RESEARCH ARTICLE

High-fidelity direct contrast synthesis from magnetic resonance fingerprinting

Ke Wang[1,2] | **Mariya Doneva**[3] | **Jakob Meineke**[3] | **Thomas Amthor**[3] | **Ekin Karasan**[1] | **Fei Tan**[4] | **Jonathan I. Tamir**[5] | **Stella X. Yu**[1,2,6] | **Michael Lustig**[3]

[1]Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, California USA

[2]International Computer Science Institute, University of California at Berkeley, Berkeley, California USA

[3]Philips Research Europe, Hamburg, Germany

[4]Bioengineering, UC Berkeley-UCSF, San Francisco, California USA

[5]Chandra Family Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas USA

[6]Computer Science and Engineering, University of Michigan, Ann Arbor, Michigan USA

Correspondence
Ke Wang, Electrical Engineering and Computer Sciences, University of California at Berkeley, CA 94720, USA.
Email: kewang@berkeley.edu

Purpose: This work was aimed at proposing a supervised learning-based method that directly synthesizes contrast-weighted images from the Magnetic Resonance Fingerprinting (MRF) data without performing quantitative mapping and spin-dynamics simulations.

Methods: To implement our direct contrast synthesis (DCS) method, we deploy a conditional generative adversarial network (GAN) framework with a multi-branch U-Net as the generator and a multilayer CNN (PatchGAN) as the discriminator. We refer to our proposed approach as **N-DCSNet**. The input MRF data are used to directly synthesize T1-weighted, T2-weighted, and fluid-attenuated inversion recovery (FLAIR) images through supervised training on paired MRF and target spin echo-based contrast-weighted scans. The performance of our proposed method is demonstrated on in vivo MRF scans from healthy volunteers. Quantitative metrics, including normalized root mean square error (nRMSE), peak signal-to-noise ratio (PSNR), structural similarity (SSIM), learned perceptual image patch similarity (LPIPS), and Fréchet inception distance (FID), were used to evaluate the performance of the proposed method and compare it with others.

Results: In-vivo experiments demonstrated excellent image quality with respect to that of simulation-based contrast synthesis and previous DCS methods, both visually and according to quantitative metrics. We also demonstrate cases in which our trained model is able to mitigate the in-flow and spiral off-resonance artifacts typically seen in MRF reconstructions, and thus more faithfully represent conventional spin echo-based contrast-weighted images.

Conclusion: We present N-DCSNet to directly synthesize high-fidelity multicontrast MR images from a single MRF acquisition. This method can significantly decrease examination time. By directly training a network to generate contrast-weighted images, our method does not require any model-based simulation and therefore can avoid reconstruction errors due to dictionary matching and contrast simulation (code available at: https://github.com/mikgroup/DCSNet).

KEYWORDS
convolutional neural network, direct contrast synthesis, generative adversarial network, magnetic resonance fingerprinting

1 | INTRODUCTION

MRI is an effective imaging modality offering tremendous benefits to both science and medicine. The main advantage of MRI is the richness of soft tissue contrast that can be generated by simply changing the pulse sequence parameters. Image contrast in MRI is dominated by biophysical tissue properties, such as proton density (PD), longitudinal/transverse relaxation (T1/T2), magnetic susceptibility, and diffusion. These parameters provide information on the tissue composition and its micro-structure, and are excellent biomarkers for diagnosing and assessing disease. Measuring the quantitative value of tissue parameters, that is, through quantitative MRI, is desirable, because it provides a standardized metric for tissue properties.[1] However, quantitative MRI has been notoriously challenging to implement and standardize in clinical practice. Traditional mapping sequences require many lengthy scans to map a single parameter and thus are unsuitable for rapid imaging. Consequently, current diagnostic examinations are composed of a series of several scans, each qualitatively emphasizing one of the physical parameters above. For example, routine brain MRI includes PD-weighted scans, wherein brighter pixel intensities indicate a higher density of protons; T1-weighted (T1w) scans, wherein brighter intensities indicate shorter T1 recovery; T2-weighted (T2w) scans, wherein brightness indicates longer T2 relaxation; fluid-attenuated inversion recovery (T1/T2-FLAIR), wherein fluid signals are suppressed; and diffusion scans, wherein brighter intensities indicate less diffusivity. The relative contrast differences within and across these scans can aid in the assessment of disease.

Owing to the need for multiple scans to obtain multiple contrasts, the typical MRI protocol is lengthy, requiring patients to remain still for tens of minutes and hindering scanner throughput. In recent years, notable research efforts have focused on acquiring or synthesizing multicontrast images from single scans or fewer scans to shorten the total examination time.[2-7] These techniques have shown early success in clinical practice.[8,9] For example, synthetic MR methods[2,8,10-12] acquire multiple short scans and use parameter fitting and physical models to simulate a variety of contrast-weighted images. T2 shuffling[3,13] reconstructs multiple contrast-weighted images along the transverse relaxation curve by using a single volumetric fast spin echo acquisition, through randomly shuffling the phase encoding view ordering and performing subspace modeling. Similarly, multitasking[14,15] approaches use tensor low-rank constraints to reconstruct multiple contrast-weighted images from a single rapid acquisition. The above approaches all require scan parameters to be carefully chosen to limit confounding factors and isolate a small number of quantitative MRI parameters contributing to the overall image contrast.

Instead of decreasing confounding factors, an alternative approach known as magnetic resonance fingerprinting (MRF)[5,6,8] was proposed to mix many quantitative parameters by using a short acquisition with randomized scan parameters. MRF has accelerated the pace of clinical quantitative MRI by demonstrating the ability to rapidly and reliably generate multiple quantitative parameter maps from a single scan. MRF acquisition is usually based on gradient echo sequences and consists of rapid repetition times (TR) with under-sampled spiral readouts, in which the flip angle is modified for every TR, such that the steady state of spin dynamics is never achieved. MRF produces a sequence of images in which tissues with different relaxation and field properties (T1, T2, PD, B0, and B1) produce a unique time series or "fingerprint." The quantitative parameters of the tissue are then extracted by matching the resulting time series of each pixel to the closest signal in a precomputed dictionary constructed by simulating the Bloch equation for parameter combinations within a realistic range.

The fact that quantitative parameters can be extracted from MRF also indicates that the information embedded should be sufficient to synthesize contrast-weighted images. Although quantitative parameter maps provide meaningful physical tissue parameters, clinicians still rely primarily on contrast-weighted images for clinical diagnosis. Therefore, an opportunity exists for MRF to enable both parameter maps and synthetic contrast MRI to be provided by a single sequence.

One approach to synthesizing contrast-weighted images from MRF is to first fit the quantitative parameters and then simulate the contrast-weighted images.[10] Figures 1 and 2A show the spin-dynamic simulation pipeline, which uses quantitative parameter maps to synthesize different contrast-weighted images by using the Bloch equation or extended phase graphs (EPG).[16] Unfortunately, contrast-weighted images generated in this manner often exhibit artifacts because of many sources of error. Errors can arise from discrepancies between the MRF sequence and the dictionary simulation, for example, when flow, diffusion, magnetization transfer, excitation slice profile, or partial volume is not modeled appropriately. This limitation is most pronounced in FLAIR contrast, in which errors are seen along the boundaries of cerebrospinal fluid.[9]

An alternative, and relatively more straightforward, pipeline avoids explicit modeling and instead directly learns how to synthesize contrast-weighted images from the MRF data through neural networks. We refer to this approach as Direct Contrast Synthesis (DCS). Previous work[17] has proposed a supervised DCS method in which
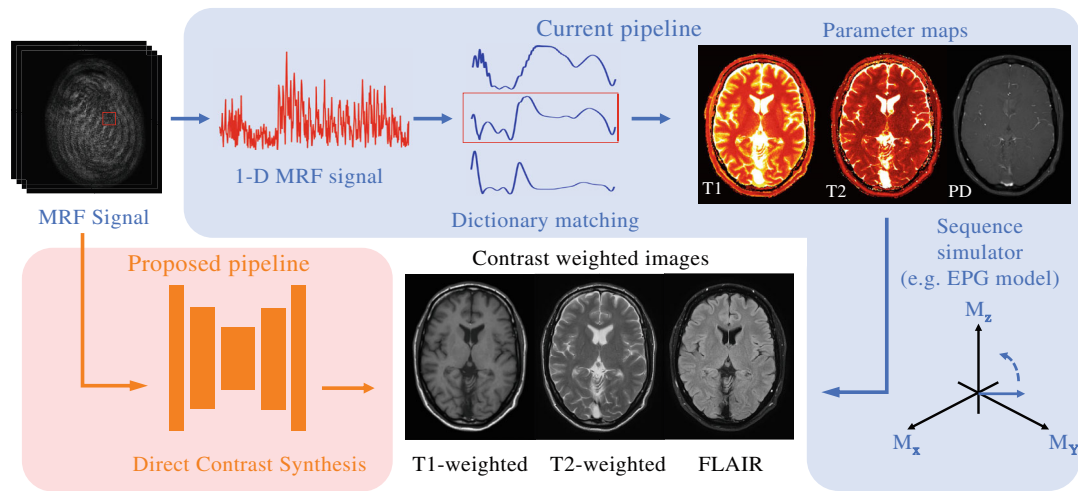
FIGURE 1 Contrast synthesis from magnetic resonance fingerprinting (MRF) via a current simulation-based pipeline and proposed direct contrast synthesis (DCS) pipeline. The simulation-based method takes the predicted quantitative parameter maps from MRF and synthesizes different contrast-weighted images by simulating the MRI physics. Our proposed DCS uses a spatial convolutional neural network to transform the MRF time series directly into different contrast-weighted images. DCS bypasses dictionary matching and contrast simulation steps, avoids modeling and acquisition imperfections, and produces high-fidelity contrast-weighted images.
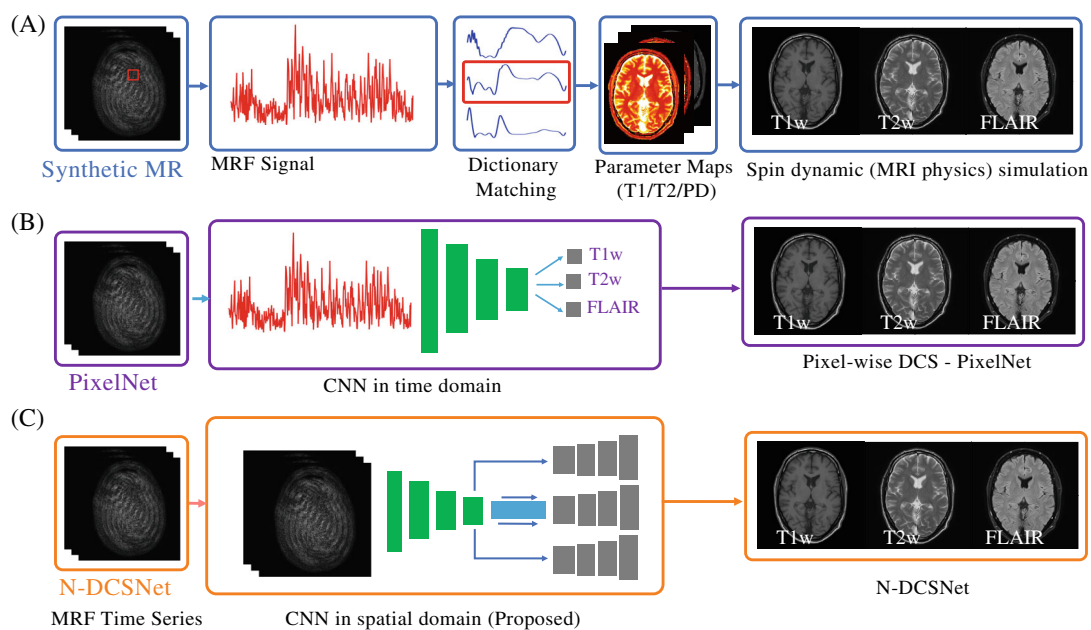

FIGURE 2 Three possible pipelines to generate contrast-weighted images from magnetic resonance fingerprinting (MRF). (A_ Synthetic MR generates multicontrast images through dictionary matching and sequence simulation (e.g., Bloch equation, extended phase graph). (B) PixelNet uses a one-dimensional pixel-wise time-domain convolutional neural network to output a qualitative contrast weighting for each voxel. (C) Our proposed N-DCSNet leverages a generative adversarial network-based architecture and spatial-convolutional network to synthesize multicontrast images.

a network was trained to take a single voxel MRF time series and map it to a specific contrast weighting (e.g., T1w, T2w, or FLAIR). This approach, which we refer to as **PixelNet**, is illustrated in Figure 2B. By training on many pairs of MRF and contrast-weighted images, PixelNet can achieve better results than dictionary mapping and simulation-based contrast synthesis.

However, by processing each pixel independently, PixelNet does not leverage the spatial structure in the data and thus can suffer from noise and spatial inconsistency. To address this issue, we propose to implement DCS as an image sequence-to-image translation task to leverage structural information. In the field of computer vision, image-to-image translation is an established problem

that aims to translate an image from a source domain to a target domain (e.g., reconstructing objects from edge maps[18] and colorizing images[19]). Recent studies have shown promising results through image-to-image convolutional neural networks (CNNs) and generative adversarial networks (GANs).[18,20] The seminal work of pix2pix[18] investigated conditional adversarial networks as a general-purpose solution to image-to-image translation problems. CycleGAN[21] improved upon the technique of learning image-to-image translation in the absence of paired examples. Image-to-image translation has also been applied in the fields of medical imaging and MRI. For example, References 22-24 learned cross-modality image synthesis between MRI and CT images; References 25 and 26 synthesized T2w images from T1w images; Reference 27 synthesized 7T high-resolution, high-SNR images from 3T input images; and Reference 28 introduced a multi-task deep learning model to synthesize multicontrast MRI images from multi-echo sequences. Recently, Reference 29 proposed a residual transformer-based deep learning model for multimodal cross-contrast MR image synthesis.

Inspired by previous works, we propose to use a conditional GAN-based architecture for DCS from MRF that enables substantial improvements in image quality and computation efficiency over simulation-based contrast synthesis and PixelNet. We refer to our approach as **N-DCSNet**, first described in Reference 30, where **N** represents N different contrasts that can be synthesized by our network (here $N = 3$). Figure 2 summarizes the three pipelines of producing synthetic, multicontrast images.

As illustrated in Figures 1 and 2C, **N-DCSNet** directly synthesizes different contrast weighted images, that is, T1w, T2w, or FLAIR, from the MRF time series data through a spatial CNN. Our generator is designed as a U-Net with a single encoder and multibranch decoders.[31] We implement a multilayer CNN (PatchGAN)[18] as the discriminator. The generator is based on spatial convolutions, thus allowing the network to learn and exploit spatial structural information. Different contrast-weighted outputs share the same encoder to exploit the shared information across contrasts. Separate decoders are designed to learn the unique features of each contrast. During the training procedure, we leverage a conditional GAN framework, wherein the time average of the MRF time series is also used as an input to the discriminator to constrain the GAN training.

In-vivo experiments on healthy volunteers show that *N-DCSNet* can generate high-fidelity, multi-contrast images from MRF time-series. Our approach outperforms contrast synthesis from parameter maps and PixelNet both qualitatively and quantitatively. Furthermore,

we demonstrate that *N-DCSNet* can inherently mitigate some artifacts that appear in MRF, such as slice in-flow artifacts and spiral off-resonance blurring. Our main contributions can be summarized as follows:

- We introduce a spatial CNN-based method to learn the mapping between MRF time series and contrast-weighted images (i.e., T1w, T2w, and FLAIR). Our approach can avoid the simulation errors typically seen in Synthetic MR.

- We use a conditional GAN-based framework to encourage finer textures and produce more faithful contrasts. Additionally, our *N-DCSNet* can inherently mitigate slice in-flow artifacts as well as spiral off-resonance blurring.

- *N-DCSNet* outperforms simulation-based contrast synthesis from parameter maps and PixelNet qualitatively and according to quantitative metrics. It also has significant computation advances. During inference, our approach is significantly faster than simulation-based contrast synthesis and PixelNet, thus improving the potential for clinical adoption.

2 | METHODS

In this section, we first describe the data acquisition protocols and the simulation-based contrast synthesis via parameters used as our baseline for comparisons (Section 2.1). Then, we introduce our GAN-based framework design for *N-DCSNet* (Section 2.2). Next, we detail the loss functions (Section 2.3) and the training process. Finally, we compare our method with previous approaches (Section 2.4).

2.1 | Data acquisition and contrast synthesis via parameters

2.1.1 | Data acquisition

After obtaining IRB approval, we scanned 21 men, ranging from 29 to 61 years of age, with a 1.5 T Philips Ingenia scanner using a 15-channel head coil. A total of 13 channels were selected by using automatic coil selection. To avoid conducting so-called "data crimes,"[32] we report our data preparation pipeline as follows. Four consecutive axial brain scans were acquired for each examination session. The participants were instructed to remain still throughout the examination so that data across scans remained registered. The scans were as follows:

- A spoiled gradient echo[33] MRF sequence with 500 time points, constant TE = 3.3 ms, TR = 20 ms (Each TR consisted of a spiral-out readout. The spirals between two consecutive TR were rotated by 9°.). The readout time is 12 ms and the undersampling factor is 20.
- T1w spin echo with TE = 15 ms, TR = 450 ms, flip angle = 69°, and two averages.
- T2w turbo spin echo with TE = 110 ms, TR = 1990–2215 ms, ETL = 16, flip angle = 90° and two averages.
- FLAIR inversion recovery turbo spin echo with TE = 120 ms, TR = 8500 ms, TI = 2500 ms, ETL = 41, flip angle = 90° and two averages.

All scans were acquired with an in-plane resolution of 0.72×0.72 mm (FOV 230×230 mm^2, matrix size 320×320) and nine to ten slices with a thickness of 5 mm.

Of the 21 participants, 17 were scanned twice (on different days), thus resulting in a total of 38 examinations. FLAIR sequences were acquired for only 26 of the 38 examinations. Only the 26 examinations with all four sequences were used in this study, of which 21 were used for training, two were used for validation, and three were used for testing. The data from participants used for testing were not included in any of the training sets.

To further minimize residual motion or misalignment between scans, we employ a two-dimensional (2D) rigid in-plane registration per slice, aligning the ground truth contrast-weighted images with the time-averaged MRF image. Moreover, we manually inspect the images and discard those exhibiting significant in-plane and through-plane movements.

2.1.2 | Preprocessing

Each of the three contrast-weighted image data was normalized with respect to the 95th percentile of the intensity values for each image. MRF time series images were reconstructed from each TR by using gridding with density compensation[34,35] followed by coil combination with Philips' CLEAR. The MRF data were then normalized as follows. For each dataset, an averaged image from the 500 time points was computed. The 95th percentile of the magnitude values from the average MRF image was then used to normalize the time series.

2.1.3 | Parameter maps and contrast simulation

The dictionary for MRF parameter mapping was simulated by using EPG.[16] The dictionary consisted of 22 031 MRF

signals with T1 parameters ranging from 4 to 3000 ms and T2 parameters ranging from 2 to 2000 ms. Each simulated signal in the dictionary was scaled to have a Euclidean norm equal to one. We used cosine similarity[6] to match the acquired MRF signal to the nearest neighbor in the simulated dictionary (Figure 2A). Additional factors, such as B1 inhomogeneity and slice profile, were not included in the simulated dictionary.

The parameter maps (T1, T2) obtained from dictionary matching were then used to simulate the contrast-weighted images. The T1w spin echo (SE) has a closed form for specific TE and TR, and PD parameters:

$$\text{SE(PD, T1, T2, TE, TR)} = \text{PD} \cdot (1 - e^{-\frac{\text{TR}-\text{TE}}{\text{T1}}}) \cdot e^{-\frac{\text{TE}}{\text{T2}}}. \quad (1)$$

PD was computed by taking the magnitude of the inner product between the acquired MRF signal and the nearest neighbor in the simulated dictionary. The T2w and FLAIR sequences are based on turbo spin echo and do not have closed forms. For these, we used EPG[16] to simulate the contrast-weighted images.

2.2 | N-DCSNet framework

Figure 3 illustrates the overall pipeline of our proposed *N-DCSNet*. Our network expects the complex-valued MRF time series $\mathbf{MRF}_{in} \in \mathbb{C}^{t \times h \times w}$ as input, where t, h, and w correspond to the number of time points, image height, and image width, respectively ($t, h, w \in \mathbb{N}$). The network outputs are real-positive (magnitude) contrast weighted images $\mathbf{T\hat{1}w}, \mathbf{T\hat{2}w}, \mathbf{FL\hat{A}IR} \in \mathbb{R}^{h \times w}$. In our experiments, $t = 500, h = w = 320$.

We designed a conditional GAN-based framework for *N-DCSNet*, the standard framework in References 18 and 21, consisting of a generator (G) and a discriminator (D).

First, for the input complex-valued MRF data with dimensions $500 \times 320 \times 320$ (time $\times$ width $\times$ height), we concatenate the real and imaginary parts along with the time dimension as channels to the network. This results in a real-valued input with dimensions $1000 \times 320 \times 320$.

Our generator is a modified U-Net,[31] which consists of one shared encoder and multiple independent decoders. The shared encoder exploits structural similarities across the multicontrast images, whereas the independent decoders learn the unique features of the different contrasts. At test time, **N-DCSNet** produces multi-contrast images with a single network. The discriminator (D) is a multilayer CNN (patchGAN)[18] that penalizes structure at a patch scale. D aims to classify whether each $N \times N$ patch in an image is real or fake. We run this discriminator convolutionally across the image, averaging all responses to provide the final output of D. To constrain the GAN training,
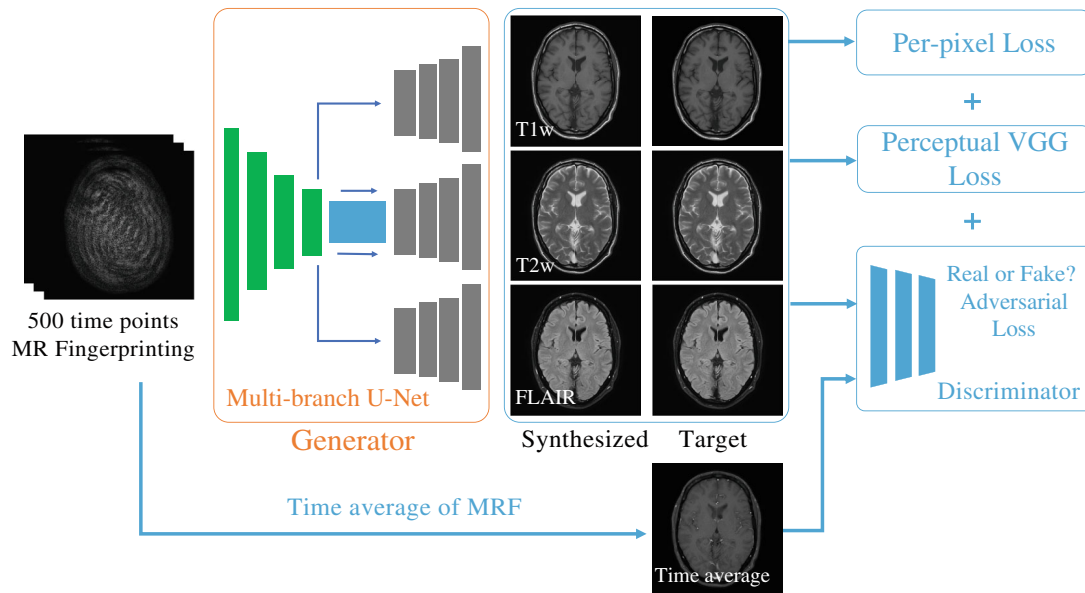
FIGURE 3 Illustration of our proposed N-DCSNet framework. Given a complex-valued magnetic resonance fingerprinting (MRF) time series $\mathbf{MRF}_{in} \in \mathbb{C}^{t \times h \times w}$, with number of time points $t \in \mathbb{N}$ and image dimensions $h, w \in \mathbb{N}$, *N-DCSNet* synthesizes three contrast-weighted images (T1w, T2w, and FLAIR) with a single network. We designed a multibranch U-Net as the generator and a multilayer convolutional neural network (CNN) as the discriminator by following the conditional generative adversarial network (GAN) training strategy. To constrain the GAN training, we additionally input the time average of MRF to the discriminator. A combination of per-pixel ℓ_1 loss, perceptual VGG loss, and adversarial loss is imposed on the network. *N-DCSNet* generates high-fidelity contrast-weighted images with sharper edges, finer textures, and more faithful contrasts than simulation-based contrast synthesis and PixelNet.

we follow Reference 18 and further input the magnitude of the MRF time-averaged image to the discriminator to provide structural guidance. This image has mixed contrast, because of averaging, and significantly fewer spiral undersampling artifacts than the MRF time-series images. We denote it $\mathbf{MRF}_{avg}$.

During training, the generator G learns to predict high-quality contrast-weighted images that cannot be distinguished from the real acquired images (ground truth) by an adversarially trained discriminator D. Meanwhile, D is simultaneously trained to distinguish the generated images (labeled as "fake") from the ground truth images (labeled as "real").

2.3 | Loss functions

Our proposed *N-DCSNet* is fully supervised, with the purpose of generating high-fidelity contrast-weighted images that are close to the ground truth real acquisitions. The loss function of our generator G is a combination of three components: (1) ℓ_1 reconstruction loss, (2) perceptual loss, and (3) adversarial loss. Given our generator G and the input MRF signal $\mathbf{MRF}_{in}$, G outputs the synthesized contrast-weighted images (T1w, T2w, and FLAIR):

$$\mathbf{T\hat{1}w}, \mathbf{T\hat{2}w}, \mathbf{FL\hat{A}IR} = G(\mathbf{MRF}_{in}). \qquad (2)$$

Then the cumulative ℓ_1 loss is formulated as:

$$L_{\ell_1} = \mathbb{E}_{\mathbf{MRF}_{in}}(\|\mathbf{T\hat{1}w} - \mathbf{T1w}\|_1 + \|\mathbf{T\hat{2}w} - \mathbf{T2w}\|_1$$

$$+ \|\mathbf{FL\hat{A}IR} - \mathbf{FLAIR}\|_1), \qquad (3)$$

where $\mathbf{T1w}, \mathbf{T2w},$ and $\mathbf{FLAIR}$ represent the real, ground-truth acquisitions of the three contrast-weighted images (Section 2.1). Per-pixel losses such as the ℓ_1 loss are known to exhibit image blurring.[18,36-38] Therefore, we incorporate additional perceptual and adversarial losses to encourage detailed reconstructions.

Perceptual losses[36,38] have been used successfully in super-resolution and image synthesis[28] tasks to improve image quality and encourage delicate structures. The underlying idea is that layer features of task-based networks, such as image classification networks, can capture high-level perceptual information in the image. Therefore, minimizing the loss in the feature space can preserve such perceptual information.[36] In this work, the perceptual loss is implemented as the ℓ_2 distance between **relu2-2** layer features of an ImageNet[39] pretrained VGG Network.[40] We denote the function used to extract these features as $\phi(\cdot)$, where $\phi(\mathbf{x})$ extracts the **relu2-2** layer features of a specific image $\mathbf{x}$. Each contrast-weighted image is scaled to $[0, 1]$, duplicated three times, and concatenated along the channel dimension (to simulate RGB channels) before feeding

into $\phi(\cdot)$. Then, the overall VGG perceptual loss term can be written as:

$$L_{\text{vgg}} = \mathbb{E}_{\mathbf{MRF}_{\text{in}}}(\|\phi(\mathbf{T\hat{1}w}) - \phi(\mathbf{T1w})\|_2 + \|\phi(\mathbf{T\hat{2}w}) - \phi(\mathbf{T2w})\|_2 + \|\phi(\mathbf{FL\hat{A}IR}) - \phi(\mathbf{FLAIR})\|_2). \quad (4)$$

The third component of our loss function is an adversarial loss. This term is used to further encourage high-frequency details and achieve more realistic synthesized outputs.[18] The generator G is trained to produce outputs that cannot be distinguished from "real" images. We concatenate the acquired images [$\mathbf{MRF}_{\text{avg}}, \mathbf{T1w}, \mathbf{T2w}, \mathbf{FLAIR}$] along the channel dimension, and treat it as the "real" sample $\mathbf{S}_{\text{real}} = [\mathbf{MRF}_{\text{avg}}, \mathbf{T1w}, \mathbf{T2w}, \mathbf{FLAIR}]$. Meanwhile, we create $\mathbf{S}_{\text{fake}} = [\mathbf{MRF}_{\text{avg}}, \mathbf{T\hat{1}w}, \mathbf{T\hat{2}w}, \mathbf{FL\hat{A}IR}]$ as the "fake" sample $\mathbf{S}_{\text{fake}}$. Then, the adversarial loss for our generator is given by:

$$L_{\text{adv}} = -\mathbb{E}_{\mathbf{S}_{\text{fake}}}[\log(D(\mathbf{S}_{\text{fake}}))]. \quad (5)$$

The overall objective function for the generator becomes:

$$L_{\text{G}} = L_{\ell_1} + \lambda_{\text{vgg}} L_{\text{vgg}} + \lambda_{\text{adv}} L_{\text{adv}}, \quad (6)$$

where λ_{vgg} and λ_{adv} are the weights of the perceptual loss and adversarial loss, respectively. In our experiments, we empirically set $\lambda_{\text{vgg}} = 0.03$ and $\lambda_{\text{adv}} = 0.015$.

Our discriminator is adversarially trained to detect the generators' outputs as "fake" images. According to Reference 20, the objective function for our discriminator L_{D} is given by:

$$L_{\text{D}} = -\mathbb{E}_{\mathbf{S}_{\text{real}}}[\log(D(\mathbf{S}_{\text{real}}))] - \mathbb{E}_{\mathbf{S}_{\text{fake}}}[\log(1 - D(\mathbf{S}_{\text{fake}}))]. \quad (7)$$

We update the parameter weights of G and D by alternatively minimizing the objectives L_{G} and L_{D}.

2.4 | Experiments

To demonstrate its effectiveness, we evaluate our *N-DCSNet* against simulation-based contrast synthesis (synthesis via parameters) and PixelNet on the same testing dataset (detailed in Section 2.1). The EPG simulation using the dictionary-matched parameters was run for all voxels in parallel by using the joblib package[41] on 24 CPUs. On the basis of the architecture introduced in Reference 17, we implemented PixelNet as a one-dimensional temporal CNN to map the MRF time series at every voxel to the corresponding three contrast-weighted scans. The PixelNet network consists of three convolutional layers followed by three fully connected layers and is trained with an ℓ_2 loss. The inference time for the different approaches is calculated by computing the average runtime of 20

separate runs of a single MRF slice. Ablation studies were also conducted to analyze the impacts of the different loss functions on the synthesized contrast-weighted images.

2.4.1 | Evaluation metrics

To quantitatively compare our results to the ground truth, we report the following evaluation metrics: normalized root mean square error (nRMSE), peak signal-to-noise ratio (PSNR), structural similarity (SSIM),[42] learned perceptual image patch similarity (LPIPS)[43] with AlexNet,[44] and Fréchet inception distance (FID) score.[45] When computing LPIPS and FID, the output images were scaled to the range [0, 255] and saved as png files.

2.4.2 | Implementation details

All the proposed algorithms and networks were implemented with PyTorch 1.8[46] on 24 GB NVIDIA 3090 graphics processing units (GPUs). Our generator and discriminator were trained by using Adam optimizer,[47] with a batch size of 4 and a learning rate of 1×10^{-4}.

We supervise the DCS with magnitude contrast weighted images. However, the MRF time series is inherently complex-valued. To reduce the sensitivity to phase, during training, we augment the phase of the MRF data on the fly by multiplying each time-series with random constant phase $e^{j\theta}$, where $j = \sqrt{-1}$ and θ is uniformly distributed between $[0, 2\pi]$.

The ablation study evaluating the loss functions' contributions was performed by comparison of the proposed combined loss function (Equation 6) against L_{ℓ_1}, $L_{\ell_1} + \lambda_{\text{vgg}} L_{\text{vgg}}$, and $L_{\ell_1} + \lambda_{\text{adv}} L_{\text{adv}}$ losses.

3 | RESULTS

3.1 | Comparisons with contrast synthesis via parameters and PixelNet

Figure 4 summarizes the results of the different contrast synthesis methods applied to a representative 2D brain slice. Compared with EPG simulation-based synthesis (synthesis via parameters),[16] and PixelNet,[17] *N-DCSNet* produces finer and cleaner structural details, sharper edges, and better perceptual agreement with the true acquisition (ground truth).

The EPG simulation-based results (synthesis via parameters) exhibit incorrect contrast and noise artifacts due to the modeling and acquisition imperfections (as expected in Section 1). PixelNet significantly improves the
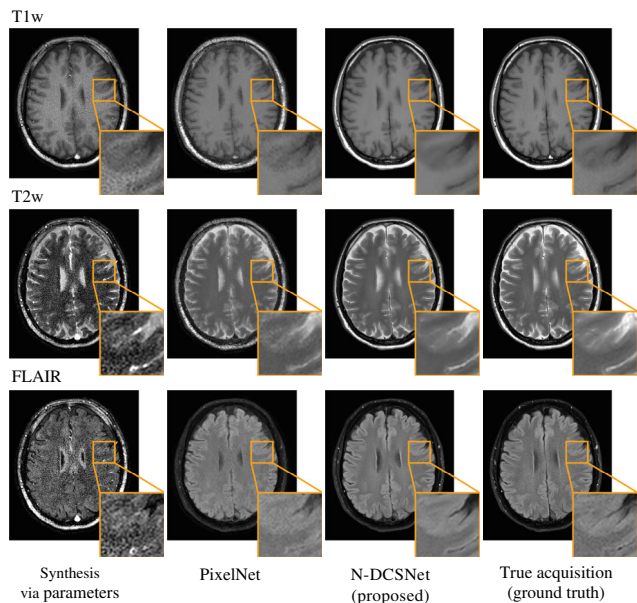
FIGURE 4 Representative contrast synthesis results of different methods (upper brain). From left to right, we compare our proposed *N-DCSNet* with simulation-based contrast synthesis via parameters,[16] PixelNet,[17] and the true acquisition. *N-DCSNet* shows better visual agreement with the true acquisition, producing finer textures and higher overall image quality than the other approaches. Zoomed-in details are displayed next to each image.

FIGURE 5 Representative contrast synthesis results of different methods (lower brain). From left to right, we compare our proposed *N-DCSNet* with simulation-based synthesis via parameters,[16] PixelNet,[17] and the true acquisition. Zoomed-in images show the inflow (vasculature) regions where parameter-based synthesis (left column) fails to deliver correct contrast, owing to the moving blood flow. In comparison, *N-DCSNet* successfully reconstructs delicate textures and produces high-quality contrast-weighted images.

synthesized image quality, but the noise artifacts persist (as shown in T1w and T2w). In comparison, *N-DCSNet* leverages both temporal and spatial information, producing more faithful contrast, preserving finer details, and showing better agreement with the ground truth images.

Figure 5 compares the results of another representative 2D slice from the lower brain. Regions of the vasculature are zoomed in and expanded at the bottom right corners. Because of the blood flow, MRF cannot retrieve accurate parameter maps by dictionary matching.[48] Therefore, synthesis via parameters fail to deliver precise contrasts in the vasculature regions (as shown in T2w images). In comparison, *N-DCSNet* produces accurate contrast and can successfully reconstruct the delicate vessel structures (as shown in T2w results). From the synthesized FLAIR images, we observe that PixelNet produces noisier images with flattened contrasts in the back of the brain. Instead, *N-DCSNet* successfully depicts the detailed textures and produces high-quality, sharper images.

Figure S1 displays an extensive collection of *N-DCSNet* synthesized images, accompanied by the corresponding parameter maps (i.e., PD, T1, T2). These parameter maps are obtained through dictionary matching. Our approach highlights the capability to produce complementary parameter maps and contrast-weighted images from a single scan.
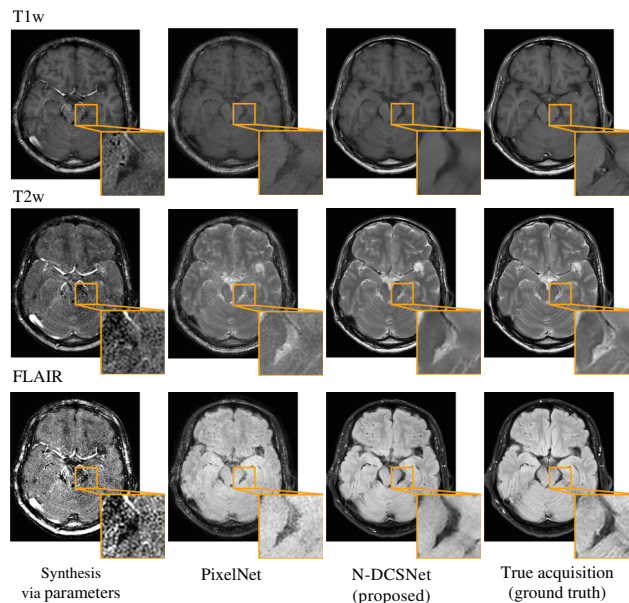
Table 1 compiles the quantitative evaluation metrics (nRMSE, PSNR, SSIM, LPIPS, and FID) of different methods (synthesis via parameters,[16] PixelNet[17] and **N-DCSNet**) for each contrast. We compute the metrics across the testing dataset and report the mean and SD. As indicated in the table, for all three contrasts (T1w, T2w, and FLAIR), our method consistently outperforms other methods in all five evaluation metrics. Of note, LPIPS and FID use learned features to measure perceptual similarity between two images or two distributions, thus resulting in better matching with human judgment than pixel-wise (nRMSE) or patch-wise (SSIM) metrics.[43,45] **N-DCSNet**, compared with PixelNet, significantly reduces the LPIPS by more than 30% and the FID by more than 50% for all three contrasts, thus demonstrating the superiority of our proposed method in terms of perceptual image quality.

Table 2 summarizes the inference times of the different approaches. As indicated in the table, simulation-based synthesis (synthesis via parameters) requires an average of 24.37 s because of the time-consuming dictionary matching and contrast simulation procedures that are repeated for each voxel across the entire image. PixelNet is more efficient, and averages 0.3421 s by leveraging parallel GPU computing (on a single NVDIA 3090). In comparison,

TABLE 1 Quantitative comparisons (nRMSE, PSNR, SSIM, LPIPS, and FID) among different contrast synthesis methods (mean ± SD).

Contrasts	Methods	nRMSE (%) ↓	PSNR (dB) ↑	SSIM ↑	LPIPS ↓ (×10⁻²)	FID ↓
T1w	Synthesis via parameters[16]	6.44 ± 1.25	24.0 ± 1.93	0.786 ± 0.030	20.1 ± 1.14	130.8
	PixelNet[17]	4.58 ± 0.83	26.9 ± 1.71	0.880 ± 0.026	11.3 ± 1.81	109.6
	N-DCSNet (ours)	**3.57 ± 0.64**	**29.1 ± 1.63**	**0.923 ± 0.019**	**6.33 ± 1.87**	**57.32**
T2w	Synthesis via parameters[16]	13.4 ± 1.68	17.5 ± 1.11	0.671 ± 0.032	21.1 ± 1.60	148.1
	PixelNet[17]	5.24 ± 0.64	25.7 ± 1.11	0.853 ± 0.027	12.6 ± 1.92	114.1
	N-DCSNet (ours)	**3.76 ± 0.59**	**28.6 ± 1.35**	**0.921 ± 0.017**	**5.77 ± 1.02**	**57.01**
FLAIR	Synthesis via parameters[16]	19.4 ± 2.75	14.3 ± 1.25	0.576 ± 0.028	20.6 ± 2.50	185.4
	PixelNet[17]	4.69 ± 0.67	26.7 ± 1.30	0.797 ± 0.025	11.3 ± 1.35	126.9
	N-DCSNet (ours)	**3.64 ± 0.65**	**29.0 ± 1.75**	**0.883 ± 0.018**	**8.63 ± 0.839**	**63.17**

Notes: We calculate the metrics for each contrast (T1w, T2w, and FLAIR) separately. *N-DCSNet* is compared with contrast synthesis via parameters[16] and PixelNet.[17] Our proposed method consistently outperforms other approaches in all five metrics for each contrast. Bold corresponds to the best results. ↑ means that higher is better, ↓ means that lower is better.

Abbreviations: FID, Fréchet inception distance; FLAIR, fluid-attenuated inversion recovery; LPIPS, learned perceptual image patch similarity; nRMSE, normalized root mean square error; peak signal-to-noise ratio (PSNR), structural similarity (SSIM).

TABLE 2 Inference times of different methods for contrast synthesis from a two-dimensional MRF time series.

	Synthesis via parameters	PixelNet	N-DCSNet (ours)
Inference time (s) ↓	24.37	0.3421	**0.01617**

Notes: N-DCSNet reduces the inference time by more than 20-fold with respect to that of PixelNet, demonstrating superior computation efficiency and the potential for clinical adoption. All experiments are implemented on a single w. Bold corresponds to the best result.

our *N-DCSNet* has 20 times faster inference time than PixelNet. *N-DCSNet* requires an average of 0.01617 s to synthesize three contrast-weighted images from a single 2D MRF time series, demonstrating superior computation efficiency and a potential for clinical translation. All experiments were run on a single NVIDIA 3090 GPU.

3.2 | Ablation study of different loss functions

To investigate and better understand the effects of loss functions (Section 2.3) on the resulting image quality, we conducted an ablation study by comparing our overall loss function L_G (Equation 6) to L_{ℓ_1}, $L_{\ell_1} + \lambda_{vgg}L_{vgg}$ and $L_{\ell_1} + \lambda_{adv}L_{adv}$ losses. We trained separate models with different objective functions and used the same training setup and datasets (i.e., training set, learning rate, epochs, etc.). Figure 6 shows the results on a representative 2D brain slice. The model trained with pure L_{ℓ_1} (left column) suffers from degraded perceptual image quality and

exhibits some blurring, in agreement with the findings in literatures.[18,28,38] Adding perceptual VGG loss (second column) encourages finer details and sharper edges. However, blurring artifacts remain (as seen in T2w and FLAIR). Adding adversarial loss on top of L_{ℓ_1} (third column) encourages even finer structures but suffers from residual blurring (T2w) and recurrent checkerboard artifacts (FLAIR). By incorporating both perceptual loss and adversarial loss, the model trained with our proposed objective (fourth column, Equation 6) further improves the synthesized image quality by reconstructing more delicate textures (T2w example) and producing more faithful contrast (FLAIR example).

Table 3 summarizes the five evaluation metrics for *N-DCSNet* trained with the different loss functions. Because the model trained with pure L_{ℓ_1} loss optimizes the pixel distances, it produces the best nRMSE and PSNR results. However, nRMSE and PSNR are known not to match human perception.[43] For perception-representative metrics (SSIM, LPIPS, and FID), *N-DCSNet* trained with our proposed full objective outperforms the other loss functions for all three contrasts (except SSIM for T1w), thus demonstrating the effectiveness of our loss functions in producing high-fidelity contrast-weighted images.

3.3 | Mitigation of spiral off-resonance artifacts

Beyond the aforementioned superior performance, we also demonstrate cases in which *N-DCSNet* effectively

FIGURE 6
Representative visual comparison of N-DCSNet with different loss functions. From left to right, our full objective (fourth column; Equation 6) is compared with L_{ℓ_1}, $L_{\ell_1} + \lambda_{vgg} L_{vgg}$, $L_{\ell_1} + \lambda_{adv} L_{adv}$ and the ground truth. Perceptual VGG loss encourages sharper edges than pure L_{ℓ_1}, whereas adversarial loss further improves the image quality. The model trained with our full objective is able to recover subtle structures and show better visual agreement with the ground truth.
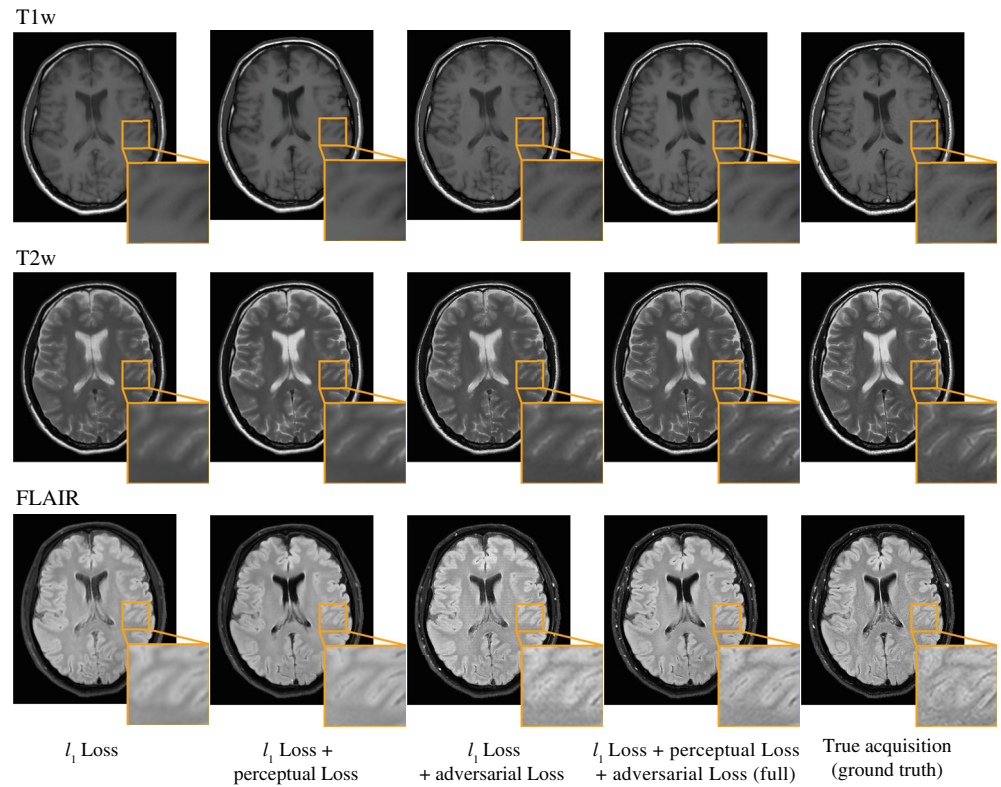

T1w

T2w

FLAIR

l_1 Loss l_1 Loss + perceptual Loss l_1 Loss + adversarial Loss l_1 Loss + perceptual Loss + adversarial Loss (full)   True acquisition (ground truth)

TABLE 3 Quantitative comparisons (nRMSE, PSNR, SSIM, LPIPS, and FID) of N-DCSNet with different loss function designs (mean ± SD).

Contrasts	Methods	nRMSE (%) ↓	PSNR (dB) ↑	SSIM ↑	LPIPS ↓ (×10⁻²)	FID ↓
T1w	L_{ℓ_1}	**3.34 ± 0.63**	**29.7 ± 1.69**	0.918 ± 0.018	8.02 ± 2.40	67.39
	$L_{\ell_1} + \lambda_{vgg} L_{vgg}$	3.43 ± 0.82	29.5 ± 2.16	**0.926 ± 0.022**	9.14 ± 2.53	66.66
	$L_{\ell_1} + \lambda_{adv} L_{adv}$	3.68 ± 0.93	28.7 ± 1.72	0.921 ± 0.020	8.04 ± 2.18	62.94
	$L_{\ell_1} + \lambda_{vgg} L_{vgg} + \lambda_{adv} L_{adv}$	3.57 ± 0.64	29.1 ± 1.63	0.923 ± 0.019	**6.33 ± 1.87**	**57.32**
T2w	L_{ℓ_1}	**3.57 ± 0.67**	**29.2 ± 1.64**	0.914 ± 0.018	10.08 ± 1.73	71.44
	$L_{\ell_1} + \lambda_{vgg} L_{vgg}$	3.67 ± 0.61	28.8 ± 1.45	0.918 ± 0.018	8.67 ± 1.46	64.55
	$L_{\ell_1} + \lambda_{adv} L_{adv}$	3.79 ± 0.72	28.4 ± 1.55	0.919 ± 0.024	7.57 ± 1.18	60.30
	$L_{\ell_1} + \lambda_{vgg} L_{vgg} + \lambda_{adv} L_{adv}$	3.76 ± 0.59	28.6 ± 1.35	**0.921 ± 0.017**	**5.77 ± 1.02**	**57.01**
FLAIR	L_{ℓ_1}	**3.44 ± 0.66**	**29.4 ± 1.72**	0.879 ± 0.017	11.1 ± 0.98	93.01
	$L_{\ell_1} + \lambda_{vgg} L_{vgg}$	3.73 ± 0.61	28.7 ± 1.55	0.878 ± 0.019	10.7 ± 1.02	96.01
	$L_{\ell_1} + \lambda_{adv} L_{adv}$	3.68 ± 0.93	28.1 ± 1.71	0.869 ± 0.021	9.62 ± 1.08	78.71
	$L_{\ell_1} + \lambda_{vgg} L_{vgg} + \lambda_{adv} L_{adv}$	3.64 ± 0.65	29.0 ± 1.75	**0.883 ± 0.018**	**8.63 ± 0.839**	**63.17**

Notes: The model trained with pure L_{ℓ_1} optimizes the per-pixel distances, producing the lowest nRMSE and highest PSNR. The model trained with our full objective outperforms other loss function designs in perceptual metrics SSIM, LPIPS, and FID. Bold corresponds to the best results. ↑ indicates that higher is better, ↓ indicates that lower is better.

Abbreviations: FID, Fréchet inception distance; FLAIR, fluid-attenuated inversion recovery; LPIPS, learned perceptual image patch similarity; nRMSE, normalized root mean square error; peak signal-to-noise ratio (PSNR), structural similarity (SSIM).

mitigates the off-resonance artifacts within the MRF time series caused by B0 inhomogeneity and the long readout time of spiral acquisitions. Previous studies have demonstrated the feasibility and potential of deep learning in off-resonance corrections.[49,50] As shown in Figures 4 and 5, parameter-based synthesis and PixelNet present blurry scalp fat signals in boundary regions of the brain, because of the MRF off-resonance effects (seen in T1w).
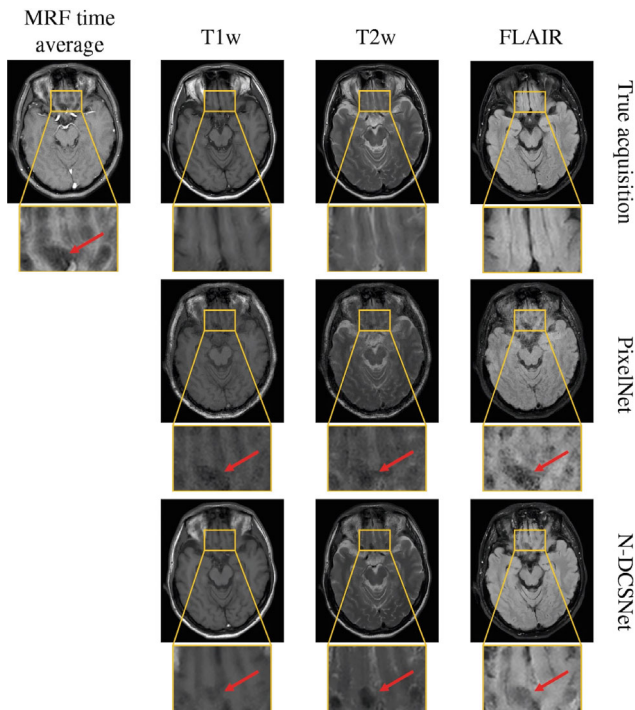
FIGURE 7 Representative N-DCSNet results in mitigating off-resonance artifacts near the nasal region. Magnetic Resonance Fingerprinting (MRF) time-averaged images display spiral off-resonance artifacts near the nasal region (as seen in zoomed-in images) due to the lengthy readout time. PixelNet also struggles to restore the structures and exhibits significant noise and distortions. N-DCSNet successfully mitigates the artifacts and produces contrast-weighted images with few residual artifacts. True acquisitions are displayed as references. Red arrows point to regions with residual artifacts.

In comparison, benefiting from spatial convolutions, N-DCSNet reconstructs a clean and sharp scalp fat signal, overcomes the off-resonance artifacts, and agrees well with the ground truth.

Figure 7 shows a representative example in which the MRF time-averaged image and PixelNet exhibit off-resonance signal loss artifacts and geometric distortion near the nasal region (indicated by the zoomed-in details). Most brain structures are blurred out or distorted, primarily because of the considerable B0 homogeneity and the long readout time for the spiral acquisition. As visualized in the figure, N-DCSNet accurately recovers most of the delicate brain structures near the nasal region. The red arrows indicate the residual artifacts.

Figure S2 presents another example in which the MRF time-averaged image and PixelNet exhibit several off-resonance signal loss artifacts in the regions close to the skull. N-DCSNet significantly reduces the artifacts and recovers the correct contrasts and structures. Some residual artifacts can be observed, as indicated by the red arrows.

4 | DISCUSSION

In this work, we present a novel high-fidelity DCS framework, N-DCSNet, for synthesizing multicontrast images from a single MRF scan. N-DCSNet directly learns a mapping between the MRF time series and the desired contrast weighted images (i.e., T1w, T2w, and FLAIR) and thus bypasses the mapping and simulation steps required for contrast synthesis from parameter maps.

As briefly introduced in Section 1, the sources of error contrast synthesis via parameter maps are attributed mainly to (1) factors that are not included in the dictionary simulation (e.g., B0/B1 homogeneity, slice profile, and flow effects), (2) approximation and error propagation in the contrast synthesis simulation (EPG algorithm),[16] and (3) artifacts (noise and aliasing) from highly undersampled MRF scans (example shown in Figure 1). As indicated by the visual comparison results (Figures 4 and 5), the parameter-based contrast synthesis method does not deliver the correct contrast and produces noisier outputs (particularly for T2w and FLAIR results). One possible way to improve the results is modeling more parameters during the dictionary simulation procedure, such as B1 inhomogeneity,[51] flow,[48] and partial volume.[52] Unfortunately, including more simulation parameters forces the dictionary to grow in size, thereby prolonging the dictionary matching time (Table 2), or severely sacrificing parameter resolution and range.

DCS leverages paired training data to learn a mapping from MRF signals to contrast-weighted images without explicitly modeling the aforementioned conditions. The previous DCS method PixelNet[17] proposed a one-dimensional temporal CNN that maps the MRF time series at each pixel to the contrast weightings for that pixel and improves synthesized image quality and inference time (Table 2). However, because PixelNet treats each pixel independently, it does not leverage the unique spatial structural information within the MRF data. In-vivo results (Figures 4 and 5) indicate that PixelNet exhibits severe noise artifacts and diminished fine textures, particularly in FLAIR scans.

Our N-DCSNet shows significant improvements by introducing a conditional GAN-based framework with a spatial convolution network as the generator. N-DCSNet produces more faithful contrasts and is able to recover finer structures with overall better image quality than the other methods examined (Figures 4 and 5). Moreover, as described in section Section 3.3 and shown in Figures 7 and S2, we demonstrate cases in which N-DCSNet effectively mitigates spiral off-resonance artifacts.

In our approach, we directly input the MRF time series to the network without performing pre-reconstruction

on the MRF data. With the current MRF undersampling factor of 20, our method generates high-fidelity synthesized contrast-weighted images. For even higher undersampling factors (or for improved quality), incorporating pre-reconstruction techniques (e.g., subspace reconstruction[53]) could be a promising direction as it may yield less-aliased inputs. However, depending on the constraints, it could also result in the removal of some information and substantially lengthen the inference time.

Despite significant improvements over previous approaches, we observe that there remains some residual oversmoothing in our results compared to the ground truth (Figures 4 and 5). This could be attributed to the following reasons:

(1) Limited training data constrains the GAN training potential, diminishes robustness against data outliers, and may potentially lead to oversmoothing.

(2) MRF and ground truth contrast-weighted scans were obtained at different times. Despite careful experimental design and in-plane registration, small-scale through-plane motion and misalignment can cause oversmoothing. Improving the experimental setup (e.g., hardware setups) to manage motion could mitigate this issue.

(3) The MRF input images are relatively noisy due to the high undersampling rate and high resolution. The network is trained to reduce noise. However, this process (training on noisy inputs) can result in oversmoothing. We believe that some of these limitations can be mitigated through improved experimental design and a larger training dataset.

Another limitation of this work is that the DCS frameworks (PixelNet and *N-DCSNet*) can generate only contrast-weighted images with fixed sequence parameters (e.g., TE or TR) and are therefore less flexible than simulation-based contrast synthesis from parameter maps. Separate networks must be trained for different MRF parameters or contrast acquisitions. Additionally, our *N-DCSNet* requires paired data; however, our approach allows each decoder branch to be trained independently, potentially relaxing this constraint, although further investigation is required. In this work, we trained *N-DCSNet* on a limited number of healthy volunteer data (21 examinations, 203 slices). To facilitate future clinical adoption, larger and more diverse clinical training data (e.g., with pathology) are necessary.

In the future, we plan to extend our framework to more diverse contrast synthesis, including but not limited to gradient echo imaging, diffusion-weighted imaging, and susceptibility-weighted imaging.

5 | CONCLUSION

In this work, we propose *N-DCSNet* to directly synthesize multicontrast MR images from a single MRF acquisition. This method significantly reduces examination time. By directly training a network to generate contrast-weighted images from MRF, our method does not require any model-based simulation and therefore avoids reconstruction errors due to simulation. In vivo experiments demonstrate that *N-DCSNet* produces high-fidelity contrast-weighted images with sharper contrast and minimal artifacts (in-flow and spiral off-resonance artifacts), and significantly outperforms simulation-based contrast synthesis and PixelNet, both visually and according to metrics. Additionally, our proposed method can inherently mitigate some off-resonance artifacts within MRF data, thereby producing high-quality contrast-weighted images with minimal residual artifacts.

CONFLICTS OF INTEREST

Mariya Doneva, Jakob Meineke, and Thomas Amthor are employees of Philips Research Europe.

DATA AVAILABILITY STATEMENT

In the spirit of reproducible research, our source code can be found at https://github.com/mikgroup/DCSNet to reproduce most of the results herein.

ORCID

Ke Wang https://orcid.org/0000-0001-5951-1727
Jakob Meineke https://orcid.org/0000-0001-8663-1468
Thomas Amthor https://orcid.org/0000-0002-1456-6131
Ekin Karasan https://orcid.org/0000-0001-5662-8145
Fei Tan https://orcid.org/0000-0003-2423-631X

TWITTER

Ke Wang @KeWangKe

REFERENCES

1. Pierpaoli C. Quantitative brain MRI. *Top Magn Reson Imaging*. 2010;21:63.

2. Tanenbaum L, Tsiouris A, Johnson A, et al. Synthetic MRI for clinical neuroimaging: results of the magnetic resonance image compilation (MAGiC) prospective, multicenter, multireader trial. *Am J Neuroradiol*. 2017;38:1103-1110.

3. Tamir J, Uecker M, Chen W, et al. T2 shuffling: sharp, multicontrast, volumetric fast spin-echo imaging. *Magn Reson Med*. 2017;77:180-195.

4. Wang K, Gong E, Zhang Y, Banerjee S, Zaharchuk G, Pauly J. OUTCOMES: rapid under-sampling optimization achieves up to 50. *arXiv:2103.04566*; 2021.

5. Ma D, Jiang Y, Chen Y, et al. Fast 3D magnetic resonance fingerprinting for a whole-brain coverage. *Magn Reson Med*. 2018;79:2190-2197.

6. Ma D, Gulani V, Seiberlich N, et al. Magnetic resonance fingerprinting. *Nature*. 2013;495:187-192.

7. Wang F, Dong Z, Reese TG, et al. Echo planar time-resolved imaging (EPTI). *Magn Reson Med*. 2019;81:3599-3615.

8. Hsieh J, Svalbe I. Magnetic resonance fingerprinting: from evolution to clinical applications. *J Med Radiat Sci*. 2020;67:333-344.

9. Vargas M, Boto J, Delatre B. Synthetic MR imaging sequence in daily clinical practice. *Am J Neuroradiol*. 2016;37:E68-E69.

10. Blystad I, Warntjes J, Smedby O, Landtblom A, Lundberg P, Larsson E. Synthetic MRI of the brain in a clinical setting. *Acta Radiol*. 2012;53:1158-1163.

11. Granberg T, Uppman M, Hashim F, et al. Clinical feasibility of synthetic MRI in multiple sclerosis: a diagnostic and volumetric validation study. *Am J Neuroradiol*. 2016;37:1023-1029.

12. Warntjes J, Leinhard O, West J, Lundberg P. Rapid magnetic resonance quantification on the brain: optimization for clinical usage. *Magn Reson Med*. 2008;60:320-329.

13. Tamir J, Taviani V, Alley M, et al. Targeted rapid knee MRI exam using T2 shuffling. *J Magn Reson Imaging*. 2019;49:e195-e204.

14. Christodoulou A, Shaw J, Nguyen C, et al. Magnetic resonance multitasking for motion-resolved quantitative cardiovascular imaging. *Nat Biomed Eng*. 2018;2:215-226.

15. Cao T, Ma S, Wang N, et al. Three-dimensional simultaneous brain mapping of T1, T2, and magnetic susceptibility with MR multitasking. *Magn Reson Med*. 2022;87:1375-1389.

16. Weigel M. Extended phase graphs: dephasing, RF pulses, and echoes-pure and simple. *J Magn Reson Imaging*. 2015;41:266-295.

17. Virtue P, Doneva M, Tamir J, Koken P, Yu S, Lustig M. Direct contrast synthesis for magnetic resonance fingerprinting. Paper presented at: Proceedings of the 26th Annual Meeting of ISMRM, Paris, France; 2018.

18. Isola P, Zhu J, Zhou T, Efros A. Image-to-image translation with conditional adversarial networks. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA; 2017:1125-1134.

19. Zhang R, Isola P, Efros A. Colorful image colorization. Paper presented at: European Conference on Computer Vision, Las Vegas, Nevada, USA; 2016:649-666.

20. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. *Commun ACM*. 2020;63:139-144.

21. Zhu J, Park T, Isola P, Efros A. Unpaired image-to-image translation using cycle-consistent adversarial networks. Paper presented at: Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy; 2017.

22. Maspero M, Savenije M, Dinkla A, et al. Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. *Phys Med Biol*. 2018;63:185001.

23. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys*. 2017;44:1408-1419.

24. Wolterink J, Dinkla A, Savenije M, Seevinck P, Berg C, Išgum I. Deep MR to CT synthesis using unpaired data. Paper presented at: International Workshop on Simulation and Synthesis In Medical Imaging, Quebec City, Canada; 2017:14-23.

25. Yu B, Zhou L, Wang L, Shi Y, Fripp J, Bourgeat P. Ea-GANs: edge-aware generative adversarial networks for cross-modality MR image synthesis. *IEEE Trans Med Imaging*. 2019;38:1750-1762.

26. Dar S, Yurt M, Karacan L, Erdem A, Erdem E, Cukur T. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Trans Med Imaging*. 2019;38:2375-2388.

27. Qu L, Zhang Y, Wang S, Yap P, Shen D. Synthesized 7T MRI from 3T MRI via deep learning in spatial and wavelet domains. *Med Image Anal*. 2020;62:101663.

28. Wang G, Gong E, Banerjee S, et al. Synthesize high-quality multi-contrast magnetic resonance imaging from multi-echo acquisition using multi-task deep generative model. *IEEE Trans Med Imaging*. 2020;39:3089-3099.

29. Dalmaz O, Yurt M, Çukur T. ResViT: residual vision transformers for multimodal medical image synthesis. *IEEE Trans Med Imaging*. 2022;41:2598-2614.

30. Wang K, Doneva M, Amthor T, et al. High fidelity direct-contrast synthesis from magnetic resonance fingerprinting in diagnostic imaging. Paper presented at: Proceedings of the 28th Annual Meeting of ISMRM, Online; 2020:867.

31. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention, Part III, 18*. 2015;234-241.

32. Shimron E, Tamir J, Wang K, Lustig M. Implicit data crimes: machine learning bias arising from misuse of public data. *Proc Natl Acad Sci*. 2022;119:e2117203119.

33. Jiang Y, Ma D, Seiberlich N, Gulani V, Griswold M. MR fingerprinting using fast imaging with steady state precession (FISP) with spiral readout. *Magn Reson Med*. 2015;74:1621-1631.

34. O'Sullivan J. A fast sinc function gridding algorithm for Fourier inversion in computer tomography. *IEEE Trans Med Imaging*. 1985;4:200-207.

35. Jackson J, Meyer C, Nishimura D, Macovski A. Selection of a convolution function for Fourier inversion using gridding (computerised tomography application). *IEEE Trans Med Imaging*. 1991;10:473-478.

36. Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. Paper presented at: European Conference on Computer Vision, Amsterdam, The Netherlands; 2016:694-711.

37. Larsen A, Sønderby S, Larochelle H, Winther O. Autoencoding beyond pixels using a learned similarity metric. Paper presented at: International Conference on Machine Learning, New York City, New York, USA; 2016:1558-1566.

38. Wang K, Tamir J, de Goyeneche A, et al. High fidelity deep learning-based MRI reconstruction with instance-wise discriminative feature matching loss. *Magn Reson Med*. 2022;88:476-491.

39. Deng J, Dong W, Socher R, Li L, Li K, Fei-Fei L. Imagenet: a large-scale hierarchical image database. Paper presented at: 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, Florida, USA; 2009:248-255.

40. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv Preprint arXiv:1409.1556*. 2014.

41. Joblib Development Team. Joblib: running Python functions as pipeline jobs. 2020. https://joblib.readthedocs.io/. Accessed August 10 2021.

42. Wang Z, Bovik A, Sheikh H, Simoncelli E. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13:600-612.

43. Zhang R, Isola P, Efros A, Shechtman E, Wang O. The unreasonable effectiveness of deep features as a perceptual metric. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, USA; 2018.

44. Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks. *Commun ACM*. 2017;60:84-90.

45. Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. Gans trained by a two time-scale update rule converge to a local Nash equilibrium. *Advances In Neural Information Processing Systems*. 2017;30.

46. Paszke A, Gross S, Massa F, et al. Pytorch: an imperative style, high-performance deep learning library. *Advances In Neural Information Processing Systems*. 2019; 32.

47. Kingma D, Ba J. Adam: a method for stochastic optimization. *arXiv Preprint arXiv:1412.6980*. 2014.

48. Flassbeck S, Schmidt S, Bachert P, Ladd M, Schmitter S. Flow MR fingerprinting. *Magn Reson Med*. 2019;81:2536-2550.

49. Zeng D, Shaikh J, Holmes S, et al. Deep residual network for off-resonance artifact correction with application to pediatric body MRA with 3D cones. *Magn Reson Med*. 2019;82: 1398-1411.

50. De Goyeneche A, Ramachandran S, Wang K, Karasan E, Yu S, Lustig M. ResoNet: physics informed deep learning based off-resonance correction trained on synthetic data. Paper presented at: Proceedings of the 30th Annual Meeting of ISMRM, London, UK; 2022:555.

51. Buonincontri G, Sawiak S. MR fingerprinting with simultaneous B1 estimation. *Magn Reson Med*. 2016;76:1127-1135.

52. Deshmane A, McGivney D, Ma D, et al. Partial volume mapping using magnetic resonance fingerprinting. *NMR Biomed*. 2019;32:e4082.

53. Zhao B, Setsompop K, Adalsteinsson E, et al. Improved magnetic resonance fingerprinting reconstruction with low-rank and subspace modeling. *Magn Reson Med*. 2018;79:933-942.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Figure S1. Gallery of N-DCSNet synthesized contrast-weighted images alongside parameter maps. N-DCSNet synthesizes high-fidelity contrast-weighted images (right three columns) from MRF data. Concurrently, the parameter maps (i.e., PD, T1, T2) can be obtained through dictionary matching (left three columns). Our approach showcases the feasibility of generating complementary parameter maps and contrast-weighted images from a single scan. **Figure S2.** Representative N-DCSNet results in mitigating spiral off-resonance artifacts in an MRF time series near the skull region. The MRF time-averaged image and PixelNet results exhibit spiral off-resonance artifacts near the skull region (zoomed-in images) because of B0 inhomogeneity and the long readout time. **N-DCSNet** recovers the structure and produces contrast-weighted images with few residual artifacts. True acquisitions are displayed as references. Red arrows point to the regions with residual artifacts.