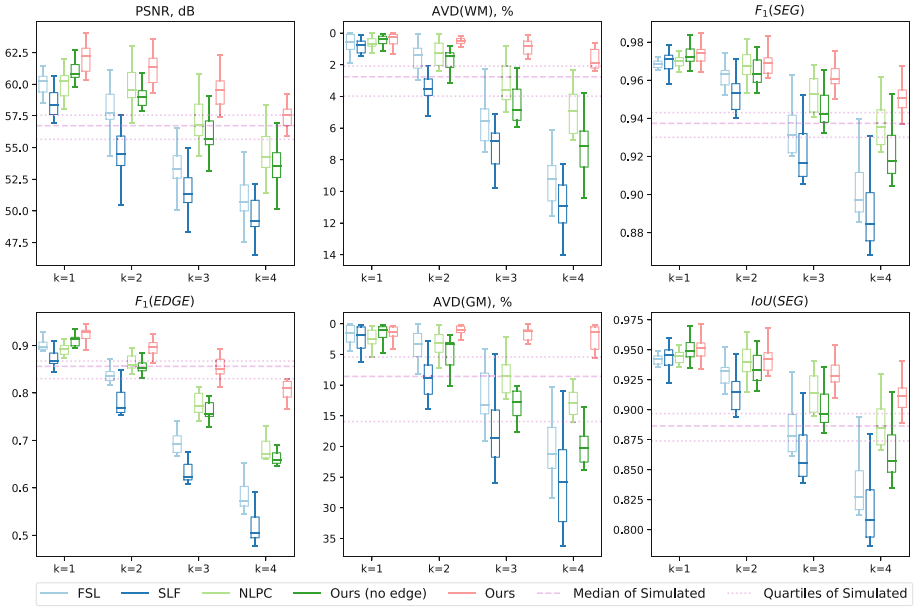

Fig. 5. Quantitative evaluation (simulated dataset). We report results for "ground truth" lesion mask input (no dilation, k = 1) and dilated lesion mask input (k = 2,3,4). The dotted lines are the medians and the quartiles of the simulated images.

of WM and GM, our method outperformed all other in-painting methods and the simulated images even when k = 4, which is not the case for any of the other methods. Interestingly, we noticed that the GM segmentation performance of our method when k = 2 is better than k = 1. We believe this is due to the discontinuity between the original lesion masks and the hypo/hyper-intense areas as discussed in the qualitative analysis. The F1 and Jaccard score over all the classes also prove the stable performance of our algorithm. Since our approach demonstrates stable performance, it allows the raters to segment in a more "aggressive" way, by deliberately oversegmenting instead of attempting to delineate precise boundaries, and our algorithm can ensure the performance of the inpainting. This can also be done by segmenting as usual and dilating the masks before inpainting.

As an ablation study, the results of our method without edge priors are also provided in Fig. 5. The no-edge experiments show similar performance to NLPC, and much lower than the with-edge results. This strongly supports our argument that the stability of our algorithm comes from the introduction of edge priors.

To provide another form of imperfect lesion mask, we used the state-of-the-art segmentation algorithm [19] to segment the lesions and inpainted using these imperfect masks. The results are in Appendix. Quantitative analysis for RLs is not possible since we don't have ground truth "lesion-free" images.

5 Conclusion and Future Work

In this work, we proposed a deep neural network for inpainting MS lesions. The network uses the edge detection results from the input T1-w image as the prior to guide the refilling operations. Both qualitative and quantitative experiments show that our method outperformed currently available methods. It is robust to inaccurate lesion mask input and the introduction of the edge prior is demonstrated to be the main driver of the stable performance. Also, as new deep learning techniques are developed, the edge prior can be easily incorporated with new architectures to achieve improved results. In future work, we will explore allowing the user to manually manipulate the edge map to guide the inpainting in the test phase. We also plan to evaluate our methods in longitudinal data.

Acknowledgements. This work was supported, in part, by NIH grant R01-NS094456 and National Multiple Sclerosis Society grant PP-1905-34001. Francesca Bagnato receives research support from Biogen Idec, the National Multiple Sclerosis Society (RG-1901-33190) and the National Institutes of Health (1R01NS109114-01). Francesca Bagnato did not receive financial support for the research, authorship and publication of this article.

References

1. Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage **54**(3), 2033–2044 (2011)
2. Battaglini, M., Jenkinson, M., De Stefano, N.: Evaluating and reducing the impact of white matter lesions on brain volume measurements. Hum. Brain Mapp. **33**(9), 2062–2071 (2012)
3. Ceccarelli, A., et al.: The impact of lesion in-painting and registration methods on voxel-based morphometry in detecting regional cerebral gray matter atrophy in multiple sclerosis. Am. J. Neuroradiol. **33**(8), 1579–1585 (2012)
4. Chard, D.T., Jackson, J.S., Miller, D.H., Wheeler-Kingshott, C.A.M.: Reducing the impact of white matter lesions on automated measures of brain gray and white matter volumes. J. Magn. Reson. Imaging **32**(1), 223–228 (2010)
5. Fischl, B.: FreeSurfer. NeuroImage **62**(2), 774–781 (2012)
6. Guizard, N., Nakamura, K., Coupé, P., Fonov, V.S., Arnold, D.L., Collins, D.L.: Non-local means inpainting of MS lesions in longitudinal image processing. Frontiers Neurosci. **9**, 456 (2015)
7. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. (2016), arXiv: 1611.07004
8. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 694–711. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_43
9. Magon, S., et al.: White matter lesion filling improves the accuracy of cortical thickness measurements in multiple sclerosis patients: a longitudinal study. BMC Neurosci. **15**, 106 (2014). https://doi.org/10.1186/1471-2202-15-106

10. Nazeri, K., Ng, E., Joseph, T., Qureshi, F.Z., Ebrahimi, M.: Edgeconnect: Generative image inpainting with adversarial edge learning (2019)
11. Pellicano, C., et al.: Relationship of cortical atrophy to fatigue in patients with multiple sclerosis. Arch. Neurol. **67**(4), 447–453 (2010)
12. Prados, F., Cardoso, M.J., MacManus, D., Wheeler-Kingshott, C.A.M., Ourselin, S.: A modality-agnostic patch-based technique for lesion filling in multiple sclerosis. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014. LNCS, vol. 8674, pp. 781–788. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10470-6_97
13. Prados, F., et al.: A multi-time-point modality-agnostic patch-based method for lesion filling in multiple sclerosis. NeuroImage. **139**, 376–384 (2016)
14. Sdika, M., Pelletier, D.: Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping. Human Brain Mapping **30**(4), 1060–1067 (2009)
15. Smith, S.M.: Fast robust automated brain extraction. Hum. Brain Mapp. **17**(3), 143–155 (2002)
16. Tustison, N.J., et al.: N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging **29**(6), 1310–1320 (2010)
17. Valverde, S., Oliver, A., Lladó, X.: A white matter lesion-filling approach to improve brain tissue volume measurements. NeuroImage Clin. **6**, 86–92 (2014)
18. Xiong, H., Tao, D.: Multiple Sclerosis Lesion Inpainting Using Non-Local Partial Convolutions.(2018), arXiv: 1901.00055
19. Zhang, H., et al.: Multiple sclerosis lesion segmentation with tiramisu and 2.5D stacked slices. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11766, pp. 338–346. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32248-9_38