ORIGINAL ARTICLE

MR-contrast-aware image-to-image translations with generative adversarial networks

Jonas Denck[1,2,3] · **Jens Guehring**[3] · **Andreas Maier**[1] · **Eva Rothgang**[2]

Abstract

Purpose A magnetic resonance imaging (MRI) exam typically consists of several sequences that yield different image contrasts. Each sequence is parameterized through multiple acquisition parameters that influence image contrast, signal-to-noise ratio, acquisition time, and/or resolution. Depending on the clinical indication, different contrasts are required by the radiologist to make a diagnosis. As MR sequence acquisition is time consuming and acquired images may be corrupted due to motion, a method to synthesize MR images with adjustable contrast properties is required.

Methods Therefore, we trained an image-to-image generative adversarial network conditioned on the MR acquisition parameters repetition time and echo time. Our approach is motivated by style transfer networks, whereas the "style" for an image is explicitly given in our case, as it is determined by the MR acquisition parameters our network is conditioned on.

Results This enables us to synthesize MR images with adjustable image contrast. We evaluated our approach on the fastMRI dataset, a large set of publicly available MR knee images, and show that our method outperforms a benchmark pix2pix approach in the translation of non-fat-saturated MR images to fat-saturated images. Our approach yields a peak signal-to-noise ratio and structural similarity of 24.48 and 0.66, surpassing the pix2pix benchmark model significantly.

Conclusion Our model is the first that enables fine-tuned contrast synthesis, which can be used to synthesize missing MR-contrasts or as a data augmentation technique for AI training in MRI. It can also be used as basis for other image-to-image translation tasks within medical imaging, e.g., to enhance intermodality translation (MRI→CT) or 7 T image synthesis from 3 T MR images.

Keywords Deep learning · Generative adversarial networks · Magnetic resonance imaging · Image synthesis

Introduction

A magnetic resonance imaging (MRI) exam typically consists of multiple sequences that yield different image tissue contrasts required for a complete and reliable diagnosis. However, the sets of sequences that are obtained vary considerably across clinical protocols, scanners, and sites. Clinical guidelines, the MR system (vendor, model, software version, field strength), internal guidelines (e.g., slot time), and radiologists' preferences determine the set of selected sequences (i.e., the MRI protocol) for a specific clinical question at a particular site. Moreover, each sequence is parameterized through multiple acquisition parameters (pulse sequence parameters) that affect image contrast, image resolution, signal-to-noise ratio, and/or acquisition time. The acquisition parameters (and sequences) can be proprietary or generic across vendors. Consequently, sequence parameterization and selection vary significantly across different radiology sites and within exams and scanners of the same site [1].

As sequence acquisition is time consuming and acquisition time is expensive, current research strives to increase MRI value [2] by, e.g., the development of optimized scan protocols [3] or the reduction of scan time by leveraging artificial intelligence (AI) [4]. AI plays a vital role in reducing scan time by either accelerating image acquisition and

✉ Jonas Denck
jonas.denck@gmail.com;
jonas.denck@siemens-healthineers.com

1 Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany

2 Department of Industrial Engineering and Health, Technical University of Applied Sciences Amberg-Weiden, Weiden, Germany

3 Siemens Healthineers, Erlangen, Germany

reconstruction [5] or synthesizing missing or corrupted image contrasts from existing ones [6–8]. The latter is typically done using (multi-)image-to-image neural networks, e.g., to synthesize T2-weighted brain images from T1-weighted images and vice versa [6, 9]. This can offer great clinical value as corrupted images due to motion or other artifacts can be replaced, or claustrophobic patients that prematurely had to end the MRI scan may avoid rescans. Current approaches synthesize a single MR-contrast with fixed or not further specified acquisition parameters from one or multiple existing MR-contrasts. These approaches have been trained and tested on different publicly available MR datasets. However, as sequence parameterizations vary in the clinical practice, these approaches are only applicable to a small share of used sequences. To truly increase the clinical value, these approaches must synthesize MR-contrasts beyond fixed sequence parameterizations.

Therefore, we have developed and trained an image-to-image generative adversarial network (GAN) that synthesizes MR images with adjustable image contrast. The acquisition parameters echo time (TE) and repetition time (TR) that influence image contrast are incorporated into the network's training. Moreover, we can translate non-fat-saturated MR images into fat-saturated MR images with the acquisition parameters TE and TR as additional inputs, which increases the overall reconstruction performance. We provide a thorough visual and quantitative evaluation of our approach and benchmark it with the commonly used pix2pix framework.

Material and methods

Generative adversarial network

In this section, we want to give a short overview of generative adversarial networks and important adaptions for medical image-to-image synthesis.

GANs are a special type of artificial neural network where two networks (generator and discriminator) are trained adversarially. For an image generation task, the generator is focused on image generation, and the discriminator learns to discriminate between real and generated (fake) images.

An important extension of GAN capabilities is pix2pix, which serves as a general-purpose solution to image-to-image translation problems. The pix2pix approach is used to learn the mapping from an input image to a paired output image under an L1 reconstruction and an adversarial loss [10].

Since paired training data are not available for many image-to-image translation tasks, cycle consistency loss was proposed (CycleGAN) [11] that allows the translation of an image from a source domain to a target domain without any

paired training examples. The original formulation includes two generative models, G and F, where G translates an image from domain A into domain B. F translates an image from domain B into domain A. The discriminators D_A and D_B learn to distinguish between real and fake images from their domains. The cycle consistency loss using the L1 distance for image g and t from domain A and B, respectively, is then defined as:

$$\mathcal{L}^{\text{Cycle}} = E\big[\|G(F(g)) - g\|_1\big] + E\big[\|F(G(t)) - t\|_1\big]. \quad (1)$$

Both pix2pix and CycleGAN are the basis for many image-to-image translation tasks in the medical imaging domain [7–9].

Image-to-image generative adversarial network

Image-to-image GANs, e.g., pix2pix, have been successfully applied to learn image-to-image translations in various domains given paired data [10]. We adopt this approach and use the non-saturating adversarial loss with R1 regularization using $\gamma = 1$ in order to produce sharp and realistic images [12].

Additionally, an L1 reconstruction loss enforces pixel-wise similarity between target ground truth and reconstructed image. However, an important characteristic for MRI is that MR images present significant intensity variation across patients and scanners, and MR image intensity standardization is an ongoing research topic [13]. Hence, a reconstruction loss that is not fully focused on pixel intensity similarities, but in addition perceptually motivated, is anticipated to improve the performance of the image-to-image GAN. Therefore, we also experimented with a weighted reconstruction loss of L1 and multi-scale structural similarity (MS-SSIM) index [14, 15]. MS-SSIM has demonstrated to better preserve image contrast at higher frequencies compared to SSIM. However, it is not sensitive to uniform biases, i.e., can produce brightness shifts, while L1 preserves pixel intensities [14]. Given two images **x** and **x**′, the MS-SSIM loss is defined as:

$$\mathcal{L}^{\text{MS-SSIM}}(x, x') = 1 - \text{MS-SSIM}(x, x') \quad (2)$$

where MS-SSIM is the multi-scale version of SSIM (7) that is defined in "Evaluation metrics" section. The MS-SSIM is a measure of structural similarity over several scales allowing to incorporate image details at different resolutions [15]. The weighted reconstruction loss is then given as:

$$\mathcal{L}^{\text{Recon}}(x, x') = \omega \cdot \mathcal{L}^{\text{MS-SSIM}}(x, x') + (1 - \omega) \cdot \mathcal{L}^{L1}(x, x') \quad (3)$$

with $\omega = 0.84$. The value for ω is set to balance the contribution of the two loss terms and proposed by [14].

Adaptive instance normalization

While conditional image-to-image GANs, e.g., pix2pix, are able to learn the mapping for categorical image-to-image translations, it cannot create MR images with fine-tuned image contrast. Therefore, we inject input and output labels (TE, TR, fat saturation) into the generator using adaptive instance normalization layers [16]. The AdaIN operation is defined as:

$$\text{AdaIN}(x, y) = \alpha(y) \cdot \left(\frac{x - \mu(x)}{\sigma(x)} \right) + \beta(y). \tag{4}$$

Each feature map x is normalized separately with its mean $\mu(x)$ and standard deviation $\sigma(x)$, then scaled and biased through learned transformations α, β, given the label set y.

AdaIN has been successfully applied for style transfer applications by injecting style encodings into the generator [16]. The style encoding is typically unknown for style transfer and is learned implicitly from an image [17]. In our case, the "style" can be understood as image tissue contrast, which is known as it is defined through the acquisition parameters. Thus, in MR imaging, the "style" is given explicitly, and it is injected via α, β, which are single layer networks with the output channels matching the channels of the network layer the acquisition parameters are injected into.

Auxiliary classifier

We also experimented with incorporating a loss term that penalizes contrast differences through the conditioning loss of an auxiliary classifier. We deviated from the conventional auxiliary classifier GAN (ACGAN) network architecture [18], which uses a classification layer in the discriminator to learn the conditions by employing a separate auxiliary classifier (AC) that is only trained on the conditions. This allowed us to pretrain the AC and break down the training complexity as the AC performance can be tuned separately from the GAN. Only a well-trained auxiliary classifier can provide useful guidance for the generative model's training, crucial for a good image reconstruction accuracy.

Network and training details

The training procedure and network architecture are depicted in Fig. 1. The generator consists of a U-Net structure with residual blocks (with filter sizes 64–128–256–512 in the encoder and vice versa in the decoder residual blocks). It injects the source image acquisition parameters into the network's encoder and the target image's acquisition parameters (i.e., target contrast) into the network's decoder via adaptive instance normalization. Through the "style" injection in the encoder, the generator is anticipated to learn an image representation independent of the acquisition parameters, thus image contrast. The target acquisition parameters are then
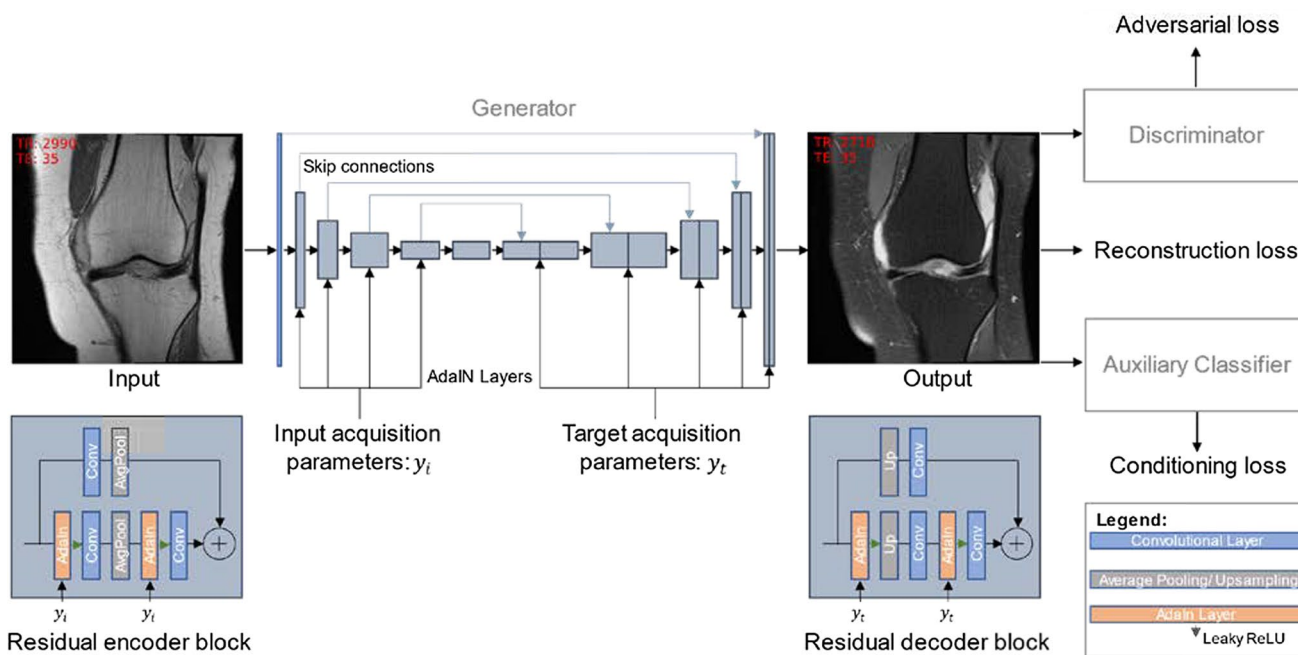

Fig. 1 Our network architecture. The generator is based on a U-Net architecture consisting of residual blocks with adaptive instance normalization. The input labels (y_i) are injected into the encoder part of the generator, and the target labels (y_t) are injected into the decoder part of the generator. The generator is trained on the reconstruction, adversarial, and conditioning loss

injected into the decoder to reconstruct the target contrast properly.

The discriminator consists of six residual blocks and a single output to discriminate between real and synthetic images (with filter sizes 64–128–256–512–512–512).

We use the recently published EfficientNet-B3 [19] architecture for the AC and train the network on the determination of TE and TR value and whether fat saturation was used. EfficientNet uses compound scaling and achieves higher accuracy and better efficiency over existing CNNs on the ImageNet and other benchmark datasets [19]. The mean squared error loss is used for TE and TR, whose values are both scaled linearly to values between 0 and 1, and binary cross-entropy for determining the use of fat saturation.

The AC is pretrained for 200 K iterations with a batch size of 64, using Adam optimizer with a learning rate of 10^{-4}, $\beta_1 = 0$, and $\beta_2 = 0.99$.

We use a batch size of eight for training the GAN and train the model for 200 K iterations. We also use the Adam optimizer with $\beta_1 = 0$ and $\beta_2 = 0.99$. The learning rates for the generator and discriminator are set to 10^{-4}. For the generator, exponential moving averages over its network parameters are employed [17, 20].

The images are scaled to intensities of $[-1, 1]$ and resized to a resolution of 256×256 pixels using bilinear interpolation to obtain an identical image resolution within the dataset. Random image shifting and zooming are applied as data augmentation during the training of the AC and GAN.

Data

We used the fastMRI dataset [4] for our training and evaluation. It contains DICOM data from 10,000 clinical knee MRI studies, each comprising a set of multiple sequence parameterizations. We applied several data filters based on the DICOM header information to get a dataset with a comparable image impression, a dense and homogenous acquisition parameter distribution (TE, TR), and high variance in anatomy. We wanted the image impression and contrast within our training set to depend on the acquisition parameters TE and TR.

Therefore, other parameters affecting the image impression were removed, such as field strength and manufacturer, by selecting the most common parameter value within the fastMRI dataset (1.5 T field strength and scanners from Siemens Healthcare, Erlangen, Germany). The MR images (fast-spin-echo sequences) from our filtered dataset were acquired on five different Siemens Healthcare scanners (MAGNETOM Aera, MAGNETOM Avanto, MAGNETOM Espree, MAGNETOM Sonata, MAGNETOM Symphony). We excluded MR images with a TR over 5000 ms and set the upper limit of TE to 50 ms to create a dataset with a dense representation of the conditions. Figure 2 shows the
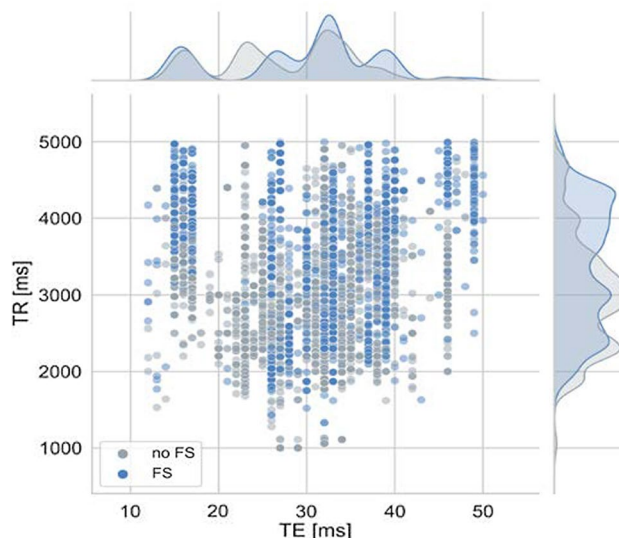

Fig. 2 Distribution of the acquisition parameters TE and TR in the training dataset, color coded by fat saturation (FS)

distribution of TE and TR values in the training dataset. Moreover, we discarded peripheral slices to limit the amount of training data, by selecting the central 14 slices for each volume.

Image pairs were then defined as images with identical DICOM attributes patient ID (0010,0020), study instance UID (0020,000D), image orientation (0020,0037), slice location (0020,1041), and slice thickness (0018,0050).

Results

Data

The training dataset consists of 237,883 MR images, with 101,328 paired images and 136,555 unpaired images, from 4815 different studies and 16,731 image series. The dataset contains 123,833 fat-saturated MR images, and the remaining images were acquired without fat saturation.

The validation and the test dataset consist of paired images from 100 disjunct, randomly selected patient IDs (3242 and 3438 images in total, respectively).

Evaluation metrics

For the evaluation of our experiments, which are described in the following, we used the normalized mean squared error (NMSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM). Given two images $\mathbf{x}$ and $\mathbf{x}'$, they are defined as:

$$\text{NMSE}(\mathbf{x}, \mathbf{x}') = \frac{\|\mathbf{x} - \mathbf{x}'\|_2^2}{\|\mathbf{x}\|_2^2} \tag{5}$$

$$\text{PSNR}(\mathbf{x}, \mathbf{x}') = 10 \cdot \log_{10}\left(\frac{\text{MAX}_{\text{range}}^2(\mathbf{x}, \mathbf{x}')}{\frac{1}{n}\|\mathbf{x} - \mathbf{x}'\|_2^2}\right) \tag{6}$$

$$\text{SSIM}(\mathbf{x}, \mathbf{x}') = \frac{\left(2 \cdot \mu_x \mu_{x'} + c_1\right) \cdot \left(2 \cdot \sigma_{xx'} + c_2\right)}{\left(\mu_x^2 + \mu_{x'}^2 + c_1\right) \cdot (\sigma_x^2 + \sigma_{x'}^2 + c_2)} \tag{7}$$

where n is the number of pixels in one image, MAX is the maximum pixel value range of two images, μ_x and $\mu_{x'}$ denote the mean values of original and translated images. σ_x and $\sigma_{x'}$ denote the standard deviation of original and translated images, and $\sigma_{xx'}$ is the covariance of both images. The variables c_1 and c_2 are added to stabilize the division with a weak denominator.

Lower values for NMSE and higher values for PSNR and SSIM demonstrate better image synthesis.

Experiments

We have conducted multiple experiments to evaluate our approach thoroughly. First, we assess the AC's performance to determine the acquisition parameters TE and TR from the MR image, as its performance is crucial for the performance evaluation and the guidance for the generator to synthesize MR images. Then, we evaluate our GAN on the task of synthesizing fat-saturated (FS) images from non-fat-saturated images and benchmark it with the pix2pix approach. Moreover, we demonstrate the capability of our approach to synthesize MR images with adaptable image contrast.

We evaluated several image-to-image GANs. Each model is based on the prior model with the additional changes as described in the following:

- Model 1: pix2pix—the training data consist of one-directional image pairs, i.e., non-FS (non-fat-saturated) to FS image pairs, under an L1-reconstruction and non-saturating adversarial loss with R1 regularization.
- Model 2: target labels are injected into the decoder with AdaIN layers.
- Model 3: source labels are additionally injected into the encoder with AdaIN layers.
- Model 4: the L1 reconstruction loss is adapted based on Formula 3.
- Model 5: image pairs with non-fat-saturated target images are added to the training target data, enabling non-FS to FS and non-FS translations.
- Model 6: unpaired non-FS image data are added to the training dataset and trained under the cycle consistency

Table 1 Evaluation of the auxiliary classifier. The reported values denote the mean and its standard deviation on the test set

Model	TE (ms)	TR (ms)	FS (%)
AC (non-FS)	1.8 ± 2.1	247 ± 262	100
AC (FS)	1.8 ± 2.2	290 ± 341	100
1	3.0 ± 3.4	447 ± 419	<u>100</u>
2	1.4 ± 1.7	354 ± 357	<u>100</u>
3	1.3 ± 1.7	298 ± 288	<u>100</u>
4	<u>1.1 ± 1.5</u>	<u>261 ± 277</u>	<u>100</u>
5	1.3 ± 2.5	<u>260 ± 270</u>	<u>100</u>
6	2.2 ± 2.0	313 ± 293	<u>100</u>

The MAE is reported for TE and TR and accuracy for FS

Statistically significant ($p \leq 0.05$) best values for each acquisition parameter are underlined.

Since for model 1–4 only FS images can be synthesized, the conditioning evaluation for all models reported in this table is based on FS images only.

Table 2 Quantitative evaluation results of the image synthesis experiments

Model	NMSE	PSNR	SSIM
1	0.13 ± 0.13	22.91 ± 2.61	0.62 ± 0.12
2	0.12 ± 0.14	23.45 ± 2.7	0.63 ± 0.11
3	0.10 ± 0.11	24.15 ± 2.58	0.64 ± 0.12
4	0.10 ± 0.11	<u>24.29 ± 2.59</u>	<u>0.66 ± 0.11</u>
5	<u>0.09 ± 0.10</u>	<u>24.48 ± 2.81</u>	<u>0.66 ± 0.11</u>
6	0.16 ± 0.21	22.14 ± 2.57	0.58 ± 0.12

Statistically significant ($p \leq 0.05$, two-sample t test) best values are underlined.

and conditioning loss (with weight $\lambda_c = 10$). The network is trained on unpaired data according to its share in the dataset. λ_c is a hyperparameter and is set heuristically such that the conditioning validation error is similar to the error of the auxiliary classifier (i.e., is not over- or underfitting w.r.t. the conditioning). This is the only model that uses the auxiliary classifier during training, which enables the generator to be trained on random acquisition parameter combinations.

The performance results on the test set are reported in Tables 1 and 2. A good performance of the auxiliary classifier is crucial for proper guidance for the generator (model 6) during training and the evaluation of all GAN models. Although the AC is trained under the MSE loss, we report the mean absolute error as it is more meaningful for the performance evaluation (Table 1). The auxiliary classifier predicts the acquisition parameters correctly with a low overall error. TE has a stronger impact on image contrast than TR (compare Fig. 4), which might indicate

why the MAE of the AC for TE is (proportionally) lower than for TR. Moreover, the fat saturation distinction works perfectly (accuracy: 100%).

We also evaluated the models' conditioning error on the test set (Table 1) to assess how well the target contrasts are synthesized. Therefore, we translated the input test images to the contrast settings of the paired target image, determined the acquisition parameters with the AC, and computed the error.

The benchmark pix2pix model (model 1) shows a high conditioning error. It is not guided by the input or target acquisition parameters and can only translate an image to the expected contrast. Incorporating the target acquisition parameters (model 2) and the input acquisition parameters (model 3) reduces the MAE on TE and TR determination, demonstrating a better generation of the target MR image contrast. Incorporating the reconstruction loss from Formula (3) and using more training data (models 4 and 5) further enhances the contrast synthesis. Adding unpaired training data (model 6) does not further improve the conditioning.

The reconstruction performance is reported in Table 2. Incorporating input and target acquisition parameters significantly increases the reconstruction performance (model 2 and 3) w.r.t. the benchmark model. Adapting the reconstruction loss and including more training data (model 4 and 5) further enhances the performance. Model 5 also extends the network's capabilities: It enables the model to interpolate the MR-contrast within non-fat-saturated images (Fig. 4) while also allowing translation of non-fat-saturated into fat-saturated images. Introducing unpaired training data and cyclic reconstruction (model 5 and 6) allows the training on a larger set of input-target acquisition parameter sets (with

randomly selected target acquisition parameters) but does not improve the reconstruction performance.

The reconstructed images in Figs. 3, 4, and 5 were generated using model 5. The reconstruction of a fat-saturated image from its non-fat-saturated image pair is presented in Fig. 3. A novel functionality of our approach is the contrast interpolation capabilities shown in Fig. 4. This capability allows image synthesis with a fine-tuned image contrast, adapted to a use case and contrast requirements. Varying TE significantly influences the muscle tissue's signal intensity, while varying TR mainly causes signal intensity differences within the joint [21].

Moreover, our approach correctly reconstructs different anatomies and acquisition parameter values, as demonstrated in Fig. 5.

Discussion

Sequence parameterizations vary considerably across different sites, scanners, and scans in clinical practice (compare Fig. 2 and [1, 22, 23]). Consequently, MR image synthesis approaches must be able to cope with this variability to be widely applicable. Therefore, we have adapted our model architecture and training process to fit these needs by injecting the acquisition parameters into the model. Acquisition parameter injection improves the reconstruction results, and our method outperforms a benchmark pix2pix approach on fat-saturated image synthesis from non-fat-saturated MR images. Additionally, we can adapt MR image contrast retrospectively and continuously using acquisition parameter injection using adaptive instance normalization. While a physics-aware U-Net was trained in [24] to increase the
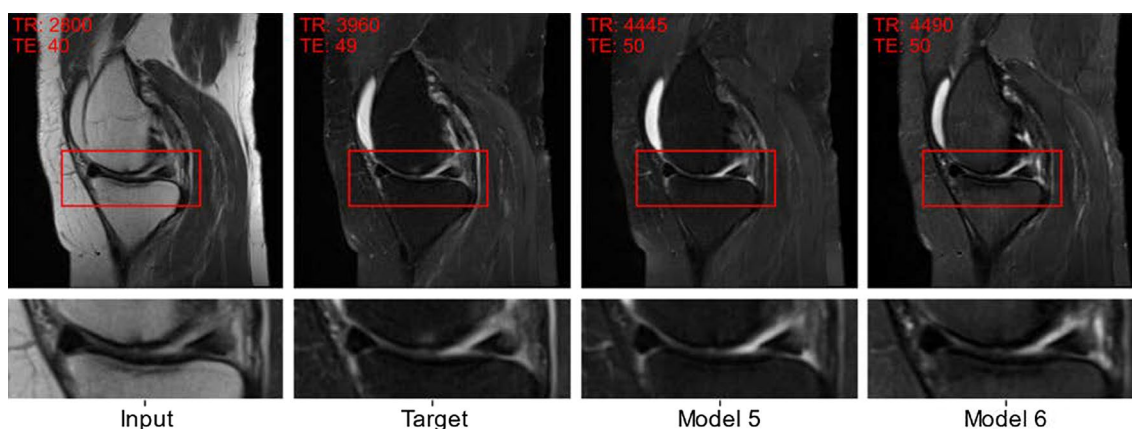

Fig. 3 Image-to-image translation example with our proposed approach. The figure shows a paired example of a real non-fat-saturated and fat-saturated MR knee image with the reconstructed MR image (prediction). The images are annotated with the true acquisition parameters (TE, TR) for the input and target, and with the acqui-sition parameters determined by the AC for the predicted image. Both models generate sharp MR images with the correct contrast (see TR/TE values). However, model 6 produces inaccurate results and contrast, e.g., in the bone structures
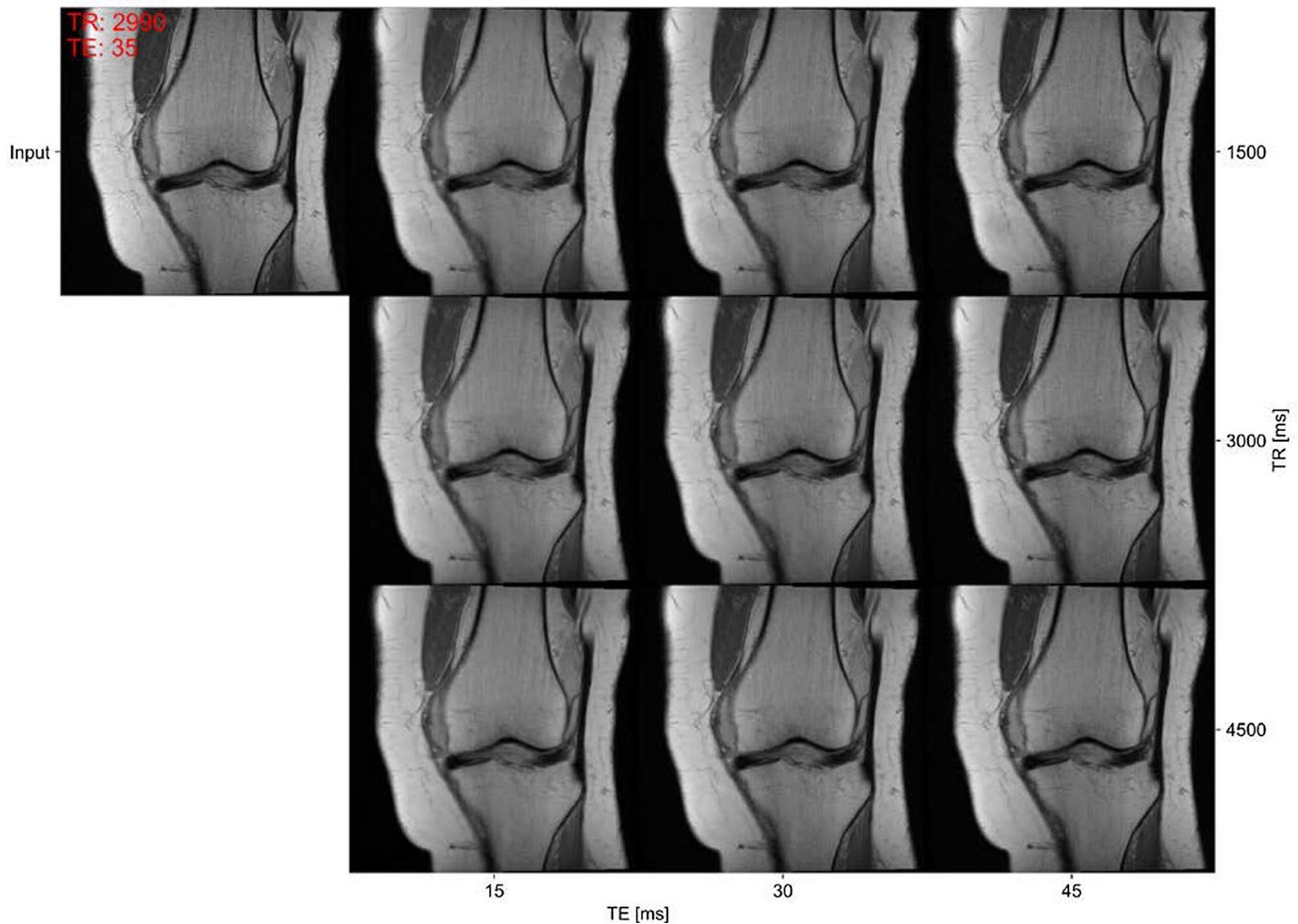
Fig. 4 Example of contrast interpolation given a real non-fat-saturated MR image (top left). The input image is annotated with the true acquisition parameters (TE, TR) for the input image. Contrast changes through signal intensity changes in the muscle tissue are visible in the synthetic images for varying TE values, and contrast changes for varying TR values are less distinct in this parameter range

robustness of segmentation results across acquisitions, our approach is the first to provide retrospective contrast adjustments using GANs.

To the best of our knowledge, our work is the second to apply an image-to-image translation network to knee MR images for contrast synthesis [8]. Comparison to published approaches is not meaningful due to the different datasets used and the training on different body regions. Training our work on other benchmark datasets, e.g., the BRATS dataset [25], was not possible as acquisition parameters are not provided in the dataset.

Model 6 receives feedback from the auxiliary classifier, a trained neural network, on how well the MR-contrast was synthesized. Incorporating the AC allows us to train the GAN under a cycle consistency loss with random target labels on unpaired data. However, a limitation of the cycle consistency loss is that it may hallucinate features in the generated images [26]. It also decreases the overall reconstruction performance, likely due to the imperfection of the auxiliary classifier. On the other hands, it also increases the generalization ability of the network. It enables the network to learn a broader range of contrast settings as a larger set of input-target label combinations can be used for training. The use of unpaired data will likely be obsolete given a more diverse dataset.

The dataset contains only PD-weighted non-fat-saturated and PD- and T2-weighted fat-saturated MR images based on the DICOM series description. Consequently, T1-weighted and T2-weighted MR images cannot be synthesized due to the lack of training data. Nonetheless, it is anticipated that our approach's capabilities can be easily extended given a more diverse (in terms of acquisition parameter value distribution) training dataset. Another important acquisition parameter influencing the MR image contrast is the flip angle. Unfortunately, the dataset does not offer the required variation in the flip angle to condition the model on different flip angles. Therefore, the incorporation of the flip angle is omitted in this approach.
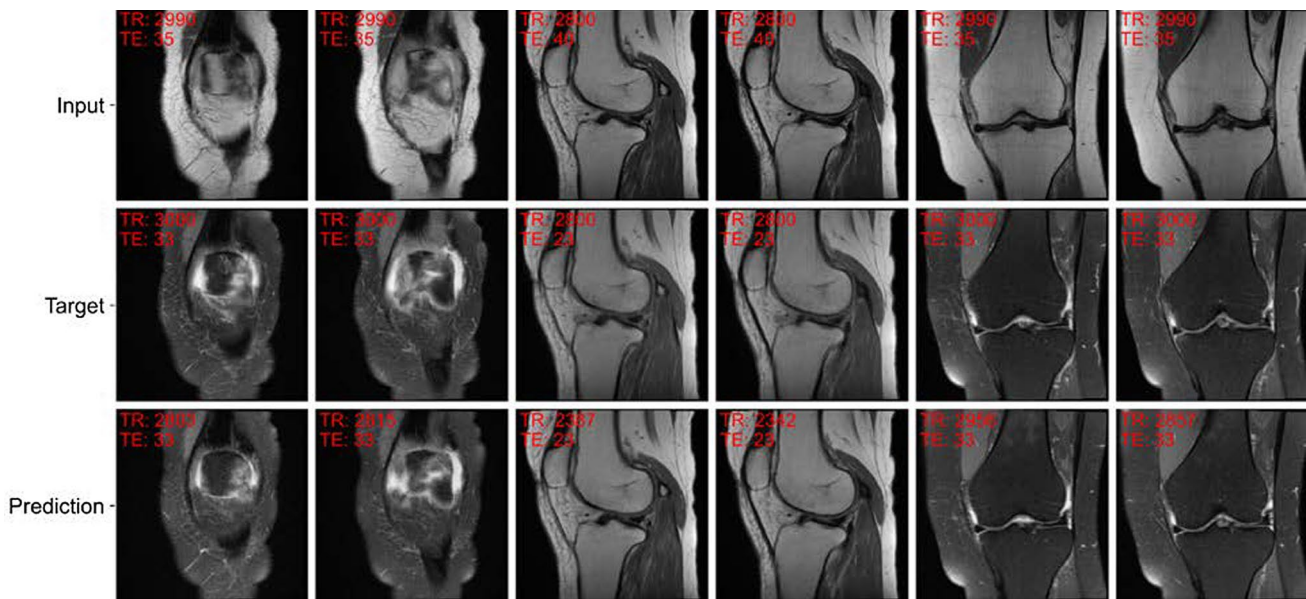
Fig. 5 Multiple examples of image pairs and the reconstructed images using our proposed approach. All images belong to a single, reference test patient, and for each image pair, an adjacent image pair is also shown. The first two columns show adjacent images of peripheral slices, while the remaining columns show central slices of sagittal and coronal image series. The images are annotated as described for the previous figures. The model properly reconstructs a wide range of anatomical structures and views, with different sets of acquisition parameters

Furthermore, image pairs within the dataset are often not appropriately registered, and through-plane motion can be observed between paired image series. 2D rigid registration methods did not improve the training performance. However, efforts to incorporate 3D non-rigid image registration may increase data quality and thus model performance.

Different MR-contrasts are typically required for a single exam as they enable the proper visualization and differentiation of the tissue characteristics. Additionally, different MR-contrasts typically provide both redundant and unique information. A general limitation of MR image-to-image contrast translation networks is that they do not reconstruct additional acquisitions but only learn to translate existing MR images into new image contrasts. Therefore, these deep learning approaches are not anticipated to be able to learn subtle changes that are visible in the target contrast but not in the source contrast. While extending the proposed approach to a multi-image-to-image translation setting is anticipated to significantly enhance the performance, the diagnostic quality of the synthesized MR images must be assessed carefully.

Our work is anticipated to be used as a method to replace missing or corrupted contrasts. This may shorten overall scan time and reduce the number of re-scans. Furthermore, it can be used as an advanced data augmentation tool to render different contrasts and increase the robustness of a trained AI application for contrast changes. It may also support the radiologists during protocoling by providing a preview of the contrast of a parameterized sequence. Additionally, this approach can also

be useful for other image-to-image translation tasks within medical imaging, e.g., to enhance intermodality translation (MRI → CT) or 7 T image synthesis from 3 T MR images [27].

Moreover, our approach can be the basis for image-to-image tasks in MRI as it is anticipated to be adaptable to any set of parameterized sequences. It is easily adaptable to additional inputs, different acquisition parameters, and applications.

In future work, we aim to improve the determination of the acquisition parameters, which will enhance contrast synthesis and enable us to evaluate our model more accurately. Moreover, the extension to additional data (e.g., the osteoarthritis initiative (OAI) dataset [28]), and multi-image-to-image contrast synthesis is anticipated to improve the performance as more data and information generally lead to an improved model. Additionally, the extension to 3D image data is anticipated to reduce prediction inconsistencies in subsequent 2D slices. Furthermore, a reader study must be conducted to quantify the diagnostic value of the synthesized images and assess our approach's clinical significance for applications such as the synthesis of additional contrasts.

Conclusion

In this paper, we proposed a novel approach to tackle MR image-to-image synthesis by guiding the learning process with the MR acquisition parameters that define the image

tissue contrast. Our approach was evaluated on the task of synthesizing fat-saturated MR images from non-fat-saturated images, outperforming a pix2pix benchmark method and demonstrated its capabilities for continuous contrast synthesis.

Our approach is easily extendible to incorporate more acquisition parameters, 3D image data, and multi-image-to-image translations that can further increase the reconstruction accuracy.

Availability of data and materials Public dataset used.

Code availability Private code repository.

Declarations

Conflict of interest Jonas Denck and Jens Guehring are employees of Siemens Healthcare GmbH. Andreas Maier has received research grants from Siemens Healthcare GmbH. Eva Rothgang was an employee of Siemens Healthcare GmbH.

Ethical approval For this type of study formal consent is not required.

Informed consent This article contains MRI data from a publicly available dataset, i.e., formal consent is not required.

References

1. Denck J, Landschütz W, Nairz K, Heverhagen JT, Maier A, Rothgang E (2019) Automated billing code retrieval from MRI scanner log data. J Digit Imaging 32:1103–1111. https://doi.org/10.1007/s10278-019-00241-z
2. van Beek EJR, Kuhl C, Anzai Y, Desmond P, Ehman RL, Gong Q, Gold G, Gulani V, Hall-Craggs M, Leiner T, Lim CCT, Pipe JG, Reeder S, Reinhold C, Smits M, Sodickson DK, Tempany C, Vargas HA, Wang M (2019) Value of MRI in medicine: more than just another test? J Magn Reson Imaging 49:e14–e25. https://doi.org/10.1002/jmri.26211
3. Chaudhari AS, Black MS, Eijgenraam S, Wirth W, Maschek S, Sveinsson B, Eckstein F, Oei EHG, Gold GE, Hargreaves BA (2018) Five-minute knee MRI for simultaneous morphometry and T2 relaxometry of cartilage and meniscus and for semiquantitative radiological assessment using double-echo in steady-state at 3T. J Magn Reson Imaging 47:1328–1341. https://doi.org/10.1002/jmri.25883
4. Knoll F, Zbontar J, Sriram A, Muckley MJ, Bruno M, Defazio A, Parente M, Geras KJ, Katsnelson J, Chandarana H, Zhang Z, Drozdzalv M, Romero A, Rabbat M, Vincent P, Pinkerton J, Wang D, Yakubova N, Owens E, Zitnick CL, Recht MP, Sodickson DK, Lui YW (2020) fastMRI: a publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning. Radiol Artif Intell 2:e190007. https://doi.org/10.1148/ryai.2020190007
5. Sriram A, Zbontar J, Murrell T, Zitnick CL, Defazio A, Sodickson DK (2020) GrappaNet: combining parallel imaging with deep learning for multi-coil MRI reconstruction. In: IEEE conference on computer vision and pattern recognition (CVPR), Seattle, WA, USA, June 13–19, pp 14303–14310.
6. Yu B, Zhou L, Wang L, Shi Y, Fripp J, Bourgeat P (2019) Ea-GANs: edge-aware generative adversarial networks for cross-modality MR image synthesis. IEEE Trans Med Imaging 38:1750–1762. https://doi.org/10.1109/TMI.2019.2895894
7. Dar SU, Yurt M, Karacan L, Erdem A, Erdem E, Cukur T (2019) Image synthesis in multi-contrast MRI with conditional generative adversarial networks. IEEE Trans Med Imaging 38:2375–2388. https://doi.org/10.1109/TMI.2019.2901750
8. Liu F (2019) SUSAN: segment unannotated image structure using adversarial network. Magn Reson Med 81:3330–3345. https://doi.org/10.1002/mrm.27627
9. Yang Q, Li N, Zhao Z, Fan X, Chang EI-C, Xu Y (2020) MRI cross-modality image-to-image translation. Sci Rep 10:3753. https://doi.org/10.1038/s41598-020-60520-6
10. Isola P, Zhu J-Y, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: IEEE conference on computer vision and pattern recognition (CVPR), Honolulu, HI. July 21–26. IEEE, pp 5967–76
11. Zhu J-Y, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: IEEE International conference on computer vision (ICCV), Venice. October 22–29. IEEE, pp 2242–51
12. Lucic M, Kurach K, Michalski M, Bousquet O, Gelly S (2018) Are GANs created equal? A large-scale study. In: Advances in neural information processing systems. December 3–8. Curran Associates Inc, Red Hook, pp 698–707
13. Madabhushi A, Udupa JK (2006) New methods of MR image intensity standardization via generalized scale. Med Phys 33:3426–3434. https://doi.org/10.1118/1.2335487
14. Zhao H, Gallo O, Frosio I, Kautz J (2017) Loss functions for image restoration with neural networks. IEEE Trans Comput Imaging 3:47–57. https://doi.org/10.1109/TCI.2016.2644865
15. Wang Z, Simoncelli EP, Bovik AC (2003) Multiscale structural similarity for image quality assessment. In: The thrity-seventh Asilomar conference on signals, systems & computers, Pacific Grove, CA, USA. Nov 9–12. IEEE, pp 1398–402
16. Huang X, Belongie S (2017) Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization. In: IEEE international conference on computer vision (ICCV), Venice. October 22–29. IEEE, pp 1510–1519
17. Choi Y, Uh Y, Yoo J, Ha J-W (2020) StarGAN v2: diverse image synthesis for multiple domains. In: IEEE conference on computer vision and pattern recognition (CVPR), Seattle, WA, USA. June 13–19, 2020, pp 8185–8194
18. Odena A, Olah C, Shlens J (2017) Conditional image synthesis with auxiliary classifier gans. In: Proceedings of the international

conference on machine learning (ICML), vol 70. Sydney, Australia, pp 2642–2651

19. Tan M, Le QV (2019) EfficientNet: rethinking model scaling for convolutional neural networks. In: International conference on machine learning (ICML), Long Beach, California, USA. June 9–15. PMLR, pp 6105–6114

20. Yazıcı Y, Foo C-S, Winkler S, Yap K-H, Piliouras G, Chandrasekhar V (2019). The unusual effectiveness of averaging in GAN training, In: International conference on learning representations (ICLR), New Orleans, LA, USA, May 6–9, 2019

21. Gold GE, Han E, Stainsby J, Wright G, Brittain J, Beaulieu C (2004) Musculoskeletal MRI at 3.0 T: relaxation times and image contrast. AJR Am J Roentgenol 183:343–351. https://doi.org/10.2214/ajr.183.2.1830343

22. Sachs PB, Hunt K, Mansoubi F, Borgstede J (2017) CT and MR protocol standardization across a large health system: providing a consistent radiologist, patient, and referring provider experience. J Digit Imaging 30:11–16. https://doi.org/10.1007/s10278-016-9895-8

23. Glazer DI, DiPiro PJ, Shinagare AB, Huang RY, Wang A, Boland GW, Khorasani R (2018) CT and MRI protocol variation and optimization at an academic medical center. J Am Coll Radiol 15:1254–1258. https://doi.org/10.1016/j.jacr.2018.06.002

24. Borges P, Sudre C, Varsavsky T, Thomas D, Drobnjak I, Ourselin S, Cardoso MJ (2019) Physics-informed brain MRI segmentation. In: Burgos N, Gooya A, Svoboda D (eds) Simulation and synthesis in medical imaging, vol 11827. Springer, Cham, pp 100–109

25. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber M-A, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp C, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SMS, Ryan M, Sarikaya D, Schwartz L, Shin H-C, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, van Leemput K (2015) The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imaging 34:1993–2024. https://doi.org/10.1109/tmi.2014.2377694

26. Cohen JP, Luck M, Honari S (2018) Distribution matching losses can hallucinate features in medical image translation. In: Medical image computing and computer-assisted intervention (MICCAI), vol 11070. Springer, Cham, pp 529–536

27. Nie D, Trullo R, Lian J, Wang L, Petitjean C, Ruan S, Wang Q, Shen D (2018) Medical image synthesis with deep convolutional adversarial networks. IEEE Trans Biomed Eng 65:2720–2730. https://doi.org/10.1109/TBME.2018.2814538

28. Osteoarthritis initiative. https://nda.nih.gov/oai/. Accessed 6 May 2021

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.