Title

Fat-saturated image generation from multi-contrast MRIs using generative adversarial networks with Bloch equation-based autoencoder regularization

Authors

Kim, Sewon
Jang, Hanbyol
Hong, Seokjun
et al.

Publication Date

2021-10-01

DOI

10.1016/j.media.2021.102198

Peer reviewed

HHS Public Access

Author manuscript

Med Image Anal. Author manuscript; available in PMC 2022 October 01.

Published in final edited form as:

Med Image Anal. 2021 October ; 73: 102198. doi:10.1016/j.media.2021.102198.

Fat-saturated Image Generation from Multi-contrast MRIs Using Generative Adversarial Networks with Bloch Equation-based Autoencoder Regularization

Sewon Kim[a], **Hanbyol Jang**[a], **Seokjun Hong**[a], **Yeong Sang Hong**[b,c], **Won C. Bae**[d,e], **Sungjun Kim**[b,c,**], **Dosik Hwang**[a,*]

[a]School of Electrical and Electronic Engineering, Yonsei University, 50, Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

[b]College of Medicine, Yonsei University, 50-1, Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

[c]Gangnam Severence Hospital, Yonsei University, 211, Eonju-ro, Gangnam-gu, Seoul 06273, Republic of Korea

[d]Department of Radiology, Veterans Affairs San Diego Healthcare System, 3350 La Jolla Village Drive, San Diego, CA 92161-0114, USA

[e]Department of Radiology, University of California-San Diego, La Jolla, CA 92093-0997, USA

Abstract

Obtaining multiple series of magnetic resonance (MR) images with different contrasts is useful for accurate diagnosis of human spinal conditions. However, this can be time consuming and a burden on both the patient and the hospital. We propose a Bloch equation-based autoencoder regularization generative adversarial network (BlochGAN) to generate a fat saturation T2-weighted (T2 FS) image from T1-weighted (T1-w) and T2-weighted (T2-w) images of human spine. To achieve this, our approach was to utilize the relationship between the contrasts using Bloch equation since it is a fundamental principle of MR physics and serves as a physical basis of

[*]Corresponding author1: Tel.: +82-2-2123-5771; dosik.hwang@yonsei.ac.kr. [**]Corresponding author2: Tel.: +82-2-2019-3510; agn70@yuhs.ac; *2000 MSC:*.

Credit Author Statement

Sewon Kim (1st author): Conceptualization, Methodology, Software, Validation, Writing original & revised draft, Writing review & editing. Hanbyol Jang: Conceptualization, Methodology. Seokjun Hong: Investigation, Writing review

Yeong Sang Hong: Data preparation, Validation.

Won C. Bae: Writing-review & editing, Funding acquisition.

Sungjun Kim (corresponding): Data preparation, Conceptualization, Writing - review & editing, Project administration, Funding acquisition.

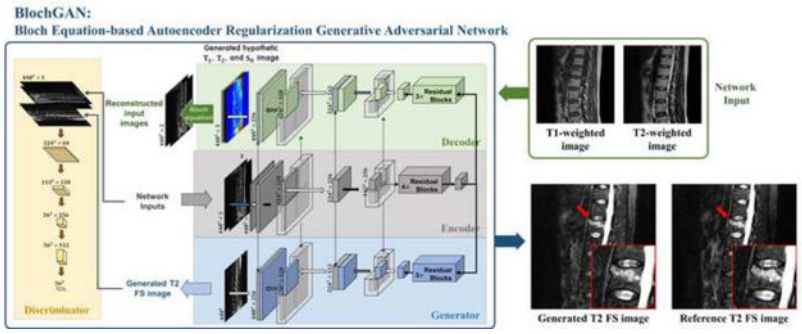
Dosik Hwang (corresponding): Supervision, Conceptualization, Writing - review & editing, Project administration, Funding acquisition.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

each contrasts. BlochGAN properly generated the target-contrast images using the autoencoder regularization based on the Bloch equation to identify the physical basis of the contrasts. BlochGAN consists of four sub-networks: an encoder, a decoder, a generator, and a discriminator. The encoder extracts features from the multi-contrast input images, and the generator creates target T2 FS images using the features extracted from the encoder. The discriminator assists network learning by providing adversarial loss, and the decoder reconstructs the input multi-contrast images and regularizes the learning process by providing reconstruction loss. The discriminator and the decoder are only used in the training process. Our results demonstrate that BlochGAN achieved quantitatively and qualitatively superior performance compared to conventional medical image synthesis methods in generating spine T2 FS images from T1-w, and T2-w images.

Graphical Abstract



Keywords

Magnetic resonance image; Image synthesis; Multi-contrast imaging; Geneartive adversarial networks; Autoencoder regularization; Bloch equation

1. Introduction

Magnetic resonance imaging (MRI) is the preferred non-invasive tool for diagnosing patients suffering from back pain. MRIs provide diagnostic information by showing a variety of contrasts between tissues. T1-weighted spin-echo (T1-w) and T2-weighted spin-echo (T2-w) images are routinely scanned to evaluate various spinal disorders (Saifuddin, 2001; Currie et al.,2011). Both sequences are essential parts of the spine MRI examination, but bright signal of the fat can obscure underlying pathology such as bone marrow edema (BME) and metastasis (Delfaut et al., 1999). Therefore, there is a consensus that adding fluid-sensitive sequences such as fat saturation (FS) T2 weighted imaging or short tau inversion recovery (STIR) which can aid in the detection of inflammation, edema, and neoplasms, although the spine MRI protocols vary depending on the in-stitutions (Saifuddin, 2001; Currie et al., 2011; Delfaut et al., 1999; Maeder et al., 2018). However, because MRI has a relatively long scanning time per pulse sequences compared to other imaging modalities, such as computed tomography (CT) and X-ray imaging, acquiring FS images in addition to T1-w and T2-w images for patients suffering from pain that makes it difficult to lie down for long periods of time places a burden on both the patient and the hospital. MR image synthesis can overcome

this problem with low cost by generating those images. This study aimed to acquire FS T2-weighted spin-echo (T2 FS) images from T1-w and T2-w images in order to obtain enough information for an accurate diagnosis without additional scanning.

Bloch equations which serve as the physical basis of MR sequences are a set of phenomenological equations used to understand nuclear magnetic resonance. Because the contrast of an MR images is based on the MR sequence, Bloch equations can be used as prior information for MR image synthesis. A model-based approach using this prior information has been proposed in studies related to MR parameter quantification map reconstruction (Maier et al., 2019; Huang et al., 2012; Sumpf et al., 2014; Zhao et al., 2015). Most MR image synthesis studies do not use the Bloch equations as prior information. Kim et al. showed that the deep learning network can learn the relationship between each contrast images by training and testing the deep learning network with the simulation dataset created on the Bloch equation basis, although, again, Bloch equations are not used as prior knowledge (Kim et al., 2020). Jog et al. synthesized the target-contrast images by inferring the imaging formula of the source images using Bloch equations and a prepared atlas. However, the atlas needed in this method requires not only MR contrast images, but also corresponding quantitative maps (Jog et al., 2015).

In this study, we proposed a Bloch equation-based autoencoder regularization generative adversarial network (BlochGAN) for MR image synthesis. BlochGAN uses multi-contrast MR images to generate other contrast images without additional scanning. The main contributions of BlochGAN are as follows.

First, our BlochGAN can efficiently learn the relationship between MR images of various contrasts given in the training process based on the Bloch equation and can properly generate target-contrast images. To our knowledge, there has been no study about medical image synthesis using the Bloch equation and generative adversarial network (GAN) or convolutional neural network (CNN) models. In this study, our BlochGAN generated spine T2 FS images from spine T1-w and T2-w images.

Second, the BlochGAN has a different network structure from other models used in image synthesis methods. BlochGAN consists of four sub-networks: an encoder, a decoder, a generator, and a discriminator. Figure 1 shows a simplified structure of the image generation part, which consists of the encoder, the decoder, and the generator. The encoder extracts important features from input images for the generator and the decoder. The generator creates a target-contrast image based on the features extracted by the encoder. The decoder generates MR parameter maps based on the features from the encoder and restores the input images like an autoencoder with the Bloch equation functions. Because the relationships between the input and the target-contrast images are based on the MR parameter maps and the Bloch equations of each sequence, the decoder in the training process could help the encoder to learn the efficient feature extraction for generating the target-contrast images. Several studies on image segmentation and classification have shown that the use of an autoencoder regularization in the training process improves a feature extraction performance of the encoder (Myronenko, 2018; Dorent et al., 2019; Wu et al., 2020; Zhang et al., 2020). However, to our knowledge, there has been no attempt to apply autoencoder regularization

in MR image synthesis. This study is the first attempt to improve MR image synthesis performance by applying the autoencoder regularization based on the Bloch equation. A preliminary work for this study was presented in 2021 at the annual meeting of the International Society for Magnetic Resonance in Medicine (Kim et al., 2021).

2. Related Works

Various medical image synthesis methods have been developed based on several basic technologies. These methods can be classified into three groups: 1) registration-based methods, 2) intensity transformation-based methods, and 3) deep learning-based methods.

2.1. Registration-based methods

The earliest medical image synthesis methods were registration-based methods, which are based on registration algorithms using atlas and deformation fields (Miller et al., 1993; Burgos et al., 2014; Lee et al., 2017). These methods construct atlas $A = \{(x, y) \mid x \in X, y \in Y\}$ based on co-registered images x and y from a source domain X and a target domain Y. Here warping function W is generated by the test image x_1 and its nearest neighbor $x_2 \in A$. The target domain image $\hat{y}_1$ which corresponds to the test image x_1 is obtained as follows: $\widehat{y_1} = W(y_2)$, where $y_2 \in A$ is a corresponding image of x_2. Because these methods use only one input image and its nearest neighbor, they are difficult to apply when the morphological difference between x_1 and x_2 is severe. Therefore, methods were developed that construct a multi-atlas to reconstruct images through spatially varying weighted summation of multiple neighbors. Methods were also developed that use one or more input images for advancing the output image quality (Burgos et al., 2014; Lee et al., 2017). However, these methods also have limitation, in that they rely heavily on the composition of the atlas, and thus they show low performance when data deviate from the pathological information of the atlas.

2.2. Intensity transformation-based methods

Intensity transformation-based methods are similar to registration-based methods in that they constitute atlases. However, there is a difference in that the process is performed in units of patches. These methods divide the input image x_1 into $n \in \mathbb{Z}$ small-sized patches and select atlas patches associated with each input patch from the composed patch-based atlas. The output patch is generated as a linear combination of selected atlas patches, and finally, the output patches are collected to become the target image $\hat{y}_1$ (Roy et al., 2011). Hofmann et al. were the first to use this method to generate pseudoCT images for attenuation correction of PET images (Hofmann et al., 2008, 2011). MR image example-based contrast synthesis (MIMECS) generated magnetization prepared rapid gradient echo images from T1-w spoiled gradient recalled images by using an example-based approach with sparse dictionary reconstruction (Roy et al., 2013). Ye et al. generated T2-w images from T1-w images by using a modality propagation (MP) method that applied a data-driven regularization approach to label propagation (Ye et al., 2013). And Cordier et al. improved the performance of the MP method by applying a multi-scale patch-based atlas and label-matching technique using tissue masks (Cordier et al., 2016). These intensity transformation-based methods

improved the performance of synthetic MRIs, but they need a lot of time for the patch-matching process. Chen et al. proposed a multi-channel registration framework to reduce the high computation cost of the single-way image synthesis method (Chen et al., 2017). To generate better images, Jog et al. proposed a new MR image synthesis framework based on random forest regression incorporating the information of pulse sequence equations and the atlas composed with MR images and the corresponding quantitative maps (Jog et al., 2015). REPLICA, regression ensembles with patch learning for image contrast agreement, that applied random forest regression based on multi-scale patches of multi-modality inputs for overcoming these temporal limitations. REPLICA showed improved results compared to other intensity transformation-based methods (Jog et al., 2017).

2.3. Deep learning-based methods

Deep learning methods, which show high performance in many image processing fields, have also been used for medical image synthesis. Van Nguyen et al. proposed a location-sensitive deep network (LSDN) that can cross-generate T1-w and T2-w images by using a multi-layer perceptron composed of four layers including two hidden layers (Van Nguyen et al., 2015). Sevetlidis et al. constructed an encoder-decoder model based on a CNN to generate fluid attenuated inversion recovery images and T2-w images from T1-w images (Sevetlidis et al., 2016). Multimodal imaging proposed by Chartsias et al. showed an MR image synthesis method that could be applied to multi-modality inputs by constructing a U-net-like encoder based on CNN and a decoder using fully convolutional layers (Chartsias et al., 2017). The contrast-conversion deep neural network (CC-DNN) generates STIR images including detailed information using T1-w, T2-w, and gradient echo images for knee diagnosis. To preserve detailed information, it is composed of only convolutional layers without any down or up-sampling, and it is trained by various loss functions for registration error compensation and representing variations of label images (Kim et al., 2020).

As a field of deep learning, GANs have shown high performance with regard to image generation. Conventional GANs use random noise as their inputs, but Nie et al. proposed a GAN-based model composed of fully convolutional networks using MR images as input to generate corresponding CT images (Nie et al., 2018). Pix2pix GAN, which is used in various image translation, uses a patch-based Markovian discriminator and pixel-wise error loss function to generate a clearer image compared to conventional GANs (Isola et al., 2017). Dar et al. showed high-quality MR image synthesis using single-modality inputs by using perceptual loss function based on pre-trained VGG-16 networks to represent the texture of target-contrast images (Dar et al., 2019). MustGAN integrates all features from both single and multiple inputs by composing a multi-stream network (Yurt et al., 2021). Unsupervised learning for medical image synthesis with GAN is performed with cyclic loss function designed to enable the network to train with unpaired data (Dar et al., 2019; Karthik et al., 2021). Several studies integrate MR image synthesis with MR image super-resolution based on GAN (Dar et al., 2020a; Yurt et al., 2020). Dar et al. designed a network that converts under-sampled MR images into full-sampled images using full-sampled images of another contrast (Dar et al., 2020a). Moreover, Yurt et al. proposed a semi-supervised super-resolution network trained with full-sampled images of another contrast (Yurt et al., 2020).

3. Method description

3.1. Theoretical background

3.1.1. Generative adversarial network for image generation—GAN is a neural network architecture designed for image generation and consists of two sub-networks based on CNN: generator (G) and discriminator (D). G receives input vector z and generates output image $\hat{y} = G(z)$ whose distribution is p_g. D distinguishes whether the output image $\hat{y}$ generated by G is a real or generated image. D is trained to perfectly distinguish $\hat{y}$ from y, and G is trained to make p_g close to the target distribution p_{target} so that D cannot distinguish $\hat{y}$ from y. The learning of GAN is the hostile training of these two sub-networks, and this is expressed as follows:

$$\min_G \min_D \mathbb{E}_{y \sim p_{target}}[\log D(y)] + \mathbb{E}_{z \sim p_z}[1 - \log D(G(z))], \tag{1}$$

where p_z denotes the distribution of z and $\mathbb{E}$ denotes expected value. Conventionally, input vector z is a random noise and p_z is normal distribution. z could be a latent variable or a latent feature and is usually the source image in image translation. However, the learning process may be unstable, because it hardly considers the distance between the decision boundary of D and the fake sample y. Mao et al. applied the least square function to D to overcome this limitation and stabilize the learning process (Mao et al., 2017).

The distribution of generated images in image translation is statistically dependent on the distribution of source images. Pix2pix GAN showed that conditional GAN is generally effective for image translation based on this statistical dependency (Isola et al., 2017). As the least square function and conditional input mentioned above are applied to the discriminator, the objective functions L_D, L_G for learning D and G are defined as follows:

$$\mathscr{L}_D = \frac{1}{2}\mathbb{E}_{y \sim p_{target}}\left[(D(x, y) - 1)^2\right] + \frac{1}{2}\mathbb{E}_{x \sim p_x}\left[D(x, G(x))^2\right], \tag{2}$$

$$\mathscr{L}_G = \frac{1}{2}\mathbb{E}_{x \sim p_x}\left[(D(x, G(x)) - 1)^2\right], \tag{3}$$

where x and p_x denote the source image and its distribution.

3.1.2. Physics for multi-contrast magnetic resonance imaging—Spin-echo pulse sequences are one of the most commonly used pulse sequences for diagnosis of musculoskeletal problems. A spin-echo pulse sequence is composed by repeating a 90 slice selective radio frequency (RF) pulse and a 180 refocusing RF pulse. The slice selective RF pulse is repeated for each repetition time, *TR*, and is followed by one (typical spin-echo pulse sequence) or more (fast or turbo spin-echo; TSE) refocusing RF pulse. The position of the refocusing RF pulse is half of the echo time, *TE*, which is MR signal sampling time. The signal obtained by the spin-echo pulse sequence is represented by the following Bloch equation:

$$S = S_0\left(1 - \exp^{-\frac{TR}{T_1}}\right)\exp^{-\frac{TE}{T_2}}, \quad (4)$$

where S_0 represents a signal amplitude without an excitation event and is proportional to the proton density value of tissue in the body, voxel size, data sampling period, number of phase-encoding steps, number of excitations, and strength of the magnetic field (Melhem et al., 1997). T_1 and T_2 refer to T_1 and T_2 relaxation times for each tissue, respectively. The MR contrast images obtained by the spin-echo pulse sequences include the T1-w image and the T2-w image. T1-w and T2-w images are determined by the TR and the TE. Images acquired by the sequences with short TR and short TE are usually considered as T1-w images, and with the long TR and long TE are usually considered as T2-w images (Helms et al., 2008; Bitar et al., 2006).

There are many reasons that fat suppression is used in MR imaging. The main purposes are to suppress the signal of normal adipose tissue to obtain visually superior contrast images, and to characterize abnormal tissues related to fat such as bone marrow infiltration (Delfaut et al., 1999). There are some various effective methods for fat suppression including the Dixon method, water excitation method, and inversion recovery-based methods (Del Grande et al., 2014). Among them, the most popular and versatile method is a chemical shift selective saturation (CHESS)/fat-saturation (Fat-Sat) method (Haase et al., 1985). Because fat and water have a chemical shift difference of 3.5 ppm (parts per million), they have a resonance frequency difference under the same external magnetic field (about 145 Hz in 1.0 T, and about 210 Hz in 1.5 T, about 420 Hz in 3.0 T). CHESS uses this difference in resonance frequency to apply a saturation RF pulse that can selectively excite only water or fat before applying an imaging pulse of the target sequence, so that a signal of only one component can be obtained during the target sequence. Therefore, it is hard to represent this sequence to an equation. For example, when the magnetization of fat is excitation into the X-Y plane due to the saturation RF pulse, a spoiling gradient is applied immediately to effectively dissipate the fat magnetization signal by dispersing the X-Y components of the fat magnetization. Then, the following imaging pulse affects only the magnetization signal of water (Del Grande et al., 2014).

3.2. Our approach

We propose BlochGAN, which generates spine T2 FS images from multi-contrast (T1-w and T2-w) spine MR images. BlochGAN can learn the relationship between T1-w, T2-w and T2 FS images and generate corresponding T2 FS images from the given T1-w and T2-w images. BlochGAN is composed of four sub-networks: an encoder, a decoder, a generator, and a discriminator. The encoder receives multi-contrast images $\mathbf{x} \in \mathbb{R}^{w \times h \times 2}$ (w and h denote the width and height of the image, respectively) as network inputs and extracts the features to be delivered to the decoder and the generator. The decoder generates MR parameter maps from the features of the encoder and restores synthetic input multi-contrast images $\hat{\mathbf{x}} \in \mathbb{R}^{w \times h \times 2}$ with the Bloch equation functions. The generator receives features from the encoder and generates synthetic T2 FS images $\hat{y} \in \mathbb{R}^{w \times h}$. Finally, the

discriminator distinguishes $\hat{y}$ generated by the generator from the acquired T2 FS image, $y \in \mathbb{R}^{w \times h}$. The detailed architecture of BlochGAN is shown in Fig. 2.

3.2.1. Bloch equation-based generative adversarial network for MR image generation—In the GANs conventional learning process, a discriminator is used to efficiently train generators. In the same way, the discriminator of BlochGAN is used to efficiently train the encoder and the generator of the BlochGAN, which are responsible for generating synthetic T2 FS images, $\hat{y}$, which is defined as below:

$$\hat{y} = Ge(En(\mathbf{x})), \tag{5}$$

where Ge and En denote the generator and the encoder of BlochGAN. According to equation (2) and (5), the loss function $\mathscr{L}_D$ for learning the discriminator of the BlochGAN is represented as follows:

$$\mathscr{L}_D = \frac{1}{2}\mathbb{E}_{\mathrm{x},y}\left[(D(\mathbf{x},y)-1)^2\right] + \frac{1}{2}\mathbb{E}_{\mathbf{x}}\left[D\left(\mathbf{x}, Ge(En(\mathbf{x}))^2\right]\right. . \tag{6}$$

According to equation (3) and (5), adversarial loss function $\mathscr{L}_{adv}$ is represented as follows:

$$\mathscr{L}_{adv} = \frac{1}{2}\mathbb{E}_{\mathbf{x}}\left[(D(\mathbf{x}, Ge(En(\mathbf{x})))-1)^2\right] \tag{7}$$

The decoder of BlochGAN was used for regularization to help the encoder avoid overfitting problem and learn efficient feature extraction during training. To do so, the decoder is trained to figure out the T_1, T_2, and S_0 values which is a basis of the relationships between T1-w, T2-w, and T2 FS images and reconstructs the input T1-w and T2-w images based on the Bloch equation. Therefore, according to equation (4), the regularization function $\mathscr{L}_R$ used in BlochGAN is defined as follows:

$$\mathscr{L}_R = \frac{1}{w \times h \times 2}\sum_{i}^{w}\sum_{j}^{h}\sum_{k}^{2}\left|\mathbf{x}_{i,j,k} - B(De(En(\mathbf{x})))_{i,j,k}\right|, \tag{8}$$

$$B(\mathbf{t};\mathbf{TR},\mathbf{TE}) = t_3\left(1-\exp^{-\frac{\mathbf{TR}}{t_1}}\right)\exp^{-\frac{\mathbf{TE}}{t_2}}, \tag{9}$$

Where $\mathbf{TR} = \{TR_{T1-w}, TR_{T2-w}\}$, $\mathbf{TE} = \{TE_{T1-w}, TE_{T2-w}\}$, and $\mathbf{t} = \{t_1, t_2, t_3\}$. t_1, t_2, and $t_3 \in \mathbb{R}^{w \times h}$ represent hypothetic quantitative T_1, T_2, and S_0 images, respectively. De denotes the decoder of BlochGAN. According to equation (8), the decoder is trained to generate hypothetic quantitative T_1, T_2, and S_0 images, and they generate $\hat{\mathbf{x}} = B(De(En(\mathbf{x})))$ together with equation (9) with the given **TR** and **TE** used to obtain **x**.

When the source and target images are perfectly registered, the training process of most conventional image translation networks relies on the pixelwise loss function between $\hat{y}$ and y. However, this pixelwise loss function aims to train the networks not for generating semantically similar images, but for generating mathematically similar images. Johnson et al. proposed a perceptual loss function that measures semantic differences by comparing high-level features obtained through a pre-trained classification network (Johnson et al., 2016). This perceptual loss function has been tried in various studies and is known to measure the similarity of images more robustly than the pixel-wise loss function (Johnson et al., 2016; Ledig et al., 2017; Yang et al., 2018). In the learning process of BlochGAN, we used both pixel-wise $l1$ loss function, $\mathscr{L}_{l1}$, and perceptual loss function, $\mathscr{L}_{vgg}$ based on a pre-trained VGG-16 (Simonyan and Zisserman, 2014). The two loss functions are represented as the following equations:

$$\mathscr{L}_{l1} = \frac{1}{w \times h} \sum_{i}^{w} \sum_{j}^{h} \left| y_{i,j} - Ge(En(\mathbf{x}))_{i,j} \right|, \tag{10}$$

$$\mathscr{L}_{vgg} = \frac{1}{n_k \times w_k \times h_k} \times \sum_{c}^{n_k} \sum_{i}^{w_k} \sum_{j}^{h_k} \left| VGG_k(y)_{c,i,j} - VGG_k(Ge(En(\mathbf{x})))_{c,i,j} \right|, \tag{11}$$

where $VGG_k(y)$ denotes the feature of the k^{th} convolutional layer in the VGG-16 network pre-trained with the ImageNet dataset for the input y, and n_k, w_k, and h_k denote the number of channels, width, and height of it, respectively.

As a result, the total loss function, $\mathscr{L}_{EDG}$, for the encoder, the decoder, and the generator, which constitute the image generation part of BlochGAN, is defined as the weighted summation of each loss function as follows.

$$\mathscr{L}_{EDG} = \lambda_{adv}\mathscr{L}_{adv} + \lambda_{R}\mathscr{L}_{R} + \lambda_{l1}\mathscr{L}_{l1} + \lambda_{VGG}\mathscr{L}_{VGG} \tag{12}$$

where λ_{adv}, λ_{R}, λ_{l1} and λ_{VGG} denote the relative weighting for each loss function.

3.2.2. Experimental details—The detailed structures of the encoder, the decoder, the generator and the discriminator are shown in Fig. 2. The encoder which generates 256-channel feature maps from the inputs, is composed of a 7 × 7 convolutional layer, a 3 × 3 convolutional layer, two down-sampling layers, and four residual blocks. The decoder is composed of three 3 × 3 convolutional layers, two up-sampling layers, and three residual blocks. The generator has the same structure as the decoder. The decoder and the generator receive the features from the encoder and generate hypothetical MR parameter maps and synthetic T2 FS images, respectively. The Bloch equation function is used to generate synthetic input images with the hypothetical MR parameter maps. For the encoder, the decoder, and the generator, every convolutional layer is followed by an instance normalization layer, and a rectified linear unit (ReLU) as an activation function.

The hyperbolic tangent is used as the last activation function following the last convolutional layer of the generator. Haar wavelet decomposition and reconstruction used as a down- and up-sampling layer, respectively, to preserve detail information. Low frequency data generated from the wavelet decomposition is delivered to the next layer as layer input, and the remaining high frequency data is delivered to the decoder and the generator at the same resolution for the wavelet reconstruction. The residual block consists of sequential connection of a 3 × 3 convolutional layer, an instance normalization layer, ReLU, a 3 × 3 convolutional layer, and an instance normalization layer. There is a residual path connecting input and output of the residual block. The discriminator is composed of four 4 × 4 convolutional layers, and each convolutional layer is followed by a leaky ReLU, and an instance normalization layer. (2, 2) strides are applied to the first three convolutional layers and the others apply (1, 1) strides.

The optimizer used in the training process of BlochGAN was Adam, and 0.002 learning rate was applied. Various λ_{adv}, λ_R, λ_{l1} and λ_{VGG} were applied to obtain optimal results: λ_{adv}=[1,2,3,4,5], λ_R=[100, 200, 300], λ_{l1}=[100, 200, 300], and λ_{VGG}=[100, 200, 300]. Based on the performance of the validation set, the optimal parameters were selected as follows: λ_{adv}, 5; λ_R, 100; λ_{l1}, 200; λ_{VGG}, 200.

3.3. Data acquisition and preprocessing

All experiments in our study were conducted under the approval of the institutional review board. We used two datasets, dataset 1 and 2. Dataset 1 was obtained from a total of 240 subjects, 40 subjects for each age group from their 20s to 70s. It was acquired from 2015 to 2016, regardless of diseases. Dataset 2 was obtained from a total of 70 subjects suspected of bone metastasis or red marrow hyperplasia. It was acquired from 2012 to 2019 regardless of age and gender. Both datasets were obtained by clinical 1.5 T MRI scanners (Avanto, Simens, Medical Solutions, Erlangen, Germany) from the same hospital (Gangnam Severance Hospital, Seoul, Korea). Seventeen slices for each contrast (T1-w TSE, T2-w TSE, T2 FS TSE images) were obtained from each subject. The pulse sequence parameters for each contrast were as follows: TR = 450 ms, TE = 9.8 ms, echo train length (ETL) = 3, matrix size = 448 × 448, acquisition matrix size = 448 × 314, field of view = 350 × 350 mm^2 (T1-w TSE); TR = 3760 ms, TE = 100 ms, ETL = 20, matrix size = 512 × 512, acquisition matrix size = 512 × 338, field of view = 350 × 350 mm^2 (T2-w TSE); TR = 4580 ms, TE = 113 ms, ETL = 17, flip angle = 150 matrix size = 448 × 448, acquisition matrix size = 448 × 224, field of view = 350 × 350 mm^2 (T2 FS TSE). To unify the matrix size, the T2-w TSE images were resized by bicubic interpolation to a size of 448 × 448 for each slice. Because the intensity values of the MR images are relative values unlike quantitative images such as CT images, all the images were not normalized by individual, by volume, or by contrast. They were divided by the maximum intensity of the entire datasets to be entered into the network as values between 0 and 1.

Our training and the testing process apply four-fold cross validation using dataset 1. In each fold, 120 subjects from 240 subjects (2040 slices from 4080 slices for each contrast) were used for the training, 60 subjects (1020 slices for each contrast) were used for the validation,

and the remaining 60 subjects (1020 slices for each contrast) were used for testing. Dataset 2 was used as an extra test set for the network trained with dataset 1.

3.4. BlochGAN model comparison

The main contribution of BlochGAN is the application of the Bloch equation-based autoencoder regularization for proper network learning. In addition to this, we used the up- and down-sampling process using the wavelet decomposition and reconstruction to generate images more accurately. In order to validate their effect, we performed two comparisons. First, we showed the performance of the Bloch equation-based autoencoder regularization by comparing three different models: BlochGAN, BlochGAN(AE), and BlochGAN(NAE). BlochGAN(NAE) uses no autoencoder regularization and BlochGAN(AE) uses conventional autoencoder regularization instead of the Bloch equation-based autoencoder regularization. For this model, the skip connection between the intermediate features of the encoder and the decoder is not used for proper learning. Secondly, we performed other comparisons to see the effect of the resizing process using wavelet transform. We compared BlochGAN with BlochGAN(NW), which uses max-pooling layers and up-sampling layers instead of wavelet decomposition and reconstruction.

3.5. Comparison with other methods

We compared the performance of BlochGAN with other conventional methods, two CNN-based methods including Multimodal imaging and CC-DNN, and two GAN-based methods including Pix2pix GAN and pGAN. All methods were implemented on an Intel Xeon 3.50 GHz octa-core CPU E5–1620 v4, 126 GB memory, and GeForce GTX 1080 Ti. Multimodal imaging was implemented on Keras 1.2.2 with Theano 1.0.2 backended based on Python 2.7, and CC-DNN and Pix2pix GAN were implemented on Keras 2.2.4 with TensorFlow 1.9.0 backended based on Python 2.7. pGAN and BlochGAN were implemented on Pytorch 0.4.1 based on Python 2.7. The hyperparameters used for training Pix2pix GAN and pGAN were test in various combinations and selected as follows based on the results of the validation set: λ, 100 in 1, 10, 100, 200, 300 (Pix2pix GAN); λ, 100 in 10, 100, 150; λ_{VGG}, 100 in 10, 100, 150 (pGAN).

3.6. Quantitative analysis

Two quantitative evaluation metrics, peak signal-to-noise ratio (PSNR) (Wang et al., 2004) and structural similarity (SSIM), were used for quantitative comparisons. PSNR evaluates image distortion and noise level between two images and SSIM evaluates the structural similarity of two images based on the human visual system (Hore and Ziou, 2010; Yu et al., 2019). All of the resulting images were normalized from 0 to 1 for applying both quantitative evaluation metrics. Paired *t*-tests (McDonald, 2009) were used to demonstrate statistical differences between the evaluation results. A *p*-value less than 0.05 indicated that the two groups were statistically different.

3.7. Expert analysis

Expert evaluation was conducted by two musculoskeletal radiologists (Y.S.H and H.Y.K, each with four years of experience in musculoskeletal radiology) for 30 subjects (15 subjects

were randomly chosen from each dataset). The reference images were provided to them without subject information, and the synthetic T2 FS images generated by five different medical image synthesis methods (Multimodal imaging, Pix2pix GAN, CC-DNN, pGAN, and BlochGAN) were also provided to them sequentially in a random order. They compared the reference and the generated T2 FS images from a clinical point of view and scored them on a scale one to five for seven different categories. The scoring criteria were as follows: 1, bad; 2, poor; 3, fair; 4, good; 5, almost perfect. The scoring categories were as follows: bone marrow, BME, disc, cerebrospinal fluid (CSF), muscle, spinal cord, and nerve root. We used intraclass correlation coefficient (ICC) to assess the level of interobserver agreement. This coefficient is interpreted as follows: 0.751.00, excellent; 0.600.75, good; 0.400.60, fair; 00.40, poor (Cicchetti, 1994).

4. Results

4.1. Model comparison results

Figure 3 shows the comparison between the generated T2 FS images by BlochGAN(NAE), BlochGAN(AE), and BlochGAN. The images of the first and the second rows are from dataset 1, and the images of the third and the fourth rows are from dataset 2. The second and the fourth row images are enlarged images of the areas indicated by the red arrows in the first and the third row images, respectively. As the images show, all three comparison methods properly reproduced the contrasts and the texture, so that they were similar to the reference images. However, the edema regions inside the vertebral bodies indicated by the red arrows in the enlarged images were not properly represented in both BlochGAN(NAE) and BlochGAN(AE). In the reference image in the second row, edema regions are in the three vertebral bodies. BlochGAN properly represented such regions brightly. On the other hand, BlochGAN(NAE) and BlochGAN(AE) represented only a part of the middle vertebral body brightly and the rest of such regions with low contrasts. In the fourth row images, BlochGAN properly represented most of the edema region in the vertebral body which has high contrast in the reference image, but BlochGAN(NAE) and BlochGAN(AE) represented only a central part of the region with high contrast.

The quantitative comparisons of the three networks is shown in Table 1. In all quantitative evaluations, BlochGAN achieved the highest scores. Compared to BlochGAN(NAE) and BlochGAN(AE), BlochGAN achieved significantly higher scores in all categories of dataset 1 and 2 ($p < 0.001$). The number of parameters used in each of BlochGAN, BlochGAN(AE), and BlochGAN(NAE) is 12,888,757 (2,764,737 for the discriminator and 10,124,020 for others), 12,516,980 (2,764,737 / 9,752,243), and 12,141,874 (2,764,737 / 9,377,137), respectively.

Figure 4 shows examples of reconstructed T1-w and T2-w images by BlochGAN(AE) and BlochGAN. As shown in the images, both BlochGAN(AE) and BlochGAN reconstructed the input multi-contrast images almost identically. Synthetic input images reconstructed by BlochGAN(AE) achieved 41.8869 PSNR, 0.9636 SSIM scores on average, and synthetic input images reconstructed by BlochGAN achieved 42.7403 PSNR, 0.9641 SSIM scores on average.

For the training set, the result images of BlochGAN(NAE), BlochGAN(AE) and BlochGAN achieved MSE of $4.0756 \pm 0.9281 \times 10^3$, $4.0942 \pm 0.9567 \times 10^3$, and $4.0456 \pm 0.9289 \times 10^3$, respectively, and Wasserstein distances of 64.0038 ± 8.3872, 64.7908 ± 8.7691, and 65.4122 ± 7.5860, respectively. For the test set, BlochGAN (NAE), BlochGAN (AE), and BlochGAN achieved MSE of $4.3082 \pm 1.0145 \times 10^3$, $4.2156 \pm 1.0321 \times 10^3$, and $3.9446 \pm 1.0220 \times 10^3$, respectively, and Wasserstein distances of 66.6301 ± 5.4938, 67.3097 ± 5.1740, and 63.9565 ± 6.3218, respectively. In both MSE and Wasserstein distance scores, both BlochGAN(NAE) and BlochGAN(AE) showed worse results in the test set than in the training set, but BlochGAN showed similar results in both the training and test sets (As a result of *t*-test, no significant difference exists between the results of BlochGAN for the training and test sets).

Figure 5 shows images generated by two different models, BlochGAN(NW) and BlochGAN. The images of the first and the second rows are from dataset 1, and the images of the third and the fourth rows are from dataset 2. The second and the fourth row images are enlarged images of the areas indicated by the red arrows in the first and the third row images, respectively. Results from the two models are similar to the reference T2 FS images. However, BlochGAN(NW) did not accurately represent the boundary of degenerated area in the vertebral body in the second row images of Fig. 5 or the closed circular structure of the spinal cord as shown in the fourth row images of Fig. 5. On the other hand, BlochGAN properly represented the clear boundary of the degenerated area in the vertebral body in the second row images of Fig. 5, and the circular structure of spinal cord at the connection point to the vertebral body in the fourth row images of Fig. 5. It also accurately identified the small circular edema region at the middle of the vertebral body in the fourth row images of Fig. 5. Table 1 shows the quantitative evaluation results of both networks. For both datasets, BlochGAN achieved significantly higher SSIM scores than BlochGAN(NW). For PSNR scores, BlochGAN and BlochGAN(NW) were not statistically different ($p \geq 0.05$) for dataset 1. On the other hand, BlochGAN achieved statistically higher scores than BlochGAN(NW) ($p < 0.01$) for dataset 2.

4.2. Comparisons with other methods

Figure 6 compares the synthetic T2 FS images generated by various MR image synthesis methods for dataset 1. The result images of pGAN and BlochGAN have a texture similar to the reference image, while other images have blurring artifacts and artificial textures. The second, fourth, and sixth rows of Fig. 6 are enlarged images of the area indicated by the red arrows in the first, third, and fifth rows, respectively. The images in the second row of Fig. 6 show the degenerated area in the vertebral body, but the area could not be identified in the result of Multimodal imaging. In the Pix2pix GAN results, the contrast of the vertebral body is lower, and the contrast of the degenerated area is much higher than in the reference image. The CC-DNN image represents the degenerated area, but it has a morphological difference and an artificial texture. The results of pGAN shows a similar texture to the reference image, but the shape of the degenerated area is morphologically different from the reference image. On the other hand, the degenerated area in the result of BlochGAN has a texture and bottleneck structure that are similar to the reference image. In the fourth row, the red and the yellow arrows indicate the blood vessel and the nucleus pulposus of

the intervertebral discs, respectively. The images of Multimodal imaging, Pix2pix GAN, and CC-DNN show the blurring artifacts in the nucleus. On the other hand, the results of pGAN and BlochGAN clearly represent the hook-like shape of the nucleus pulposus. Also, Multimodal imaging, Pix2pix GAN, and CC-DNN do not properly represented blood vessels extending toward the inside of the vertebral body, whereas pGAN represented a shape that is close to it. Moreover, BlochGAN properly represented its entire closed shape. The last row images of Fig. 6 indicate BME that plays a key role in diagnosing spine disease. Multimodal imaging and pGAN did not properly identify the edema region. Pix2pix GAN and CC-DNN identified BME, but there are blurring artifacts and contrast differences compared to the reference image. BlochGAN, on the other hand, represented the bright edema region with a contrast and texture similar to the reference image.

Figure 7 shows the comparison of T2 FS images generated by different MR image synthesis methods for dataset 2. The second and fourth rows of Fig. 7 are enlarged images of the area indicated by the red arrows in the first and third rows, respectively. In the first row of Fig. 7, the images, except for BlochGAN, show entirely different contrast compared to the reference. Only BlochGAN represents the suppressed contrasts like the reference image. The red arrows in row 2 images in Fig. 7 indicate BME regions. pGAN failed to identify the edema region. Multimodal imaging and CC-DNN identified the edema region, but they has different contrasts from the reference image. Pix2pix GAN properly represents the edema region but failed to represent the proper texture. On the other hand, in the BlochGAN image, the edema region and the overall contrasts and textures are similar to the reference image. The yellow and red arrows of row 4 images in Fig. 7 indicate tiny BME regions. Only BlochGAN properly identified them with contrast and texture similar to the reference images.

Figure 8 shows the quantitative evaluation results of the compared MR image synthesis methods. A and B are for dataset 1, and C and D are for dataset 2. A and C show PSNR scores, and B and D show SSIM scores. For both datasets, BlochGAN achieved the highest quantitative evaluation scores among all the compared methods. For all categories of dataset 1 and 2, the quantitative scores of BlochGAN and those of other methods were significantly different ($p < 0.001$).

Table 2 shows the expert evaluation results. Interobserver agreement for the expert evaluations was excellent (ICC, 0.7912). BlochGAN achieved the highest scores and the lowest standard deviation among the compared methods for all categories. There was a significantly large difference between the BlochGAN results and the results of other comparative methods ($p < 0.001$). Multimodal imaging, Pix2pix GAN, and CC-DNN showed large difference for all categories compared to BlochGAN. In the case of pGAN, the average difference was less than 0.1 point from BlochGAN in most categories, but the difference was 0.56 points for BME which is the most clinically important in fat suppression images. BlochGAN is the only method that scored 4 points or more in all categories among the compared methods.

5. Discussion

Compared to other non-invasive diagnostic modalities, advantage of MRI is that it can provide multiple contrasts. However, acquiring images with different contrasts adds scan time, which can be a burden for both patient and hospital. To solve this problem, studies have been conducted to develop a fast MRI reconstruction that can generate an MRI using less data (Eo et al., 2018; Zhu et al., 2018; Eo et al., 2020; Dar et al., 2020b). Another way to shorten the time it takes to acquire multi-contrast images is to generate them without scanning. That is, it is possible to shorten the total scan time through image translation, which generates images of different contrasts from existing images. In this study, we evaluated the performance of BlochGAN when generating T2 FS images from T1-w and T2-w images for spine MRI. Our proposed network, BlochGAN, is an artificial neural network model that combines GAN and the Bloch equation, which is a basic principle of MR physics. Compared to other conventional image synthesis methods, BlochGAN showed a markedly better performance.

As we shown in Section 4.2, conventional methods such as Multimodal imaging and Pix2pix GAN, have limitations when it comes to representing texture and detailed information of target-contrast images. In 2019, the CC-DNN proposed improved restoration of detailed information by adding custom loss functions for compensating misregistration error and representing image variation (Kim et al., 2020), but it could not completely resolve the problems of blurring artifacts and artificial textures. pGAN, proposed in 2019, can represent the proper textures in a way that is similar to the target-contrast MR images (Dar et al., 2019). It achieves this by taking adversarial loss function and perceptual loss function which can help the network learning for high-level features using the pre-trained VGG-16 network. This combination of the perceptual loss function and GAN shows some advantages in representing the textures of acquired MR images but has difficulty in representing detailed information used for diagnosis. Our proposed BloclhGAN represents the textures of the acquired MR images and restores detailed information by using perceptual loss function, adversarial loss function, and the Bloch equation-based autoencoder regularization. Figure 6 and 7 show that BlochGAN properly represents the contrasts and textures similar to the reference images and also properly restores detailed structures of tissues necessary for diagnosis.

This study demonstrated the performance of the structural elements introduced by model comparison. The most important element of BlochGAN is the Bloch equation-based autoencoder regularization. The distribution covering the Bloch equation is followed because the MR image data is obtained based on the Bloch equation and other various image processing algorithms. Therefore, network models are trained in automatic consideration of the Bloch equation when they trained with the MR image data. However, the dataset used for the training is sampled from the distribution covering not only the Bloch equation but also noise. Moreover, when network models are trained with only the dataset, they do not learn the real distribution but only the sampled distribution. These can cause an overfitting problem. Several studies on image classification and segmentation have already demonstrated a decrease of the overfitting problem and performance improvement by regularizing features from the encoder using an autoencoder (Myronenko, 2018; Dorent et

al., 2019; Wu et al., 2020; Zhang et al., 2020). However, applying proper regularization using the conventional autoencoder regularization technique is difficult for MR image translation, because it does not consider the physical basis of the MR image data and only focuses on preserving the key information of input images to restore themselves. On the other hand, the BlochGAN regularization technique is based on the MR physics. The decoder of BlochGAN aims to generate hypothetical quantitative T_1, T_2, and S_0 maps and combine them using the Bloch equation to restore the input images. Therefore, the decoder regulates the encoder to generate the appropriate features to make the proper T_1, T_2, and S_0 maps. This regularization is also effective for the generator because target-contrast MR image is also based on the same T_1, T_2, and S_0 maps. The results of Section 4.1 show that BlochGAN is more robust to overfitting compared with BlochGAN(NAE) and BlochGAN(AE). Although both BlochGAN(AE) and BlochGAN reconstructed the input multi-contrast images almost identically, quantitative and qualitative differences exist between the T2 FS results of BlochGAN(AE) and BlochGAN. In addition, BlochGAN(NAE) and BlochGAN(AE) showed worse performance in the test than in the training, whereas BlochGAN showed similar performance in both the test and training.

In Section 4.1, BlochGAN achieved higher quantitative evaluation scores than BlochGAN(NAE) and BlochGAN(AE). However, a difference between BlochGAN and the other two comparative models exists beyond the numerical difference. Figure S1 in the Supplementary materials shows that BlochGAN has a lower risk of misrepresenting the critical regions compared to the other two comparative models by comparing the histograms of the intensity difference between the results of the three models and the reference T2 FS images. Using the model in clinical use is difficult if such risk is high because the purpose of acquiring T2 FS image is to obtain information necessary for accurate diagnosis. In this regard, BlochGAN can provide more reliable images than BlochGAN(NAE) and BlochGAN(AE). In addition, the network should be applicable to various devices for clinical use. Table S1 and Figure S2 in the Supplementary materials show the performance of BlochGAN(NAE), BlochGAN(AE), and BlochGAN on the external device dataset. In these experiments, BlochGAN showed better performance compared to BlochGAN(NAE) and BlochGAN(AE). This indicates that BlochGAN is less susceptible to the device differences compared with the other comparative models.

In a conventional U-net structure, a pooling layer such as max-pooling improves the learning efficiency of the network by discarding some feature information that the network regards as not important (Goodfellow et al., 2016). However, information, such as high frequency information, discarded in this process could be important for successful image synthesis. Therefore, loss of such information may lead to the omission of a detailed structure from the generated image. It is possible to restore detailed information without abandoning the learning efficiency of the network by using wavelet decomposition and reconstruction because wavelet transformation can reduce and restore the image size without discarding any information. Ye et al. theoretically supported this and showed performance improvement in image segmentation (Ye et al., 2018). In this study, we compared BlochGAN(NW), which applies the max-pooling layer and interpolations for the down- and up-sampling process, and BlochGAN. The latter one generated more detailed outputs. The difference of quantitative evaluation scores between the two models was higher in SSIM scores than in PSNR scores.

SSIM and PSNR are quantitative evaluation metrics used to evaluate the structural difference between two images. SSIM quantifies the structural difference between two images based on the human visual system, PSNR is mainly used to determine the noise level difference between two images because it is calculated based on the pixelwise mean squared error and maximum intensity of the images. That is, PSNR gives more focus on noise and SSIM gives more focus on structural difference (Hore and Ziou, 2010; Yu et al., 2019). Therefore, it can be seen that the images generated by BlochGAN(NW) and BlochGAN have a similar noise level, but the structural difference between the reference images and BlochGAN(NW) results is statistically bigger than the difference between the reference images and BlochGAN results. This indicates that the downand up-sampling process by wavelet decomposition and reconstruction preserves more detailed structural information than the max-pooling layer and the up-sampling layer.

In the current study, when radiologists evaluated the similarity of the generated T2 FS images to the acquired T2 FS images used as standard reference, pGAN and BlochGAN showed higher agreement than other conventional image synthesis methods in all evaluated structures and tissues. In particular, BlochGAN showed highest degree of agreement for BME, which is the main clinical goal to be obtained through the acquisition of fat-suppressed T2-weighted images. Although it should be further verified by using more meticulously obtained data later whether the images synthesized using BlochGAN are not inferior to the fat-suppressed T2-weighted images in the evaluation of the lesion of bone marrow, we believe that this study was sufficient at least to verify the feasibility of BlochGAN for the evaluation of BME.

The use of 3D data instead of 2D data is a good way to improve inter-slice consistency and solve the misalignment in zaxis. And many studies of medical image processing showed the use of 3D data produces better results than using only 2D data (Dou et al., 2016; Mzoughi et al., 2020). However, a hardware limitation exists in the implementation because 3D-based networks require several times bigger memory than 2D-based networks. Therefore, the current experiments were carried out for 2D data. Table S2 shows the results of the additional experiments for BlochGAN using multi-slice inputs. BlochGAN using multi-slice inputs achieved higher quantitative evaluation scores than BlochGAN using single-slice inputs for both dataset 1 and 2. Increasing the dimension of the input is expected to improve BlochGAN performance based on these results. This can be realized through network optimization and hardware development and will be covered in our future works.

This study showed the performance of BlochGAN for MR image synthesis. In various experiments, synthetic T2 FS images generated by BlochGAN were quantitatively and qualitatively superior to the synthetic images generated by other comparative methods. However, the T1-w and T2-w spin-echo sequences used in this study are the most basic and simple pulse sequences which do not consider many additional variables (e.g., field and RF inhomogeneities). Therefore, BlochGAN should be verified with more complicated pulse sequences and various anatomies to be used clinically. Figure S3 and Table S3 in the Supplementary materials show that BlochGAN can be applied to knee STIR MRIs. Based on these results, BlochGAN is expected to be applied to more complicated pulse sequences

and different anatomies although more validation is still needed. Thus, our future work will focus on the validation of BlochGAN under various conditions.

6. Conclusion

In this paper, we demonstrated the feasibility of our proposed network, a Bloch equation-based autoencoder regularization generative adversarial network, termed BlochGAN, which generates T2 FS spine MR images from T1-w and T2-w images without additional scanning. It is a first attempt to train the network in medical image synthesis using a Bloch equation. Our network achieved quantitatively and qualitatively higher performance compared to other conventional image synthesis methods. Our method can potentially be used to shorten the scanning time for acquiring multi-contrast MRIs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (2019R1A2B5B01070488), Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Science and ICT (2018M3A9H6081483), and the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: 202011C04-03). Dr. Bae received grant support (Grant Number: R01 AR066622, P30 AR073761) from the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health.

References

Bitar R, Leung G, Perng R, Tadros S, Moody AR, Sarrazin J, Roberts TP, 2006. Mr pulse sequences: what every radiologist wants to know but is afraid to ask. Radiographics 26(2), 513–537. [PubMed: 16549614]

Burgos N, Cardoso MJ, Thielemans K, Modat M, Pedemonte S, Dickson J, Duncan JS, 2014. Attenuation correction synthesis for hybrid petmr scanners: application to brain studies. IEEE transactions on medical imaging 33(12), 2332–2341. [PubMed: 25055381]

Chartsias A, Joyce T, Giuffrida MV, Tsaftaris SA, 2017. Multimodal mr synthesis via modality-invariant latent representation. IEEE transactions on medical imaging 37(3), 803–814. [PubMed: 29053447]

Chen M, Carass A, Jog A, Lee J, Roy S, Prince JL, 2017. Cross contrast multi-channel image registration using image synthesis for mr brain images. Medical Image Analysis 36, 2–14. [PubMed: 27816859]

Cicchetti DV, 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological assessment 6(4), 284.

Cordier N, Delingette H, Lł M, Ayache N, 2016. Extended modality propagation: image synthesis of pathological cases. IEEE transactions on medical imaging 35(12), 2598–2608. [PubMed: 27411217]

Currie S, Galea-Soler S, Barron D, Chandramohan M, Groves C, 2011. Mri characteristics of tuberculous spondylitis. Clinical radiology 66(8), 778–787. [PubMed: 21570065]

Dar SUH, Yurt M, Karacan L, Erdem A, Erdem E, ukur T, 2019. Image synthesis in multi-contrast mri with conditional generative adversarial networks. IEEE transactions on medical imaging 38(10), 2375–2388. [PubMed: 30835216]

Dar SUH, Yurt M, Shahdloo M, Ildz ME, Tnaz B, ukur T, 2020a. Prior-guided image reconstruction for accelerated multi-contrast mri via generative adversarial networks. IEEE Journal of Selected Topics in Signal Processing 14(6), 1072–1087.

Dar SUH, zbey M, atl AB, ukur T, 2020b. A transferlearning approach for accelerated mri using deep neural network. Magnetic resonance in medicine 84(2), 663–685. [PubMed: 31898840]

Del Grande F, Santini F, Herzka DA, Aro MR, Dean CW, Gold GE, Carrino JA, 2014. Fat-suppression techniques for 3-t mr imaging of the musculoskeletal system. Radiographics 34(1), 217–233. [PubMed: 24428292]

Delfaut EM, Beltran J, Johnson G, Rousseau J, Marchandise X, Cotten A, 1999. Fat suppression in mr imaging: techniques and pitfalls. Radiographics 19(2), 373–382. [PubMed: 10194785]

Dorent R, Joutard S, Modat M, Ourselin S, Vercauteren T, 2019. Heteromodal variational encoder-decoder for joint modality completion and segmentation, in: In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 74–82.

Dou Q, Chen H, Yu L, Zhao L, Qin J, Wang D, Mok VCT, Shi L, Heng PA, 2016. Automatic detection of cerebral microbleeds from mr images via 3d convolutional neural networks. IEEE transactions on medical imaging 35(5), 1182–1195. [PubMed: 26886975]

Eo T, Jun Y, Kim T, Jang J, Lee HJ, Hwang D, 2018. Kikinet: crossdomain convolutional neural networks for reconstructing undersampled magnetic resonance images. Magnetic resonance in medicine 80(5), 2188–2201. [PubMed: 29624729]

Eo T, Shin H, Jun Y, Kim T, Hwang D, 2020. Accelerating cartesian mri by domain-transform manifold learning in phase-encoding direction. Medical Image Analysis, 101689.

Goodfellow I, Bengio Y, Courville A, Bengio Y, 2016. Deep learning. Cambridge: MIT press.

Haase A, Frahm J, Hanicke W, Matthaei D, 1985. 1h nmr chemical shift selective (chess) imaging. Physics in Medicine & Biology 30(4), 341. [PubMed: 4001160]

Helms CA, Major NM, Anderson MW, Kaplan P, Dussault R, 2008. Musculoskeletal mri e-book. Elsevier Health Sciences.

Hofmann M, Bezrukov I, Mantlik F, Aschoff P, Steinke F, Beyer T, Pichler BJ, Schlkopf B, 2011. Mri-based attenuation correction for wholebody pet/mri: quantitative evaluation of segmentation-and atlas-based methods. Journal of nuclear medicine 52(9), 1392–1399. [PubMed: 21828115]

Hofmann M, Steinke F, Scheel V, Charpiat G, Farquhar J, Aschoff P, Brady M, Schlkopf B, Pichler BJ, 2008. Mri-based attenuation correction for pet/mri: A novel approach combining pattern recognition and atlas registration. Journal of nuclear medicine 49(11), 1875–1883. [PubMed: 18927326]

Hore A, Ziou D, 2010. Image quality metrics: Psnr vs. ssim, in: In 2010 20th international conference on pattern recognition, pp. 2366–2369.

Huang C, Graff CG, Clarkson EW, Bilgin A, Altbach MI, 2012. T2 mapping from highly undersampled data by reconstruction of principal component coefficient maps using compressed sensing. Magnetic resonance in medicine 67(5), 1355–1366. [PubMed: 22190358]

Isola P, Zhu JY, Zhou T, Efros AA, 2017. Image-to-image translation with conditional adversarial networks, in: In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1125–1134.

Jog A, Carass A, Pham DL, Prince JL, 2017. Random forest regression for magnetic resonance image synthesis. Medical Image Analysis 35, 475–4888. [PubMed: 27607469]

Jog A, Carass A, Roy S, Pham DL, Prince JL, 2015. Mr image synthesis by contrast learning on neighborhood ensembles. Medical Image Analysis 24(1), 67–76.

Johnson J, Alahi A, Fei-Fei L, 2016. Perceptual losses for real-time style transfer and super-resolution, in: In European conference on computer vision, pp. 694–711.

Karthik EMN, Laporte C, Cheriet F, 2021. Three-dimensional segmentation of the scoliotic spine from mri using unsupervised volume-based mr-ct synthesis, in: In Medical Imaging 2021: Image Processing, p. 115961H.

Kim S, Jang H, Hong S, Hong YS, Bae WC, Kim S, Hwang D, 2021. Fat saturated mr image generation using gan with bloch equation-based autoencoder regularization, in: Proceedings of the 29th Annual Meeting of International Society of Magnetic Resonance in Medicine, p. 0810.

Kim S, Jang H, Jang J, Lee YH, Hwang D, 2020. Deeplearned short tau inversion recovery imaging using multicontrast mr images. Magnetic resonance in medicine.

Ledig C, Theis L, Huszr F, Caballero J, Cunningham A, Acosta A, Shi W, 2017. Photo-realistic single image super-resolution using a generative adversarial network, in: In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4681–4690.

Lee J, Carass A, Jog A, Zhao C, Prince JL, 2017. Multi-atlas-based ct synthesis from conventional mri with patch-based refinement for mri-based radiotherapy planning, in: In Medical Imaging 2017: Image Processing. International Society for Optics and Photonics, p. 101331.

Maeder Y, Dunet V, Richard R, Becce F, Omoumi P, 2018. Bone marrow metastases: T2-weighted dixon spin-echo fat images can replace t1-weighted spin-echo images. Radiology 286(3), 948–959. [PubMed: 29095674]

Maier O, Schoormans J, Schloegl M, Strijkers GJ, Lesch A, Benkert T, Stollberger R, 2019. Rapid t1 quantification from high resolution 3d data with modelbased reconstruction. Magnetic resonance in medicine 81(3), 2072–2089. [PubMed: 30346053]

Mao X, Li Q, Xie H, Lau RY, Wang Z, Paul Smolley S, 2017. Least squares generative adversarial networks, in: In Proceedings of the IEEE international conference on computer vision, pp. 2794–2802.

McDonald JH, 2009. Handbook of biological statistics. MD: sparky house publishing.

Melhem ER, Israel DA, Eustace S, Jara H, 1997. Mr of the spine with a fast t1-weighted fluid-attenuated inversion recovery sequence. American journal of neuroradiology 18(3), 447–454. [PubMed: 9090401]

Miller MI, Christensen GE, Amit Y, Grenander U, 1993. Mathematical textbook of deformable neuroanatomies. Proceedings of the National Academy of Sciences 90(24), 11944–11948.

Myronenko A, 2018. 3d mri brain tumor segmentation using autoencoder regularization, in: In International MICCAI Brainlesion Workshop, pp. 311–320.

Mzoughi H, Njeh I, Wali A, Slima MB, BenHamida A, Mhiri C, Mahfoudhe KB, 2020. Deep multi-scale 3d convolutional neural network (cnn) for mri gliomas brain tumor classification. Journal of Digital Imaging 33, 903–915. [PubMed: 32440926]

Nie D, Trullo R, Lian J, Wang L, Petitjean C, Ruan S, Shen D, 2018. Medical image synthesis with deep convolutional adversarial networks. IEEE Transactions on Biomedical Engineering 65(12), 2720–2730. [PubMed: 29993445]

Roy S, Carass A, Prince J, 2011. A compressed sensing approach for mr tissue contrast synthesis, in: In Biennial International Conference on Information Processing in Medical Imaging, pp. 371–383.

Roy S, Carass A, Prince JL, 2013. Magnetic resonance image example-based contrast synthesis. IEEE transactions on medical imaging 32(12), 2348–2363. [PubMed: 24058022]

Saifuddin A, 2001. Mri of acute spinal trauma. Skeletal radiology 30(5), 237–246. [PubMed: 11407714]

Sevetlidis V, Giuffrida MV, Tsaftaris SA, 2016. Whole image synthesis using a deep encoder-decoder network, in: In International Workshop on Simulation and Synthesis in Medical Imaging, pp. 127–137.

Simonyan K, Zisserman A, 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint 1409, 1556.

Sumpf TJ, Petrovic A, Uecker M, Knoll F, Frahm J, 2014. Fast t2 mapping with improved accuracy using undersampled spin-echo mri and model-based reconstructions with a generating function. IEEE transactions on medical imaging 33(12), 2213–2222. [PubMed: 24988592]

Van Nguyen H, Zhou K, Vemulapalli R, 2015. Cross-domain synthesis of medical images using efficient location-sensitive deep network, in: In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 677–684.

Wang Z, Bovik AC, Sheikh HR, Simoncelli EP, 2004. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 13(4), 600–612. [PubMed: 15376593]

Wu Z, Hou B, Jiao L, 2020. Multiscale cnn with autoencoder regularization joint contextual attention network for sar image classification. IEEE Transactions on Geoscience and Remote Sensing.

Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, Wang G, 2018. Lowdose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. IEEE transactions on medical imaging 37(6), 1348–1357. [PubMed: 29870364]

Ye DH, Zikic D, Glocker B, Konukoglu E, 2013. Modality propagation: coherent synthesis of subject-specific scans with data-driven regularization, in: In International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 606–613.

Ye JC, Han Y, Cha E, 2018. Deep convolutional framelets: A general deep learning framework for inverse problems. SIAM Journal on Imaging Sciences 11(2), 991–1048.

Yu Z, Xiang Q, Meng J, Kou C, Ren Q, Lu Y, 2019. Retinal image synthesis from multiple-landmarks input with generative adversarial networks. Biomedical engineering online 18(1), 1–15. [PubMed: 30602383]

Yurt M, Dar SUH, Erdem A, Erdem E, Oguz KK, ukur T, 2021. mustgan: Multi-stream generative adversarial networks for mr image synthesis. Medical Image Analysis 70, 101944. [PubMed: 33690024]

Yurt M, Dar SUH, Tnaz B, zbey M, ukur T, 2020. Semi-supervised learning of mutually accelerated multi-contrast mri synthesis without fullysampled ground-truths. arXiv preprint 2011, 14347.

Zhang X, Hu W, Cai J, 2020. Adaptive weighted loss makes brain tumors segmentation more accurate in 3d mri volume, in: In Proceedings of the 2020 2nd International Conference on Big Data Engineering, pp. 29–35.

Zhao B, Lu W, Hitchens TK, Lam F, Ho C, Liang ZP, 2015. Accelerated mr parameter mapping with lowrank and sparsity constraints. Magnetic resonance in medicine 74(2), 489–498. [PubMed: 25163720]

Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS, 2018. Image reconstruction by domain-transform manifold learning. Nature 555(7697), 487–492. [PubMed: 29565357]

Highlights

- Bloch Equation-based Autoencoder Regularization GAN (BlochGAN).
- BlochGAN uses multi-contrast MR images to generate other contrast images.
- BlochGAN can learn the multi-contrast MR relationship based on the Bloch equation.
- Performance of BlochGAN is quantitatively and qualitatively demonstrated.
- BlochGAN can reduce the scan time of multi-contrast MR images.

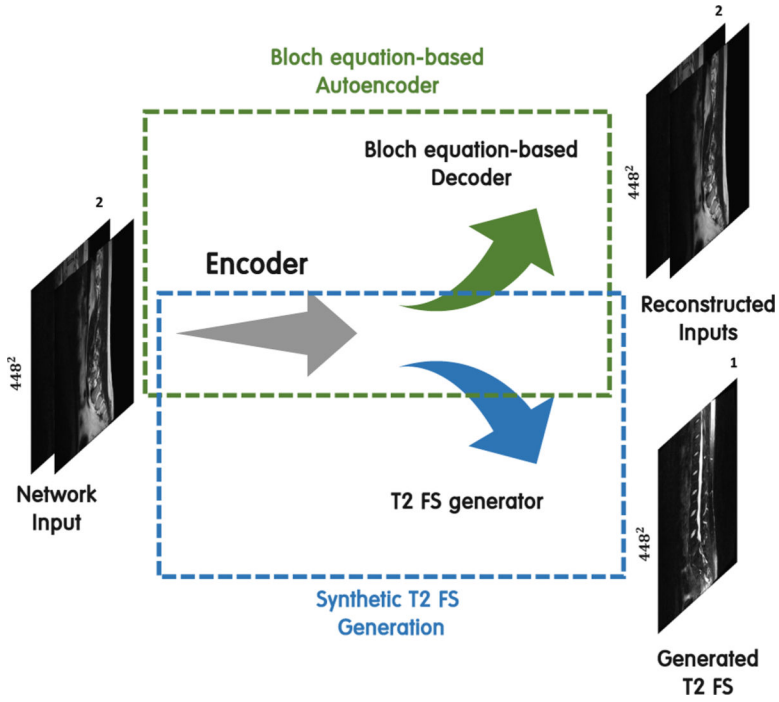


Fig. 1.

A simplified structure of the image generation part in BlochGAN. The network input images are fed into the encoder, which generates features from them. Moreover, the features are fed into the Bloch equation-based decoder and the fat saturation T2-weighted (T2 FS) image generator, and the reconstructed inputs and the synthetic T2 FS images are generated.

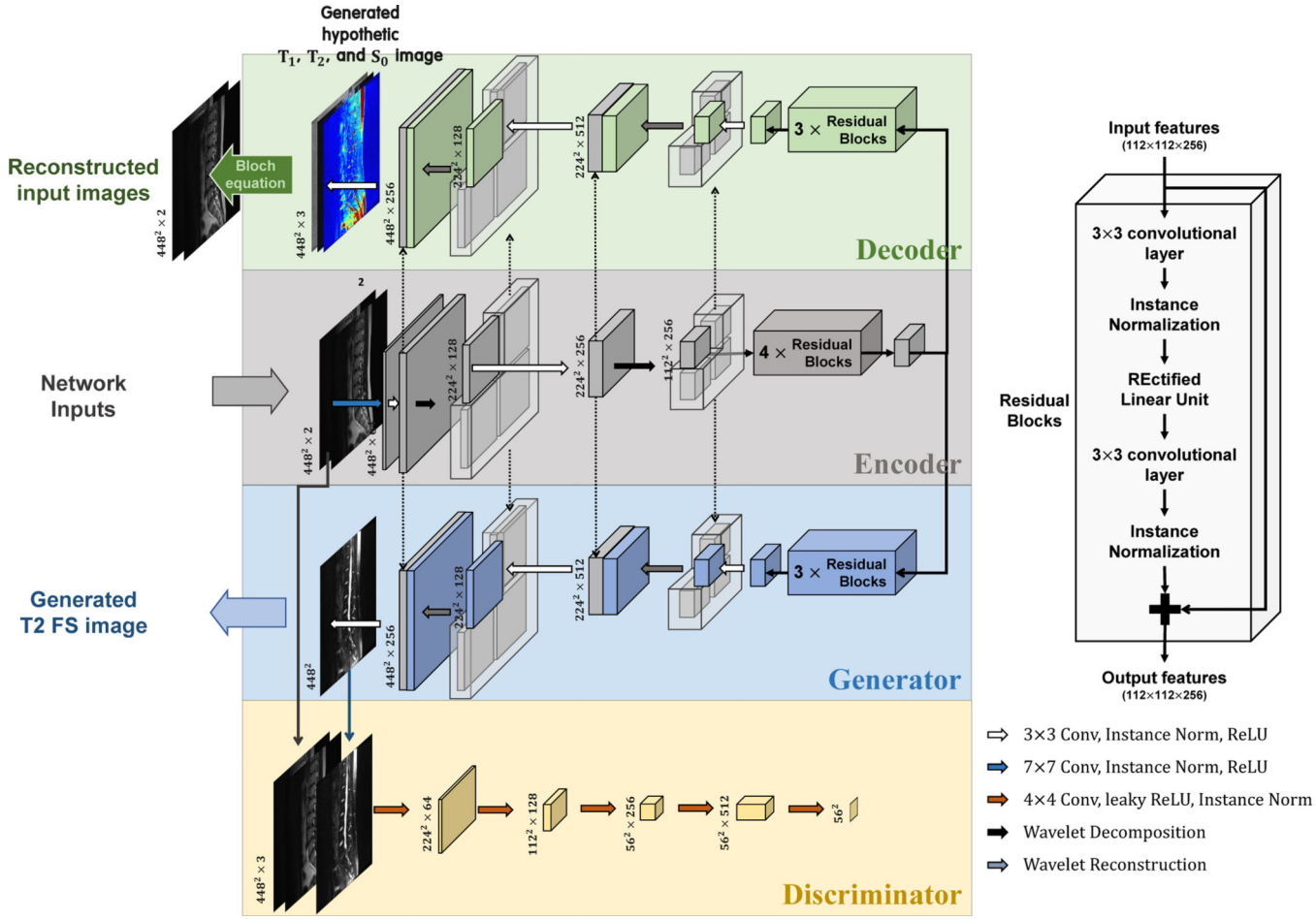


Fig. 2.

The entire network structure of BlochGAN, which consists of four subnetworks: 1) The encoder which generates features from the multi-contrast magnetic resonance (MR) images. 2) The decoder which generates hypothetical MR parameter maps from the features provided by the encoder. The reconstructed input images are generated from the hypothetical MR parameter maps by applying the Bloch equations. 3) The generator generates fat saturation T2-weighted (T2 FS) images from the features provided by the encoder. 4) The discriminator classifies both real T2 FS images and generated T2 FS images from the generator.

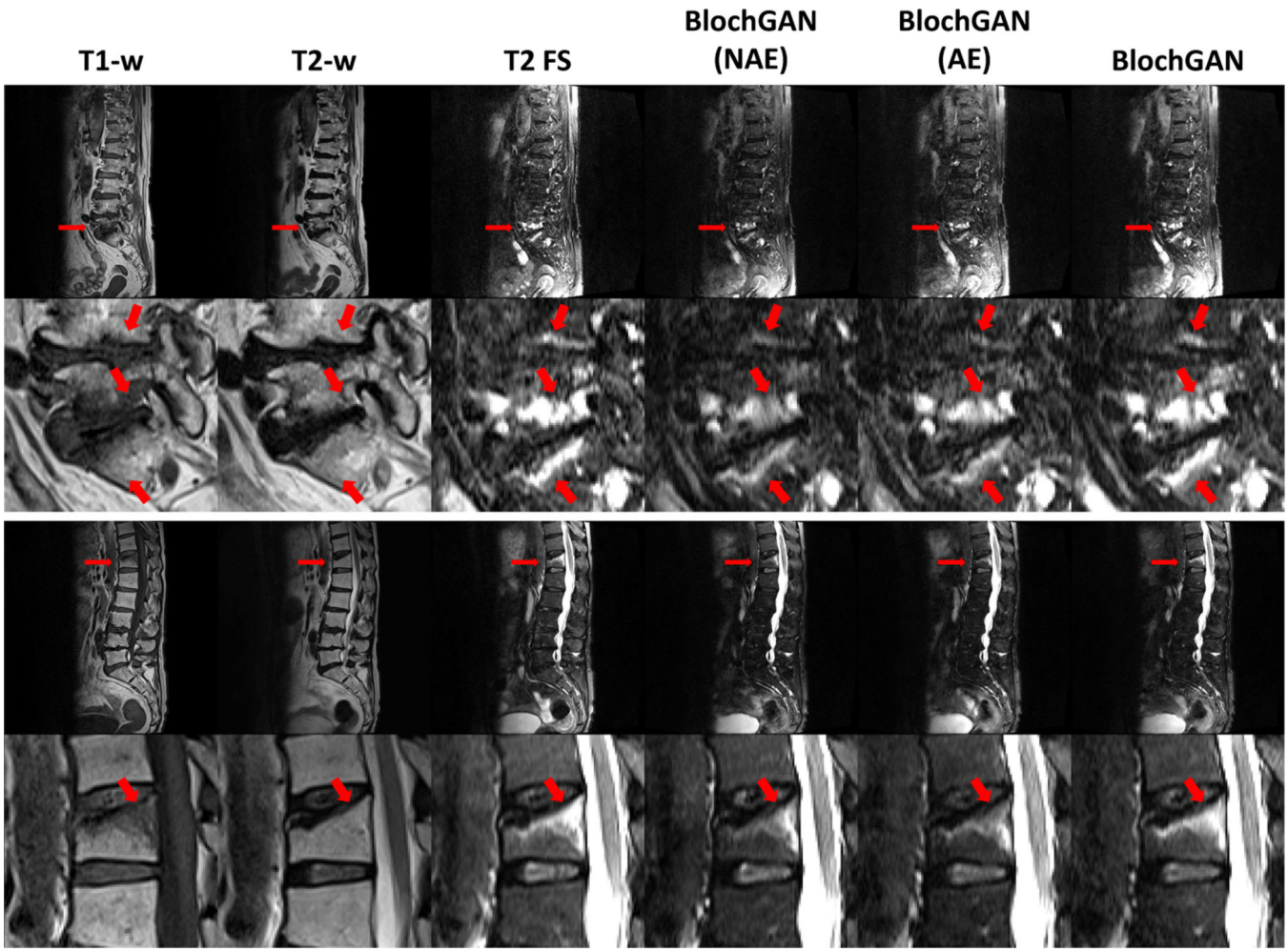


Fig. 3.

Acquired T1-weighted, T2-weighted, and fat saturation T2-weighted (T2 FS) images and generated T2 FS images using BlochGAN(NAE), BlochGAN(AE), and BlochGAN. BlochGAN(NAE) uses no autoencoder regularization, and BlochGAN(AE) uses conventional autoencoder regularization instead of the Bloch equation-based autoencoder regularization of BlochGAN. Images in rows 1 and 2 are from dataset 1, and images from rows 3 and 4 are from dataset 2. Rows 2 and 4 are the magnified images of the areas in rows 1 and 3 indicated by the red arrows. Red arrows in rows 2 and 4 indicate edema regions.

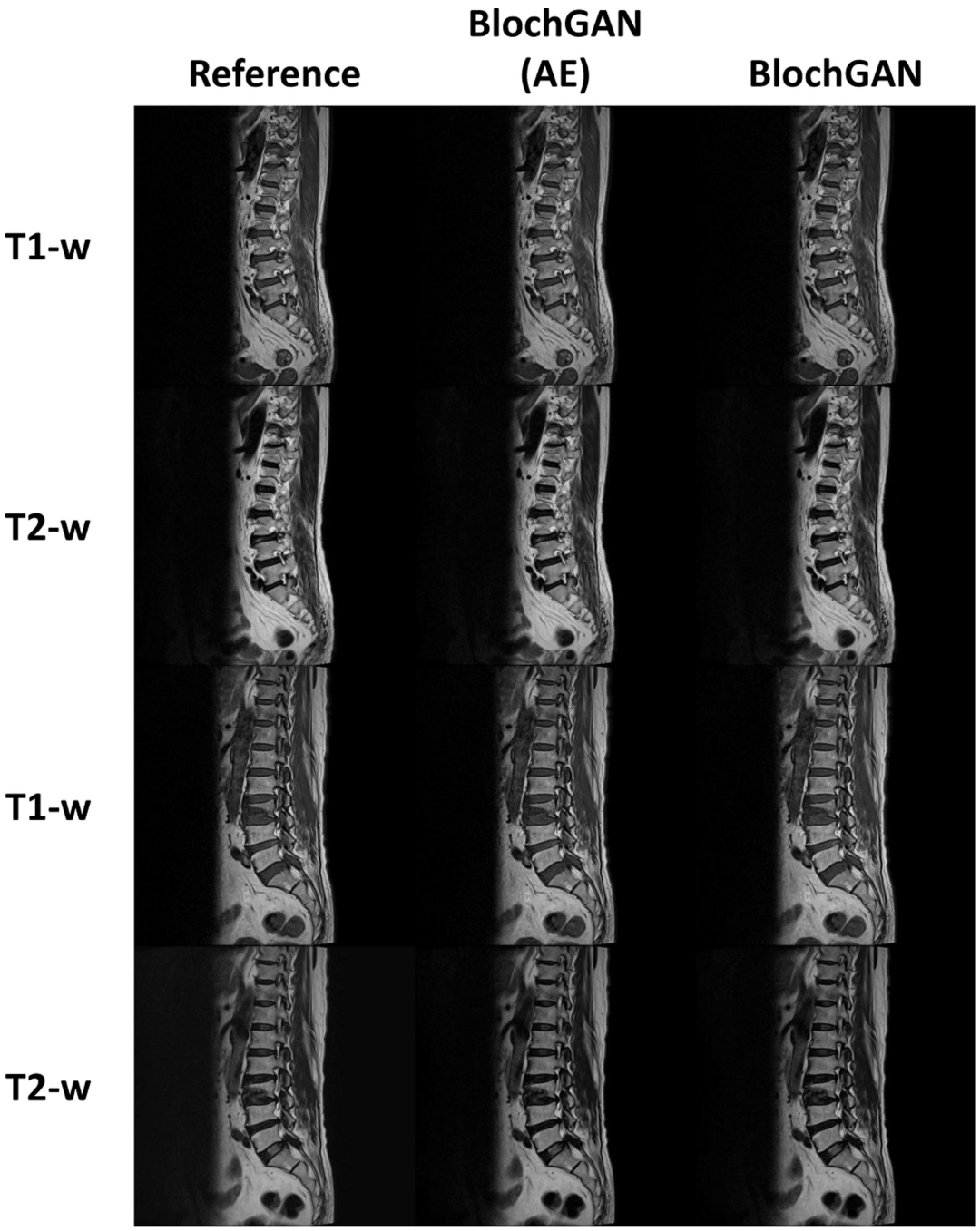


Fig. 4.

Comparison between reference and reconstructed T1- and T2-weighted images by BlochGAN(AE) and BlochGAN from dataset 1 (rows 1 and 2) and 2 (rows 3 and 4). BlochGAN(AE) uses conventional autoencoder regularization instead of the Bloch equation-based autoencoder regularization of BlochGAN.

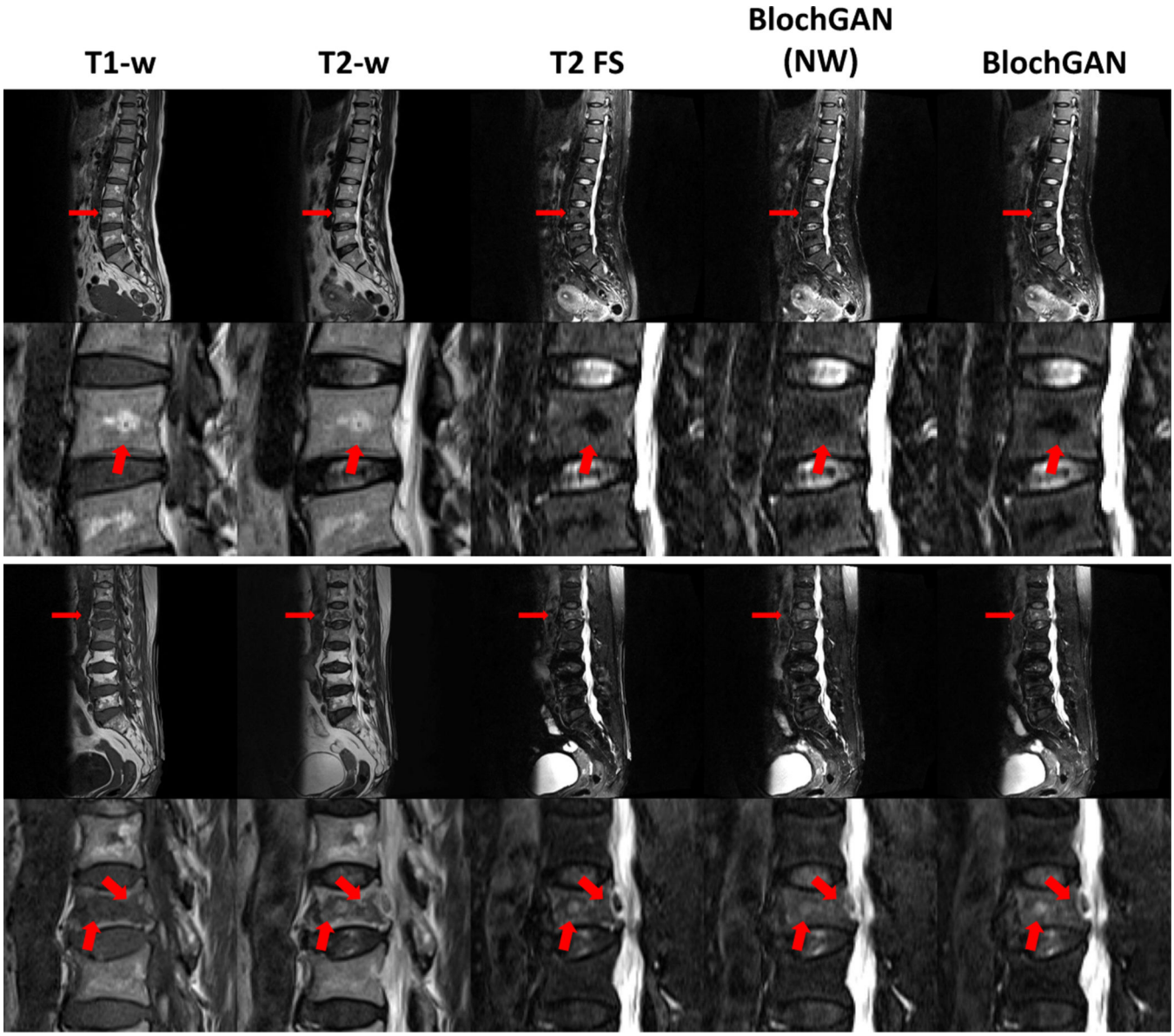


Fig. 5.

Acquired T1-weighted, T2-weighted, and fat saturation T2-weighted (T2 FS) images and generated T2 FS images using BlochGAN(NW) and BlochGAN. BlochGAN(NW) uses max-pooling layers and up-sampling layers instead of wavelet decomposition and reconstruction. Rows 1 and 2 are from dataset 1 and rows 3 and 4 are from dataset 2. Rows 2 and 4 are the magnified images of the areas indicated by red arrows in rows 1 and 3. Red arrows in row 2 indicate degenerated areas, and red arrows in row 4 indicate bone marrow edema regions and portions of spinal cords.

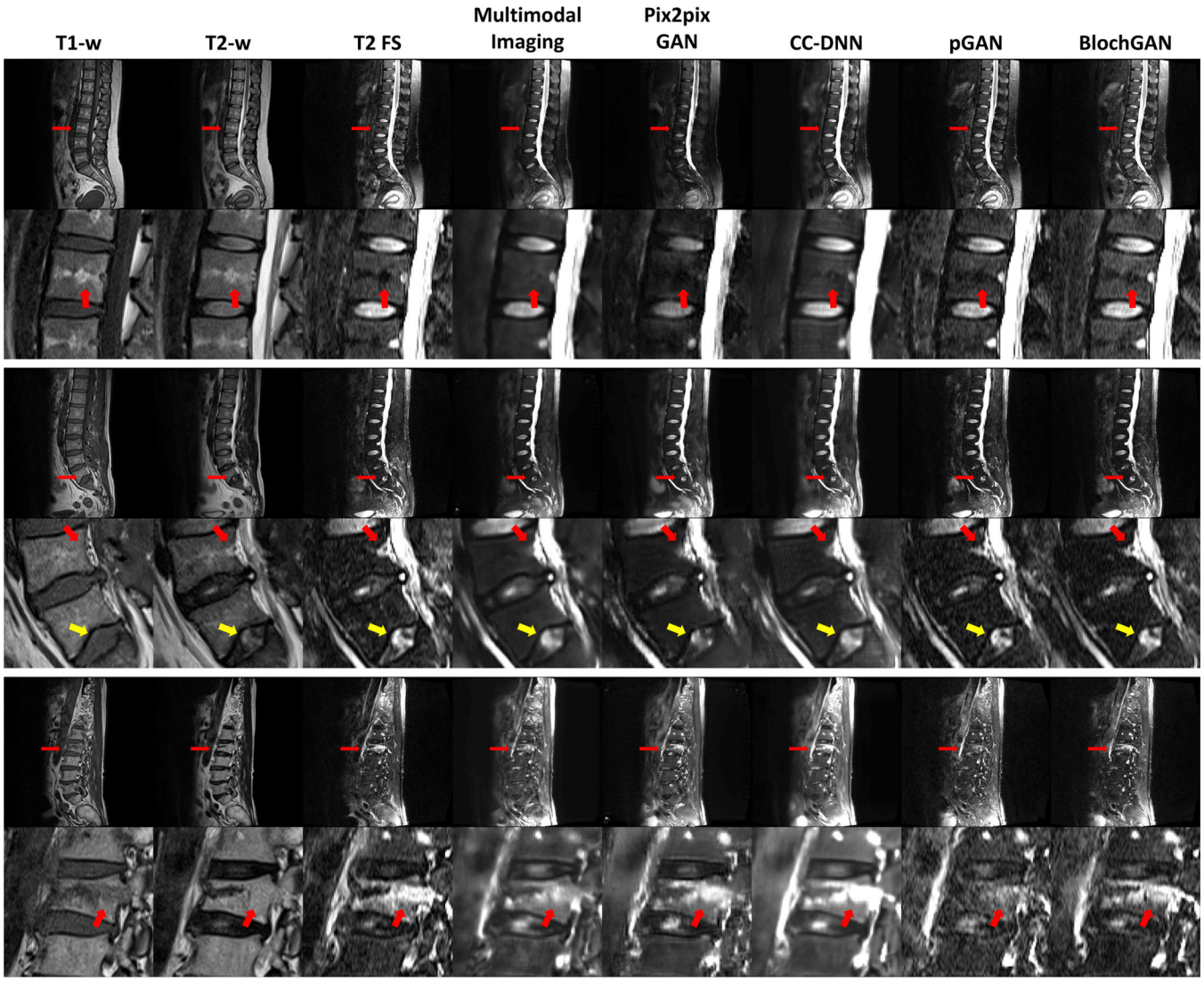


Fig. 6.

Acquired T1-weighted, T2-weighted, and fat saturation T2-weighted (T2 FS) images and generated T2 FS images using Multimodal Imaging, Pix2pix GAN, CC-DNN, pGAN, and BlochGAN from dataset 1. Rows 2, 4, and 6 are the magnified images of the areas indicated by red arrows in rows 1, 3, and 5. Red arrows in row 2 indicate degenerated areas. Red and yellow arrows in row 4 indicate the blood vessel and the nucleus pulposus of the intervertebral discs, respectively. Red arrows in row 6 indicate bone marrow edema regions.

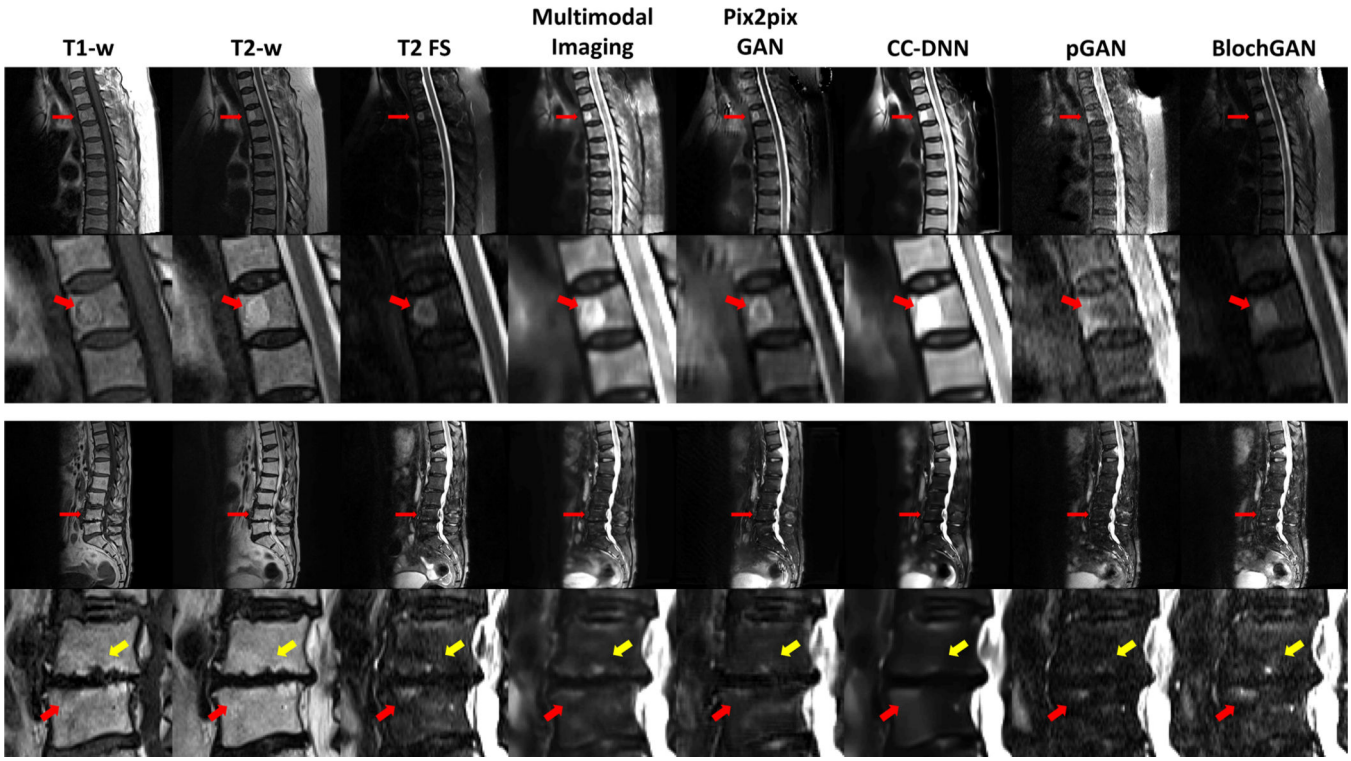


Fig. 7.

Acquired T1-weighted, T2-weighted, and fat saturation T2-weighted (T2 FS) images and generated T2 FS images using Multimodal Imaging, Pix2pix GAN, CC-DNN, pGAN, and BlochGAN from dataset 2. Rows 2, 4, and 6 are the magnified images of the areas indicated by red arrows in rows 1, 3, and 5. Red arrows in row 2 indicate bone marrow edema (BME) regions. Red and yellow arrows in row 4 indicate tiny BME regions.

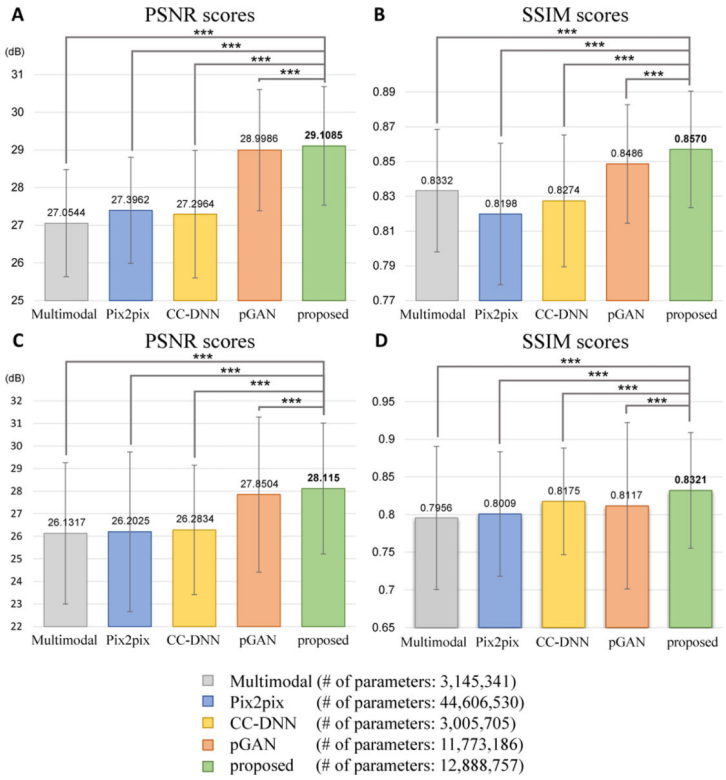


Fig. 8.

Quantitative evaluation results of the proposed method and other comparative methods. (A, B) Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) scores for dataset 1. (C, D) PSNR and SSIM scores for dataset 2. The numbers in parenthesis indicate the number of parameters used in each network.

Table 1.

Quantitative evaluation results of model comparisons

		BlochGAN(NAE)	BlochGAN(AE)	BlochGAN(NW)	BlochGAN
Datasetl	PSNR (dB)	28.8951 ± 1.5887 (***)	28.7660 ± 1.5216 (***)	29.0641 ± 1.5960 ()	**29.1085 ± 1.5725**
	SSIM	0.8446 ± 0.0354 (***)	0.8412 ± 0.0359 (***)	0.8527 ± 0.0337 (***)	**0.8570 ± 0.0335**
Dataset2	PSNR (dB)	27.9736 ± 3.1481 (***)	27.9541 ± 2.9810 (***)	27.9724 ± 3.3548 (**)	**28.1150 ± 2.8986**
	SSIM	0.8067 ± 0.1148 (***)	0.8082 ± 0.0999 (***)	0.8162 ± 0.1072 (***)	**0.8321 ± 0.0766**

Abbreviations: PSNR, peak signal-to-noise ratio; SSIM, structural similarity.

Notes: *p*-values between the proposed method and other methods were calculated by paired *t*-tests.

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$

Table 2.

Expert evaluations of synthetic T2 fat saturation images generated by Multimodal imaging, Pix2pix GAN, CC-DNN, pGAN, and our proposed Bloch equation-based autoencoder regularization generative adversarial network (BlochGAN) for 30 subjects randomly selected from dataset 1 and 2.

Categories	Multimodal imaging	Pix2pix GAN	CC-DNN	pGAN	BlochGAN
Bone marrow	2.80 ± 0.63	2.85 ± 0.66	2.60 ± 0.72	4.63 ± 0.64	**4.65 ± 0.55**
BME	2.21 ± 0.77	2.65 ± 0.69	2.29 ± 0.87	3.76 ± 0.89	**4.32 ± 0.53**
Disc	3.83 ± 0.58	3.28 ± 0.64	3.17 ± 0.49	4.78 ± 0.52	**4.85 ± 0.40**
CSF	4.15 ± 0.71	3.67 ± 0.86	4.05 ± 0.62	4.88 ± 0.59	**4.92 ± 0.33**
Muscle	2.42 ± 0.59	2.85 ± 0.48	2.45 ± 0.72	4.68 ± 0.62	**4.75 ± 0.44**
Spinal cord	3.72 ± 0.72	3.32 ± 0.68	3.55 ± 0.59	4.77 ± 0.70	**4.85 ± 0.40**
Nerve root	3.10 ± 0.64	2.62 ± 0.67	2.74 ± 0.76	4.40 ± 0.65	**4.50 ± 0.54**
Total	3.17 ± 0.94 (***)	3.06 ± 0.76 (***)	3.02 ± 0.89 (***)	4.60 ± 0.71 (***)	**4.72 ± 0.49**

Abbreviations: BME, bone marrow edema; CSF, cerebrospinal fluid.

Notes: *p*-values between the proposed method and other methods were calculated by paired *t*-tests.

* $p < 0.05$

** $p < 0.01$

*** $p < 0.001$