Deep-Learning-Based Contrast Synthesis From MRF Parameter Maps in the Knee Joint

Olli Nykänen, PhD, Mika Nevalainen, MD, PhD, Victor Casula, PhD, Antti Isosalo, MSc (Tech), Satu I. Inkinen, PhD, Marko Nikki, MD, PhD, Riccardo Lattanzi, PhD, Martijn A. Cloos, PhD, Mikko J. Nissi, PhD, and Miika T. Nieminen, PhD

Background: Magnetic resonance fingerprinting (MRF) is a method to speed up acquisition of quantitative MRI data. However, MRF does not usually produce contrast-weighted images that are required by radiologists, limiting reachable total scan time improvement. Contrast synthesis from MRF could significantly decrease the imaging time.

Purpose: To improve clinical utility of MRF by synthesizing contrast-weighted MR images from the quantitative data provided by MRF, using U-nets that were trained for the synthesis task utilizing L1- and perceptual loss functions, and their combinations.

Study Type: Retrospective.


Field Strength and Sequence: A 3 T, multislice-MRF, proton density (PD)-weighted 3D-SPACE (sampling perfection with application optimized contrasts using different flip angle evolution), fat-saturated T2-weighted 3D-space, water-excited double echo steady state (DESS).

Assessment: Data were divided into training, validation, test, and radiologist’s assessment sets in the following way: 136 subjects to training, 3 for validation, 3 for testing, and 42 for radiologist’s assessment. The synthetic and target images were evaluated using 5-point Likert scale by two musculoskeletal radiologists blinded and with quantitative error metrics.

Statistical Tests: Friedman’s test accompanied with post hoc Wilcoxon signed-rank test and intraclass correlation coefficient. The statistical cutoff \( P < 0.05 \) adjusted by Bonferroni correction as necessary was utilized.

Results: The networks trained in the study could synthesize conventional images with high image quality (Likert scores 3–4 on a 5-point scale). Qualitatively, the best synthetic images were produced with combination of L1- and perceptual loss functions and perceptual loss alone, while L1-loss alone led to significantly poorer image quality (Likert scores below 3). The interreader and intrareader agreement were high (0.80 and 0.92, respectively) and significant. However, quantitative image quality metrics indicated best performance for the pure L1-loss.

Data Conclusion: Synthesizing high-quality contrast-weighted images from MRF data using deep learning is feasible. However, more studies are needed to validate the diagnostic accuracy of these synthetic images.

Evidence Level: 4.

Technical Efficacy: Stage 1.

J. MAGN. RESON. IMAGING 2022.
MRI is the imaging modality of choice for the diagnostic imaging of various soft tissues. In standard MRI, anatomical or qualitative images with differently weighted contrasts are acquired and used by radiologists to assist in diagnosis. In some cases, semi-quantitative scoring is applied to estimate the state of the pathology. In addition to gathering anatomical images, MRI can also be used for acquiring quantitative data, such as relaxation time maps. Often, this quantitative information potentially allows more precise diagnoses compared to qualitative anatomical imaging. Traditionally, quantitative MRI (qMRI) has been held back by long scan times. Over the years, many approaches to speed up qMRI have been studied. These approaches include, but are not limited to, compressed sensing- and machine learning-based approaches to reconstruct images from sparsely sampled MR data, parallel imaging, and magnetic resonance fingerprinting (MRF).

Unlike traditional qMRI methods, which use series of anatomical images to estimate quantitative parameters, MRF prioritize the time dynamics of the MR signal over image quality in the traditional sense of the word. Specifically, MRF sequences continuously vary acquisition parameters to gather information rich time-varying signals for each voxel. These signals are then matched with a pregenerated dictionary to find the underlying quantitative parameters for each voxel. Even though MRF vastly decreases imaging times for achieving qMRI information, standard contrast-weighted MR images are required for clinical work. Thus, speeding up clinical MRI protocols even further would require achieving both quantitative and anatomical images using a single sequence. MRF could be an excellent candidate for such a sequence as it provides a wealth of information within a single scan.

One way to provide anatomical and quantitative images from a single scan is to utilize synthetic MRI, in which qMRI data are used to synthesize conventional contrast-weighted anatomical images. Traditionally, image synthesis has been performed on a voxel-by-voxel basis using Bloch-simulations to predict contrast-weighted images. However, this approach is limited by either simplified models that fail to capture the complex dynamics found in organic samples or difficulty of modeling certain important aspects such as partial-volume effects or noise characteristics. Some of these problems may be overcome utilizing deep convolutional neural networks (DCNN) in the synthesis instead of model-based synthesis. Previously, DCNNs have been successfully utilized for many different image synthesis tasks in medical imaging, including cross-modality synthesis between MRI and CT, synthesizing contrast-weighted images from a set of other contrast-weighted images and synthesizing fat-saturated contrast-weighted knee images from corresponding nonsaturated images.

As DCNNs have been successfully utilized in image synthesis from one contrast to another, they are naturally a tool of choice for synthesizing contrast-weighted images from quantitative parameter maps acquired by MRF. However, DCNN-based synthetic MRI from MRF data to anatomical images has not been extensively studied and existing studies have only focused on brain imaging. Moreover, in these studies, the whole image sequences from the MRF scan were utilized as the input of the neural networks.

In this study, the aim was to utilize qMRI parameters acquired by MRF (proton density, T1 relaxation time, T2 relaxation time, and B1+ field) to synthesize various contrast-weighted images that are acquired during routine clinical MRI scans of the knee joint. Hypothesis was that the developed method can be used to produce conventional MR images that match those acquired by dedicated pulse sequences thus making MRF clinically more appealing.

Methods

Data Acquisition

The study was conducted under relevant ethical permission and subjects had given informed consent (Northern Ostrobothnia Hospital District Ethical Committee, permission numbers 88/2019 and 144/2019). MRI data were acquired at 3 T (Siemens MAGNETOM Skyra, Erlangen, Germany) (hereafter referred to as scanner #1) from 142 knee joints of subjects from Northern Finland Birth Cohort 1986. The data were acquired from healthy volunteers and contained only few pathologies. Conventional contrast-weighted images were collected using 3-D sequences (proton density [PD]-weighted SPACE sequence, fat-saturated T2-weighted SPACE sequence and water excitation DESS sequence) (sequence parameters found in Table 1). Data from 42 knee joints from same cohort, acquired using another 3 T MRI scanner (Siemens MAGNETOM Vida, Erlangen, Germany) (scanner #2) were incorporated to obtain a large enough dataset for broader radiological evaluation of the image quality of the synthesized contrast-weighted images. The same subject was never imaged with both scanners. An MRF sequence previously introduced for articular cartilage evaluation was utilized to reconstruct the PD-, T1-, T2- and B1-maps (Fig. 1).

Data Processing

Preprocessing was done by re-slicing and co-registering (with rigid transform) the conventional data to match the MRF-slice geometry using Slicer version 4.11.20200930, using linear interpolation as interpolation method and zero-filling the areas near image edges that were not covered by the conventional sequences (Fig. 1). Subsequently, voxels having zero values (due to zero filling the image edges) in the conventional (target) images were also set to zero in the MRF data. After the preprocessing, the MRF-data and target contrast images were divided into training, validation, and test sets as
follows: out of 1420 slices imaged with scanner #1, 1360 were used for training, 30 for validation, and 30 for testing of the trained networks. The unusually small validation and test sets were used to allow as much data as possible for the network training. The division of the data was performed in such a way that data from single subject belonged in either training, validation, or test set to avoid information leakage. The 420 slices of data acquired with the scanner #2 were used only for the radiological evaluation. This dataset did not contain images from the same subjects imaged using scanner #1.

**Network Training**

DCNN-models having U-Net\(^\text{28}\) model architecture (Fig. 2) were trained to generate the desired MRI contrasts from the MRF data. For each conventional contrast (Table 1), a separate U-Net model was trained to predict the images (Fig. 1) in supervised setting, using training pairs \(\{X, Y\}\), where \(X\) is the MRF data input \(X = (4 \times 288 \times 288)\) consisting of the quantitative maps \((S_0, T_1, T_2, B_1)\) in the channel dimension, and \(Y\) was the corresponding contrast target image of size \(288 \times 288\). The values in the input and target data were normalized between zero and one prior to training.

The U-Nets were generated and trained using Tensorflow\(^\text{29}\) (version 2.5.1). The network weights were initialized using Kaiming He-uniform initialization and trained using RMSProp algorithm with a batch size of 8 for 240 epochs using learning rate of 0.001.

L1-loss, perceptual loss using the outputs of the third convolution of the third layer of ImageNet pre-trained VGG19 network\(^\text{30,31}\) and their combinations were utilized as loss-functions in model training and validation. The combined loss functions were realized as a weighted sum of the L1- and perceptual loss functions. The utilized weights were 0.01, 0.05, and 0.20 for perceptual loss and 0.99, 0.95, and 0.80 for L1-loss, referred to as combined loss 1, 2, and 3, respectively. To utilize the perceptual loss function based on the VGG19-network that is originally utilized for RGB-images, the target images were replicated to three channels.

**Data Analysis**

The performance of the trained networks was assessed by computing mean structural similarity indices (SSIM) and peak-signal-to-noise-ratios (PSNR) between the target images and the predicted contrast-weighted images for the test data sets of both scanners. SSIM maps were calculated between targets and predictions with combined loss function 2 (0.05 weight for perceptual and 0.95 weight for L1-loss) to point out regions that were the most erroneously predicted. The resulting images that were evaluated using SSIM and PSNR were also visually inspected by the senior authors of the manuscript to gain quick qualitative insight about the results.

### Table 1. The Acquisition Parameters for Each Sequence Utilized in the Study

<table>
<thead>
<tr>
<th>Sequence</th>
<th>TR (msec)</th>
<th>TE (msec)</th>
<th>FA (deg)</th>
<th>FOV (mm × mm)</th>
<th>Acquired Image Size</th>
<th>Reconstructed Image Size</th>
<th>Slice Thickness (mm)</th>
<th>Bandwidth/pixel (Hz)</th>
<th>Acceleration</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD SPACE</td>
<td>900</td>
<td>76</td>
<td>120</td>
<td>160 × 160</td>
<td>256 × 256</td>
<td>256 × 256</td>
<td>0.6</td>
<td>425</td>
<td>None</td>
</tr>
<tr>
<td>T2 SPACE FS</td>
<td>1000</td>
<td>123</td>
<td>120</td>
<td>160 × 160</td>
<td>256 × 256</td>
<td>256 × 256</td>
<td>0.6</td>
<td>415</td>
<td>None</td>
</tr>
<tr>
<td>DESS</td>
<td>14.1</td>
<td>5</td>
<td>25</td>
<td>160 × 160</td>
<td>256 × 256</td>
<td>256 × 256</td>
<td>0.6</td>
<td>250</td>
<td>None</td>
</tr>
<tr>
<td>MRF</td>
<td>Varying</td>
<td>3.5</td>
<td>Varying</td>
<td>160 × 160</td>
<td>288 × 288</td>
<td>288 × 288</td>
<td>3</td>
<td>495</td>
<td>None</td>
</tr>
</tbody>
</table>
Furthermore, the data acquired with scanner #2 (42 subjects) were utilized for qualitative blinded evaluation by a fellowship-trained musculoskeletal radiologist with 8 years of experience (reader #1) and a fellowship-trained emergency radiologist with more than 20 years of experience (reader #2). Reader #1 also performed another reading 6 months after the initial reading. Before independent image evaluation by each radiologist, 10 blinded cases were assessed as a training session. A 5-point Likert scale was used: 1 = very poor, 2 = poor, 3 = average, 4 = good, and 5 = very good. The overall quality of each image stack was assessed, including the signal intensity, noise, and anatomic accuracy of the following structures: cartilage, bone marrow, menisci, Hoffa’s fat pad, and muscles surrounding the knee joint. Data evaluated by the radiologist were limited to PD- and T2-weighted fat-saturated contrasts and predictions by networks trained with three different loss functions (L1-loss, perceptual loss, and combined loss 2, i.e., 0.05 weight for the perceptual loss) due to time constraints. The diagnostic utility of the synthesized images could not be studied since the utilized data contained too few pathologies and any evidence to one direction or another would have been only anecdotal.

**Statistical Analysis**

The differences in the qualitative scores between the target and predicted images were evaluated using Friedman test accompanied with post hoc Wilcoxon signed-rank test. Interreader and intrareader agreements were inspected using intraclass correlation coefficient (ICC2). For the comparison of qualitative results, the limit of statistical significance ($P < 0.05$) was adjusted with Bonferroni correction and thus the limit of statistical significance was set to $P < 0.003$. All
statistical tests were performed using IBM SPSS statistics version 27.

Results

Visual Inspection

It was observed that predictions of all target contrasts (PD-weighted, T2-weighted Fat-saturated, and water-excited DESS) had a quality comparable to the target contrast-weighted images acquired with the dedicated sequences (Fig. 3). More careful visual inspection revealed that the predictions of PD-weighted contrast were not entirely accurate in regions where most of the observed signal was coming from fat-protons, such as bone marrow (Fig. 3). The lowest SSIM-values were observed at the bone marrow regions (Fig. 4). Moreover, visual inspection of the predicted images suggested that utilizing perceptual loss, in addition to, or instead of pure L1-loss function led to predicted contrasts that were less smooth and more realistic compared to target images than the predictions acquired by utilizing L1-loss alone (Fig. 3). Utilizing L1-loss alone lead to overly smooth images (Fig. 3).

Quantitative Evaluation

The SSIM-analysis supported the initial visual inspection in the sense that SSIM was higher between the targets and predictions of fat-saturated contrasts (T2w and DESS), whereas marginally lower SSIMs and PSNRs were observed for the PD-weighted contrast (Table 2). The quantitative analysis also demonstrated that the best performance was achieved utilizing L1-loss as a loss function, which somewhat contradicts the visual inspection (Table 2, Fig. 3). However, the differences between observed SSIMs and PSNRs were relatively small between most of the different loss functions (Table 2).

Blinded Qualitative Evaluation

The target MR images had very good image quality (4.83 ± 0.51 for PD-weighted contrast and 4.44 ± 0.67 for fat-saturated T2-weighted contrast, combining the averages and standard deviations from all individual readings) and the quality was significantly (P < 0.003) higher than for the UNet synthesized images (Table 3, Supplementary Table S1). The quality of the synthetic images was still good when applying combined (3.76 ± 0.68 for PD-weighted contrast and 3.56 ± 0.65 for fat-saturated T2-weighted contrast) or perceptual (3.85 ± 0.63 for PD-weighted contrast and 3.76 ± 0.68 for fat-saturated T2-weighted contrast) losses, whereas only average image quality was achieved with L1-loss (2.78 ± 0.60 for PD-weighted contrast and 2.97 ± 0.59 for fat-saturated T2-weighted contrast). The difference between the combined loss and L1-loss was also significant (P < 0.003) (Table 3, Supplementary Table S1). This finding confirmed the earlier notion that despite their better performance as measured by SSIM, the images synthesized utilizing only L1-loss were visually too smooth.

Discussion

The results suggest that synthesizing conventional contrast-weighted MR images (even fat-saturated ones) from the MRF parameter maps of the knee is feasible and leads to images that closely resemble those obtained with dedicated clinical sequences. While quantitative analysis indicated that best synthesis performance was reached for fat-saturated contrasts using L1-loss alone, the qualitative assessment by experienced radiologists contradicted with this and instead pointed that adding perceptual loss or utilizing it alone led to better image quality while L1-loss lead to overly smooth solutions.

The difference in quality between the target and synthetic images might be partially caused by the network training being conducted with data purely acquired using scanner #1 and testing data was acquired using scanner #2, which was necessitated by the limited amount of data. Other possibility would have been to mix up the data sets and reserve enough data for the quantitative evaluation. However, such approach might have compromised the network training with the limited data in the first place.

Likely explanation for the smooth results especially when using L1-loss alone is that even after preprocessing, the slices from MRF and conventional sequences did not completely overlap and thus the voxelwise L1-loss favored overly smooth solutions. Somewhat smooth solutions when utilizing L1-loss alone have also been reported earlier.31

The observed discrepancy between quantitative and qualitative analysis regarding fat-saturation can be partially explained by the different effect of the bone marrow fat on quantitative and qualitative metrics. In the nonsaturated images, there is high signal from fatty tissues and thus errors in synthesizing fat signal are likely to lead into worse performance as evaluated by quantitative error metrics. Also, the used MRF-sequence was not tailored for imaging fatty tissues, which may have caused difficulties in predicting the bone marrow contrast. Meanwhile, in the qualitative analysis
FIGURE 3: A comparison of the target contrast-weighted images and synthetic images generated with the different loss functions. Left Column: water excited DESS-contrast (DESS), middle column: proton density-weighted (PDw) contrast, right column: T2-weighted fat-saturated (T2w) contrast.
performed by experienced radiologist, even small errors in the rendition of fat saturation may be a more glaring problem than, for example, slight over or underestimation of the bone marrow fat signal. This effect, however, is likely subjective as the reader #2 gave higher scores for the fat-saturated images. The lower quality of the fat-saturated target images indicates that the fat-saturated sequence itself may have performed sub-optimally.

Even though synthetic MRI has a relatively long history and the lack of diagnostic contrast-weighted image data is a major drawback, only a few studies have aimed to synthesize contrast-weighted images from MRF data, none of which focused on musculoskeletal imaging. However, a couple of studies examining contrast synthesis from MRF data utilizing DCNNs have been conducted in brain. Compared to these previous studies, there are several key differences. First, our application here needs to deal with a more prominent fat signal and even produce fat-saturated images. Moreover, fatty tissues usually have distinct spectral peaks with different relaxation properties not accurately represented in the MRF dictionaries. Second, rather than using raw MRF time series data as an input to the network, we trained the network to predict contrast-weighted images from the MRF parameter maps, which means that our method does not need to store large amounts of MRF time series data after the scanning. Naturally, utilizing the raw MRF-time series data might have been beneficial if the raw signal would contain residual information that is not explained by the quantitative parameter maps from MRF. However, testing this was not possible since we did not have the raw MRF-data available. Last, our study incorporated image quality analysis by a radiologist showing that while the synthetic images are of good quality, they are not yet as good as the dedicated state-of-the-art clinical MRI sequences. Another popular approach to speed up simultaneous acquisition of qualitative and quantitative data has been synthesizing quantitative parameter maps from the results of routine MRI scans. The key issue in this reverse approach compared to our method is that one might need to measure unnecessary contrast to retrieve enough information to synthesize quantitative parameter maps, and often clinical sequences are poorly optimized for qMRI purposes.

Limitations
First, the dataset was not initially collected for this kind of study and thus the MRF data were obtained with a different slice plan compared to the conventional (target) data. Due to this, the target images were constructed by re-slicing the

### TABLE 2. SSIM and PSNR Between the Target and Synthetic Images for Each Loss Function Used for the scanner #1 test set (30 slices)

<table>
<thead>
<tr>
<th></th>
<th>SSIM L1-loss</th>
<th>Combined loss 1</th>
<th>Combined loss 2</th>
<th>Combined loss 3</th>
<th>Perceptual loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD SPACE</td>
<td>0.81 ± 0.04</td>
<td>0.79 ± 0.05</td>
<td>0.78 ± 0.05</td>
<td>0.79 ± 0.05</td>
<td>0.77 ± 0.04</td>
</tr>
<tr>
<td>T2 SPACE FS</td>
<td>0.95 ± 0.01</td>
<td>0.94 ± 0.01</td>
<td>0.94 ± 0.01</td>
<td>0.94 ± 0.01</td>
<td>0.92 ± 0.01</td>
</tr>
<tr>
<td>DESS</td>
<td>0.94 ± 0.01</td>
<td>0.94 ± 0.01</td>
<td>0.94 ± 0.01</td>
<td>0.93 ± 0.01</td>
<td>0.92 ± 0.01</td>
</tr>
<tr>
<td>PSNR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PD SPACE</td>
<td>30.5 ± 1.2</td>
<td>30.0 ± 1.2</td>
<td>29.7 ± 1.0</td>
<td>30.4 ± 1.4</td>
<td>30.4 ± 1.1</td>
</tr>
<tr>
<td>T2 SPACE FS</td>
<td>42.3 ± 1.4</td>
<td>41.7 ± 0.9</td>
<td>41.9 ± 1.4</td>
<td>41.4 ± 1.1</td>
<td>40.6 ± 0.7</td>
</tr>
<tr>
<td>DESS</td>
<td>43.7 ± 1.6</td>
<td>42.9 ± 1.3</td>
<td>43.1 ± 1.5</td>
<td>42.2 ± 1.1</td>
<td>42.3 ± 1.1</td>
</tr>
</tbody>
</table>

The values are mean values for the error metric in the test set ± standard deviation.
TABLE 3. The Average Image Quality Likert-scores and Their Standard Deviations Given by the Musculoskeletal Radiologists for Both the Target and Synthesized PD- and t2-Weighted Fat-Saturated Images With Respective Loss Functions

<table>
<thead>
<tr>
<th>PD SPACE</th>
<th>Target</th>
<th>L1-loss</th>
<th>Combined loss 2</th>
<th>Perceptual loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reader #1</td>
<td>4.92 ± 0.47</td>
<td>2.95 ± 0.51</td>
<td>4.19 ± 0.39</td>
<td>3.99 ± 0.55</td>
</tr>
<tr>
<td>Reader #2</td>
<td>4.64 ± 0.59</td>
<td>2.60 ± 0.64</td>
<td>3.10 ± 0.67</td>
<td>3.60 ± 0.77</td>
</tr>
<tr>
<td>Reader #1 re-read</td>
<td>4.92 ± 0.41</td>
<td>2.77 ± 0.58</td>
<td>4.00 ± 0.31</td>
<td>3.99 ± 0.46</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>T2 SPACE FS</th>
<th>Target</th>
<th>L1-loss</th>
<th>Combined loss 2</th>
<th>Perceptual loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reader #1</td>
<td>4.46 ± 0.66</td>
<td>3.29 ± 0.55</td>
<td>3.90 ± 0.45</td>
<td>3.58 ± 0.62</td>
</tr>
<tr>
<td>Reader #2</td>
<td>4.07 ± 0.67</td>
<td>2.85 ± 0.55</td>
<td>3.43 ± 0.76</td>
<td>4.14 ± 0.69</td>
</tr>
<tr>
<td>Reader #1 re-read</td>
<td>4.76 ± 0.45</td>
<td>2.79 ± 0.54</td>
<td>3.36 ± 0.55</td>
<td>3.56 ± 0.54</td>
</tr>
</tbody>
</table>

Data from 42 subjects (i.e. 84 condyles) were evaluated. Differences in the score distributions were deemed statistically significant by Friedman’s test and Wilcoxon signed rank post hoc test (P < 0.003).

This study shows that it is possible to use deep convolutional neural networks to synthetize typical contrast-weighted images of the knee from quantitative parameter maps obtained with MRF sequence. Even though our reported synthetic images did not reach the excellent image quality of the state-of-the-art high-resolution 3D contrast-weighted images, the achieved quality warrants further studies to evaluate whether they nevertheless provide sufficient diagnostic utility to be translated into the clinics.
Data Availability Statement
The underlying codes are available from corresponding author upon reasonable request. The training data were achieved from Northern Finland Birth Cohort 1986 and cannot be made publicly available.

References