Deep-learned short tau inversion recovery imaging using multi-contrast MR images

Sewon Kim | Hanbyol Jang | Jinseong Jang | Young Han Lee
Dosik Hwang

1School of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea
2Department of Radiology and Center for Clinical Imaging Data Science (CCIDS), Yonsei University College of Medicine, Seoul, Korea

Purpose: To generate short tau, or short inversion time (TI), inversion recovery (STIR) images from three multi-contrast MR images, without additional scanning, using a deep neural network.

Methods: For simulation studies, we used multi-contrast simulation images. For in-vivo studies, we acquired knee MR images including 288 slices of T_1-weighted (T1-w), T_2-weighted (T2-w), gradient-recalled echo (GRE), and STIR images taken from 12 healthy volunteers. Our MR image synthesis method generates a new contrast MR image from multi-contrast MR images. We used a deep neural network to identify the complex relationships between MR images that show various contrasts for the same tissues. Our contrast-conversion deep neural network (CC-DNN) is an end-to-end architecture that trains the model to create one image from three (T1-w, T2-w, and GRE images). We propose a new loss function to take into account intensity differences, misregistration, and local intensity variations. The CC-DNN-generated STIR images were evaluated with four quantitative evaluation metrics, including mean squared error, peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and multi-scale SSIM (MS-SSIM). Furthermore, a subjective evaluation was performed by musculoskeletal radiologists.

Results: Our method showed improved results in all quantitative evaluations compared with other methods and received the highest scores in subjective evaluations by musculoskeletal radiologists.

Conclusion: This study suggests the feasibility of our method for generating STIR sequence images without additional scanning that offered a potential alternative to the STIR pulse sequence when additional scanning is limited or STIR artifacts are severe.

KEYWORDS
deep learning, image synthesis, knee, magnetic resonance imaging, neural network, short tau inversion recovery, short-TI inversion recovery
Magnetic Resonance in Medicine

1 | INTRODUCTION

MRI is an important diagnostic tool, whereby a variety of pulse sequences are used to acquire images that emphasize specific tissues, for detecting anatomical abnormalities. Musculoskeletal radiologists use various MR contrast images depending on the specific pathology under investigation.

A common MR method, T1-weighted fast spin-echo (T1-w) imaging, is used to evaluate bone marrow in musculoskeletal imaging. Proton density (PD) or T2-weighted fast spin-echo (T2-w) images show the anatomical details of cartilage, ligaments, and menisci. Conventional or fast spin-echo long echo time (TE) sequence or T2 gradient-recalled echo (GRE) sequence imaging are designed to diagnose ligament pathologies using water-sensitive images. Finally, fat suppression sequence imaging are designed to diagnose ligament pathologies developed from the trained trees of different scales depending on the specific pathology under investigation.

There are various approaches for medical image synthesis used in association with synthetic MRIs. Location-sensitive deep network architecture based on a pixel-based multi-layer perceptron generates target modality images using pixel location and intensity. The deep encoder-decoder image synthesis method generates a target contrast image using an encoder-decoder model based on convolutional neural networks (CNNs). In addition, multimodal imaging demonstrates visually and quantitatively improved performance using a U-net based encoder and decoder model with multi-contrast inputs. However, this method also leaves room for improvement toward reconstructing high frequency components.

Recently, deep learning-based methods have been studied in association with synthetic MRIs. Location-sensitive deep network architecture based on a pixel-based multi-layer perceptron generates target modality images using pixel location and intensity. The deep encoder-decoder image synthesis method generates a target contrast image using an encoder-decoder model based on convolutional neural networks (CNNs). In addition, multimodal imaging demonstrates visually and quantitatively improved performance using a U-net based encoder and decoder model with multi-contrast inputs. However, this method also leaves room for improvement toward reconstructing high frequency components.

There are various approaches for medical image synthesis based on generative adversarial network (GAN) architecture. Conditional GAN-based methods train generator models with paired datasets using adversarial and pixel-wise loss functions. For unpaired datasets, the cyclic loss is used to train the model. These GAN-based models yielded quantitatively better results than state-of-the-arts (Replica and Multimodal imaging).

Because most deep learning-based methods learn the distribution of datasets, they are sensitive to the proportion of pathological data. Therefore, from a clinical point of view, these methods require more careful investigation. In particular, GAN-based methods are more sensitive to data distribution and have higher feature hallucination risks that may cause more misdiagnosis than other learning-based methods.

The main goal of this study was to use CNNs to generate synthetic STIR images of the knee using CNNs from acquired multi-contrast images (T1-w, T2-w, and GRE), which are routinely acquired during knee MR protocols to reduce the scanning time required. In this paper, we will show that the deep learning-based image synthesis approach can be applied to knee MR images, which feature fine tissues such as articular cartilage or meniscus. Using our method, we could subtract 4 min and 38 sec of a knee MR imaging protocol from the total scan time of 14 min and 28 sec. We will hereafter refer to the proposed method as a contrast conversion-deep neural network (CC-DNN). A preliminary study of the proposed method was partially presented in 2018 at the annual meeting of the International Society for Magnetic Resonance in Medicine.

2 | METHODS

2.1 | Magnetic resonance physics for multi-contrast images

MRI is an imaging modality that exploits the phenomenon of nuclear magnetic resonance. Various pulse sequences in MRI can acquire signals by adjusting the radio frequency...
where signal (S) is the acquired signal, repetition time (TR) is the interval of the repeated RF pulse, echo time (TE) is the interval from the moment the RF pulse is applied to signal acquisition, and \( S_0 \) indicates the signal amplitude without relaxation decay, which is proportional to the effective PD. Based on (1), we can acquire \( T_1 \)-w or \( T_2 \)-w images by adjusting TR and TE.

Gradient-echo images are acquired using a gradient-echo pulse sequence based on Equation 2:

\[
S = S_0 \frac{\sin(\theta) \left(1 - \exp\left(-\frac{\theta}{T_1}\right)\right) \exp\left(-\frac{\theta}{T_2}\right)}{\left(1 - \cos(\theta) \exp\left(-\frac{\theta}{T_1}\right)\right)}
\]

where \( \theta \) denotes the flip angle of the RF and \( T_1^* = T_1 T_2^*/(T_2 + T_1^*) \) denotes the relaxation time due to a non-uniform magnetic field \( T_2^* \) and magnetic susceptibility. Finally, STIR images are acquired using the inversion recovery (IR) measurement method, as shown in Equation 3:

\[
S = S_0 \left(1 - 2\exp\left(-\frac{\theta}{T_1}\right) + \exp\left(-\frac{\theta}{T_2}\right)\right) \exp\left(-\frac{\theta}{T_1}\right)
\]

where \( TI \) is the time at which the longitudinal magnetic moment of fat becomes zero by \( T_1 \) relaxation after the inversion RF pulse is applied.\(^{30}\)

### 2.2 Image generation for simulation data

If the signals reported by the American College of Radiology 3 can be derived from the spin-echo (Equation 1) and gradient-echo (Equation 2) signals, a STIR image can be generated by utilizing the spin- and gradient-echo images. To test the feasibility of this technique, we performed experiments with simulation data to verify that our method can derive the output signal of Equation (3) from three different input signals. Two of the input signals are from Equation (1), whose TR and TE are set differently for \( T_1 \)-w and \( T_2 \)-w images, respectively, and the other is from Equation (2). For the simulation experiments, we generated four intrinsic MR parameters \((T_1, T_2, T_2^*, \text{ and } S_0)\) whose pixel values are all possible combinations of \( T_1, T_2, T_2^* \) and \( S_0 \). \( S_0 \) ranged from 0.01 to 1 at intervals of 0.01, \( T_1 \) ranged from 10 to 4,000 ms at intervals of 10, \( T_2 \) ranged from 10 to 2,000 ms at intervals of 10, and \( T_2^* \) ranged from 10 to 1,000 ms at intervals of 10. Additionally, a constraint of \( T_1 > T_2 > T_2^* \) was considered. These combinations \((T_1, T_2, T_2^* \text{ and } S_0)\) were applied to Equations (1), (2), and (3) with various scan parameters \((TR, TE, TI, \text{ and } \theta)\) used in clinical imaging settings to produce signal values for \( T_1 \)-w, \( T_2 \)-w, GRE, and STIR. The generated values were reconstructed into four different 128 × 128 two-dimensional knee simulation images, as shown in Figure 1. The number of randomly selected tuples in the four contrast images for training and test was 10,000; 70% for the training, 20% for the validation, and 10% for the test in our neural network development.

### 2.3 Image acquisition for in-vivo data

This study was approved by an institutional review board and written informed consent was obtained from all volunteers. For data acquisition, a 3.0T MRI scanner (Ingenia CX, Philips Healthcare, Best, The Netherlands) was used and the pulse sequence parameters were as follows: \( TR = 522 \text{ ms}, TE = 11 \text{ ms} \text{ (T1-w)}; TR = 3,000 \text{ ms}, TE = 80 \text{ ms} \text{ (T2-w)}; TR = 500 \text{ ms}, TE = 10 \text{ ms}, \text{ flip angle} = 25^\circ \text{ (GRE)}; TR = 4,478 \text{ ms}, TE = 60 \text{ ms}, TI = 220 \text{ ms} \text{ (STIR)}. \) The common scan parameters included image resolution, 384 × 384; slice thickness, 3 mm; voxel size, 0.45 × 0.57 × 0.3 mm³; and field of view, 160 mm. 24 slices per person were obtained for each of \( T_1 \)-w, T2-w, GRE, and STIR imaging from 12 healthy volunteers (11 males, 1 female, aged 27 ± 3 years). A set of 288 in vivo knee MR images were composed for each contrast. The volunteers were requested to minimize their movement during MR scanning for alignment of multi-contrast images. No additional image registration process was done. The total acquisition time required to obtain all multi-contrast images for each subject was 14 min 28 s, which includes 2:39 min on \( T_1 \)-w imaging, 3:12 min on \( T_2 \)-w imaging, 3:44 min on GRE imaging, and 4:38 min on STIR imaging.

### 2.4 Data processing

Because the intensity values of MR images we got were relative values, we did min-max normalization using the maximum and minimum values of the entire dataset for every slice to preserve the intensity ratios between contrasts. Four-fold cross-validation (CV) was used, and each experiment included nine subjects (216 slices of each contrast) for training and validation; and three subjects (72 slices of each contrast) were used for testing. Images from the same subject belonged to the same group.

### 2.5 Deep-learning architecture

Our CC-DNN model for generating STIR images from three different images \((T_1 \text{-w, T}_2 \text{-w, and GRE})\) was trained and
applied to the image domain. Let \( I_{T1}, I_{T2}, \) and \( I_{GRE} \in \mathbb{R}^{w \times h} \) be three input images (T1-w, T2-w, and GRE, respectively), where \( w \) and \( h \) indicate the width and height of the image. The chosen output STIR image that is used as the label to train the CC-DNN is denoted as \( ISTIR \in \mathbb{R}^{w \times h} \). Figure 2 shows the overall architecture of our CC-DNN model. The \( ISTIR \) estimation process of the CC-DNN model can be expressed by the following equation:

\[
\hat{I}_{STIR} = C_1 \left[ H_{\text{concat}} \left( \text{concat} \left( [H_{T1}(I_{T1}), H_{T2}(I_{T2}), H_{GRE}(I_{GRE})] \right) \right) \right] \tag{4}
\]

where \( \hat{I}_{STIR} \) is an estimated \( I_{STIR} \), and \( C_1 \) is a convolutional layer where kernel size is \( a \) and the number of feature maps is \( b \). \( H \) is a sub-hypothesis function consisting of three residual blocks and an input skip connection; and \( \text{concat} \) denotes a concatenation layer. Residual block, \( R \), is expressed as follows:

\[
R(x) = x + C_1(\sigma(C_2^{256}(\sigma(C_3^{256}(x))))) , \quad x \in \mathbb{R}^{w \times h \times c} \tag{5}
\]

where \( \sigma \) denotes an activation function, and \( c \) denotes the number of input feature maps. The activation function \( \sigma \) used in our CC-DNN is the leaky-rectified linear unit (\( \alpha = 0.1 \)) function.

Each convolutional layer except the final one is followed by a batch normalization layer. The Adam optimizer\(^3\) was used to train our CC-DNN model.

### 2.6 Loss function

The CC-DNN model estimates \( I_{STIR} \) from \( I_{T1}, I_{T2}, \) and \( I_{GRE} \) and is optimized by our proposed loss function, which consists of three loss components. The first one is the mean squared error (MSE) function, which is expressed by the following equation:

\[
c_1 \left( I_{STIR}, \hat{I}_{STIR} \right) = \frac{1}{w \times h} \sum_{i=1}^{w} \sum_{j=1}^{h} \left( I_{STIR}[i,j] - \hat{I}_{STIR}[i,j] \right)^2 \tag{6}
\]

This loss component optimizes the CC-DNN model by reducing the intensity differences of all pixels of \( I_{STIR} \) and \( \hat{I}_{STIR} \).

Multi-contrast images obtained from the same subject may have slight positional differences (misregistration), because they are scanned sequentially rather than by a single
sequence. To minimize these differences, subjects were asked not to move, but various unavoidable movements due to breathing, blood vessel phase changes, etc., may have caused small errors of one to two pixels in alignment that cannot be easily compensated for with even the most advanced registration algorithms. However, this slight misregistration can cause blurring or loss of detail in deep-learning processes.

To compensate for this misregistration, and preserve as much detail as possible, the second loss component considers the surroundings when it calculates the intensity difference of each pixel. It calculates intensity differences between the output pixel and its $k \times k$ surrounding pixels of the label image and keeps only the minimum differences. The second loss component is expressed by the following equation:

$$c_2 \left( \mathbf{I_{STIR}} - \mathbf{\hat{I}_{STIR}} \right) = \frac{1}{w \times h} \sum_{i=\frac{h}{2}+1}^{w-\frac{1}{2}} \sum_{j=\frac{h}{2}+1}^{h-\frac{1}{2}} \min \left( M_{i,j} - \mathbf{\hat{I}_{STIR}}[i,j] \times J \right)^2,$$

where $J \in \mathbb{R}^{k \times k}$ is an all-ones matrix.

In STIR images, pixels with various values and complex distribution represent various structural shapes. The contrasts of STIR images are difficult to represent perfectly using the MSE function because it calculates the average value of the squared differences of all pixels and tends to follow a Gaussian distribution.

The third loss component compensates for this by minimizing the differences in local average and variation between the output and label images. This allows the generated $\mathbf{\hat{I}_{STIR}}$ to learn the complex structural diversity of $\mathbf{I_{STIR}}$. The third loss component is expressed by the following equation:

$$c_3 \left( \mathbf{\hat{I}_{STIR}} \right) = \frac{1}{w \times h} \sum_{i=\frac{h}{2}+1}^{w-\frac{1}{2}} \sum_{j=\frac{h}{2}+1}^{h-\frac{1}{2}} \left( \mathbf{var} \left( M_{i,j} \right) - \mathbf{var} \left( \mathbf{N}_{i,j} \right) \right)^2,$$

where $\mathbf{N}_{i,j} = \begin{bmatrix} \mathbf{\hat{I}_{STIR}}[i-\frac{k}{2},j-\frac{k}{2}] & \cdots & \mathbf{\hat{I}_{STIR}}[i-\frac{k}{2},j+\frac{k}{2}] \\ \vdots & \ddots & \vdots \\ \mathbf{\hat{I}_{STIR}}[i+\frac{k}{2},j-\frac{k}{2}] & \cdots & \mathbf{\hat{I}_{STIR}}[i+\frac{k}{2},j+\frac{k}{2}] \end{bmatrix}$.

The ratio between the $c_1$, $c_2$, and $c_3$ loss functions for training and the hyperparameter $k$ were selected based on the performance of the validation set in each CV.

### 2.7 Comparison with other methods

The performance of the proposed CC-DNN model was compared with other recently developed image synthesis techniques, including Replica, Multimodal imaging,
and Pix2pix GAN. 36 We implemented Replica in Python 3.6 with an Intel Xeon CPU E5-1620 v4 at 3.50 GHz octa core and 126 GB memory. Python 2.7 on an Intel Core i7-6700K CPU at 4.00 GHz octa core, with 32 GB of memory, and GeForce GTX 1080Ti were used to implement the Multimodal imaging, Pix2pix, and our CC-DNN model codes were based on Keras 2.2.4 using Tensorflow 1.8.0 backended.

2.8 | Quantitative image analysis

For quantitative evaluation, four different metrics were used: MSE, peak signal-to-noise ratio (PSNR), structural similarity

FIGURE 3  The results of simulation tests with 0%, 2% (signal-to-noise ratio (SNR) 26), and 4% (SNR 21) Rician noise
(SSIM), and multi-scale SSIM (MS-SSIM). MSE and PSNR have been widely used for the quantitative evaluation of medical images and SSIM and MS-SSIM are known to be more related to the perception of human visual systems than MSE and PSNR. All images were normalized to the intensity range 0-1 before evaluation, and every evaluation was conducted by subject. To demonstrate the statistical significance of improvement compared to other methods, P-values were calculated by Wilcoxon signed rank test; P-values lower than .05 were considered to be statistically significant.

2.9 Expert analysis

Three musculoskeletal radiologists (H.S.L., I.H.J., and O.K.S., each with a year of experience in musculoskeletal radiology) evaluated the similarity between the generated STIR images and the acquired STIR images from a clinical point of view. They were blinded to the subject information and independently evaluated the generated STIR images by subject. The acquired T1-w, T2-w, GRE, and STIR images were provided to them and the four STIR images generated by the four methods (Multimodal imaging, Replica, Pix2pix, and our proposed method) were randomly placed and provided. Intraclass correlation coefficient (ICC) was calculated to assess the level of interobserver agreement. This coefficient is interpreted as excellent: 0.75-1.00; good: 0.60-0.75; fair: 0.40-0.59; and poor: 0.40. Subjective scoring was carried out using a score of 1 to 5 for how similar the synthetic STIR image was compared to the acquired STIR image for 19 categories. The criteria of scoring were as follows: 1 point: bad; 2 point: poor; 3 point: fair; 4 point: good; 5 point: excellent.

![Figure 5](image-url)

**FIGURE 5** Four examples of comparison between the acquired short tau, or short inversion time (TI), inversion recovery (STIR) images and the four different synthetic STIR images generated by Multimodal imaging, Replica, Pix2pix, and the proposed method. The inset images are magnifications of the yellow boxes. The black and red arrows in A indicate blood vessels and cartilage, respectively. The black and blue arrows in B and C indicate bright artifact regions. The red arrows in B and C indicate regions in which flow-related artifacts appear.
2.10 | Contrast-to-noise ratio

We obtained contrast-to-noise-ratios (CNRs) for tissues of interest in the diagnosis of knee pathology. CNR for foreground tissue $f$ and background tissue $b$ is expressed as follows:

$$\text{CNR}(f, b) = \frac{\text{mean}_{r \in f}(S(r)) + \text{mean}_{r \in b}(S(r))}{\text{stddev}_{r \in b}(S(r))}$$  \hspace{1cm} (11)$$

where $S$ denotes the acquired signal.

3 | RESULTS

3.1 | Simulation studies

In this experiment, we demonstrated that CC-DNN can learn the relationships among different MR contrasts based on Bloch equations, which govern the basic physical principles of MRI. Figure 3 shows CC-DNN-generated STIR images. The simulation results achieved 37.39 dB PSNR and 0.957 SSIM scores. The results show that our CC-DNN model has the ability to learn the complex relationships between three Equations (1), (2), and (4) and reproduce synthetic STIR images from T1-w, T2-w, and GRE images. Additional experiments with various levels of Rician noise added were performed to reflect real MR scanning conditions. More information about Rician noise is provided in Supporting Information Equation S1, which is available online. For cases with 2% Rician noise added, the results of our CC-DNN model achieved a PSNR score of 31.28 dB and SSIM score of 0.864. For cases with 4% Rician noise added, the results of our CC-DNN model achieved a PSNR score of 26.39 dB and SSIM score of 0.687. The ground-truth simulation STIR images with Rician noise look noisier than the output images, and the scores of PSNR and SSIM metrics which compare the two images are low.

The results of two additional simulation experiments show the effects of the c2 and c3 loss function. First, we randomly shifted the multi-contrast images in the simulation dataset by 0 to 2 pixels, vertically and horizontally, to show that the c2 loss function can reduce the misregistration-caused artifacts. Figure 4B,C show the results of this experiment when only the c1 loss function is used for training and when the c1 and c2 loss functions are used together, respectively. When only the c1 loss component is used, blurring due to misregistration is severe. On the other hand, this blurring is corrected and the boundary becomes more clearly distinguished when the c2 loss function is used together with the c1 loss component. When both loss functions are used, PSNR is increased from 21.54 to 26.81 and SSIM is increased from 0.66 to 0.74 compared to the when only the c1 loss component is used. Figure 4E,F show the results of additional noise simulation. When only the c1 loss component is used for training, noise makes intensity distributions between tissues unclear and

**FIGURE 6** The results of magnetic resonance (MR) synthesis for the pathological data. The inset images are magnifications of the yellow boxes. The red arrows in A indicate the regions of bone marrow edema (BME), and the red arrows in B indicate the ganglions.
tissues hard to distinguish. On the other hand, when the c3 and c1 loss components were used together for training, the difference in intensity between tissues can be clearly distinguished. In this experiment, PSNR increased from 32.87 to 34.07 and SSIM increased from 0.90 to 0.92 when the c1 and c3 loss components were used together for training compared to when only the c1 loss component was used.

3.2 | In vivo studies

We achieved best results when we trained for the first 200 epochs using the c1 and c2 loss components with a ratio of 1:0.3 with a k of 5 and then performed an additional 200 epochs of training using the c1 and c3 loss components with a ratio of 1:0.3. Our initial learning rate was 0.0001 and the decay of learning rate for each epoch was 0.000001. Details about the training process and hyperparameters can be found in the Supporting Information Figures S1, S2, and S3.

A comparison of the CC-DNN model with other MR synthesis methods is shown in Figure 5: acquired STIR images and synthetic STIR images generated by Multimodal imaging, Replica, Pix2pix, and our proposed method, respectively. The overall contrasts of the synthetic STIR images look similar to those of the acquired STIR images. However, the inset images indicated the yellow boxes in Figure 5A show that some of the finer details of tissues were missing when Multimodal imaging, Replica, and Pix2pix were used. On the other hand, these details were retained in the image generated by our proposed method. For example, the cartilage indicated by red arrows in Figure 5A which is thin, but clinically important, is clearly restored in our results. The regions indicated by red arrows in Figure 5B,C are overly bright compared to the acquired STIR image. Moreover, Replica and Pix2pix generated misleading contrasts that mimicked prepatellar bursitis, as indicated by the black arrow in Figure 5C. The STIR sequence has a disadvantage in that it appears to be more prone to flow-related artifacts, which were reduced in our results. The regions indicated by red arrows in Figure 5B,C have a low SNR and few artifacts caused by blood flow. However, those artifacts were reduced in the deep-learned STIR imaging, which clearly show fat and muscle tissues with tiny blood vessels.

STIR sequences are useful for observing bone marrow edema (BME). Therefore, it needs to be confirmed how well synthesized STIR images can show BME. We tested our CC-DNN model using additional data obtained from a subject suffering from BME. Figure 6A shows STIR images with BME generated by various synthesis methods. In all cases, the training dataset did not include images with BME. In Figure 6A, Multimodal imaging, Replica, and Pix2pix failed to generate BME or restored it faintly. On the other hand, the result of our method clearly shows the existence of BME although the shape and size of it are different to that of the acquired STIR image. Figure 6B shows the test results from pathological data obtained from a patient who suffers from ganglions. The results also show that our CC-DNN model reproduced the contrast of ganglion closely, as well as the boundaries and shapes.

All the quantitative evaluation results are reported in Table 1. For all quantitative evaluation metrics, the proposed method demonstrated a significantly higher level of performance than Replica and Multimodal imaging (P < .001). In comparison to Pix2pix, the proposed method achieved higher scores for PSNR (P < .05) and significantly higher scores for SSIM and MS-SSIM (P < .001). The MSE of the proposed method was 0.00166, which was 1.2% of the average intensity of the image excluding the background (0.13333).

<p>| TABLE 1 | Quality evaluation of synthetic short tau, or short inversion time (TI), inversion recovery (STIR) images generated by Multimodal imaging, Replica, Pix2pix, and the proposed method |</p>
<table>
<thead>
<tr>
<th>Evaluation metrics</th>
<th>Multimodal imaging</th>
<th>Replica</th>
<th>Pix2pix</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE (× 10−3)</td>
<td>3.230 ± 0.501***</td>
<td>3.141 ± 0.556 ***</td>
<td>1.895 ± 0.454</td>
<td>1.558 ± 0.482</td>
</tr>
<tr>
<td>PSNR</td>
<td>26.29 ± 0.600***</td>
<td>26.43 ± 0.652***</td>
<td>27.45 ± 0.967*</td>
<td>28.34 ± 1.113</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.776 ± 0.021***</td>
<td>0.778 ± 0.024***</td>
<td>0.796 ± 0.021***</td>
<td>0.822 ± 0.023</td>
</tr>
<tr>
<td>MS-SSIM</td>
<td>0.837 ± 0.018***</td>
<td>0.845 ± 0.022***</td>
<td>0.861 ± 0.021***</td>
<td>0.881 ± 0.020</td>
</tr>
</tbody>
</table>

Notes: Wilcoxon signed rank tests were used to calculate P values between the proposed method and other methods.
Abbreviations: MSE, mean squared error; MS-SSIM, multi-scale structural similarity; PSNR, peak signal-to-noise ratio; SSIM, structural similarity.
*P < .05; **P < .01; ***P < .001.
The proposed method achieved the highest scores in four quantitative evaluation metrics. The metrics used for the quantitative evaluations in this study may not, however, directly reflect diagnostic quality. Radiologists generally focus on clinically important regions of MRIs of the knee to make diagnostic decisions. They consider the shapes, contrasts, thicknesses, lengths, and patterns of target tissues. Table 2 shows the subjective evaluation results of the similarities between the generated and acquired STIR images for clinically important categories. Three radiologists determined that the deep-learned STIR imaging better represented the acquired STIR images than the images generated by the other methods. The ICC for Multimodal imaging is 0.66, for Replica 0.61, for our proposed method 0.63, which indicated good agreement. The ICC of 0.43 in Pix2pix indicated fair agreement. Our proposed method achieved the highest score for all 19 categories that radiologists consider most important for diagnosing knee abnormalities using STIR imaging. The magnified images in Figure 7B compare regions of cartilage, lateral meniscus, tendon, and anterior cruciate ligament (ACL) between the acquired and deep-learned STIR images. For those regions, it is hard to find a noticeable difference between the two images. Figure 7C shows the segmentation results for those four diagnostic regions from both STIR images. All segmentation, except for the lateral meniscus, was manually performed by an author and reviewed by a musculoskeletal radiologist (Y.H.L., with 10 years of musculoskeletal experience). For the lateral meniscus, a simple threshold was applied to both images with a value of 0.1. As shown in Figure 7C, segmented regions from both STIR images overlap almost identically. Furthermore, line profiles at the yellow line position in Figure 7D indicate that the graphs from both STIR images match well.

One of the advantages of the STIR sequence is that the resulting images have high CNRs. We compared the CNRs of deep-learned STIR images to those of acquired STIR images (Figure 8). The CNRs of the deep-learned STIR images were higher than those for the acquired STIR images for cartilage-bone, cartilage-meniscus, muscle-bone, muscle-fat, and muscle-meniscus. For cartilage-bone and cartilage-meniscus,
FIGURE 7  Comparison between acquired STIR images (left, red) and deep-learned STIR images (right, blue). A, Acquired and deep-learned STIR images. Clinically important regions (cartilage, meniscus, tendon, and anterior cruciate ligament) are indicated by the boxes. B, Magnified images from (A) with overlaid segmented areas that are clinically important tissues. C, Segmented areas from the acquired STIR images and deep-learned STIR images. The white area indicates overlapping. D, Line profile graph for the yellow lines of (B). The red line indicates the line profile of the acquired STIR and the blue line indicates the line profile of the deep-learned STIR.
which require a clear distinction, the average CNRs from the deep-learned STIR images were 3.6 and 1.9 times higher than those from the acquired STIR images, respectively. The average CNRs from the deep-learned STIR images for muscle-bone and muscle-fat, which have small intensity differences, were 3.88 and 2.46 times higher than those from the acquired STIR images, respectively. Additionally, the average CNR for muscle-meniscus was 1.85 times higher for the deep-learned STIR than that for the acquired STIR. Overall, the contrast between tissues of deep-learned STIR images was clearer than that for the acquired STIR images.

4 | DISCUSSION

One of the major advantages of MRI as a diagnostic imaging modality is that it can show various contrasts between tissues. However, multiple acquisitions of various contrast images require substantial scan time, which may burden both patients and clinics. There are some studies trying to reduce scan time of single contrast MRI. In this study, we show that an image generation method with a different contrast can be obtained from several other contrast images without the need for additional scanning, which could reduce costs. This approach potentially shortens the overall acquisition time required to obtain enough contrast images. The time required to generate an image using the CC-DNN model is just 0.14 seconds.

The CC-DNN model has the ability to synthetically restore the finer details of knee MRIs better than other MR synthesis methods. This is because the CC-DNN model consists of convolutional layers and skip connections with no pooling layer. The pooling layer that is used in Multimodal imaging and Pix2pix is an important component of VGG-net or U-net. It tends to remove relatively less important information and keep the key features of the image, which allows efficient classification or segmentation of images. However, distinguishing key diagnostic information from all the features found in MR image synthesis is a difficult task. Information that is considered to be noise and thus removed during the pooling process could be key to restoring diagnostically important shapes or textures of the target image. Replica uses resizing to perform a multi-resolution process. This can increase the efficiency of the random forest method, but the details lost through resizing may overwrite the result. The results of our experiments show that some detailed features such as blood vessels and bone texture were lost when other methods were used, but remained noticeable in our deep-learned STIR images. In our additional experiment, when U-net architecture was used instead of CC-DNN architecture, while other conditions remained the same (such as the optimizer, loss function, etc.), some feature details were lost in a manner similar to when using other methods. A comparison between the CC-DNN model results and the U-net results is shown in Supporting Information Figure S4. This suggests that the use of a pooling layer can cause loss of detailed information.

One of the disadvantages of the STIR sequence is that it appears more prone to flow-related artifacts. Due to blood flow, some unsaturated blood follows saturated blood into a slice, which has undergone a prior RF pulse, when a 90-degree inversion pulse is applied. This unsaturated blood creates a stronger signal than expected. However, our CC-DNN model could reduce flow-related artifacts, because it is based on information from $T_1$-w, $T_2$-w, and GRE images where flow-related artifacts are reduced due to the absence of an inversion pulse. Additionally, in CNR measurements of various tissues, deep-learned STIR images show higher CNRs than acquired STIR images. Therefore, our CC-DNN...
model could potentially provide a clear contrast with reduced noise compared to acquire STIR imaging.

In the deep-learned STIR images, edge-like structures appear in areas such as in the adipose tissue between the patellar and tibia as shown in Supporting Information Figure S5. These are the expressions of the tissues appearing at the same location in the T1-w, T2-w, and GRE images. However, they are more emphasized in the deep-learned STIR images than in the acquired STIR images in which the tissues disappear or fade due to noise. This could be a limitation of the current implementation of the deep-learned STIR imaging in generating images as close to the acquired STIR imaging as possible. However, these edge-like structures were evaluated as having little influence on the clinical quality of the generated image because they were classified by the radiologists as parts having little effect on diagnosis. One of the biggest problems for medical image synthesis is blurring artifacts caused by misregistration of multi-contrast images or partial volume effects of each image. We reduced blurring artifacts with the use of multiple loss components and by removing the down-sampling layers. But, despite these improvements, our results are not completely free of blurring artifacts.

The goal of our study was to test the feasibility of using deep learning to generate STIR images from three multi-contrast inputs. Our results show the possibility of deep learning-based contrast conversion. In addition, we showed that our method could be applicable to pathological data for specific pathological cases. However, in order to be used in practice, our study has to be validated for applicability on various pathological data covering a wide range of contrasts. Our future work will aim to resolve the artifacts still present in our results and verify our method with varying and large pathology dataset.

In conclusion, we have demonstrated the feasibility of CC-DNN-generated STIR imaging without additional scanning. Our deep-learned STIR images show higher CNR than acquired STIR images and offers a potential alternative to the STIR pulse sequence when additional scanning ability is limited or STIR artifacts are severe.

ACKNOWLEDGMENTS

This research was supported by a National Research Foundation of Korea (NRF) grant (No. 2019R1A2B5B01070488, 2018M3A9H6081483, 2018R1A2B6009076) funded by the Ministry of Science, ICT and Future Planning (MSIP). The authors thank Hong Seon Lee, Inha Jung and Ok Kyu Song for their detailed analysis as musculoskeletal experts. This research was partially supported by the Graduate School of YONSEI University Research Scholarship Grants in 2019.

REFERENCES


ORCID

Sewon Kim https://orcid.org/0000-0002-3893-252X
Hanbyol Jang https://orcid.org/0000-0001-9573-2586
Jinseong Jang https://orcid.org/0000-0002-0042-9304


**SUPPORTING INFORMATION**

Additional Supporting Information may be found online in the Supporting Information section.
EQUATION S1 Rician noise addition

FIGURE S1 Simulation results using c2 and c3 loss functions. (A) results of using c1 and c2 loss functions (B) results of using c1, c2 and c3 loss functions together (C) sequential training results using (c1 + c2) and (c1 + c3) loss functions, (D) ground truth simulation STIR images with 7% Rician noise (SNR 18)

FIGURE S2 The validation MSE graphs according to the changes of α, β and k. α and β indicates the ratio of c1 to c2 loss component, and the ratio of c1 to c3 loss component, respectively

FIGURE S3 The training loss and the validation loss graphs for three deep learning-based methods in each cross-validation fold

FIGURE S4 The comparison between the CC-DNN model and the U-net model. Red circles show that the detailed structures disappeared in the U-net results compared to the CC-DNN model

FIGURE S5 Edge-like structures appear in the areas such as adipose tissue between patellar and tibia

https://doi.org/10.1002/mrm.28327