



HAL
open science

Predicting PET-derived Myelin Content from Multisequence MRI for Individual Longitudinal Analysis in Multiple Sclerosis

Wen Wei, Emilie Poirion, Benedetta Bodini, Matteo Tonietto, Stanley
Durrleman, Olivier Colliot, Bruno Stankoff, Nicholas Ayache

► **To cite this version:**

Wen Wei, Emilie Poirion, Benedetta Bodini, Matteo Tonietto, Stanley Durrleman, et al.. Predicting PET-derived Myelin Content from Multisequence MRI for Individual Longitudinal Analysis in Multiple Sclerosis. *NeuroImage*, 2020, 223C (117308), 10.1016/j.neuroimage.2020.117308 . hal-02922534

HAL Id: hal-02922534

<https://inria.hal.science/hal-02922534>

Submitted on 26 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Predicting PET-derived Myelin Content from Multisequence MRI for Individual Longitudinal Analysis in Multiple Sclerosis

Wen Wei^{a,b,c,*}, Emilie Poirion^c, Benedetta Bodini^{c,d}, Matteo Tonietto^c, Stanley Durrleman^{b,c}, Olivier Colliot^{b,c}, Bruno Stankoff^{c,d}, Nicholas Ayache^a

^aUniversité Côte d'Azur, Inria, Epione Project-Team, Sophia Antipolis, France

^bInria, Aramis project-team, Paris, France

^cInstitut du Cerveau, ICM, Inserm U 1127, CNRS UMR 7225, Sorbonne Université, F-75013, Paris, France

^dAPHP, Hôpital Saint Antoine, Neurology Department, Paris

Abstract

Multiple sclerosis (MS) is a demyelinating and inflammatory disease of the central nervous system (CNS). The demyelination process can be repaired by the generation of a new sheath of myelin around the axon, a process termed remyelination. In MS patients, the demyelination-remyelination cycles are highly dynamic. Over the years, magnetic resonance imaging (MRI) has been increasingly used in the diagnosis of MS and it is currently the most useful para-clinical tool to assess this diagnosis. However, conventional MRI pulse sequences are not specific for pathological mechanisms such as demyelination and remyelination. Recently, positron emission tomography (PET) with radio-tracer [¹¹C]PIB has become a promising tool to measure *in-vivo* myelin content changes which is essential to push forward our understanding of mechanisms involved in the pathology of MS, and to monitor individual patients in the context of clinical trials focused on repair therapies. However, PET imaging is invasive due to the injection of a radioactive tracer. Moreover, it is an expensive imaging test and not offered in the majority of medical centers in the world. In this work, by using multisequence MRI, we thus propose a method to predict the parametric map of [¹¹C]PIB PET, from which we derived the myelin content changes in a longitudinal analysis of patients with MS. The method is based on the proposed conditional flexible self-attention GAN (CF-SAGAN) which is specifically adjusted for high-dimensional medical images and able to capture the relationships between the spatially separated lesional regions during the image synthesis process. Jointly applying the sketch-refinement process and the proposed attention regularization that focuses on the MS lesions, our approach is shown to outperform the state-of-the-art methods qualitatively and quantitatively. Specifically, our method demonstrated a superior performance for the prediction of myelin content at voxel-wise level. More important, our method for the prediction of myelin content changes in patients with MS shows similar clinical correlations to the PET-derived gold standard indicating the potential for clinical management of patients with MS.

Keywords: Conditional GANs, Attention Mechanism, PET Imaging, Multisequence MRI, Demyelination and Remyelination, Deep Learning, Multiple Sclerosis

1. Introduction

Multiple sclerosis (MS) is a demyelinating and inflammatory disease of the central nervous system (CNS). The principal hallmark of MS is the presence of focal demyelinating lesions, consisting of a loss of myelin surrounding the axon, leading to the degeneration of the axon to an extent that is variable between patients and between plaques. These lesions appear abun-

dantly in the white matter (WM). However, the demyelination process can be repaired by the generation of a new sheath of myelin around the axon, a process termed remyelination. These pathological features of multiple sclerosis might be highly dynamic over time but largely heterogeneous across patients [1, 2]. Therefore, a reliable measure of the tissue myelin changes is essential to push forward our understanding of mechanisms involved in the pathology of MS, and to monitor individual patients in the clinical setting or in the context of clinical trials focused on repair therapies.

*Corresponding author: Email: wen.wei@inria.fr;

Over the years, magnetic resonance imaging (MRI) has been increasingly used in the diagnosis of MS and it is currently the most useful paraclinical tool to assess this diagnosis. Although conventional MRI pulse sequences, such as T2-weighted or fluid attenuated inversion recovery (FLAIR) sequences, are sensitive techniques to detect WM lesions of MS, they lack the specificity for the underlying pathological process, and especially have limitations in differentiating between inflammation, axonal loss, demyelination and remyelination. Semi-quantitative MRI techniques, such as magnetization transfer ratio map (MTR), diffusion weighted imaging or T2 relaxometry, also have potential for the measurement of myelin content, but their ability to do so is only partially characterized, and furthermore MTR is affected not only by myelin, but also by water content and inflammation [3].

Positron emission tomography (PET) is an alternative imaging modal, which can target specific tissue substrates and detect tissue changes at the cellular and molecular level. In a recent longitudinal study [1], PET imaging with amyloid radiotracer [^{11}C]PIB could detect a decreased tracer uptake in WM lesions compared to normal-appearing WM (NAWM), which paralleled myelin content. Furthermore, longitudinal data presented by the authors support the ability of [^{11}C]PIB to capture demyelination and remyelination in lesions over time. Note that PET imaging is invasive due to the injection of a radioactive tracer. In addition, it cannot be used in all clinical centers as it is an expensive imaging technique and not available in the majority of medical centers in the world. Therefore, it would be of high interest to predict the individual PET-derived myelin dynamic changes from multisequence MRI.

1.1. Related work

To the best of our knowledge, this is, to date, the first work to predict PET-derived demyelination and remyelination for individual longitudinal analysis in MS. On the contrary, there has been several works focusing on medical image modality prediction and synthesis for various applications, e.g., image denoising and artifact reduction [4, 5, 6, 7, 8, 9], image super-resolution [10, 11, 12], and inter-modal conversion [13, 14, 15, 16, 17]. Extensive efforts have been dedicated to decrease the influence from noise and artifacts in low-dose CT. Dictionary learning based approaches were developed for low dose X-ray CT reconstruction [5, 6, 7]. The works in [8, 9] proposed iterative algorithms by minimizing the total variation to reduce noise and artifacts in CT images. Nevertheless, these algorithms still lost some

anatomical details and suffered from remaining artifacts. Atlas-based methods have also been used for MRI-to-CT synthesis [18, 19]. Although the CT images generated by these methods [18, 19] demonstrated a good ability for attenuation correction, the results highly rely on the registration accuracy and the synthesized image quality may also depend on the prior knowledge for tuning large amounts of parameters in registration step. Moreover, they may be unable to accurately predict subtle lesional features, whose location can be highly variable between patients. For the same application, a structured random forest and auto-context model has been proposed to synthesize CT image from MRI for attenuation correction [20]. In this approach, the input MRI is first partitioned into patches and fed into structured random forest to generate corresponding CT patch. An auto-context model is then used to refine the prediction. However, the input images should be presented by the crafted features which will influence the image synthesis result.

Deep learning based methods have recently shown remarkable ability to automatically learn the underlying features with better descriptive power. The authors in [21] proposed a location-sensitive deep network (LSDN) to synthesize T2-w MRI from T1-w MRI. The target model is predicted by fusing both image features and spatial information. There are also several works trying to do image super-resolution using deep learning models, such as estimating full-dose PET image from low-dose image [11], improving the synthetic FLAIR image quality [12]. Moreover, the works in [16, 17] demonstrated that using multisequence MRI can improve the MRI-to-MRI synthesis performance. It is thus suggested using multisequence as inputs when it is possible. In recent years, generative adversarial networks (GANs) have achieved promising results in nature image synthesis [22, 23, 24] because of the ability to learn the messy and complicated representations of data. As the nonlinear function is more complex in multimodal synthesis, many works investigated the possibility to use GANs to do so. The work in [25] used GANs to synthesize the MR images by using the PET for the quantification of cortical amyloid load. The authors in [26] used multi-channel GANs to synthesize PET images from CT images. There are also several studies working on other modal synthesis, such as retinal images [27, 28], ultrasound images [29] and endoscopy images [30].

By using various radiotracers, PET images can detect different physiological activities, such as metabolism, blood flow, regional chemical composition. As different radiotracers can cause different PET activities, there is no existing clue showing the ability to predict generic

PET activity from MRI. Some methods [31, 32, 33, 34] have already been proposed to predict FDG PET from conventional MRI for AD classification. However, different from these FDG PET synthesis works, our goal is to predict myelin content as defined by [^{11}C]PIB PET from specific MRI which are sensitive but without direct information about myelin content. Moreover, predicting myelin content (as defined by [^{11}C]PIB PET) is a more difficult task because the signal is more subtle and with weaker relationship to anatomical information that could be found in MR images. In our recent works [35, 36], we proposed Sketcher-Refiner GANs to predict PET-derived demyelination from multisequence MRI. Though our previous method demonstrated good ability for static demyelination prediction, the dynamic demyelination-remyelination process which is more important for clinical usage was not predicted.

Moreover, none of the methods discussed above have considered the relationships between spatially separated lesional regions. To model the long-range dependencies between different regions, the recent idea of self-attention mechanism has been first introduced in [37] for machine translation. The core idea behind the self-attention mechanism is to compute a weight distribution on the input data and assign higher values to more relevant elements. Based on this idea, lots of methods have been developed and improved on this basis for natural language processing [38, 39] and various image/video problems in computer vision [40, 41, 42, 22]. The work [22] introduced the self-attention GAN (SAGAN) to improve the modeling of long range dependencies for image synthesis. However, unlike their low-dimensional datasets, medical images are often high-dimensional which makes the model easily reach the memory constraints when calculating the attention maps used in self-attention mechanism.

1.2. Contributions

The novelties and contributions of our paper are as follows:

1. Through MS individual *longitudinal analysis*, we have shown that the PET-derived dynamic myelin changes can be predicted by using multisequence MR images which are not specific but sensitive to myelin content. To the best of our knowledge, this is, to date, the first work to so.
2. Inspired by the work in [22], a conditional flexible self-attention GAN (CF-SAGAN) is used to model the relationships between spatially separated lesional regions, and it is further improved and

specifically adjusted for high-dimensional medical images.

3. Demyelination and remyelination are quantified within MS lesions. An attention regularization for MS lesions is designed so that the neural networks can pay more attention on the MS lesions during the 3D image generation process.
4. Compared with the state-of-the-art methods, our method is shown to outperform these methods qualitatively and quantitatively. Specifically, our method demonstrated a superior performance for the prediction of myelin content at voxel-wise level.
5. Importantly, our method for the prediction of myelin content changes in MS patients shows similar clinical correlations to the PET-derived gold standard indicating the potential for clinical management of patients with MS.

2. Method

2.1. Overview

MS lesions vary in size, location and MRI signal intensity, but they may display the same basic features of pathology, such as demyelination in the WM. Under this scenario, to better predict the myelin changes reflected by the [^{11}C]PIB PET DVR parametric map from multisequence MRI, it is necessary to model the relationships between spatially separated lesional regions. The classic GAN represents both generator G and discriminator D as convolutional networks. During the image synthesis process, the convolutional operations can only process a local neighborhood information because of the limited local receptive field. For example, in a convolution operation, it is hard to model the correlation between top-left and bottom-right positions. These long-range dependencies across different regions in the image can be modeled by stacking several convolutional layers to result in large receptive fields. But doing so, the earlier layers can be almost negligible because of the vanishing gradient, and the optimization algorithms may have trouble on huge parameter space and thus make the GANs training more unstable [43]. Moreover, the medical images are usually high-dimensional. Increasing the depth of the networks and the size of the convolution kernels can dramatically increase the computational cost.

To address this situation, the self-attention mechanism presented in [37, 22] can be used. In detail, it

is modeled by a *Transformer* model with three major concepts: *Key*, *Value* and *Query*. As presented in [22], the C convolutional feature maps $x \in \mathbb{R}^{C \times L}$ is branched out into three copies, corresponding to the three components: *Key* $f(x)$, *Value* $h(x)$ and *Query* $g(x)$ with $f(x) = W_f x$, $g(x) = W_g x$, and $h(x) = W_h x$. The above weight matrices $W_f \in \mathbb{R}^{\bar{C} \times C}$, $W_g \in \mathbb{R}^{\bar{C} \times C}$, $W_h \in \mathbb{R}^{C \times C}$ are part of the model parameters which are implemented by $1 \times 1 \times 1$ convolutions. Each key, value and query are reduced from the high dimensional features to the dimension of $\bar{C} = C/8$ in our implementation for computational efficiency.

Then we transpose the key $f(x)$ and matrix-multiply it by the query $g(x)$ and take the Softmax on all the rows to calculate the attention map:

$$\beta_{j,i} = \text{Softmax}(f(x_i)^T g(x_j)) \quad (1)$$

Where $\beta_{j,i}$ indicates the extent to which the model attends to the i_{th} location when synthesizing the j_{th} region. After integrating the attention map into the self-attention layer, the output of the attention layer $o \in \mathbb{R}^{C \times L}$ is defined as follows:

$$o_j = \sum_{i=1}^N \beta_{j,i} h(x_i) \quad (2)$$

2.2. Conditional Flexible Self-Attention GAN (CF-SAGAN)

Theoretically the self-attention layer [22] is able to capture the long-range dependencies across different image regions which are not covered by the convolution kernels. However, as the medical images are generally high-dimensional, the storage of the attention map can easily reach the memory limits. To adapt to our needs, we simply insert Pooling layers to decrease the size of the input feature maps and then use UpPooling layers to resize the image to match the shape of the input as shown in Fig. 1. By doing so, the self-attention layer can reduce flexibly the size of the attention map by the cubic of the pool size p , making it possible to perform attention on high-dimensional medical images and also other high-dimensional data, such as video and texts. The output of the attention layer o is then multiplied by a scale parameter γ and added back to the original input feature maps:

$$y = \gamma o + x \quad (3)$$

While the scaling parameter γ is increased gradually from 0 during the training, the network is configured to first rely on the cues in the local regions and then gradually learn to assign more weight to the regions that are

further away. In our experiments, we used MaxPooling as pooling operations. The other pooling operations can also be explored, such as AveragePooling, GlobalMax-Pooling, etc.

As the input and the output have the same size, our flexible self-attention layer can be inserted between any two convolutional layers. In our CF-SAGAN, the proposed flexible self-attention layer has been used to train both the generator G and the discriminator D , optimizing the traditional cross-entropy loss function.

2.3. Attention Regularization for MS Lesions

The underlying image synthesis process for the prediction of PET-derived myelin changes follows our previous work [36] which proposed a sketch-refinement process to improve final image quality. We extended this approach by using two CF-SAGANs to improve the prediction performance. Our first CF-SAGAN (Sketcher) aims to sketch the anatomy and physiology information from multisequence MR images. The L1 loss is used to regularize the model globally and encourage less blurring. The other CF-SAGAN used as Refiner takes both the output of the sketcher and the input multisequence MR images to refine and generate the final image reflecting the tissue myelin content. Almost all of the state-of-the-art methods, such as [13, 18, 14, 44], aimed to synthesize the whole image with no local attentions. As the myelin changes are mainly quantified within MS lesions, we introduced an attention regularization to make the Refiner focus more on MS lesions during the image generation process.

Dividing the whole image into three regions of interest (ROIs) from n subjects: lesions R_{Les} , normal appearing white matter (NAWM) R_{NAWM} defined as the white matter outside visible lesions, and ‘‘other’’ R_{other} , with the number of voxels in each region m_{Les} , m_{NAWM} , m_{other} respectively, the proposed attention regularization is described as follows:

$$\mathcal{L}_{L1}(G_R) = \frac{1}{2nm} \sum_{i=1}^n \sum_{j=1}^m \omega_j |I^{i,j} - \hat{I}^{i,j}|, \quad (4)$$

$$\omega_j = \begin{cases} 1 - \frac{m_{Les}}{m}, & j \in R_{Les} \\ 1 - \frac{m_{NAWM}}{m}, & j \in R_{NAWM} \\ 1 - \frac{m_{other}}{m}, & j \in R_{other} \end{cases}$$

where I and \hat{I} are the true image and the prediction output from the Refiner, and i, j is the index of a subject and a voxel respectively.

While our prior works [35, 36] used a weighted L1 loss to change model’s attention towards MS lesions by giving more weight on MS lesions, our proposed

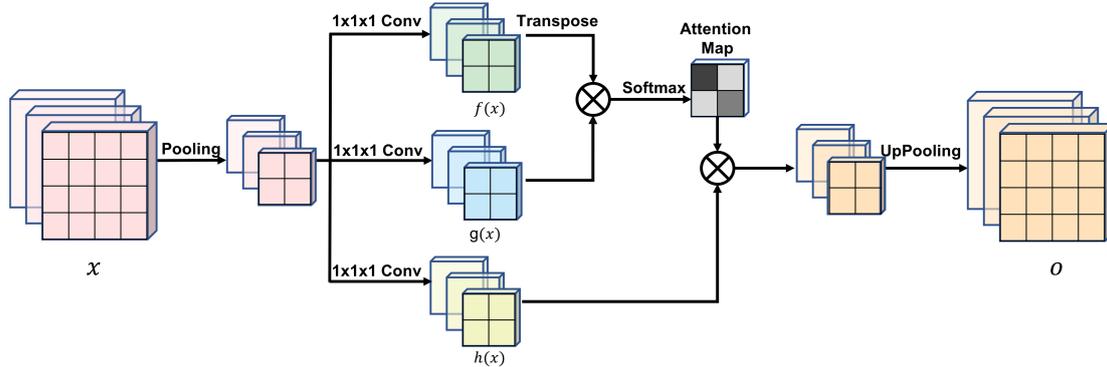


Figure 1: The proposed flexible self-attention layer. The input feature maps x and the output o have the same size. A Pooling and an UpPooling operations have been added to meet the high-dimensional medical image usage.

attention regularization, also regarded as normalized weighted L1 loss, can alleviate the influence from MS lesion size variety across patients and penalize proportionally for different regions.

2.4. Clinical Longitudinal Dataset

Our clinical dataset consists of a longitudinal collection of 18 MS patients (12 women, mean age 31.4 years, sd 5.6) which were clinically assessed and scored using the expanded disability status scale (EDSS) [45] and the multiple sclerosis severity score (MSSS) [46], and 10 age-matched healthy volunteers (8 women, mean age 29.4, sd 6.3). All the patients first underwent MRI and PET scan at baseline (t_0). Then they all repeated the whole protocol after either 1-2 months or 3-4 months (t_1) to explore the best time interval for dynamic remyelination and demyelination quantification. The healthy volunteers only underwent one scan. All participants signed written informed consent to participate in the study, which was approved by the local ethics committee of the Pitié-Salpêtrière hospital. The clinical and demographic information is detailed in [1]. At t_1 , because of missing MR images, 3 patients were excluded from the t_1 dataset. Finally, for each participant, we used the following data:

a) **PET IMAGES:** PET examinations were performed on a high-resolution research tomograph (HRRT; CPS Innovations, Knoxville, TN) which achieves an intraslice spatial resolution of 2.5mm, with 25-cm axial and 31.2-cm transaxial fields of view. The 90-minute emission scan was initiated with a 1-minute intravenous bolus injection of [^{11}C]PIB (mean = 358 ± 34 MBq). The Logan graphical reference method [47] was applied at the voxel level on PET scans in native space to obtain

[^{11}C]PIB PET DVR parametric map ($1.22 \times 1.22 \times 1.22\text{mm}^3$).

b) **MR IMAGES:** MR images were collected using a 3 Tesla Siemens TRIO 32-channel TIM system including Magnetisation Transfer Ratio map (MTR) ($1 \times 1 \times 1.1\text{mm}^3$), and three measures derived from Diffusion Tensor Imaging (DTI): Fractional Anisotropy (FA), Radial Diffusivity (RD) and Axial Diffusivity (AD) ($2 \times 2 \times 2\text{mm}^3$). The three ROIs (lesions, NAWM and “other”) used in Eq. 4 were delineated as follows. WM lesions of MS patients were manually contoured by an expert rater on T2-w scans with reference to FLAIR images. The corresponding lesion masks were generated and aligned to the individual T1-w scan using FLIRT algorithm in the FSL package [48]. After performing a “lesion-filling” procedure in patients only, T1-w scans were segmented using FreeSurfer [49] to obtain a WM mask. The NAWM is then defined as the WM outside visible lesions on T2-w scans.

2.5. Indices of Myelin Content Changes

Following a validated procedure [1], voxels characterized as demyelinated were identified as those whose DVR value fall below one standard deviation of the mean DVR value of all the voxels in healthy controls that were localized at the same distance from the CSF. This step returned individual maps of demyelinated voxels inside WM lesions in patients, which were generated for each of the 2 time-points (baseline- t_0 , follow-up- t_1). In each patient, individual maps of remyelinating and demyelinating voxels inside WM lesions were computed based on the trajectory of each voxel.

From the demyelinated map at both time points (t_0 and t_1), the percentage of demyelinated voxels over the

total lesion load measured at baseline (t_0) was calculated for each patient. To further measure the myelin content changes, we defined the following indices:

- **Global index of myelin content change:** It is the difference between the derived percentage at t_1 and the corresponding percentage at t_0 . This index reflects the subject-specific prevalence of either myelin loss or myelin repair over the follow-up interval.
- **Index of dynamic demyelination:** It is defined as the ratio of lesional voxels which are normally myelinated at baseline t_0 and then classified as demyelinated voxels at t_1 to the total lesion load. This index reflects the ongoing myelin loss.
- **Index of dynamic remyelination:** It is defined as the ratio of lesional voxels which are classified as demyelinated at baseline t_0 and recovered a normal myelin level at t_1 to the total lesion load. This index reflects ongoing myelin repair.

2.6. Network Architectures

In our work, the two CF-SAGAN used as Sketcher and Refiner are both three-dimensional to fully capture the spatial information between slices. Both of them share the same architectures for their generators and discriminators. Instead of using the whole image, we applied large overlapped patches ($64 \times 64 \times 64$) to train the models, which can save computational resource and also provide sufficient training samples. We detailed the architectures of the generator and the discriminator as follows.

- **Generator architecture:** The generator is critical to the quality of the generated images. We used an encoder-decoder as a backbone architecture. The encoder is used to extract the spatial features from the input image, and the decoder aims to construct the final output from the encoded features. To help the network to retrieve the lost spatial information during the down-sampling operations and boost the information flow between the end and the start of the network, we introduced the long skip connection to the network architecture which can also be regarded as U-Net [50]. In addition, we added the short connection into the generator as shown in *ResDown Block* and *ResUp Block* in Fig. 2 (a) and (b) respectively. This improvement, also called residual connection [51], is known for its ability to mitigate the problem of

vanishing gradient by allowing this alternate short-cut path for gradient to flow through, and enhance the feature exchanges across layers. Taking the advantages of our proposed flexible self-attention unit and long/short skip connections, our generator is illustrated in Fig. 2. The encoder includes a sequence of a convolutional layer, four ResDown blocks and a flexible self-attention unit between. The number of feature maps starts from 16 and doubles after each convolutional layer or ResDown block. The decoder inversely involves four ResUp blocks with a flexible self-attention in between and a convolutional layer to output the final image. All of the different levels in the encoder are transferred to the corresponding levels in the decoder by using long connections.

- **Discriminator architecture:** The discriminator is used as a classifier to distinguish if the input is in class “True” or “Fake”. We adopted the *3D PatchDiscriminator* from the work of [35]. Each downsampling operation is realized by a ResDown block. In addition, the flexible self-attention unit is added to reinforce the discriminator to capture complicated long-range constraints on the global image structure. The general architecture is demonstrated in Fig. 2. Note that the long connections in ResDown block are not used in the discriminators.

3. Experiments and Evaluation

3.1. Implementation and Training Details

We used the clinical longitudinal dataset described in the Section 2.4 for the prediction of dynamic myelin changes. The model is trained, validated and tested on baseline (t_0) dataset by using 3-fold cross validation. For each cross-validation split, the dataset is divided into a training set with 2/3 subjects (including 3 subjects for validation), and a testing set consisting of 1/3 subjects. The fixed model is then directly applied to the images of these 1/3 subjects from t_1 dataset. This can also be considered as a way to evaluate model generalization. For both time points (t_0 and t_1), the preprocessing steps mainly consist of brain extraction [52], intensity inhomogeneity correction [53] and affine intra-subject registration of MR data onto [^{11}C]PIB PET DVR image space using FLIRT algorithm in the FSL package [48]. Finally, we cropped the images into 50% overlapped $64 \times 64 \times 64$ patches with a resolution of

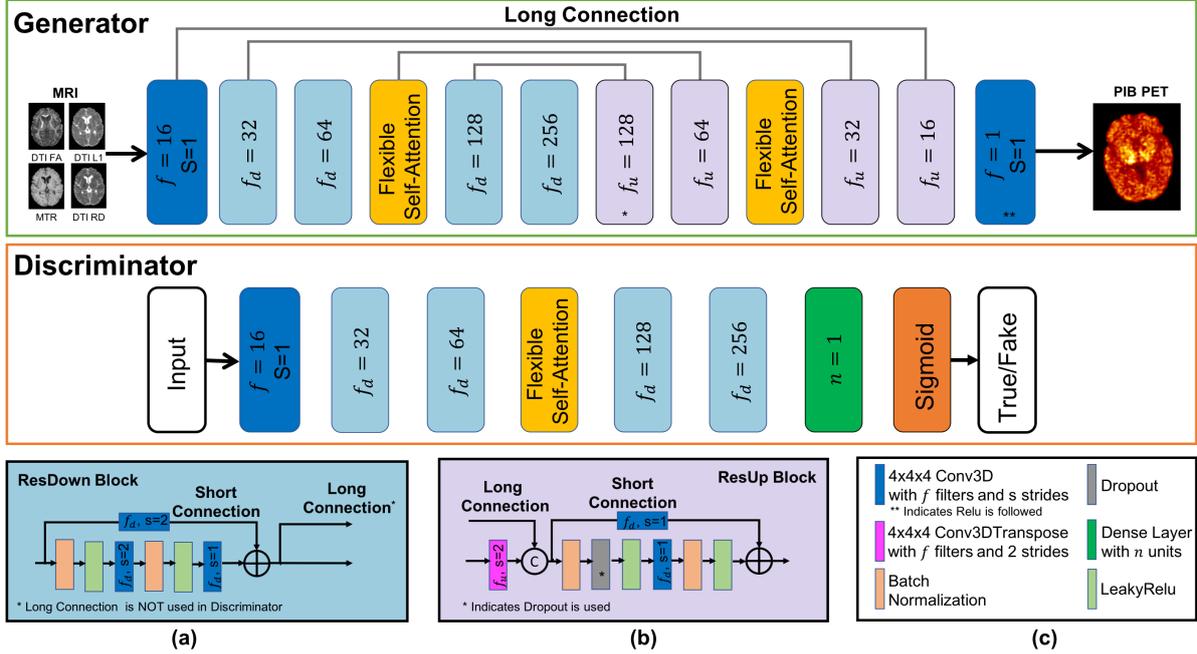


Figure 2: The overall framework of the proposed CF-SAGAN. (a-c) Illustration of the different units and layers used in the Generator and the Discriminator. The values f_d, f_u, f indicate the number of feature maps in each unit and s is the stride number. The parameter n refers to the units number in the Dense layer. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$1.22 \times 1.22 \times 1.22 \text{mm}^3$. The details of acquisition parameters and PET data quantification are described in [1] and [54].

Our model was implemented with Tensorflow [55]. The convolution kernel size is $3 \times 3 \times 3$ and the rate for dropout layer is 50%. To train the model, the whole data underwent the zero mean and unit variance normalization. Data augmentation is also applied including three rotations (90, 180 and 270 degrees), scaling up by 1.25 and scaling down by 0.75. During the training process, we first iteratively trained the Sketcher for 370 epochs by fixing our Refiner. Then we trained the Refiner from scratch for another 370 epochs by fixing our Sketcher. The optimization was performed with the ADAM solver with $10^{-4}, 5 \times 10^{-5}$ as initial learning rates for the Sketcher and the Refiner respectively.

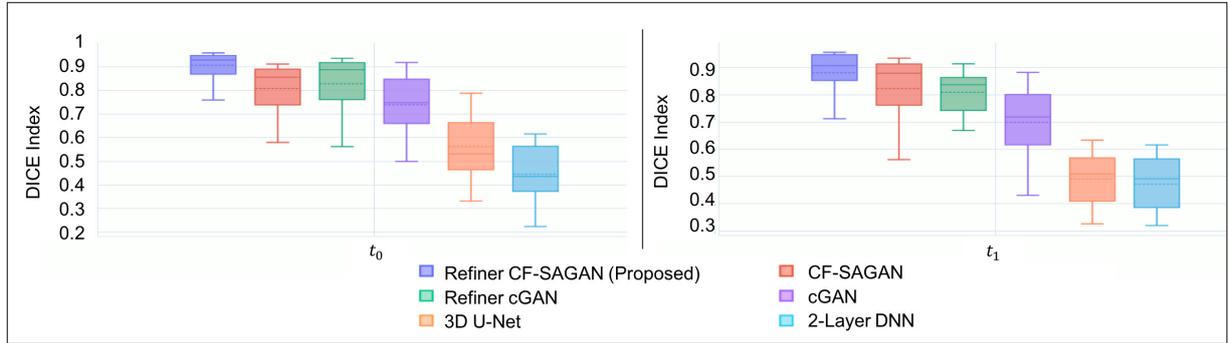
It is known that training a GAN model can become unstable and even produce an early model collapse. In our work, we used several techniques to improve the stability of training our CF-SAGAN models. First, we used strided convolutions for downsampling as shown in Fig. 2. By doing so, the network can learn its own spatial downsampling. Similarly, the upsampling operation is done by strided deconvolutional layer. Second, as suggested in [56], the labels used by the discriminator for true/fake samples are smoothed to reduce the vulner-

ability of neural networks to adversarial examples. Instead of setting hard labels (0 and 1), the label was set as a random number between 0 and 0.1 for 0 labels, and a random value between 0.9 and 1.0 to represent 1 labels. In this way, the task for the discriminator becomes more challenging and it can match the difficulty of the generator task, so that the adversarial training becomes balanced. For the same reason, the third technique is imbalanced learning rate. In our work, the discriminator was updated two times per generator update step during the training process. Last, LeakyReLU is used to allow a stable training of CF-SAGANs with 0.2 as slope coefficient.

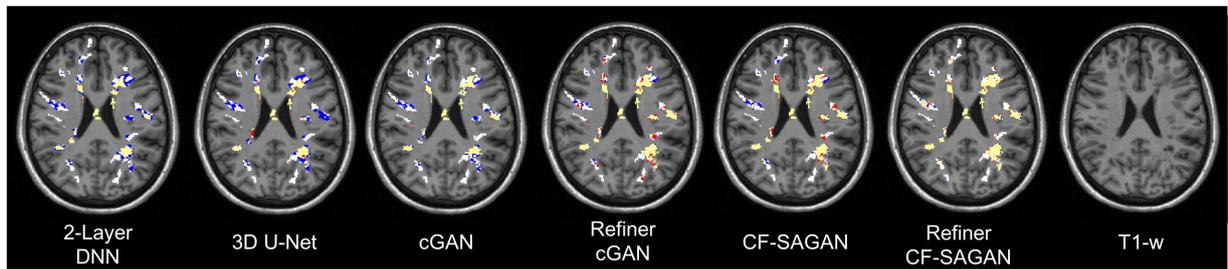
To evaluate our method, we compare it with several state-of-the-art methods including a 2-layer DNN [32], a 3D U-Net [31], a cGAN [26] and a Sketcher-Refiner framework by using two cGANs (denoted as Refiner cGAN) [35]. All state-of-the-art methods were kept the same architecture and parameter settings as described in the work of [35]. ReLU was used as the last layer to keep positive output for all the methods. We describe the evaluation details in the following.

3.2. Evaluation of Static Demyelination Prediction

As mentioned in Section 2.5, within MS lesions each voxel can be classified as demyelinated or normal. Ap-



(A)



(B)

Figure 3: (A) The box plots show the median (middle solid line), mean (middle dotted line) and min-max (below and above line) DICE index from all the methods for both time points. (B) Demyelinated voxels classified from the ground truth and the predicted PET images within MS lesions. Agreement between methods is marked in yellow (both true and predicted PET indicated demyelination) and white (both methods did not indicate demyelination). Disagreement is marked in red (demyelination only with the predicted PET) and blue (only with the true PET). The DICE coefficients in this case are 0.53 (2-Layer DNN), 0.56 (3D U-Net), 0.76 (cGAN), 0.85 (Refiner cGAN), 0.83 (CF-SAGAN) and 0.89 (Refiner CF-SAGAN). The corresponding T1-w MR images are also shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

plying the previous clinical procedure described in [1] for both time points (t_0 and t_1), we calculated the static demyelination maps and used DICE index to measure, for each patient, the agreement between the masks of demyelinated voxels classified within MS lesions from both the ground truth and the PET images predicted by all the methods. The comparison results are shown in Fig. 3 (A). Among all the methods, our proposed Refiner CF-SAGAN gets the best static demyelination prediction for both time points. Precisely, we got 0.91 ± 0.06 and 0.89 ± 0.08 as mean DICE for t_0 and t_1 respectively. Comparing with our previous work Refiner cGAN [35] which got 0.83 ± 0.12 and 0.80 ± 0.08 for both time points, our method demonstrated a better ability to predict the demyelination in MS lesions at the voxel-wise level. Moreover, CF-SAGAN integrating our proposed flexible self-attention unit demonstrates a superior performance than a simple cGAN. This improvement consistently indicates the ability of the proposed flexible self-attention layer to learn long-range spatial dependencies and thus to have better prediction results. It is worth noting that the performance is boosted af-

ter our refinement process, i.e. from cGAN to Refiner cGAN and from CF-SAGAN to Refiner CF-SAGAN. It can also be clearly seen that the GAN-based methods generally outperform the 2-layer-DNN and 3D U-Net. This also highlights that the relationship between myelin content and multimodal MRI data is complex, and a single simple network is not powerful enough to model it. Examples of demyelinated voxel masks generated from all the methods are shown in Fig. 3 (B).

Furthermore, we measured the percentage of demyelinated voxels over total lesion load of each patient for both the ground truth and the predicted PET. It can be seen from the Fig. 4 that our prediction results are nearly the same as the ground truth for all of the patients for both time points showing a very good accuracy of our method.

3.3. Evaluation of Dynamic Demyelination and Remyelination Prediction

To quantify the myelin changes in MS lesions, we calculated the three indices proposed in Section 2.5 for

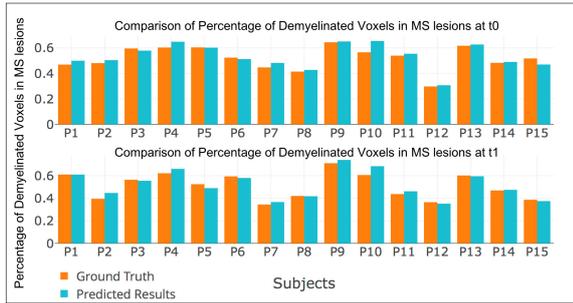


Figure 4: Percentage of demyelinated voxels in white matter MS lesions for each patient computed from the ground truth (orange) and from our method (blue) for t_0 (top) and t_1 (bottom).

both the ground truth and all the methods. We first calculated the *global index of myelin content change* which is the difference between two time points. This index reflects a balance between a predominant demyelination and remyelination. After sorting by the myelin changes derived from the ground truth, the comparison result is shown in Fig. 5 (A). It shows that although there is a high between-patient variability ranging from -0.13 to +0.14, the PET-derived overall myelin changes can be better reproduced by our predicted images. Specifically, most of the patients can be well classified into the right category (demyelination or remyelination). The only false classification is due to the tiny change which is only 0.006 according to the ground truth (see P8 in Fig. 5 (A)). On the contrary, Refiner cGAN and CF-SAGAN have 6 and 4 false classifications respectively and some subjects’ predictions are even far from the ground truth. To have an easily interpretable figure, the prediction results of other methods (cGAN, 3D U-Net and 2-Layer-DNN) are not shown in Fig. 5, as their miss-classification rates are all greater than 0.5 that corresponds to random guessing.

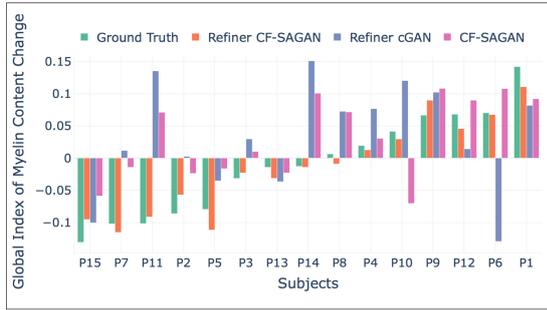
For each patient, we then calculated the *index of dynamic demyelination* and *index of dynamic remyelination* for both the ground truth and all the prediction results. These two indices which are defined in Section 2.5 are able to reflect the ongoing myelin loss and repair. From the comparison results shown in Fig. 5 (B)(C), it can be found that our prediction results obtain the best approximation for almost all the patients. In particular, comparing with the ground truth, we achieved almost the same level for the majority of patients. We furthermore compared, in each patient, the masks of dynamic demyelination and remyelination voxels classified from both the true and the predicted PET within MS lesions. The average DICE indices are 0.71 ± 0.11 and 0.69 ± 0.12 for the dynamic demyelination and remyelination maps derived from the ground

truth and our predicted [^{11}C]PIB PET. While Refiner cGAN and CF-SAGAN only get 0.62 ± 0.18 (0.56 ± 0.12) and 0.59 ± 0.17 (0.58 ± 0.09) as mean DICE for demyelination (remyelination) maps respectively. Considering the slight myelin changes between two time points and the high variability between patients, the agreement demonstrates a promising ability of our method to predict the dynamic demyelination and remyelination in MS lesions. Three examples are illustrated in Fig. 6, showing a small prediction discrepancy from our method and a high between-patient variability.

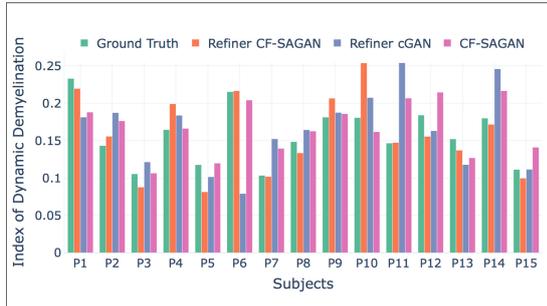
3.4. Clinical Correlation

It is concluded in the recent clinical research work [1] that there is no significant association between the index of dynamic demyelination and EDSS ($p = 0.72$), but the index of remyelination is a significant explanatory factor for EDSS with -0.67 as beta-coefficient ($p = 0.006$). We thus calculate the correlation between the synthesized-PET-derived indices of myelin content change and the clinical score EDSS. In our work, because all the mandatory MRI data were not available for the initial patients list, the clinical correlation have been recomputed for both the synthesized-PET-derived and true-PET-derived indices based on a subset of subject to check the consistency of our method. As no significant change was found in EDSS during follow-up, the EDSS measured at baseline was used for clinical correlation which is calculated using multiple linear regression with EDSS as response variable and age, gender and T2-w lesion load as additional covariates.

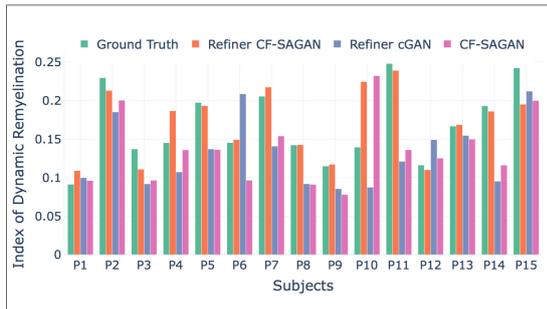
As shown in Fig. 7 and Appendix Table 2(a), no significant effect was found on EDSS for the true-PET-derived index of dynamic demyelination ($p = 0.578$), whereas a significant association was detected between the true-PET-derived index of dynamic remyelination and EDSS ($p = 0.021$; beta-coefficient= -0.703) showing patients with lower disability presenting a higher proportions of remyelinating voxels. A similar finding can be observed between EDSS and the synthesized-PET-derived indices (Fig. 7 and Appendix 2 Table 2(b)) with no significant correlation for the index of dynamic demyelination ($p = 0.676$) and a significant inverse correlation for the index of dynamic remyelination ($p = 0.012$; beta-coefficient= -0.734). This observation demonstrates the consistency and the ability of our method for the prediction of PET-derived myelin content in MS. The correlation between the index of remyelination and EDSS has also been calculated for the other methods (2-layer DNN, 3D U-Net, cGAN, Refiner cGAN and CF-SAGAN), but none of them can reproduce this inverse correlation ($p > 0.05$) showing that



(A)



(B)



(C)

Figure 5: The patient-level comparison for three myelin change indices computed from the ground truth and all the methods. (A) Global index of myelin content change values for each patient. Patients with positive values indicate a predominant demyelination process. Patients with negative values means a predominant process of remyelinating. (B) Index of dynamic demyelination for each patient. (C) Index of dynamic remyelination for each patient.

their predictions of myelin content changes are not accurate enough.

3.5. Evaluation of Attention Regularization

To study the contribution of the proposed attention regularization term, we conduct comparison experiments for different combinations of each of three models (3D U-Net, cGAN and CF-SAGAN) using one of regularization terms including L1, weighted L1 (denoted as WL1) used in [35] and the proposed attention

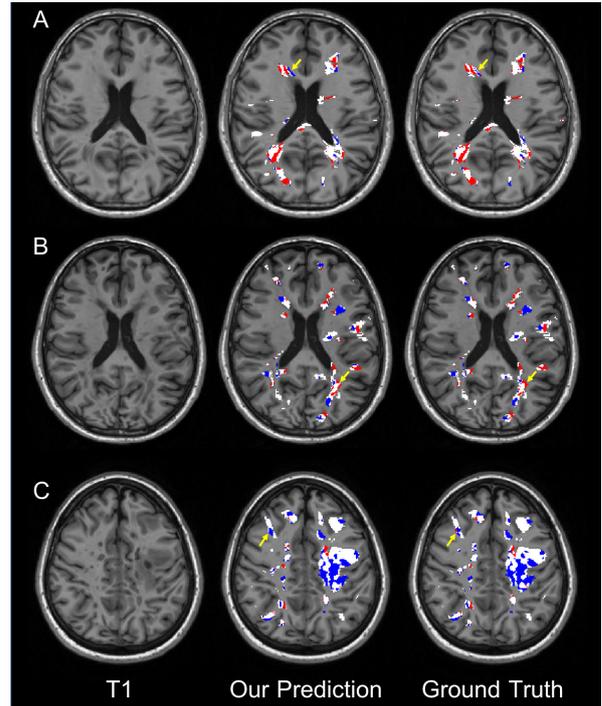


Figure 6: Three examples of lesional myelin content changes showing demyelinating (in red) and remyelinating (in blue) voxels derived from the true longitudinal $[^{11}\text{C}]\text{PIB}$ PET (right) and our predictions (middle), localized inside white matter (WM) lesions (in white), overlaid onto the corresponding T1-w MR image (left). These three patients respectively show a clear prevalence of demyelination over remyelination (Patient A), an active demyelination together with moderate remyelination in all visible WM lesions (Patient B) and an extensive process of remyelination (Patient C). The yellow arrows indicate some prediction discrepancies. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

regularization, also regarded as normalized weighted L1, denoted as NWL1. As demyelination and remyelination are quantified within MS lesions, the performance is evaluated by myelin content prediction discrepancy within MS lesions (defined as mean absolute difference between the ground truth and the predicted PET) for local image quality and also by MSE for global image quality.

The comparison results are provided in Table 1 with Table 1(a) showing the performance on global image quality measured by MSE and Table 1(b) showing the performance on local image quality measured by myelin content prediction discrepancy within MS lesions. We can clearly see that the models using L1 loss presented a performance superior to the other combinations on MSE, but they achieved the worst results on myelin content prediction inside MS lesions. Com-

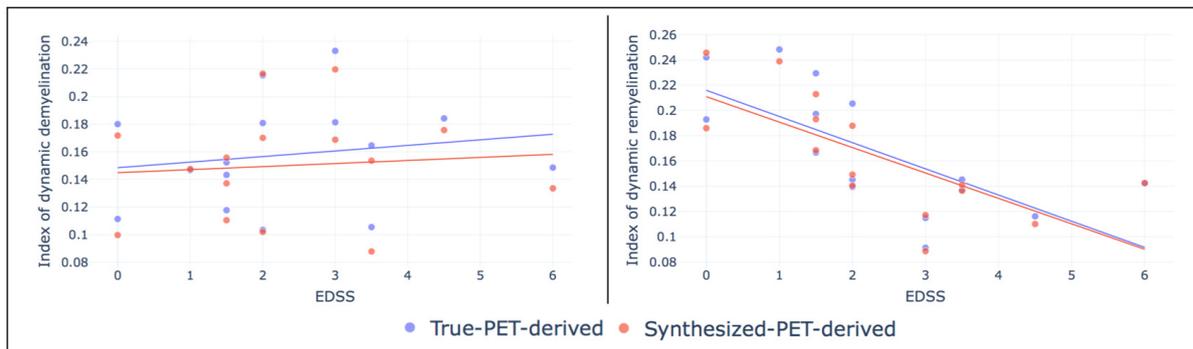


Figure 7: Scatter plot diagrams and fitting lines represent the correlation between true-PET-derived (blue) or synthesized-PET-derived (red) index of dynamic demyelination(left)/remyelination(right) and EDSS. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

pared with the models using WL1 which got 0.0113, 0.0103 and 0.0097 as MSE for 3D U-Net, cGAN and CF-SAGAN respectively, the methods using NWL1 are shown to outperform by a large margin on MSE with 0.0111, 0.0098 and 0.0089 for 3D U-Net, cGAN and CF-SAGAN respectively. The NWL1-based methods have even shown slightly better performance on myelin content prediction discrepancy inside MS lesions (two of three models using NWL1 got best results and cGAN with NWL1 obtained 0.028 which is very close the best value 0.027). In our work, the two CF-SAGAN are respectively used as Sketcher for the global generation of anatomy and physiology information, and as Refiner for the local refinement for MS lesions. The applications of L1 for the Sketcher and NWL1 for the Refiner exactly take full advantage of their different characteristics. It is also found in experiments that the NWL1 allows faster convergence and stable training with 0.8 days as overall training time instead of 1.2 days by using WL1 on GPU NVIDIA GTX 1080 Ti.

Both the qualitative and quantitative experimental results demonstrate that our method can synthesize high quality image. In the next section, our proposed method will be evaluated clinically for the prediction of myelin content for MS individual longitudinal analysis.

4. Discussion

In this work, we proposed a method to predict PET-derived demyelination and remyelination for individual longitudinal analysis in MS from multisequence MR images. The method is based on our proposed conditional flexible self-attention GAN (CF-SAGAN) which is specifically adjusted for high-dimensional medical images and able to capture the relationships between

Table 1: Comparison of different regularization terms

(a) Comparison of MSE obtained from different methods using different regularization terms. Results are displayed as mean (standard deviation).

	L1	WL1	NWL1
3D U-Net	0.0107 (0.0041)*	0.0113 (0.0043)*	0.0111 (0.0047)*
cGAN	0.0094 (0.0038)	0.0103 (0.0042)	0.0098 (0.0044)
CF-SAGAN	0.0085 (0.0042)	0.0097 (0.0039)	0.0089 (0.0041)

*indicates CF-SAGAN is significantly better with $p < 0.05$ by two-sided T-test.

(b) Comparison of MS lesion myelin content prediction discrepancy (defined as mean absolute difference between the ground truth and the predicted PET) obtained from different methods using different regularization terms. Results are displayed as mean (standard deviation).

	L1	WL1	NWL1
3D U-Net	0.035 (0.027)	0.032 (0.029)	0.031 (0.032)
cGAN	0.030 (0.017)	0.027 (0.022)	0.028 (0.019)
CF-SAGAN	0.027 (0.020)	0.024 (0.016)	0.022 (0.017)

the spatially separated lesional regions during the image synthesis process. Our result is further improved dramatically by following the sketch-refinement process with the second CF-SAGAN as the Refiner and using our proposed attention regularization to make the network focus on the MS lesions.

Through the clinical evaluation with the previous clinical procedure described in [1], Our method demon-

strates superior performance compared to the state-of-the-art approaches including a 2-layer DNN [32], a 3D U-Net [31], a cGAN [26] and a Sketcher-Refiner framework by using two cGANs (denoted as Refiner cGAN) [35]. From the aspect of static demyelination prediction, our prediction results approximate the true-PET-derived percentage of demyelinated voxels individually for both time points. In particular, a better agreement, between demyelination maps derived from the true and the predicted PET, has been achieved than our prior work [35]. Regarding dynamic demyelination and remyelination prediction, the three indices of myelin content change derived from our predicted PET images are very similar to those derived from the ground truth. Moreover, the same clinical correlation between the index of dynamic remyelination and EDSS can be found from both the true and our predicted PET images. All these findings indicate the potential of our method for clinical management of patients with MS, although it still needs to be validated and confirmed in large populations. Regarding the global image quality (Appendix .1), our method can generate objectively more realistic images which are the most similar to the ground truth. By using MSE and PSNR as image quality metrics, our method is shown to outperform the other approaches on all metrics.

Several studies have already explored the possibility to synthesize FDG-PET from MR images [31, 32, 33, 34]. But none of them considered the local attentions and the output images were synthesized directly without any local focus. Our previous work in [35] applied a weighted L1 loss (WL1) to make the network pay more attention on MS lesions where demyelination and remyelination are quantified. On this basis, we extended and proposed an attention regularization term in this work (denoted as NWL1 in Section 3.5). It can be found in Table 1 that L1 loss can yield the best global image quality but with the worst local image quality. Both the WL1 and the NWL1 are good at local image synthesis, but the NWL1 can achieve superior performance on global image quality than WL1. The main reason is that these three regularization terms play different roles during the image generation process. The L1 loss drives the network to output images which are only globally close to the ground truth, but without any specific attention on some regions of interest. The WL1 is designed to transfer more attention on the pre-defined ROIs. However, the larger the weight is assigned to these regions, the less attention on the other regions will be paid during the generation process. Especially, in some patients, when MS lesions are extremely small, the network would only focus on these tiny regions and

cannot output an anatomically and structurally plausible image. Nevertheless, our NWL1 can not only make the network pay more attention on these specifically pre-defined ROIs leading to a remarkable local image prediction, but can also find a balance and take consideration of other regions generating a competitive global image quality.

In fact, the tracer [^{11}C]PIB was initially used for β -amyloid plaques in Alzheimer’s disease (AD). But, myelin signal quantified by [^{11}C]PIB PET is more subtle than amyloid plaques and with weaker relationship to the anatomical information found in MR images making this synthesis problem more difficult. Multiple GANs can be used to improve synthesis quality as proved in several works [57, 58, 36]. Inspired by this idea and given the ability of our CF-SAGAN to capture the relationships between the spatially separated lesional regions, we used two CF-SAGANs to improve the prediction performance. Unlike the traditional cascade GANs used in [57, 58], our two CF-SAGANs named as Sketcher and Refiner act as different roles for sketching anatomy and physiology information (Sketcher) and refining myelin content (Refiner) due to the use of NWL1 mentioned above. In practice, the number of GANs can be increased depending on different tasks and the gained performance after each iteration. We found that the performance will not be improved after two iterations by using three CF-SAGANs and the performance is nearly the same between two and three CF-SAGANs. The number of CF-SAGANs is thus fixed to two to save computational resource. The same conclusion is also conducted in [58, 36].

The proposed method might be further improved by considering several limitations of our work. First, like the weighted L1 loss [36, 35], the needs of the masks of different ROIs still remains for our proposed NWL1. In practice, these masks cannot be always available. In the future work, instead of using manually predefined attention, a self-learned attention could be helpful to solve this problem. Second, multisequence MR images are used as inputs to provide as much information as possible to the network. However, the subjects with incomplete MR images cannot be used and have to be excluded from the dataset. This is a great loss especially for small medical image datasets. It would be helpful for future work to discover the way to deal with multisequence images independently, so that every incomplete MR series can be used for training. Furthermore, the intra-subject registration used in our work as a preprocessing step for multisequence MR images may also influence the synthesized image quality because of image noise and different selections of parameters in the regis-

tration step. The use of combined MR-PET systems can also avoid this preprocessing step. Last, our method is only evaluated on a small, single-center dataset. Further experiments on larger, multi-center, datasets, will thus be needed to assess the generalizability of the approach more in depth. Such further validation is crucial before translation to the clinic can be considered.

5. Conclusion

In this paper, we proposed a method to predict PET-derived demyelination and remyelination for individual longitudinal analysis in MS. The method is based on our proposed conditional flexible self-attention GAN (CF-SAGAN) which is specifically adjusted for 3D medical images and able to capture the relationships between the spatially separated lesional regions during the 3D image synthesis process. We also introduced an attention regularization which can not only lead a remarkable local image quality on MS lesions, but can also take consideration of other regions generating a competitive global image quality. Jointly applying the sketch-refinement process, our approach is shown to outperform the state-of-the-art methods qualitatively and quantitatively. Importantly, the clinical evaluations of our method for the prediction of myelin content from multisequence for MS individual longitudinal analysis show similar results to the PET-derived gold standard.

Acknowledgments

The first author is funded by an Inria fellowship. This work has been supported by the French government, through the 3IA Côte d’Azur Investments in the Future project managed by the National Research Agency (ANR) with the reference number ANR-19-P3IA-0002. The research leading to these results has received funding from the French government under management of Agence Nationale de la Recherche as part of the "Investissements d’avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute) and reference ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6) and from the "Contrat d’Interface Local" program (to Dr Colliot) from Assistance Publique-Hôpitaux de Paris (AP-HP). The clinical study was funded by specific grants from ELA (European Leukodystrophy Association, grant 2007-0481), and INSERM-DHOS (grant 2008-recherche clinique et translationnelle) and has been sponsored by APHP (Assistance Publique des Hôpitaux de Paris). Emilie Poirion has been funded by

IUIS (Institut Universitaire d’Ingenierie pour la Santé, Sorbonne Université, and both Benedetta Bodini and Emilie Poirion received funding from fondation AR-SEP).

Appendix .1. Evaluation of Global Image Quality

In this Appendix, we provide the evaluation of global image quality for all the methods. The image quality is evaluated by taking into account both the qualitative differences through human perception as well as quantitative aspect measured by mean square error (MSE) and peak signal-to-noise ratio (PSNR) which are defined as follows:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^N \|(I^i, \hat{I}^i)\|_2 \quad (.1)$$

$$\text{PSNR} = 20 \cdot \log_{10}(\text{MAX}_I) - 10 \cdot \log_{10}(\text{MSE}) \quad (.2)$$

where MAX_I is the maximum voxel value of the image.

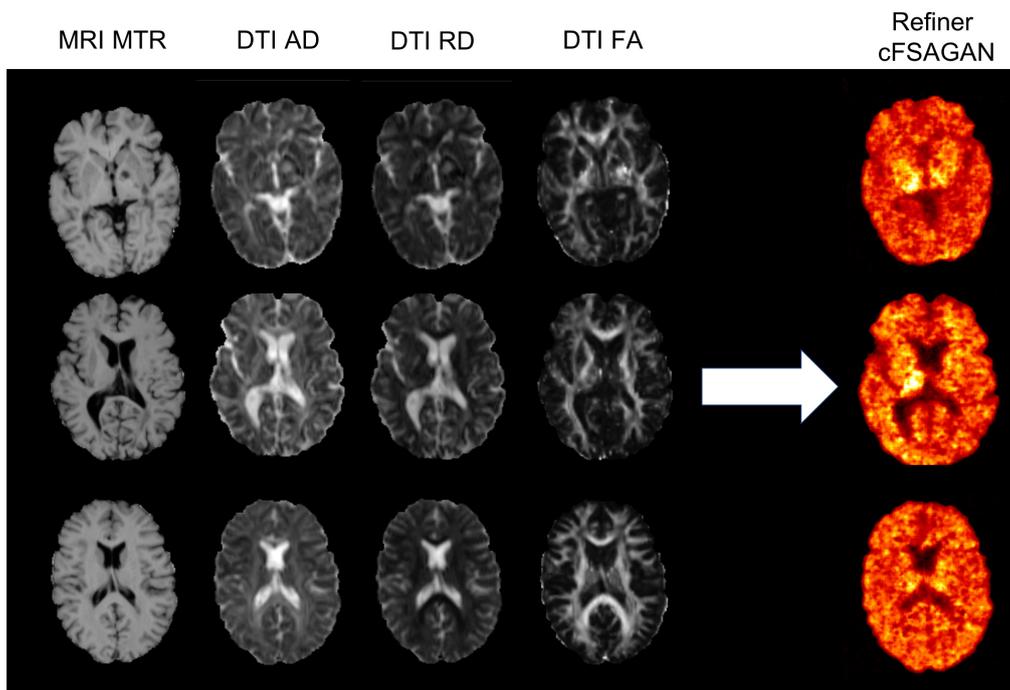
The quantitative evaluation is summarized in Table .2. The proposed Refiner CF-SAGAN ranks the best among all the methods in both metrics. Comparing the proposed Refiner CF-SAGAN and the 2-layer-DNN, the Refiner CF-SAGAN outperformed by 42.65% in terms of MSE ($p < 0.05$) and 9.07% in PSNR ($p < 0.05$). Our Refiner CF-SAGAN also achieved better results than 3D U-Net, in all metrics, with improvements of 27.10% in MSE ($p < 0.05$) and 3.37% in PSNR ($p < 0.05$). It can also be found that the image quality of CF-SAGAN is further improved when the sketch-refinement process is applied to output the final image (denoted as Refiner CF-SAGAN in Table .2).

Table .2: Image quality metrics obtained with our method and the other methods. MSE: mean square error; PSNR: peak signal-to-noise ratio. Results are displayed as mean (standard deviation).

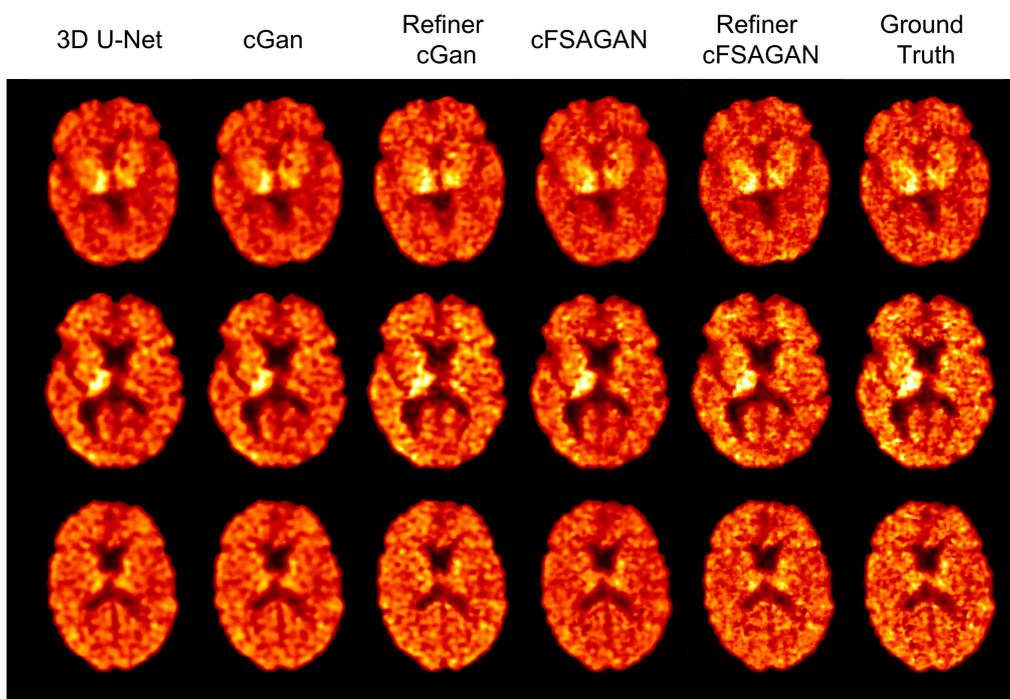
	MSE	PSNR
2-Layer DNN	0.0136 (0.0048)*	27.767 (1.214)*
3D U-Net	0.0107 (0.0041)*	29.297 (0.986)*
cGAN	0.0094 (0.0038)	29.475 (0.981)
Refiner cGAN	0.0083 (0.0037)	30.044 (1.095)
CF-SAGAN	0.0085 (0.0042)	29.942 (1.065)
Refiner CF-SAGAN (Proposed)	0.0078 (0.0038)	30.285 (0.0993)

* indicates our method is significantly better with $p < 0.05$ by two-sided T-test

The similar conclusion can be drawn from the qualitative comparison shown in Fig. .8(B) with the corresponding ground truth (the right-most column). As observed, the image quality of the 2-Layer-DNN is obviously the worst. In particular, some anatomical or structural traces which are not present in the ground truth can still be found in the 2-layer-DNN predicted image. GAN-based methods generally outperform the 2-layer-DNN and 3D U-Net with more shape information. These enhancements can be attributed to the use of adversarial loss to generate a more realistic output. The performance is further boosted after the refinement process. Therefore, taking the advantages of adversarial training, the sketch-refinement process and the proposed flexible self-attention, our Refiner CF-SAGAN can generate objectively sharper, less blurry images which are closest to the ground truth.



(A)



(B)

Figure .8: (A) Illustration of multisequence MRI as inputs and our prediction results. (B) Qualitative comparison of the results of our proposed framework (CF-SAGAN), of the refined version of our proposed method (“Refiner CF-SAGAN”), and of the other state-of-the-art methods.

Appendix .2. Clinical Correlation

(a) Correlation between **true-PET-derived** index of dynamic demyelination/remyelination and EDSS with age, gender and T2 lesion load as covariates.

Dependent Variable: EDSS Score	Coefficient	Standard Error	t	p	Beta-coefficient
Index of demyelination	0.090	0.157	0.58	0.578	0.216
Age	-0.019	0.090	-0.21	0.840	-0.065
Gender	-0.893	1.126	-0.79	0.446	-0.269
T2 Lesion Load	1.01e-5	1.75e-5	0.58	0.577	0.207
Index of remyelination	-0.233	0.085	-2.75	0.021*	-0.703
Age	-0.047	0.067	-0.70	0.498	-0.164
Gender	-0.062	0.869	-0.07	0.944	-0.019
T2 Lesion Load	5.22e-6	1.16e-5	0.45	0.661	0.107

*Tests significant at significance level $p = 0.05$.

(b) Correlation between **synthesized-PET-derived** index of dynamic demyelination/remyelination and EDSS with age, gender and T2 lesion load as covariates.

Dependent Variable: EDSS Score	Coefficient	Standard Error	t	p	Beta-coefficient
Index of demyelination	0.071	0.165	0.43	0.676	0.173
Age	-0.014	0.091	-0.15	0.882	-0.048
Gender	-0.980	1.120	-0.87	0.402	-0.295
T2 Lesion Load	9.99e-6	1.92e-5	0.52	0.614	0.205
Index of remyelination	-0.258	0.084	-3.07	0.012*	-0.734
Age	-0.050	0.064	-0.78	0.452	-0.017
Gender	-0.063	0.816	-0.08	0.940	-0.019
T2 Lesion Load	7.47e-6	1.1e-5	0.68	0.513	0.153

*Tests significant at significance level $p = 0.05$.

References

- [1] B. Bodini, M. Veronese, D. García-Lorenzo, M. Battaglini, E. Poirion, A. Chardain, L. Freeman, C. Louapre, M. Tchikviladze, C. Papeix, F. Dollé, B. Zalc, C. Lubetzki, M. Bottlaender, F. Turkheimer, B. Stankoff, Dynamic imaging of individual remyelination profiles in multiple sclerosis, *Annals of Neurology* 79 (5) (2016) 726–738.
- [2] P. Patrikios, C. Stadelmann, A. Kutzelnigg, H. Rauschka, M. Schmidbauer, H. Laursen, P. S. Sorensen, W. Brück, C. Lucchinetti, H. Lassmann, Remyelination is extensive in a subset of multiple sclerosis patients, *Brain* 129 (12) (2006) 3165–3172. doi:10.1093/brain/awl217.
- [3] A. Petiet, I. Adanyeguh, M.-S. Aigrot, E. Poirion, B. Nait-Oumesmar, M. Santin, B. Stankoff, Ultrahigh field imaging of myelin disease models: Toward specific markers of myelin integrity?, *Journal of Comparative Neurology* 527 (13) (2019) 2179–2189. doi:10.1002/cne.24598.
- [4] Z. Wang, C. Vandersteen, T. Demarcy, D. Gnansia, C. Raffaelli, N. Guevara, H. Delingette, Deep Learning Based Metal Artifacts Reduction in Post-operative Cochlear Implant CT Imaging, in: D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, A. Khan (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Springer International Publishing, Cham, 2019, pp. 121–129.
- [5] Y. Chen, F. Yu, L. Luo, C. Toumoulin, Improving abdomen tumor low-dose CT images using dictionary learning based patch processing and unsharp filtering, in: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013, pp. 4014–4017.
- [6] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, G. Wang, Low-Dose X-ray CT Reconstruction via Dictionary Learning, *IEEE Transactions on Medical Imaging* 31 (9) (2012) 1682–1697.
- [7] Y. Zhang, X. Mou, G. Wang, H. Yu, Tensor-Based Dictionary Learning for Spectral CT Reconstruction, *IEEE Transactions on Medical Imaging* 36 (1) (2017) 142–154.
- [8] Y. Liu, J. Ma, Y. Fan, Z. Liang, Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction, *Physics in Medicine and Biology* 57 (23) (2012) 7923–7956.
- [9] Z. Tian, X. Jia, K. Yuan, T. Pan, S. B. Jiang, Low-dose CT reconstruction via edge-preserving total variation regularization, *Physics in Medicine and Biology* 56 (18) (2011) 5949–5967.
- [10] K. Bahrami, F. Shi, X. Zong, H. W. Shin, H. An, D. Shen, Reconstruction of 7T-Like Images From 3T MRI, *IEEE Transactions on Medical Imaging* 35 (9) (2016) 2085–2097.
- [11] S. Kaplan, Y.-M. Zhu, Full-Dose PET Image Estimation from Low-Dose PET Image Using Deep Learning: a Pilot Study, *Journal of Digital Imaging* 32 (5) (2019) 773–778.
- [12] A. Hagiwara, Y. Otsuka, M. Hori, Y. Tachibana, K. Yokoyama, S. Fujita, C. Andica, K. Kamagata, R. Irie, S. Koshino, T. Maekawa, L. Chougar, A. Wada, M. Takemura, N. Hattori, S. Aoki, Improving the Quality of Synthetic FLAIR Images with Deep Learning Using a Conditional Generative Adversarial Network for Pixel-by-Pixel Image Translation, *American Journal of Neuroradiology*.
- [13] S. Roy, A. Carass, N. Shiee, D. L. Pham, J. L. Prince, MR contrast synthesis for lesion segmentation, in: *Proc IEEE Int Symp Biomed Imaging*, 2010, pp. 932–935.
- [14] D. H. Ye, D. Zikic, B. Glocker, A. Criminisi, E. Konukoglu, Modality propagation: coherent synthesis of subject-specific scans with data-driven regularization, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2013, pp. 606–613.
- [15] V. Sevetlidis, M. V. Giuffrida, S. A. Tsiftaris, Whole Image Synthesis Using a Deep Encoder-Decoder Network, in: *Simulation and Synthesis in Medical Imaging*, Springer International Publishing, Cham, 2016, pp. 127–137.
- [16] A. Chatsias, T. Joyce, M. V. Giuffrida, S. A. Tsiftaris, Multimodal MR Synthesis via Modality-Invariant Latent Representation, *IEEE Transactions on Medical Imaging* 37 (3) (2018) 803–814.
- [17] W. Wei, E. Poirion, B. Bodini, S. Durrleman, O. Colliot, B. Stankoff, N. Ayache, Fluid-attenuated inversion recovery MRI synthesis from multisequence MRI using three-dimensional fully convolutional networks for multiple sclerosis, *Journal of Medical Imaging* 6 (1) (2019) 1–9.
- [18] N. Burgos, M. J. Cardoso, K. Thielemans, M. Modat, S. Pedemonte, J. Dickson, A. Barnes, R. Ahmed, C. J. Mahoney, J. M. Schott, J. S. Duncan, D. Atkinson, S. R. Arridge, B. F. Hutton, S. Ourselin, Attenuation Correction Synthesis for Hybrid PET-MR Scanners: Application to Brain Studies, *IEEE Transactions on Medical Imaging* 33 (12) (2014) 2332–2341.
- [19] J. Lee, A. Carass, A. Jog, C. Zhao, J. L. Prince, Multi-atlas-based CT synthesis from conventional MRI with patch-based refinement for MRI-based radiotherapy planning, in: M. A. Styner, E. D. Angelini (Eds.), *Medical Imaging 2017: Image Processing*, Vol. 10133, International Society for Optics and Photonics, SPIE, 2017, pp. 434–439.
- [20] T. Huynh, Y. Gao, J. Kang, L. Wang, P. Zhang, J. Lian, D. Shen, Estimating CT Image From MRI Data Using Structured Random Forest and Auto-Context Model, *IEEE Trans Med Imaging* 35 (1) (2016) 174–183.
- [21] H. Van Nguyen, K. Zhou, R. Vemulapalli, Cross-domain synthesis of medical images using efficient location-sensitive deep network, in: N. Navab, J. Hornegger, W. M. Wells, A. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer International Publishing, Cham, 2015, pp. 677–684.
- [22] H. Zhang, I. J. Goodfellow, D. N. Metaxas, A. Odena, Self-Attention Generative Adversarial Networks, arXiv:1805.08318.
- [23] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, Image-to-Image Translation with Conditional Adversarial Networks, arxiv.
- [24] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, in: *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.
- [25] H. Choi, D. S. Lee, Generation of Structural MR Images from Amyloid PET: Application to MR-Less Quantification, *Journal of Nuclear Medicine* 59 (7) (2018) 1111–1117.
- [26] L. Bi, J. Kim, A. Kumar, D. Feng, M. Fulham, Synthesis of Positron Emission Tomography (PET) Images via Multi-channel Generative Adversarial Networks (GANs), in: M. J. Cardoso, T. Arbel, F. Gao, B. Kainz, T. van Walsum, K. Shi, K. K. Bhatia, R. Peter, T. Vercauteren, M. Reyes, A. Dalca, R. Wiest, W. Niessen, B. J. Emmer (Eds.), *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*, Springer International Publishing, Cham, 2017, pp. 43–51.
- [27] P. Costa, A. Galdran, M. I. Meyer, M. Niemeijer, M. Abràmoff, A. M. Mendonça, A. Campilho, End-to-End Adversarial Retinal Image Synthesis, *IEEE Transactions on Medical Imaging* 37 (3) (2018) 781–791.
- [28] H. Zhao, H. Li, S. Maurer-Stroh, L. Cheng, Synthesizing retinal and neuronal images with generative adversarial nets, *Medical Image Analysis* 49 (2018) 14–26.
- [29] Y. Hu, E. Gibson, L.-L. Lee, W. Xie, D. C. Barratt, T. Vercauteren, J. A. Noble, Freehand Ultrasound Image Simulation with Spatially-Conditioned Generative Adversarial Networks, in: M. J. Cardoso, T. Arbel, F. Gao, B. Kainz, T. van Wal-

- sum, K. Shi, K. K. Bhatia, R. Peter, T. Vercauteren, M. Reyes, A. Dalca, R. Wiest, W. Niessen, B. J. Emmer (Eds.), *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment*, Springer International Publishing, Cham, 2017, pp. 105–115.
- [30] F. Mahmood, R. Chen, N. J. Durr, Unsupervised Reverse Domain Adaptation for Synthetic Medical Images via Adversarial Training, *IEEE Transactions on Medical Imaging* 37 (12).
- [31] A. Sikka, S. V. Peri, D. R. Bathula, MRI to FDG-PET: Cross-Modal Synthesis Using 3D U-Net for Multi-modal Alzheimer’s Classification, in: A. Gooya, O. Goksel, I. Oguz, N. Burgos (Eds.), *Simulation and Synthesis in Medical Imaging*, Springer International Publishing, Cham, 2018, pp. 80–89.
- [32] R. Li, W. Zhang, H.-I. Suk, L. Wang, J. Li, D. Shen, S. Ji, Deep Learning Based Imaging Data Completion for Improved Brain Disease Diagnosis, in: *MICCAI 2014*, Vol. 8675 of LNCS, Springer, 2014, pp. 305–312.
- [33] Y. Pan, M. Liu, C. Lian, T. Zhou, Y. Xia, D. Shen, Synthesizing Missing PET from MRI with Cycle-consistent Generative Adversarial Networks for Alzheimer’s Disease Diagnosis, in: A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, G. Fichtinger (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, Springer International Publishing, Cham, 2018, pp. 455–463.
- [34] Y. Wang, L. Zhou, L. Wang, B. Yu, C. Zu, D. S. Lalush, W. Lin, X. Wu, J. Zhou, D. Shen, Locality Adaptive Multi-modality GANs for High-Quality PET Image Synthesis, in: *MICCAI 2018*, Springer, Cham, pp. 329–337.
- [35] W. Wei, E. Poirion, B. Bodini, S. Durrleman, N. Ayache, B. Stankoff, O. Colliot, Predicting PET-derived demyelination from multimodal MRI using sketcher-refiner adversarial training for multiple sclerosis, *Medical Image Analysis* 58 (2019) 101546.
- [36] W. Wei, E. Poirion, B. Bodini, S. Durrleman, N. Ayache, B. Stankoff, O. Colliot, Learning Myelin Content in Multiple Sclerosis from Multimodal MRI Through Adversarial Training, in: *MICCAI 2018*, Springer, Cham, pp. 514–522.
- [37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, I. Polosukhin, Attention is all you need, in: *Advances in Neural Information Processing Systems* 30, Curran Associates, Inc., 2017, pp. 5998–6008.
- [38] T. Shen, J. Jiang, T. Zhou, S. Pan, G. Long, C. Zhang, DiSAN: directional self-attention network for RNN/CNN-free language understanding, in: *The Thirty-Second AAAI Conference on Artificial Intelligence*, Association for the Advancement of Artificial Intelligence (AAAI), 2018, pp. 5446–5455.
- [39] A. Ambartsoumian, F. Popowich, Self-Attention: A Better Building Block for Sentiment Analysis Neural Network Classifiers, in: *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Association for Computational Linguistics, 2018, pp. 130–139.
- [40] X. Wang, R. Girshick, A. Gupta, K. He, Non-local neural networks, in: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7794–7803.
- [41] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, H. Lu, Dual attention network for scene segmentation, in: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [42] H. Tang, D. Xu, N. Sebe, Y. Wang, J. J. Corso, Y. Yan, Multi-Channel Attention Selection GAN With Cascaded Semantic Guidance for Cross-View Image Translation, in: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [43] N. Kodali, J. Abernethy, J. Hays, Z. Kira, On Convergence and Stability of GANs (2017). [arXiv:1705.07215](https://arxiv.org/abs/1705.07215).
- [44] L. Xiang, Q. Wang, D. Nie, L. Zhang, X. Jin, Y. Qiao, D. Shen, Deep embedding convolutional neural network for synthesizing CT image from t1-weighted MR image, *Medical Image Analysis* 47 (2018) 31–44.
- [45] J. F. Kurtzke, Rating neurologic impairment in multiple sclerosis, *Neurology* 33 (11) (1983) 1444–1444.
- [46] R. H. Roxburgh, S. R. Seaman, T. Masterman, A. E. Hensiek, S. J. Sawcer, S. Vukusic, I. Achiti, C. Confavreux, M. Coustans, E. le Page, G. Edan, G. V. McDonnell, S. Hawkins, M. Trojano, M. Liguori, E. Cocco, M. G. Marrosu, F. Tesser, M. A. Leone, A. Weber, F. Zipp, B. Mitterski, J. T. Epplen, A. Oturai, P. S. Sørensen, E. G. Celius, N. T. Lara, X. Montalban, P. Villoslada, A. M. Silva, M. Marta, I. Leite, B. Dubois, J. Rubio, H. Butzkueven, T. Kilpatrick, M. P. Mycko, K. W. Selmaj, M. E. Rio, M. Sá, G. Salemi, G. Savettieri, J. Hillert, D. Compston, Multiple sclerosis severity score, *Neurology* 64 (7) (2005) 1144–1151.
- [47] J. Logan, J. S. Fowler, N. D. Volkow, G.-J. Wang, Y.-S. Ding, D. L. Alexoff, Distribution Volume Ratios without Blood Sampling from Graphical Analysis of PET Data, *Journal of Cerebral Blood Flow & Metabolism* 16 (5) (1996) 834–840.
- [48] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, S. M. Smith, Fsl, *NeuroImage* 62 (2) (2012) 782 – 790, 20 YEARS OF fMRI.
- [49] B. Fischl, Freesurfer, *Neuroimage* 62 (2) (2012) 774–781.
- [50] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, in: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Vol. 9351 of LNCS, Springer, 2015, pp. 234–241.
- [51] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, in: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [52] S. M. Smith, Fast robust automated brain extraction, *Human Brain Mapping* 17 (3) (2002) 143–155.
- [53] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, J. C. Gee, N4itk: Improved n3 bias correction, *IEEE Transactions on Medical Imaging* 29 (6) (2010) 1310–1320.
- [54] M. Veronese, B. Bodini, D. García-Lorenzo, M. Battaglini, S. Bongarzone, C. Comtat, M. Bottlaender, B. Stankoff, F. E. Turkheimer, Quantification of [11C]PIB PET for Imaging Myelin in the Human Brain: A Test—Retest Reproducibility Study in High-Resolution Research Tomography, *Journal of Cerebral Blood Flow & Metabolism* 35 (11) (2015) 1771–1782.
- [55] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, software available from tensorflow.org (2015).
- [56] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, X. Chen, Improved techniques for training gans, in: *Advances in Neural Information Processing Systems* 29, Curran Associates, Inc., 2016, pp. 2234–2242.
- [57] Y. Wang, L. Zhang, J. van de Weijer, Ensembles of Generative Adversarial Networks, *CoRR* abs/1612.00991. [arXiv:1612.00991](https://arxiv.org/abs/1612.00991).
- [58] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, D. Shen, Medical Image Synthesis with Deep Convolutional Adversarial Networks, *IEEE Transactions on Biomedical Engineering* 65 (12) (2018) 2720–2730.