Synthetic Generation and Latent Projection Denoising of Rim Lesions in Multiple Sclerosis

Alexandra G. Roberts
agr78@cornell.edu

Ha M. Luu
hal4025@med.cornell.edu

Mert Şişman
ms2893@cornell.edus

Dominick J. Romano
djr327@cornell.edu

Alexey V. Dimov
avd34@cornell.edu

Ceren Tozlu
ct595@cornell.edu

Pascal Spincemille
pas2018@med.cornell.edu

Ilhami Kovanlikaya
ilk2002@med.cornell.edu

Susan A. Gauthier
sag2015@med.cornell.edu

Thanh D. Nguyen
tdn2001@med.cornell.edu

Yi Wang
yiwang@med.cornell.edu

Abstract

Quantitative susceptibility maps from magnetic resonance images can provide both prognostic and diagnostic information in multiple sclerosis, a neurodegenerative disease characterized by the formation of lesions in white matter brain tissue. In particular, susceptibility maps provide adequate contrast to distinguish between "rim" lesions, surrounded by deposited paramagnetic iron, and "non-rim" lesion types. These paramagnetic rim lesions (PRLs) are an emerging biomarker in multiple sclerosis. Much effort has been devoted to both detection and segmentation of such lesions to monitor longitudinal change. As paramagnetic rim lesions are rare, addressing this problem requires confronting the class imbalance between rim and non-rim lesions. We produce synthetic quantitative susceptibility maps of paramagnetic rim lesions and show that inclusion of such synthetic data improves classifier performance and provide a multi-channel extension to generate accompanying contrasts and probabilistic segmentation maps. We exploit the projection capability of our trained generative network to demonstrate a novel denoising approach that allows us to train on ambiguous rim cases and substantially increase the minority class. We show that both synthetic lesion synthesis and our proposed rim lesion label denoising method best approximate the unseen rim lesion distribution and improve detection in a clinically interpretable manner. We release our code and generated data at https://github.com/agr78/PRLx-GAN *upon publication.*

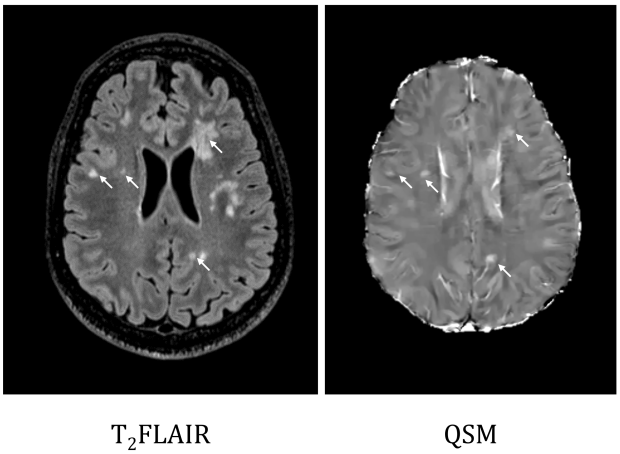

Figure 1. Example of MS patient with lesions depicted on qualitative T_2FLAIR and whole brain QSM. Lesions appearing on both contrasts are indicated by arrows.

1. Introduction

1.1. Background

Multiple sclerosis (MS) is a debilitating neurodegenerative disease [40] with a rising global presence [60]. MS is the most common demyelinating disorder [62] and there is currently no cure [15], though a variety of disease-modifying therapies exist to slow symptom progression and improve quality of life [48].

The efficacy of these therapies is measured by the appearance of gadolinium-enhancing lesions in brain white matter as identified on magnetic resonance imaging

(MRI) [19]. The importance of MRI in both MS diagnosis and prognosis [30] motivates the development of automatic lesion detection and segmentation algorithms to address this growing clinical need [10].

1.2. Imaging techniques

Much progress on the understanding of MS has been gained through the qualitative $\mathrm{T_1w}$ and $\mathrm{T_2FLAIR}$ MRI contrasts [13]. However, longitudinal studies increasingly rely on quantitative susceptibility maps (QSM) as robust biomarkers in MS disease progression [7, 14, 58, 68]. Of critical importance is the visualization of paramagnetic rim lesions linked to symptom severity [55].

Due to the iron depositions surrounding the rim of these lesions [21], susceptibility contrasts like QSM are required to differentiate between lesion subtypes [22]. Co-registered $\mathrm{T_2FLAIR}$ and whole brain QSM [46] are shown in an example MS patient in Figure 1, with lesions indicated by the white arrows. Iron is involved in a variety of neuroinflammatory diseases and deposition often increases with inflammatory response [61]. Activated immune cell microglia at the edge of the lesion are the primary source of iron, generating contrast between the paramagnetic rim and diamagnetic lesion core [25].

As QSM directly measures small changes in the applied magnetic field arising from paramagnetic or diamagnetic tissue content, it is a reliable method to detect rim lesions [45]. Figure 2 illustrates the need for QSM to enable rim identification as compared to $\mathrm{T_2FLAIR}$.

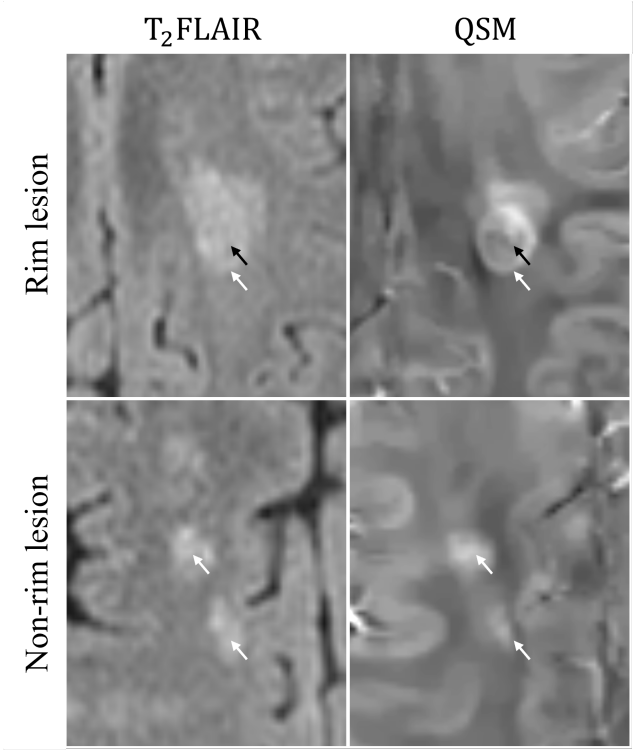

Figure 2. Example of rim (top) and non-rim (bottom) lesion visualization on $\mathrm{T_2FLAIR}$ (left) and QSM (right). Note that rim lesions can only be differentiated from non-rim lesions on QSM. The rim lesion is composed of a hypointense core (black arrow) and hyperintense rim (white arrow), contrast to a hyperintense non-rim lesion.

1.3. Lesion class imbalance

Though paramagnetic rim lesions differentiate between MS and other neurodegenerative disorders with high specificity [38], only 10% of all MS lesions are estimated to be rim lesions [43], introducing a class imbalance problem for detection and segmentation algorithms.

Despite consensus [4] on rim lesion characteristics, ambiguous cases remain when expert readers disagree on these criteria. It is desirable to "denoise" such lesions in order to make use of all possible rim lesion data.

1.4. Contributions

We present a novel sample denoising approach based on generated synthetic quantitative susceptibility maps of paramagnetic rim lesions. We exploit the projection capability [8] in our trained generative network to "denoise" ambiguous samples by recovering unambiguous synthesized samples. We evaluate the effect of training with said data in the rim classification problem and demonstrate improvement in trained detectors when denoised synthetic data is included. This approach enables us to train with noisy or ambiguous labels by recovering their synthetic analog from the trained generator network. To our knowledge, our proposed method is the first to allow ambiguous or contested rim lesion labels to augment the rare, unambiguous rim lesion class during rim detection. Finally, we provide a multi-contrast extension to enable generation of susceptibility maps, $\mathrm{T_2FLAIR}$ images, and probabilistic rim lesion segmentations, broadening the utility of our framework.

2. Related Works

2.1. Synthetic MRI

Synthetic MRI data ranges from classical representations generated via signal models [2, 5] to more recent deep learning approaches aiming to estimate mappings over synthetic data that are applicable to *in vivo* data [17, 42] and across various contrasts and resolutions [24]. Given pretrained models and approximate source and target distributions [63], transfer learning approaches can address the challenges of gathering sufficient MRI data [33, 39, 57].

Another subset of solutions focus on generating synthetic data for small datasets [53] and label scarcity arising from new [59] or lethal [1] pathologies. Rim segmentation efforts have addressed the class imbalance problem by

oversampling the minority class in the latent space [9, 66], included in our comparison.

Synthetic lesions in multiple sclerosis have been generated [49] using a variational autoencoder on qualitative $\mathrm{T_2FLAIR}$. Like our work, this model aims to generate synthetic MS lesions, differing in that the whole brain learned mapping is between healthy cases and MS patients on qualitative $\mathrm{T_2FLAIR}$.

We train a generative adversarial network [16] (GAN) with the goal of learning the mapping from random noise initializing the generator model to synthetic, quantitative, paramagnetic rim lesions. We do this using susceptibility images as rim lesions are differentiable from non-rim lesions only on QSM.

2.2. Latent projection denoising

From our choice in architecture arises an opportunity to recover ambiguous or "noisy" rim lesions by projecting their latent vectors into the latent space of a pretrained GAN synthesizing only unambiguous rim lesions, closely related to GAN inversion [64]. Namely, these ambiguous cases are lesions where expert raters disagree on the label. The presence of label noise has been addressed by conditional GANs [41, 54] (cGAN), which are trained on both majority and minority class labels.

Given the data imbalance between rim and non-rim lesions, we focus on the minority class rather than the majority class and we show its inclusion nearly doubles the required training time. We seek to make use of ambiguous, noisy real rim lesions by training the generator only the unambiguous rim lesion minority class. We use the projection into the learned latent space from unambiguous (or noiseless) rim lesions to recover denoised lesions. Other works related to this effort include modeling label noise as a latent space shift [23], a technique applied to correct classifications rather than augment data.

Also related is estimation of the noise transition matrix [3] to calibrate classifiers trained on noisy labels, which requires some understanding of the noise distribution. Perhaps most relevant is the use of GAN inversion for undersampled MRI reconstruction [31], which deals with the presence of instrument noise rather than mitigation of more subjective label noise.

3. Method

3.1. Dataset

A group of 256 MS subjects (mean age, $46.2 \pm 11.8\ years$, 79 men (30.8%), 177 women (69.2%)) were imaged on a $3T$ Magnetom Skyra scanner. An axial $3D$ multi-echo GRE sequence was used to acquire phase data for QSM with $FOV = 24.0\ cm$, $\frac{TE_1}{dTE} = 6.28/4.06\ ms$, 8 echos, $T_R = 40\ ms$, voxel size $0.4 \times 0.4 \times 1\ mm^3$.

QSM was reconstructed using referenced morphology-enabled dipole inversion (MEDI+0) [34, 35]. The dataset was prepared as follows [37]. Each lesion on QSM was cropped to an image patch of $64 \times 64 \times 24$ voxels. Two expert readers independently created rim lesion ground truth labels according to the recent consensus statement [4].

Lesions were classified as "rim" only if both readers agreed on their paramagnetic rim lesion status, otherwise they were classified as "non-rim". Hyperintense rim areas of each identified lesion were manually traced and checked by the same two readers. Critical to our denoising approach is the definition of an **ambiguous rim lesion**, where only one of the two expert raters described the lesion as a paramagnetic rim lesion, and the label is noisy.

3.2. Radiologist assessment

An expert radiologist guided by the recent rim lesion consensus [4] reviewed 110 uncurated example slices of real and synthetic paramagnetic rim lesions. In two separate experiments, the lesions were classified as "real" or "synthetic" and categorized as "rim" or "non-rim".

3.3. Classifier network

To evaluate the improvement in classification with added synthetic data, a simple convolutional neural network classifier for binary classification was implemented [32]. The network consisted of 6 convolutional layer units including pooling and batch normalization operations, followed by a rectified linear unit (ReLU) activation function. The classifier was trained for 25 epochs with the Adam optimizer with learning rate 10^{-3} using the cross-entropy loss function.

3.4. Generative network

StyleGAN2 with adaptive discriminator augmentation (StyleGAN2-ADA, shortened in this work for brevity to "ADA-GAN") [28] was trained via Frechet Inception Distance [20] (FID) minimization with the Adam optimizer (learning rate 2.5×10^{-3}, first and second moment decay rates $\beta_1 = 0.9$, $\beta_2 = 0.99$, respectively) using overfitting heuristic $r_t = 0.6$. The training dataset contained 200 rim lesions and 400 non-rim lesions. For testing, 60 rim lesions and 120 non-rim lesions were withheld from both the generator and classifier. Training required 32 hours with 8 NVIDIA GeForce RTX 2080 graphics cards to generate $25,000$ synthetic rim lesions. An overview of the architecture is given in Figure 3.

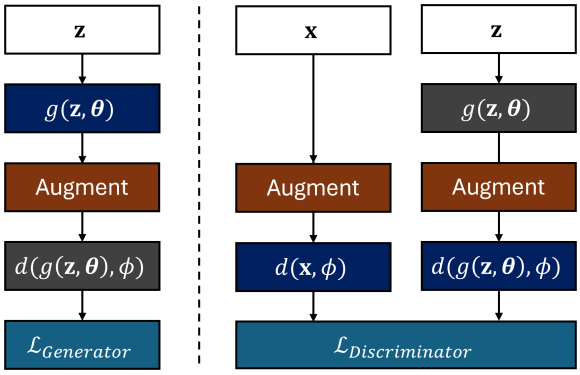

Figure 3. Schematic of the GAN with adaptive discriminator augmentation blocks used to generate synthetic rim lesion data from limited datasets. The architecture consists of generative and discriminative networks $g(\mathbf{z}, \theta)$ and $d(., \phi)$ and receives inputs real images $\mathbf{x}$ and latent variable $\mathbf{z} \sim \mathcal{N}(0, 1)$. Adapted from [28].

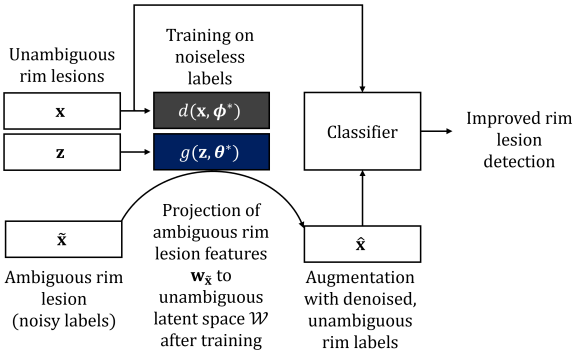

Figure 4. Outline of denoising approach, beginning with generative model training on unambiguous rim lesion $\mathbf{x}$. After training, features of ambiguous rim lesions (with noisy labels) are extracted to latent variable $\mathbf{w}_{\tilde{\mathbf{x}}}$ and projected onto the unambiguous latent space $\mathcal{W}$. Finally, the real rim lesion training data is augmented with these "denoised" rim lesions and the classifier performance is seen to improve.

3.5. Denoising approach

We exploit the projection capability of the chosen network [27] to create synthetic denoised samples from ambiguous or noisy samples $\tilde{\mathbf{x}}$. We note that this particular application is not aiming to remove instrument or random noise within the image, but noise in the labeling process that results in ambiguous lesions where raters disagree. As such, the ambiguous rim lesion image is transformed to be unambiguous and the label is "denoised". Denoised samples $\hat{\mathbf{x}} = s(\mathbf{w}^*, \theta_s^*)$ are computed from the synthesis network s module in the generator network g. The latent projection $\mathbf{w}^*$ is the "closest" intermediate latent space vector to noisy image $\tilde{\mathbf{x}}$ features, $\mathbf{w}_{\tilde{\mathbf{x}}}$, supported by trained generator g, described in Figure 4. This optimal latent space vector $\mathbf{w}^*$ is obtained as in [29]

$$\mathbf{w}^* = \text{argmin}_{\mathbf{w}} L_P(\tilde{\mathbf{x}}, s(\mathbf{w}, \theta_s^*)) + \alpha \sum_{i,j} L_N(n_i, n_j) \quad (1)$$

Where L_P is the perceptual loss [67] over extracted features [51] at each layer and L_N is the noise regularization term (scaled by parameter α) from added noise vector $\mathbf{n}$ at original resolution i and downsampled resolution j, $n_i \sim \mathcal{N}(0, I)$, and n_j.

Equation 1 is solved using the Adam optimizer over 1000 iterations with regularization $\alpha = 10^5$, first and second moment decay rates $\beta_1 = 0.9$, $\beta_2 = 0.99$, and a scheduled learning rate initialized at 10^{-1}. Rather than augmenting the dataset with synthetic rim lesions $\mathbf{x}'$ mapped from random noise, we augment with the denoised projections $\hat{\mathbf{x}}$ of noisy samples $\tilde{\mathbf{x}}$, related by $\hat{\mathbf{x}} = s(\mathbf{w}^*, \theta_s^*)$ from the generator g with weights θ^*. An overview is depicted in Figure 5. We term this augmentation to increase the minority sample class via latent projection denoising as "ADA-GAN-LD".

3.6. Ambiguous rim lesion denoising

We apply the aforementioned denoising projection in the latent space of ADA-GAN to transform 100 ambiguous rim lesions - after training only on unambiguous rim lesions. We compare the resulting ADA-GAN-LD dataset to a number of other possible augmentations described in later sections. We also combine synthetic data and denoised rim lesions in an augmentation referred to as "ADA-GAN+LD.

3.7. Conditional generative network

The standard network was retrained with a conditional GAN with "rim" and "non-rim" labels. The training dataset above was expanded to include 400 non-rim lesions and required 48 hours of computing time with 8 NVIDIA GeForce RTX 2080 graphics cards to generate $25,000$ synthetic rim and non-rim lesions.

3.8. Augmentation comparison

The generative model, ADA-GAN, was compared to simple affine transformation augmentation ("Affine") and the learned synthetic minority oversampling technique, "DeepSMOTE" [9] by training separate classifier networks on each augmentation. An ablation study was also performed to find the optimal combination of synthetic, conditioned synthetic and/or denoised data. The FID was computed between each augmented training dataset distribution and the unseen test rim lesion distribution.

3.9. Clinical interpretation

To improve detection explainability, we produce class activation maps [50, 69] (CAMs) from the unaugmented classi-

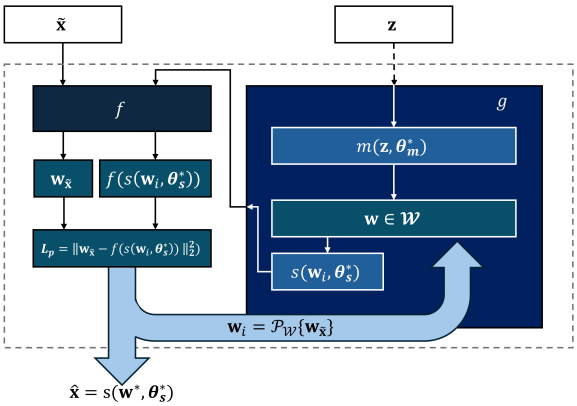

Figure 5. A simplified representation (adapted from [27]) of the mapping m and synthesis s modules used to enable denoising via the trained generative network g. Given an ambiguous rim $\tilde{\mathbf{x}}$ lesion, denoising occurs via latent projection $\mathcal{P}$ after feature extraction by pretrained network f. The denoised projection $\mathbf{w}^*$ results from minimizing the perceptual loss L_2 term (L_P) with added noise regularization (Equation 1). After training with the architecture in Figure 3, the synthesis module s decodes the denoised projection $\hat{\mathbf{x}}$ corresponding to the "closest" unambiguous rim lesion in the latent space $\mathcal{W}$, enabling the augmentation in Figure 4.

fier and the classifier augmented with our proposed denoising method.

3.10. Multi-contrast generalization

As paramagnetic rim lesion detection methods often rely on both susceptibility contrasts and T_2FLAIR [36], we extend the generative model and denoising approach with a contrast dimension comprised of susceptibility, rim mask, and T_2FLAIR channels.

In particular, the inclusion of a rim lesion mask or probabilistic map allows the synthetic and denoised data to be used in training segmentation algorithms. The standard network was retrained with this extension and required 34 hours of computing time with 8 NVIDIA GeForce RTX 2080 graphics cards to generate $25,000$ multi-contrast maps.

4. Results

4.1. Lesion distribution

From the cohort, a total of 260 rim lesions (3.3%) and 7720 non-rim lesions (96.7%) were identified. Another 177 lesions were identified as "rim" by merely one of the two expert readers. Cohen's kappa agreement between the two readers is 0.73 (substantial agreement). Out of 256 patients, 92 (35.9%) had at least one rim lesion - 35 (13.7%) had 1 rim lesion, 18 (7%) had 2 rim lesions, and 36 (14.1%) had from 3 to 12 rim lesions. The ambiguous rim lesions were defined as the 177 paramagnetic rim lesions identified

Dataset	Rim lesion fraction	Real image fraction
Real rims	0.31	0.55
Synthetic rims	0.4	0.29

Table 1. Our expert radiologist estimated nearly half (0.4) of the synthetic rim lesions to be a true rim lesion as opposed to about one third (0.31) of the real rim lesions. In a separate experiment, and nearly a third (0.29) of the uncurated synthetic lesions were estimated to be real images alongside 0.55 of the real rim lesions.

Augmentation	FID from test rim lesions
Real non-rim only	46.35
DeepSMOTE rims	36.48
Real rim only	34.49
ADA-GAN rims	34.36
ADA-GAN+LD rims (ours)	34.24
ADA-GAN-LD rims (ours)	**34.17**

Table 2. Computed FID between each training dataset distribution after augmentation and the unseen test rim lesion distribution. As expected, the majority non-rim lesion distribution is the furthest from the minority test rim lesion distribution. Real rim lesions, DeepSMOTE augmentation, and ADA-GAN augmentation all bring the training dataset of rim lesions closer to the unseen test rim lesion distribution, with our denoising approach minimizing the FID.

where one of the two expert raters classified the lesion as a rim lesion.

4.2. Radiologist grading

From 55 real and 55 synthetic lesions, an expert radiologist, estimated nearly half (0.4) of the synthetic rim lesions to be true rim lesions as opposed to about one third (0.29) of the real rim lesions. In a separate experiment, nearly a third (0.31) of the uncurated synthetic lesions were estimated to be real images alongside just over half (0.55) of the real rim lesions. We remark that the overall fraction of rim lesions identified is a result of grading lesions over a single slice rather than the entire $3D$ volume.

We interpret these findings to suggest that convincing synthetic rims capture the less ambiguous features in rim definition, leading to a higher fraction of synthetic cases being classified as rim lesions. An example of an realistic synthetic rim (identified as real by an expert radiologist) and a similar real rim lesion is given in Figure 6.

4.3. Rim lesion denoising

Our denoising method allowed us to recover 177 additional rim lesions to expand the minority class from 260 to 437 cases. This augmentation increased our minority class labels by 68%. We projected 100 noisy rim lesions $\tilde{\mathbf{x}}$ into

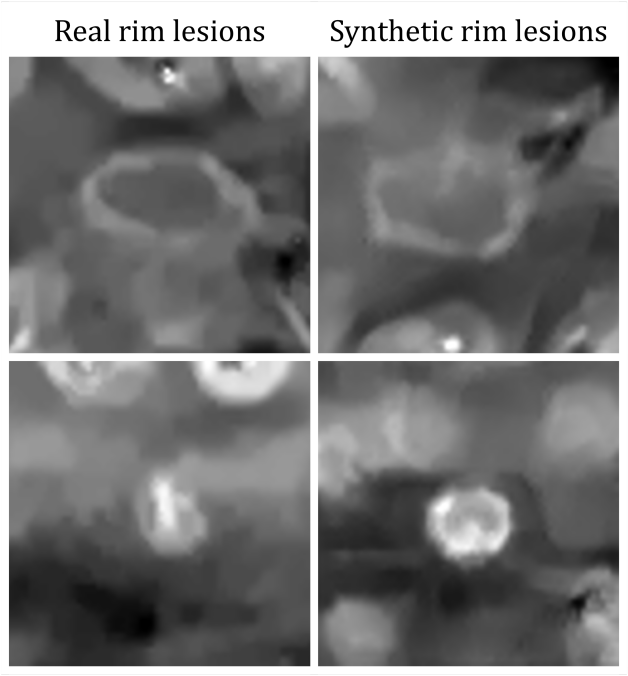

Figure 6. Typical real rim lesions compared to examples of synthetic rim lesions classified as real rim lesions.

the latent space $\mathcal{W}$ of the generator g trained on only unambiguous rim lesions $\mathbf{x}$. We augment the classifier rim lesion training data with these denoised rim lesion samples and see optimal performance. Denoised example rim lesions $\hat{\mathbf{x}}$ are given in Figure 7.

We notice clinically interpretable changes to the denoised image $\hat{\mathbf{x}}$ when decoded from the unambiguous latent space by the synthesizer module s, such as defined hypointense lesion cores and the removal of obscuring artifacts.

4.4. Classifier performance

Evaluating the FID in Table 2 reveals the intuitive result that the majority non-rim lesion distribution is the furthest from the minority test rim lesion distribution. Real rim lesions without augmentation, DeepSMOTE augmentation, and ADA-GAN augmentation draw the training distribution of rim lesions closer to the unseen test rim lesion distribution, with our denoising augmentation approach minimizing the FID. Augmenting the training data with 100 synthetic rim lesions ("ADA-GAN", "DeepSMOTE") improved the classifier performance as seen in Table 3. In particular, accuracy and sensitivity were increased while the precision remained comparable to the dataset with no augmentation during training ("None"). Note that augmenting by simple random affine transformations ("Affine") to increase the number of rim lesions (and replace an equal number of non-rim lesions) slightly improves classifier accuracy. We include our denoising approach ("ADA-GAN-LD") to transform "ambiguous rims", which degrade classifier performance, into the unambiguous rim latent space, resulting in optimal classifier results.

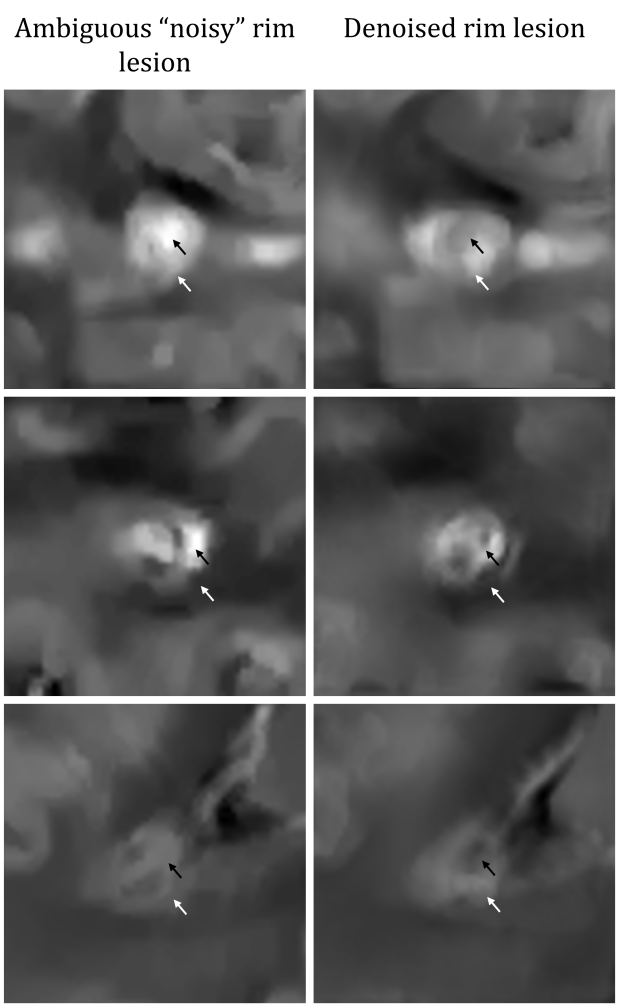

Figure 7. An ambiguous, "noisy" rim lesion $\mathbf{x}$ and the denoised projection $\hat{\mathbf{x}}$. Note the hypointensity (black arrows) at the denoised rim core not present in the original ambiguous rim lesion. Hyperintense rims are annotated with white arrows.

4.5. Ablation study

We compare the standard ADA-GAN architecture to its conditional variant (ADA-cGAN) by comparing training loss curves to determine the effect of training the generator on both rim and non-rim lesions. We note that the conditional generative model, ADA-cGAN, converges to a slightly lower FID (8.79 compared to 14.20) during training as seen in Figure 8. We also compare our denoising method (ADA-GAN-LD) and combine the denoised and synthetic data (ADA-GAN+LD). We find that all variations of im-

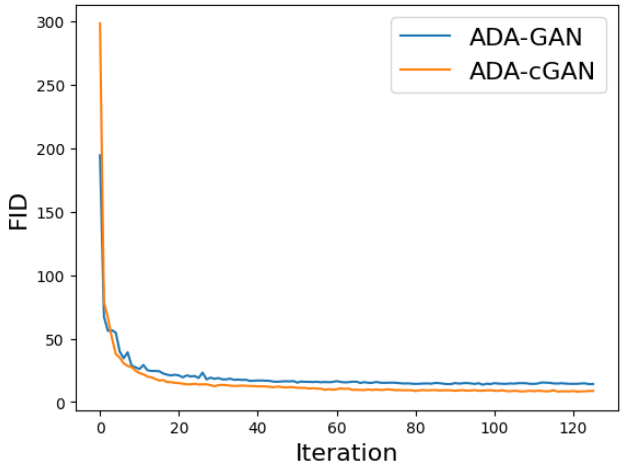

Figure 8. Conditional and standard GAN training curves. Inclusion of the non-rim label allows the training loss to converge to a slightly smaller FID.

Augmentation	Accuracy	Precision	Sensitivity
Ambiguous rims	0.78	0.93	0.83
None	0.79	0.93	0.85
Affine	0.83	0.92	0.83
DeepSMOTE	0.84	0.92	0.91
ADA-GAN-LD (ours)	**0.87**	0.91	**0.95**

Table 3. Classifier performance using various training datasets. Including the noisy label ambiguous rim in the minority rim class results in a slight decrease in performance. All augmentation strategies improve test results, with our denoising approach improving accuracy and sensitivity and offering comparable precision.

Augmentation	Accuracy	Precision	Sensitivity
None	0.79	0.93	0.85
ADA-GAN	0.85	0.91	0.93
ADA-GAN+LD (ours)	0.85	0.92	0.92
ADA-cGAN	0.86	0.91	0.93
ADA-GAN-LD (ours)	**0.87**	0.91	**0.95**

Table 4. Classifier performance during ADA-GAN ablation study. We compare the addition of synthetic data, synthetic and denoised data, conditional synthetic data, and our proposed denoising approach as possible classifier augmentations. All augmentations improve classifier performance, particularly the conditional augmentation and the inclusion of denoised ambiguous rim lesions.

prove classifier performance, particularly our proposed projection denoising augmentation. Including the denoised rim lesions improves classifier accuracy and sensitivity and offers comparable precision as seen in Table 4.

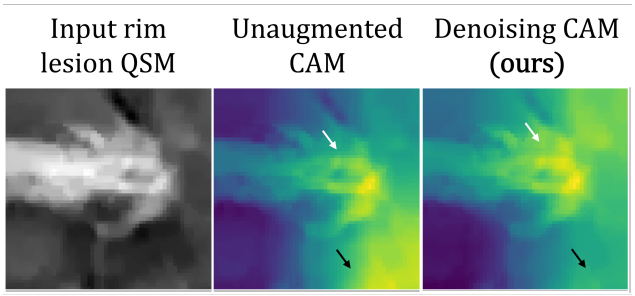

Figure 9. Improved interpretability of class activation maps when using the denoised rim lesion data augmentation, note the shift in intensity from the right corner to the lesion rim (black arrow) and an increase in intensity near the lesion rim (white arrow).

4.6. Class activation maps

Both CAMs in Figure 9 show an emphasis on surrounding white matter, which creates contrast in comparison to demylinated lesions. We notice shift in intensity from the right corner to the lesion rim and note our latent projection denoising augmentation helps identify clinically relevant areas of the lesion. In the map from the classifier trained with our augmentation, both the lesion itself and the hyperintense rim region are highlighted.

4.7. Multi-contrast extension and segmentation

Combining the susceptibility map, T_2FLAIR and paramagnetic rim lesion mask allows realistic generation of new samples as seen in Figure 10. We note the realistic lesion depiction across each contrast as T_2FLAIR lesions typically bound the appearance of the lesion on susceptibility maps. Further, the T_2FLAIR lesion is uniformly hyperintense, as expected. The generated probabilistic map correctly avoids the hypointense lesion core and correctly identifies the paramagnetic rim. Additionally, a variety of rim lesions are well-represented - both rims encircling around a half of the core circumference (right column) and its entirety (left column) are realistically depicted in both susceptibility map and T_2FLAIR images with accurate probabilistic segmentation maps.

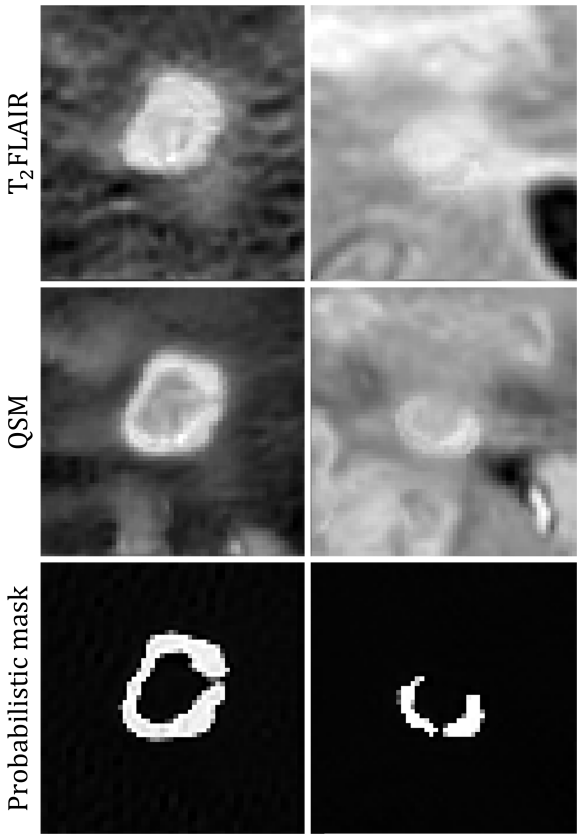

Figure 10. Multi-contrast synthetic example including the lesion susceptibility, probabilistic map, and T_2FLAIR images.

5. Discussion

5.1. Label denoising and class imbalance

In the medical domain, prognostic tools are commonly marred by some form of rater noise, resulting in class uncertainty [26] or varied numerical outcomes [47]. The projection denoising approach we propose only requires access to some unambiguous or noiseless cases to learn a latent space upon which the problematic sample can be projected and increases our minority class dataset by nearly 70%. Though this is an active area of research [56], the variation in quality of snythesized data can require extensive curation [18]. Our latent projection denoising approach produced an augmented distribution closer to the unseen test rim lesion distribution than the purely or combined synthetic augmentations, which may alleviate curation requirements. Future investigation may focus on latent projection denoising into the conditional latent space, which demonstrated improved convergence during training.

5.2. Class rebalancing

We supplement training datasets here with 100 additional lesions for each augmentation method compared but note that the optimal extent of rebalancing with synthetic data is a current topic of interest [65]. Future investigation may focus on the optimal fraction of synthetic training data for the rim lesion detection problem.

5.3. Multi-constrast extension

Jointly generating susceptibility, T_2FLAIR, or other contrasts alongside probabilistic segmentation masks is beneficial beyond the imbalanced class problem. When used in conjunction with augmentation techniques, our method can enable deep learning approaches on small and/or incomplete medical imaging datasets where data collection is slow, laborious and expensive.

5.4. Applications

Other susceptibility contrasts such as source separation [12], myelin imaging [70], and oxygen extraction fraction [52] have clinical value in the treatment of multiple sclerosis. We note that extension to these maps using our proposed multi-contrast method is feasible. Our proposed framework can be generalized into higher dimensions to accommodate additional spatial and/or temporal data, and future work should focus on addressing the need [6, 11, 44] for additional training data required for such applications.

6. Conclusion

We examine the quality of synthetic samples via expert radiologist assessment and show that realistic rim lesions can be acquired from generative modeling on multiple contrasts. We demonstrate the effectiveness of our proposed latent projection denoising of ambiguous rim lesions by comparing the FID between different training and unseen test datasets. We further show the improvement in paramagnetic rim lesion detection on QSM with the inclusion of this denoised data and observe increased interpretability of the classifier class activation maps, indicating clinically relevant predictions resulting from augmentation with realistic synthetic data.

References

[1] Bilal Ahmad, Jun Sun, Qi You, Vasile Palade, and Zhongjie Mao. Brain tumor classification using a combination of variational autoencoders and generative adversarial networks. *Biomedicines*, 10(2):223, 2022. 2

[2] Joachim André, Sami Barrit, and Patrice Jissendi. Synthetic mri for stroke: a qualitative and quantitative pilot study. *Scientific Reports*, 12(1), 2022. 2

[3] HeeSun Bae, Seungjae Shin, Byeonghu Na, JoonHo Jang, Kyungwoo Song, and Il-Chul Moon. From noisy predic-

tion to true label: Noisy prediction calibration via generative model, 2022. 3

[4] F. Bagnato, P. Sati, C. C. Hemond, C. Elliott, S. A. Gauthier, D. M. Harrison, C. Mainero, J. Oh, D. Pitt, R. T. Shinohara, S. A. Smith, B. Trapp, C. J. Azevedo, P. A. Calabresi, R. G. Henry, C. Laule, D. Ontaneda, W. D. Rooney, N. L. Sicotte, D. S. Reich, and M. Absinta. Imaging chronic active lesions in multiple sclerosis: a consensus statement. *Brain*, 147(9):2913–2933, 2024. 1460-2156 Bagnato, Francesca Orcid: 0000-0001-9544-9019 Sati, Pascal Hemond, Christopher C Elliott, Colm Orcid: 0000-0002-3004-523x Gauthier, Susan A Orcid: 0000-0002-4015-1781 Harrison, Daniel M Mainero, Caterina Oh, Jiwon Pitt, David Shinohara, Russell T Smith, Seth A Trapp, Bruce Azevedo, Christina J Calabresi, Peter A Orcid: 0000-0002-7776-6472 Henry, Roland G Laule, Cornelia Ontaneda, Daniel Orcid: 0000-0002-2838-9148 Rooney, William D Sicotte, Nancy L Reich, Daniel S Orcid: 0000-0002-2628-4334 Absinta, Martina Orcid: 0000-0003-0276-383x Roche-Genentech/ MS-1610-37047/Patient Centered Outcomes Research Institute/ RG-1802-30140/National MS Society/ R01 NS104283/NS/NINDS NIH HHS/United States 1750327/Fondazione Regionale per la Ricerca Biomedica Early Career Award/ RG-1707-28775/Atara/ K23 NS126718/NS/NINDS NIH HHS/United States KL2TR001454/TR/NCATS NIH HHS/United States 1r01ns122980-01/nih-ninds/ PA-2107-38081/International Progressive MS Alliance/ MS Society of Canada/ U01NS116776-01/NH/NIH HHS/United States R01 NS122980/NS/NINDS NIH HHS/United States I01CX002160-01A1/Veterans Health Administration/ I01 CX002160/CX/CSRD VA/United States 1U01NS116776-01/Brain Canada/ TL1 TR001454/TR/NCATS NIH HHS/United States R01 NS104403/NS/NINDS NIH HHS/United States R21 NS116434/NS/NINDS NIH HHS/United States 17313/Conrad N. Hilton Foundation/ R01 NS082347/NS/NINDS NIH HHS/United States 2019-1677/Cariplo Foundation/ R35 NS097303/NS/NINDS NIH HHS/United States Natural Sciences and Engineering Research Council of Canada/ R56 NR019306/NR/NINR NIH HHS/United States R01 NS 112274/Erwin Rautenberg Foundation/ 1R21NS1226737-01/International Collaboration on Repair Discoveries/ MS210103/Department of Defense/ Z01 NS003119/Intramural Research Program of NIH-NINDS/ RFA-2203-39325/NMSS Society/ Craig H. Neilsen Foundation/ U01 NS116776/NS/NINDS NIH HHS/United States R01NS082347/Voros Innovation Impact Funds/ Principia/ Journal Article Review England 2024/01/16 Brain. 2024 Sep 3;147(9):2913-2933. doi: 10.1093/brain/awae013. 2, 3

[5] Martina F. Callaghan, Siawoosh Mohammadi, and Nikolaus Weiskopf. Synthetic quantitative mri through relaxometry modelling. *NMR in Biomedicine*, 29(12):1729–1738, 2016. 2

[6] Eric R. Chan, Connor Z. Lin, Matthew A. Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas Guibas, Jonathan Tremblay, Sameh Khamis, Tero Karras, and Gordon Wetzstein. Efficient geometry-aware 3d generative adversarial networks, 2022. 8

[7] W. Chen, S. A. Gauthier, A. Gupta, J. Comunale, T. Liu, S. Wang, M. Pei, D. Pitt, and Y. Wang. Quantitative susceptibility mapping of multiple sclerosis lesions at various ages. *Radiology*, 271(1):183–92, 2014. 1527-1315 Chen, Weiwei Gauthier, Susan A Gupta, Ajay Comunale, Joseph Liu, Tian Wang, Shuai Pei, Mengchao Pitt, David Wang, Yi R01 EB013443/EB/NIBIB NIH HHS/United States R01 NS072370/NS/NINDS NIH HHS/United States Journal Article Research Support, Non-U.S. Gov't United States 2014/01/31 Radiology. 2014 Apr;271(1):183-92. doi: 10.1148/radiol.13130353. Epub 2013 Nov 18. 2

[8] Antonia Creswell and Anil Anthony Bharath. Inverting the generator of a generative adversarial network. *IEEE Transactions on Neural Networks and Learning Systems*, 30(7): 1967–1974, 2019. 2

[9] Damien Dablain, Bartosz Krawczyk, and Nitesh V. Chawla. Deepsmote: Fusing deep learning and smote for imbalanced data, 2021. 3, 4

[10] Alessandro Pasquale De Rosa, Marco Benedetto, Stefano Tagliaferri, Francesco Bardozzo, Alessandro D'Ambrosio, Alvino Bisecco, Antonio Gallo, Mario Cirillo, Roberto Tagliaferri, and Fabrizio Esposito. Consensus of algorithms for lesion segmentation in brain mri studies of multiple sclerosis. *Scientific Reports*, 14(1), 2024. 2

[11] Kangle Deng, Gengshan Yang, Deva Ramanan, and Jun-Yan Zhu. 3d-aware conditional image synthesis, 2023. 8

[12] Alexey V. Dimov, Thanh D. Nguyen, Kelly M. Gillen, Melanie Marcille, Pascal Spincemaille, David Pitt, Susan A. Gauthier, and Yi Wang. Susceptibility source separation from gradient echo data using magnitude decay modeling. *Journal of Neuroimaging*, 32(5):852–859, 2022. 8

[13] Massimo Filippi, Maria A Rocca, Olga Ciccarelli, Nicola De Stefano, Nikos Evangelou, Ludwig Kappos, Alex Rovira, Jaume Sastre-Garriga, Mar Tintorè, Jette L Frederiksen, and et al. Mri criteria for the diagnosis of multiple sclerosis: Magnims consensus guidelines. *The Lancet Neurology*, 15 (3):292–303, 2016. 2

[14] Cristiana Fiscone, Leonardo Rundo, Alessandra Lugaresi, David Neil Manners, Kieren Allinson, Elisa Baldin, Gianfranco Vornetti, Raffaele Lodi, Caterina Tonon, Claudia Testa, Mauro Castelli, and Fulvio Zaccagna. Assessing robustness of quantitative susceptibility-based mri radiomic features in patients with multiple sclerosis. *Scientific Reports*, 13(1):16239, 2023. 2

[15] K. Gohil. Multiple sclerosis: Progress, but no cure. *P t*, 40 (9):604–5, 2015. Gohil, Kunj Journal Article United States 2015/09/30 P T. 2015 Sep;40(9):604-5. 1

[16] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014. 3

[17] Karthik Gopinath, Andrew Hoopes, Daniel C. Alexander, Steven E. Arnold, Yael Balbastre, Benjamin Billot, Adrià Casamitjana, You Cheng, Russ Yue Zhi Chua, Brian L. Edlow, Bruce Fischl, Harshvardhan Gazula, Malte Hoffmann, C. Dirk Keene, Seunghoi Kim, W. Taylor Kimberly, Sonia Laguna, Kathleen E. Larson, Koen Van Leemput, Oula

Puonti, Livia M. Rodrigues, Matthew S. Rosen, Henry F. J. Tregidgo, Divya Varadarajan, Sean I. Young, Adrian V. Dalca, and Juan Eugenio Iglesias. Synthetic data in generalizable, learning-based neuroimaging. *Imaging Neuroscience*, 2:1–22, 2024. 2

[18] Shuang Hao, Wenfeng Han, Tao Jiang, Yiping Li, Haonan Wu, Chunlin Zhong, Zhangjun Zhou, and He Tang. Synthetic data in ai: Challenges, applications, and ethical implications, 2024. 8

[19] C. C. Hemond and R. Bakshi. Magnetic resonance imaging in multiple sclerosis. *Cold Spring Harb Perspect Med*, 8 (5), 2018. 2157-1422 Hemond, Christopher C Bakshi, Rohit Journal Article Review United States 2018/01/24 Cold Spring Harb Perspect Med. 2018 May 1;8(5):a028969. doi: 10.1101/cshperspect.a028969. 2

[20] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium, 2018. 3

[21] Annika Hofmann, Nik Krajnc, Assunta Dal-Bianco, Christian J. Riedl, Tobias Zrzavy, Celia Lerma-Martin, Gregor Kasprian, Claudia E. Weber, Francesco Pezzini, Fritz Leutmezer, and et al. Myeloid cell iron uptake pathways and paramagnetic rim formation in multiple sclerosis. *Acta Neuropathologica*, 146(5):707–724, 2023. 2

[22] W. Huang, E. M. Sweeney, U. W. Kaunzner, Y. Wang, S. A. Gauthier, and T. D. Nguyen. Quantitative susceptibility mapping versus phase imaging to identify multiple sclerosis iron rim lesions with demyelination. *J Neuroimaging*, 32(4):667–675, 2022. 1552-6569 Huang, Weiyuan Sweeney, Elizabeth M Kaunzner, Ulrike W Wang, Yi Orcid: 0000-0003-1404-8526 Gauthier, Susan A Orcid: 0000-0002-4015-1781 Nguyen, Thanh D Orcid: 0000-0002-1411-7694 R01 NS090464/NS/NINDS NIH HHS/United States R01 NS104283/NS/NINDS NIH HHS/United States UL1 TR002384/TR/NCATS NIH HHS/United States Journal Article Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't United States 2022/03/10 J Neuroimaging. 2022 Jul;32(4):667-675. doi: 10.1111/jon.12987. Epub 2022 Mar 9. 2

[23] Weipeng Huang, Qin Li, Yang Xiao, Cheng Qiao, Tie Cai, Junwei Liao, Neil J. Hurley, and Guangyuan Piao. Correcting noisy multilabel predictions: Modeling label noise through latent space shifts, 2025. 3

[24] Juan E. Iglesias, Benjamin Billot, Yaël Balbastre, Colin Magdamo, Steven E. Arnold, Sudeshna Das, Brian L. Edlow, Daniel C. Alexander, Polina Golland, and Bruce Fischl. Synthsr: A public ai tool to turn heterogeneous clinical brain scans into high-resolution t1-weighted images for 3d morphometry. *Science Advances*, 9(5):eadd3607, 2023. 2

[25] Jinhee Jang, Yoonho Nam, Yangsean Choi, Na-Young Shin, Jae Young An, Kook-Jin Ahn, Bum-Soo Kim, Kwang-Soo Lee, and Woojun Kim. Paramagnetic rims in multiple sclerosis and neuromyelitis optica spectrum disorder: A quantitative susceptibility mapping study with 3-t mri. *Journal of Clinical Neurology*, 16(4):562, 2020. 2

[26] L. Ju, X. Wang, L. Wang, D. Mahapatra, X. Zhao, Q. Zhou, T. Liu, and Z. Ge. Improving medical images classification with label noise using dual-uncertainty estimation. *IEEE Trans Med Imaging*, 41(6):1533–1546, 2022. 1558-254x Ju, Lie Wang, Xin Wang, Lin Mahapatra, Dwarikanath Zhao, Xin Zhou, Quan Liu, Tongliang Ge, Zongyuan Journal Article Research Support, Non-U.S. Gov't United States 2022/01/08 IEEE Trans Med Imaging. 2022 Jun;41(6):1533-1546. doi: 10.1109/TMI.2022.3141425. Epub 2022 Jun 1. 8

[27] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks, 2019. 4, 5

[28] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data, 2020. 3, 4

[29] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan, 2020. 4

[30] Ulrike W. Kaunzner and Susan A. Gauthier. Mri in the assessment and monitoring of multiple sclerosis: an update on best practice. *Therapeutic Advances in Neurological Disorders*, 10(6):247–261, 2017. 2

[31] Varun A Kelkar and Mark Anastasio. Prior image-constrained reconstruction using style-based generative models. In *Proceedings of the 38th International Conference on Machine Learning*, pages 5367–5377. PMLR, 2021. 3

[32] Yann LeCun and Yoshua Bengio. *Convolutional networks for images, speech, and time series*, page 255–258. MIT Press, Cambridge, MA, USA, 1998. 3

[33] Wei Li, Yifei Zhao, Xi Chen, Yang Xiao, and Yuanyuan Qin. Detecting alzheimer's disease on small dataset: A knowledge transfer perspective. *IEEE Journal of Biomedical and Health Informatics*, 23(3):1234–1242, 2019. 2

[34] Tian Liu, Jing Liu, Ludovic De Rochefort, Pascal Spincemaille, Ildar Khalidov, James Robert Ledoux, and Yi Wang. Morphology enabled dipole inversion (medi) from a single-angle acquisition: Comparison with cosmos in human brain imaging. *Magnetic Resonance in Medicine*, 66(3):777–783, 2011. 3

[35] Zhe Liu, Pascal Spincemaille, Yihao Yao, Yan Zhang, and Yi Wang. Medi+0: Morphology enabled dipole inversion with automatic uniform cerebrospinal fluid zero reference for quantitative susceptibility mapping. *Magnetic Resonance in Medicine*, 79(5):2795–2803, 2018. 3

[36] C. Lou, P. Sati, M. Absinta, K. Clark, J. D. Dworkin, A. M. Valcarcel, M. K. Schindler, D. S. Reich, E. M. Sweeney, and R. T. Shinohara. Fully automated detection of paramagnetic rims in multiple sclerosis lesions on 3t susceptibility-based mr imaging. *Neuroimage Clin*, 32:102796, 2021. 2213-1582 Lou, Carolyn Sati, Pascal Absinta, Martina Clark, Kelly Dworkin, Jordan D Valcarcel, Alessandra M Schindler, Matthew K Reich, Daniel S Sweeney, Elizabeth M Shinohara, Russell T R01 MH112847/MH/NIMH NIH HHS/United States R01 NS060910/NS/NINDS NIH HHS/United States KL2 TR001879/TR/NCATS NIH HHS/United States R01 NS112274/NS/NINDS NIH HHS/United States S10 OD023495/OD/NIH HHS/United States R01 NS085211/NS/NINDS NIH HHS/United States

Journal Article Research Support, N.I.H., Extramural Research Support, N.I.H., Intramural Research Support, Non-U.S. Gov't Netherlands 2021/10/14 Neuroimage Clin. 2021;32:102796. doi: 10.1016/j.nicl.2021.102796. Epub 2021 Aug 27. 5

[37] Ha Luu, Mert Sisman, Ilhami Kovanlikaya, Tam Vu, Pascal Spincemaille, Yi Wang, Francesca Bagnato, Susan Gauthier, and Thanh Nguyen. Qsm-rimds: A detection and segmentation tool for paramagnetic rim lesions in multiple sclerosis, 2025. 3

[38] P. Maggi, P. Sati, G. Nair, I. C. M. Cortese, S. Jacobson, B. R. Smith, A. Nath, J. Ohayon, V. van Pesch, G. Perrotta, C. Pot, M. Théaudin, V. Martinelli, R. Scotti, T. Wu, R. Du Pasquier, P. A. Calabresi, M. Filippi, D. S. Reich, and M. Absinta. Paramagnetic rim lesions are specific to multiple sclerosis: An international multicenter 3t mri study. *Ann Neurol*, 88(5):1034–1042, 2020. 1531-8249 Maggi, Pietro Orcid: 0000-0003-1697-5585 Sati, Pascal Nair, Govind Cortese, Irene C M Jacobson, Steven Orcid: 0000-0003-3127-1287 Smith, Bryan R Nath, Avindra Orcid: 0000-0003-0927-5855 Ohayon, Joan van Pesch, Vincent Perrotta, Gaetano Orcid: 0000-0002-9795-1018 Pot, Caroline Théaudin, Marie Orcid: 0000-0002-3026-3595 Martinelli, Vittorio Scotti, Roberta Wu, Tianxia Du Pasquier, Renaud Calabresi, Peter A Orcid: 0000-0002-7776-6472 Filippi, Massimo Orcid: 0000-0002-5485-0479 Reich, Daniel S Orcid: 0000-0001-6313-9884 Absinta, Martina Orcid: 0000-0003-0276-383x R01 NS082347/NS/NINDS NIH HHS/United States ZIA NS003119/ImNIH/Intramural NIH HHS/United States Journal Article Multicenter Study Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't United States 2020/08/18 Ann Neurol. 2020 Nov;88(5):1034-1042. doi: 10.1002/ana.25877. Epub 2020 Sep 9. 2

[39] Sandeep Kumar Mathivanan, Sridevi Sonaimuthu, Sankar Murugesan, Hariharan Rajadurai, Basu Dev Shivahare, and Mohd Asif Shah. Employing deep learning and transfer learning for accurate brain tumor detection. *Scientific Reports*, 14(1), 2024. 2

[40] Gabrielle M. Mey, Kedar R. Mahajan, and Tara M. Desilva. Neurodegeneration in multiple sclerosis. *WIREs Mechanisms of Disease*, 15(1), 2023. 1

[41] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *CoRR*, abs/1411.1784, 2014. 3

[42] Elisa Moya-Sáez, Óscar Peña-Nogales, Rodrigo de Luis-García, and Carlos Alberola-López. A deep learning approach for synthetic mri based on two routine sequences and training with synthetic data. *Computer Methods and Programs in Biomedicine*, 210:106371, 2021. 2

[43] K. C. Ng Kee Kwong, D. Mollison, R. Meijboom, E. N. York, A. Kampaite, M. J. Thrippleton, S. Chandran, and A. D. Waldman. The prevalence of paramagnetic rim lesions in multiple sclerosis: A systematic review and meta-analysis. *PLoS One*, 16(9):e0256845, 2021. 2

[44] Xingang Pan, Bo Dai, Ziwei Liu, Chen Change Loy, and Ping Luo. Do 2d gans know 3d shape? unsupervised 3d shape reconstruction from 2d image gans. In *International Conference on Learning Representations*, 2021. 8

[45] J. A. Reeves, M. Mohebbi, R. Zivadinov, N. Bergsland, M. G. Dwyer, F. Salman, F. Schweser, and D. Jakimovski. Reliability of paramagnetic rim lesion classification on quantitative susceptibility mapping (qsm) in people with multiple sclerosis: Single-site experience and systematic review. *Mult Scler Relat Disord*, 79:104968, 2023. 2211-0356 Reeves, Jack A Mohebbi, Maryam Zivadinov, Robert Bergsland, Niels Dwyer, Michael G Salman, Fahad Schweser, Ferdinand Jakimovski, Dejan R01 NS114227/NS/NINDS NIH HHS/United States UL1 TR001412/TR/NCATS NIH HHS/United States Journal Article Review Systematic Review Netherlands 2023/09/17 Mult Scler Relat Disord. 2023 Nov;79:104968. doi: 10.1016/j.msard.2023.104968. Epub 2023 Sep 14. 2

[46] Alexandra G. Roberts, Dominick J. Romano, Mert Şişman, Alexey V. Dimov, Thanh D. Nguyen, Ilhami Kovanlikaya, Susan A. Gauthier, Yi Wang, and Pascal Spincemaille. Maximum spherical mean value filtering for whole-brain qsm. *Magnetic Resonance in Medicine*, 91(4):1586–1597, 2024. 2

[47] Alexandra Grace Roberts, Jinwei Zhang, Ceren Tozlu, Dominick Romano, Sema Akkus, Heejong Kim, Mert Rory Sabuncu, Pascal Spincemaille, Jianqi Li, Yi Wang, Xi Wu, and Brian Harris Kopell. Quantitative susceptibility mapping radiomics with label noise compensation for predicting deep brain stimulation outcomes in parkinson's disease. *medRxiv*, 2024. 8

[48] D. Robertson and N. Moreo. Disease-modifying therapies in multiple sclerosis: Overview and treatment considerations. *Fed Pract*, 33(6):28–34, 2016. 1945-337x Robertson, Derrick Moreo, Natalie Journal Article United States 2016/06/01 Fed Pract. 2016 Jun;33(6):28-34. 1

[49] Mostafa Salem, Sergi Valverde, Mariano Cabezas, Deborah Pareto, Arnau Oliver, Joaquim Salvi, Àlex Rovira, and Xavier Lladó. Multiple sclerosis lesion synthesis in mri using an encoder-decoder u-net. *IEEE Access*, 7:25171–25184, 2019. 3

[50] Ramprasaath R. Selvaraju, Abhishek Das, Ramakrishna Vedantam, Michael Cogswell, Devi Parikh, and Dhruv Batra. Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization. *CoRR*, abs/1610.02391, 2016. 4

[51] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015. 4

[52] Mert Şişman, Alexandra Roberts, Hangwei Zhuang, Renjiu Hu, Junghun Cho, Shun Zhang, Pascal Spincemaille, Thanh Nguyen, and Yi Wang. Simulating brain gradient-echo magnetic resonance images through microstructural modeling. In *13th International Workshop on Innovative Simulation for Healthcare*, 2024. 8

[53] Vajira Thambawita, Pegah Salehi, Sajad Amouei Sheshkal, Steven A. Hicks, Hugo L. Hammer, Sravanthi Parasa, Thomas de Lange, Pål Halvorsen, and Michael A. Riegler. Singan-seg: Synthetic training data generation for medical image segmentation. *PLOS ONE*, 17(5):1–24, 2022. 2

[54] Kiran Koshy Thekumparampil, Ashish Khetan, Zinan Lin, and Sewoong Oh. Robustness of conditional gans to noisy labels, 2018. 3

[55] Adon Toru Asahina, Joe Lu, Pooja Chugh, Srishti Sharma, Prakriti Sharma, Sheryn Tan, Joshua Kovoor, Brandon Stretton, Aashray Gupta, Annabel Sorby-Adams, Rudy Goh, Adil Harroud, Margareta A. Clarke, Nikos Evangelou, Sandy Patel, Andrew Dwyer, Marc Agzarian, Stephen Bacchi, and Mark Slee. Prognostic significance of paramagnetic rim lesions in multiple sclerosis: A systematic review. *Journal of Clinical Neuroscience*, 129:110810, 2024. 2

[56] Uddeshya Upadhyay and Suyash Awate. A mixed-supervision multilevel gan framework for image quality enhancement, 2021. 8

[57] Juan Miguel Valverde, Vandad Imani, Ali Abdollahzadeh, Riccardo De Feo, Mithilesh Prakash, Robert Ciszek, and Jussi Tohka. Transfer learning in magnetic resonance brain imaging: A systematic review. *Journal of Imaging*, 7(4):66, 2021. 2

[58] C. C. Voon, T. Wiltgen, B. Wiestler, S. Schlaeger, and M. Mühlau. Quantitative susceptibility mapping in multiple sclerosis: A systematic review and meta-analysis. *Neuroimage Clin*, 42:103598, 2024. 2213-1582 Voon, Cui Ci Wiltgen, Tun Wiestler, Benedikt Schlaeger, Sarah Mühlau, Mark Journal Article Meta-Analysis Review Systematic Review Netherlands 2024/04/07 Neuroimage Clin. 2024;42:103598. doi: 10.1016/j.nicl.2024.103598. Epub 2024 Mar 25. 2

[59] Abdul Waheed, Muskan Goyal, Deepak Gupta, Ashish Khanna, Fadi Al-Turjman, and Plácido Rogerio Pinheiro. Covidgan: Data augmentation using auxiliary classifier gan for improved covid-19 detection. *IEEE Access*, 8:91916–91923, 2020. 2

[60] Clare Walton, Rachel King, Lindsay Rechtman, Wendy Kaye, Emmanuelle Leray, Ruth Ann Marrie, Neil Robertson, Nicholas La Rocca, Bernard Uitdehaag, Ingrid van der Mei, Mitchell Wallin, Anne Helme, Ceri Angood Napier, Nick Rijke, and Peer Baneke. Rising prevalence of multiple sclerosis worldwide: Insights from the atlas of ms, third edition. *Multiple Sclerosis Journal*, 26(14):1816–1821, 2020. PMID: 33174475. 1

[61] R. J. Ward, D. T. Dexter, and R. R. Crichton. Iron, neuroinflammation and neurodegeneration. *Int J Mol Sci*, 23 (13), 2022. 1422-0067 Ward, Roberta J Dexter, David T Crichton, Robert R Journal Article Review Switzerland 2022/07/10 Int J Mol Sci. 2022 Jun 30;23(13):7267. doi: 10.3390/ijms23137267. 2

[62] Paula Wildner, Mariusz Stasiołek, and Mariola Matysiak. Differential diagnosis of multiple sclerosis and other inflammatory cns diseases. *Multiple Sclerosis and Related Disorders*, 37:101452, 2020. 1

[63] Jake Williams, Abel Tadesse, Tyler Sam, Huey Sun, and George D. Montanez. Limits of transfer learning, 2020. 2

[64] Weihao Xia, Yulun Zhang, Yujiu Yang, Jing-Hao Xue, Bolei Zhou, and Ming-Hsuan Yang. Gan inversion: A survey, 2022. 3

[65] Moon Ye-Bin, Nam Hyeon-Woo, Wonseok Choi, Nayeong Kim, Suha Kwak, and Tae-Hyun Oh. Synaug: Exploiting synthetic data for data imbalance problems, 2024. 8

[66] Hang Zhang, Thanh D. Nguyen, Jinwei Zhang, Melanie Marcille, Pascal Spincemaille, Yi Wang, Susan A. Gauthier, and Elizabeth M. Sweeney. Qsmrim-net: Imbalance-aware learning for identification of chronic active multiple sclerosis lesions on quantitative susceptibility maps. *NeuroImage: Clinical*, 34:102979, 2022. 3

[67] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric, 2018. 4

[68] S. Zhang, T.D. Nguyen, S.M. Hurtado Rúa, U.W. Kaunzner, S. Pandya, I. Kovanlikaya, P. Spincemaille, Y. Wang, and S.A. Gauthier. Quantitative susceptibility mapping of time-dependent susceptibility changes in multiple sclerosis lesions. *American Journal of Neuroradiology*, 2019. 2

[69] Bolei Zhou, Aditya Khosla, Àgata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. *CoRR*, abs/1512.04150, 2015. 4

[70] M. Şişman, T. D. Nguyen, A. G. Roberts, D. J. Romano, A. V. Dimov, I. Kovanlikaya, P. Spincemaille, and Y. Wang. Microstructure-informed myelin mapping (mimm) from routine multi-echo gradient echo data using multiscale physics modeling of iron and myelin effects and qsm. *Magn Reson Med*, 93(4):1499–1515, 2025. 1522-2594 Şişman, Mert Orcid: 0000-0003-0862-4478 Nguyen, Thanh D Orcid: 0000-0002-1411-7694 Roberts, Alexandra G Orcid: 0000-0002-4000-2102 Romano, Dominick J Orcid: 0000-0002-8691-406x Dimov, Alexey V Orcid: 0000-0003-0460-6635 Kovanlikaya, Ilhami Orcid: 0000-0001-6339-9862 Spincemaille, Pascal Orcid: 0000-0002-8821-1341 Wang, Yi Orcid: 0000-0003-1404-8526 R01 NS136864/NS/NINDS NIH HHS/United States R01NS095562/NH/NIH HHS/United States R01NS090464/NH/NIH HHS/United States RG-1602-07671/MS Society/ R01HL151686/NH/NIH HHS/United States R01NS105144/NH/NIH HHS/United States S10OD021782/NH/NIH HHS/United States Journal Article United States 2024/11/18 Magn Reson Med. 2025 Apr;93(4):1499-1515. doi: 10.1002/mrm.30369. Epub 2024 Nov 17. 8