Article

Enhanced Magnetic Resonance Image Synthesis with Contrast-Aware Generative Adversarial Networks

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

1 Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander Universität Erlangen-Nürnberg, 91058 Erlangen, Germany; andreas.maier@fau.de

2 Siemens Healthcare GmbH, 91052 Erlangen, Germany; jens.guehring@siemens-healthineers.com

3 Department of Industrial Engineering and Health, Technical University of Applied Sciences Amberg-Weiden, 92637 Weiden, Germany; e.rothgang@oth-aw.de

* Correspondence: jonas.denck@fau.de

Abstract: A magnetic resonance imaging (MRI) exam typically consists of the acquisition of multiple MR pulse sequences, which are required for a reliable diagnosis. With the rise of generative deep learning models, approaches for the synthesis of MR images are developed to either synthesize additional MR contrasts, generate synthetic data, or augment existing data for AI training. While current generative approaches allow only the synthesis of specific sets of MR contrasts, we developed a method to generate synthetic MR images with adjustable image contrast. Therefore, we trained a generative adversarial network (GAN) with a separate auxiliary classifier (AC) network to generate synthetic MR knee images conditioned on various acquisition parameters (repetition time, echo time, and image orientation). The AC determined the repetition time with a mean absolute error (MAE) of 239.6 ms, the echo time with an MAE of 1.6 ms, and the image orientation with an accuracy of 100%. Therefore, it can properly condition the generator network during training. Moreover, in a visual Turing test, two experts mislabeled 40.5% of real and synthetic MR images, demonstrating that the image quality of the generated synthetic and real MR images is comparable. This work can support radiologists and technologists during the parameterization of MR sequences by previewing the yielded MR contrast, can serve as a valuable tool for radiology training, and can be used for customized data generation to support AI training.

Keywords: adversarial learning; deep learning; magnetic resonance imaging; image synthesis

1. Introduction

In magnetic resonance imaging (MRI), multiple contrasts are usually acquired within a single exam that are required to make a reliable diagnosis. Each contrast is based on the parameterization of an MR sequence through multiple acquisition (pulse sequence) parameters. In general, the acquisition parameters for an MR sequence affect image contrast, image resolution, signal-to-noise ratio, and scan time. Important acquisition parameters that affect the image contrast are the repetition time (TR) and echo time (TE). The parameterization of a sequence is subject to clinical guidelines, the MR system (vendor, model, software version, and field strength), the clinical protocol (i.e., the set of parameterized sequences used for the exam), internal guidelines (e.g., slot time), and radiologists' preferences. Clinical guidelines (e.g., ACR-SPR-SSR Practice Parameters [1]) offer recommendations but "are not inflexible rules or requirements of practice and are not intended, nor should they be used, to establish a legal standard of care" [1]. Thus, the guidelines do not specify exact acquisition parameter settings as they will depend on the field strength and desired contrast weighting.

Consequently, sequence parameterizations differ significantly across different radiology sites [2,3]. However, due to the various degrees of freedom in protocol configuration

and sequence parameterization, this can also be true for a single radiology site. These differences additionally increase the complexity of MRI protocoling (i.e., selecting a set of parameterized sequences for an MRI exam), image interpretation and diagnostics, and the development of AI (artificial intelligence) applications for MRI. Since most AI applications are trained and evaluated on a limited set of MR sequences with fixed or narrowly defined acquisition parameter values [4], the applicability of AI-based applications to sequences with different parameterizations is not guaranteed. Consequently, retraining with a new set of MR images may be necessary. However, since the abundance of clinical images for AI training is limited, this is not always possible.

To mitigate the problem of the availability of MR images for varying contrast settings, we developed an approach for MR image synthesis that can be parameterized with acquisition parameters. Our method can generate customized training data for AI applications and serve as an additional data augmentation tool through contrast augmentation. Moreover, it can be used as a tool for MRI training for technologists and radiologists and can support protocoling and sequence parameterization by visualizing the yielded contrast of a parameterized sequence.

We developed a generative adversarial network (GAN) that can generate synthetic MR images of the knee conditioned on the repetition and echo time. The model can additionally synthesize MR images conditioned on the image orientation. In contrast to existing AI-based approaches for the synthesis of MR images, our model is not conditioned on different sets of contrasts (e.g., T1w, T2w, or PDw) but on the acquisition parameters that determine the contrast weighting. This allows the generation of MR images with fine-tuned contrast, adjusted to the application's specific needs.

The rest of our paper is structured as follows: first, we shortly review the current literature of generative adversarial networks with the focus on MR image contrast synthesis (Section 2). Then, we describe our approach, including the data used to train the GAN in Section 3. We evaluate our approach qualitatively and quantitatively (Section 4), discuss our results (Section 5), and then conclude our work (Section 6).

2. Background

Generative adversarial networks [5] learn to generate images through the adversarial training of a generator network G, which is trained to produce realistic samples and tries to fool the discriminator network D that learns to distinguish between real and synthetic samples. Commonly, the generator takes as input a latent or noise vector z, randomly sampled from a normal distribution. The original GAN formulation is adapted in several ways, e.g., introducing conditional GANs [6,7], adapting the input for image-to-image [8], and text-to-image [9] translations, enhancing the network architecture [10], the loss formulation [11,12], and training procedure [13]. GANs are successfully applied to various domains, and a detailed overview of GANs and their variants are given in [14].

GANs have also been successfully applied to the field of medical imaging (e.g., [15,16]), with applications that can mainly be divided into seven categories: synthesis, segmentation, reconstruction, detection, denoising, registration, and classification, whereas the majority of publications address synthesis applications [17].

In the following, we review publications focused on synthesizing MR images with noise-to-image GANs (Table 1). A detailed general overview of GANs in the field of medical imaging is given in [18].

Most publications in the field of MR image synthesis with noise-to-image GANs address the generation of different contrasts (e.g., T1w, T2w, and PDw) in a categorical manner, without the incorporation of the acquisition parameter values. Their main target application is enhanced data augmentation for deep learning applications (see Table 1).

In ref. [19], a Laplacian GAN [20] was trained on 2D image slices from sagittal brain MR 3D-T1w slices to enhance the augmentation of biomedical datasets.

A semi-coupled GAN was used as a data generation method for deep learning-based detection of incomplete left-ventricle coverage [21]. Additionally, a Wasserstein-GAN was

used to synthesize brain MR images of different contrasts (T1w, T1c, T2w, and FLAIR) with a resolution of 128 × 128 pixels [22] and a progressive growing GAN (PGGAN) [13] to both synthesize brain MR images and place brain metastases on synthetic MR images (256 × 256 pixels) for enhanced data augmentation for AI training [23,24]. Multi-modal MR images were synthesized with a PGGAN in [25]. Moreover, GANs were used for enhanced image denoising [26] for brain MR images and the generation of additional training data for brain tissue segmentation networks [27].

Performance benefits through additional GAN-based data augmentation are reported for different medical deep learning applications [28,29]. However, the improvement gain depends on the amount of real data seen during training, with the greatest improvements observed for training procedures with a limited amount of real data [27].

Current GAN-based MR image and contrast synthesis methods only generate single categorical contrasts with no scale of similarity between the different contrasts [19,23,24,27,30]. Consequently, the GAN does not learn to disentangle the underlying anatomy from the contrast specifications, limiting the GAN's capability for MR image synthesis. We solve this by conditioning the GAN on the acquisition parameters that determine the MR image contrast, therefore disentangling anatomy and contrast synthesis.

Table 1. Literature review: MR image synthesis using noise-to-image GANs.

Ref.	Anatomy	Method/Network Architecture	Sequence Specification	Resolution [Pixels]	Application
[19]	Brain	LAPGAN	T1w (4/2000 ms)	128 × 64	DA
[21]	Heart	SCGAN	Cine	120 × 120	DA
[22]	Brain	DCGAN/WGAN	T1w, T1c, T2w, FLAIR (BRATS 2016 [31])	128 × 128	DA
[23]	Brain	CPGGAN	T1c (BRATS 2016)	256 × 256	DA
[24]	Brain	PGGAN + MUNIT/ SimGAN	T1c (BRATS 2016)	256 × 256	DA
[25]	Brain	PGGAN	T1w, T1c, T2w, FLAIR (BRATS 2016)	256 × 256	DA, unsupervised classification of pathology
[26]	Brain	DCGAN	T1w	220 × 172	Image denoising
[27]	Brain	PGGAN	FLAIR	128 × 128	Segmentation

Legend: DA: data augmentation. FLAIR: fluid-attenuated inversion recovery sequence [32]. T1c: T1-weighted contrast-enhanced sequence. LAPGAN: Laplacian GAN. SCGAN: semi-coupled GAN. DCGAN: deep convolutional GAN [10]. WGAN: Wasserstein GAN [12]. PGGAN: progressive growing GAN. CPGGAN: conditional progressive growing GAN. MUNIT: multimodal unsupervised image-to-image translation framework [30]. SimGAN: semantic image manipulation using generative adversarial networks [33].

3. Materials and Methods

3.1. Progressive Growing WGAN-GP

In order to train a network on the synthesis of MR images, we used a progressive growing GAN proposed by [13] that started with learning low-resolution images and progressively increases the resolution by stacking additional layers to the network. Progressive growing allowed the network to learn large-scale features of the data distribution first and refined structures when adding additional layers and progressing in training, which stabilized GAN training significantly. Our network architecture was trained with 800 k images before doubling the resolution (following the training procedure proposed in [13]). A new layer was faded in during training with additional 800 k images until we reached the final resolution of 256 × 256. We used a Wasserstein GAN with gradient penalty loss (WGAN-GP) [12] following the proposed training procedure in [13]:

$$L_{WGAN-GP} = \mathbb{E}_{\tilde{x} \sim P_g}[D(\tilde{x})] - \mathbb{E}_{x \sim P_r}[D(x)] + \lambda_{gp} \cdot \mathbb{E}_{\hat{x} \sim P_{\hat{x}}}\left[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2\right] \quad (1)$$

where $\mathbb{E}[\cdot]$ denotes the expected value, P_r is the real data distribution, P_g is the data distribution generated through $\widetilde{x} = G(z, c)$, and c denotes the target labels (conditions). A gradient penalty term, weighted by λ_{gp} ($\lambda_{gp} = 10$), was added for the random sample $\hat{x} \sim P_{\hat{x}}$, with $\nabla_{\hat{y}}$ denoting the gradient operator toward the generated samples. $P_{\hat{x}}$ described the distribution of points uniformly sampled along straight lines from pairs of points from P_r and P_g [11]. The utilization of a gradient penalty term resulted in a more stable training process than weight clipping of the discriminator as proposed in [12].

3.2. Separate Auxiliary Classifier

We deviated from the conventional auxiliary classifier GAN (ACGAN) network architecture [7], 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 use data augmentation on the training data for the AC. In general, this enhances the classification performance of a trained network by avoiding overfitting [34,35], and only a well-trained auxiliary classifier can provide good guidance for the training of the generator. In contrast to training the AC, heavy data augmentation diminishes the GAN performance of generating sharp, realistic images. Consequently, no data augmentation was used for training the discriminator that learns to score the realness of a given image in a WGAN-GP.

We used the DenseNet-121 architecture [36] for the AC and jointly trained the network to determine TR, TE, and imaging orientation of the patient (*IOP*) from MR images only. The categorical cross-entropy loss (*CCE*) was used for the image orientation and the mean squared error (*MSE*) loss for TR and TE, which were both scaled to values between 0 and 1. The AC was trained to minimize the following loss:

$$L_{AC} = \lambda_{IOP} \cdot CCE_{IOP}(\text{c}, \widetilde{c}) + \lambda_{TE} \cdot MSE_{TE}(\text{c}, \widetilde{c}) + \lambda_{TR} \cdot MSE_{TR}(\text{c}, \widetilde{c}) \tag{2}$$

where $\widetilde{c} = C(x)$ is the output of the auxiliary classifier for an image x with labels c. We heuristically set $\lambda_{IOP} = 1$ and $\lambda_{TE} = \lambda_{TR} = 10$.

3.3. Controllable GAN

To avoid overfitting the generator to the conditioning, we incorporated the training procedure of ControlGAN [37]. Overfitting can occur if the auxiliary classifier reaches imperfect classification performance and consequently an issue for many machine learning-based classification and regression tasks. We used an adaptive loss weighting to avoid overfitting the generator on the conditions and balanced the GAN training to produce realistic images. An adaptive weight loss parameter γ_t for time step t for each condition c of the GAN was introduced:

$$\gamma_{c, t} = \min\left[\tau_{c,}, \max\left[0, \gamma_{c, t-1} + r \cdot \left\{L_c(c, C(G(z, c))) - \hat{E} \cdot L_c(c, C(x))\right\}\right]\right] \tag{3}$$

where r is a learning rate parameter for γ_t, γ_0 is set to zero, and τ is a maximum constraint for γ_t. The parameter τ was set to 100 and the learning rate r to 0.01 for our training procedure. $\hat{E}$ balances the ratio between the classification loss on the real training images and the generated images and was set to one. L_c denotes the condition-specific loss (*CCE* for the image orientation or *MSE* for *TR* and *TE*). Thus, the *GAN* was trained to minimize the conditioning loss:

$$L_{AC-GAN} = \gamma_{IOP, t} \cdot CCE_{IOP}(c, C(G(z, c))) + \gamma_{TE, t} \cdot MSE_{TE}(c, C(G(z, c))) + \gamma_{TR, t} \cdot MSE_{TR}(c, C(G(z, c))) \tag{4}$$

Minimizing the sum of the WGAN-GP loss (Equation (1)) and the auxiliary classifier loss with adaptive weights (Equation (4)) led to the generation of realistic MR images with the intended MR contrast. Thus, the overall GAN loss function was given as the sum of L_{GAN-GP} and L_{AC-GAN}.

3.4. Data

For training and evaluation of our GAN, we used the fastMRI dataset [38]. It contained DICOM data from 10,000 clinical knee MRI studies, each comprising a set of multiple pulse sequence parameterizations. We applied several data filters based on the DICOM header information to obtain a dataset with a comparable image impression, a dense and homogenous acquisition parameter distribution (TR, TE), and a high variance in anatomy. We wanted the image impression and contrast within our training set to depend on the acquisition parameters TR and TE. Therefore, other parameters affecting the image impression, such as field strength and manufacturer, were removed by selecting the most common parameter value within the fastMRI dataset (1.5T field strength and scanners from Siemens Healthcare, Erlangen, Germany). Several attributes were missing in the DICOM header, but we could deduce the manufacturer for DICOM images with the missing manufacturer by adopting the manufacturer of the used receiver coil. The MR images from our filtered dataset were acquired on five different Siemens Healthcare scanners (MAGNETOM Aera, MAGNETOM Avanto, MAGNETOM Espree, MAGNETOM Sonata, MAGNETOM Symphony). To create a dense data distribution for the conditioning parameters TR and TE, we only took image series with TR values between 1800 ms and 5000 ms, set the upper limit of TE to 50 ms, and discarded fat saturated images. Then, we took the six central slices from each volume to discard peripheral slices. The final dataset contained MR images from 5387 different studies and 8535 image series. The joint distribution of TR and TE values in the training dataset is shown in Figure 1.
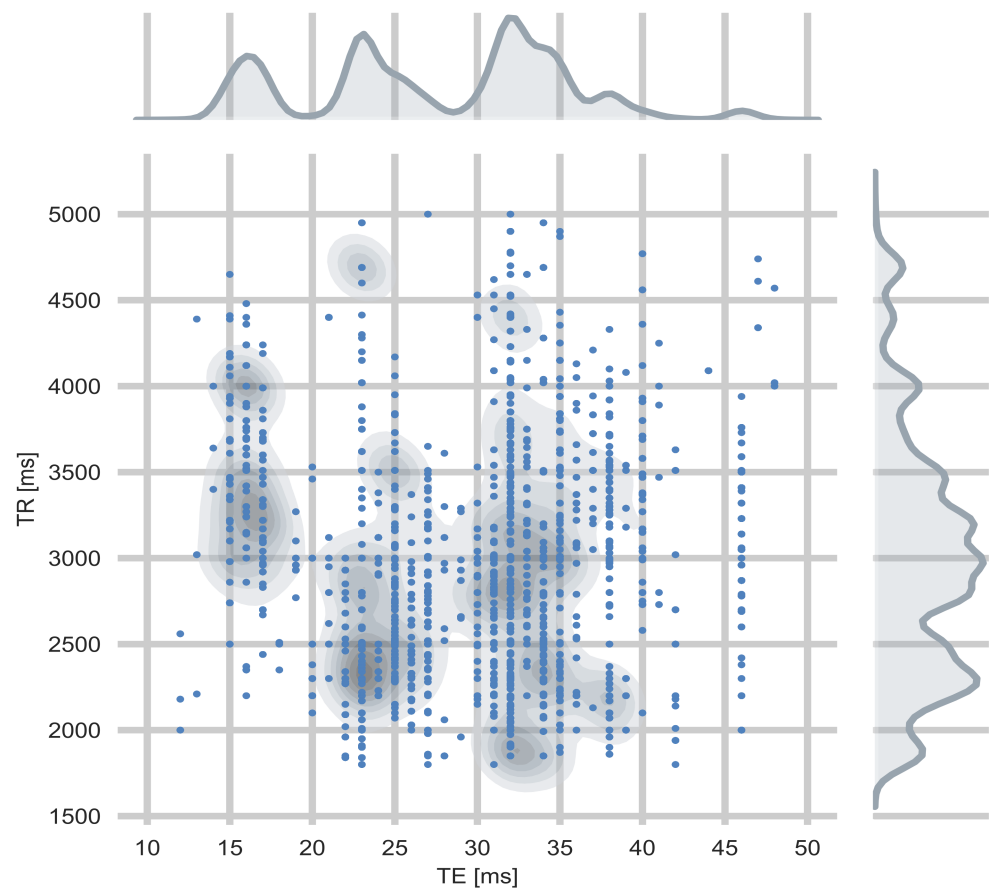

Figure 1. Distribution of the acquisition parameters TR and TE in the training dataset. The kernel density estimate plot visualizes the density of the bivariate target distribution. The multiple modes of the multimodal distribution (compared with distribution of TE values), arise from varying sequence parameterizations used at different scanners in the dataset.

The dataset of 51,205 images was split into a training, validation, and test dataset randomly by study IDs, with 2000 images each used for validation and testing. The remaining images were used for training. Each slice was normalized to intensity values between -1 and 1 and resized to the smallest common resolution within the dataset (256×256 pixels) using bilinear interpolation to obtain identical image resolution. Thus, the complete dataset could be used for training without discarding images due to insufficient resolution and image upsampling was avoided, which reduced image quality.

3.5. Training Details

We trained the discriminator and generator networks with a balanced number of weight updates (Figure 2). No conditioning was applied during the progressive growing of the networks until the final resolution was reached. The AC loss was incorporated with adaptive loss weights as defined in Equation (4). The AC was pre-trained on the same dataset as the GAN for the final resolution for 200 epochs (batch size of 64, Adam optimizer [39] with learning rate 0.001, $\beta_1 = 0$, $\beta_2 = 0.99$). We trained the GAN (batch size of 16, Adam optimizer with learning rate 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.99$) until it observed ten million images, at which point no further improvement for the conditioning loss could be observed.
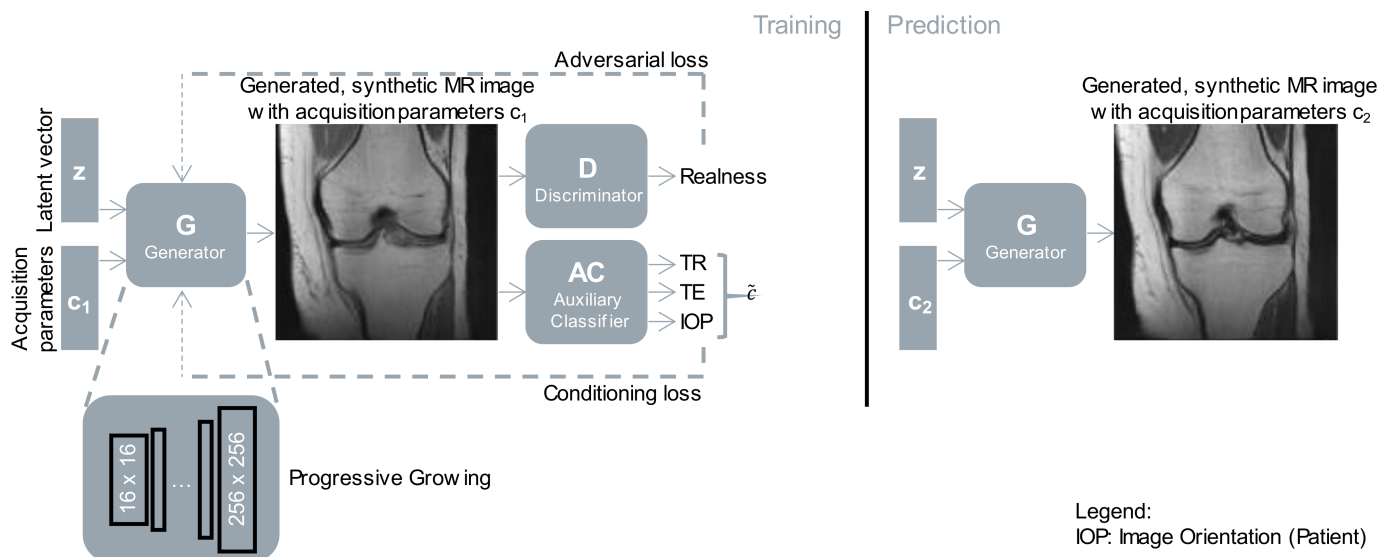
Figure 2. Training and inference phase of our GAN. The generator was trained to synthesize MR images for a given latent vector z and a set of acquisition parameters c_1, guided by two networks, the discriminator, and the auxiliary classifier. After training, a synthetic MR image can be generated for a given latent vector z with any acquisition parameters c_2 (prediction phase). The shown images were generated for a random latent vector with two different sets of acquisition parameters (represented by c_1, c_2), corresponding to coronal imaging orientation, and a TR of 3000 ms, TE of 15 ms, and 45 ms for c_1 and c_2.

4. Results

The evaluation of our model consisted of a qualitative and quantitative component: we evaluated how indifferentiable the synthetic MR images were from real MR images and how well the intended contrast settings (through the acquisition parameters TR and TE) were reflected within the synthetic MR images.

4.1. Qualitative Evaluation

While different measures have been proposed for the assessment of GANs without reference images (i.e., without corresponding ground truth image pairs), a human observer study remains the gold standard for image quality evaluation [18]. The so-called visual Turing tests are commonly used in computer vision and medical imaging to evaluate how indistinguishable generated images are from real ones [19,23,24,40,41]. However,

when asking medical experts to label randomly displayed images as either real or synthetic in a visual Turing test, the experiment may be subject to bias toward labeling more images as synthetic [23,24]. Thus, the reported (accuracy) metrics are also biased and less conclusive. Therefore, we adapted the commonly used, biased experiment by using a grid of images (3 × 2 images) with an equal number of real and synthetic images displayed. The evaluator or expert labels each image as real or synthetic (by either clicking the left or the right mouse button) and must mark the same number of images as real and synthetic within each grid. Additionally, for the images marked as synthetic, the expert can explain why the image appears synthetic. This experiment setup removes potential labeling bias by enforcing the true label distribution for the predicted labels.

We asked two experts to label 150 images (75 synthetic, 75 real images; displayed in random, but the same order for both experts) as either synthetic or real with the experiment mentioned above. The experts had more than 15 years' experience in MRI as an MRI technologist (expert 1) and five years' experience as a radiologist (expert 2). No prior information (e.g., examples of synthetic MR images) was provided before the experiment, and only the MR images and the TR and TE values were shown. The TR and TE values and the imaging orientation of the displayed synthetic images matched the values from the real images. No feedback was provided during the experiment on whether the labeling was correct. The confusion matrix of the visual Turing test is presented in Table 2.

**Table 2.** Confusion matrix of visual Turing test. IRA: inter-reader agreement.

			True Label	
			Real	Synthetic
Predicted label	Real	Expert 1	53	22
		Expert 2	36	39
		IRA	29	10
	Synthetic	Expert 1	22	53
		Expert 2	39	36
		IRA	15	24

The experts reached an accuracy score of 71% (expert 1) and 48% (expert 2) on the identification of real and synthetic MR images. The experts were unable to distinguish a significant share of real and synthetic images correctly, which shows the generator's ability to synthesize MR images indistinguishable from real images with reference to anatomy and contrast. The low inter-reader agreement (IRA) for true positives (29/75 images) and true negatives (24/75 images) shows that only the minority of cases can clearly be identified as either real or synthetic.

According to the experts, image quality impediments of synthetic MR images were mainly attributed to overly smooth tissue (muscles, fat tissue, and bones) compared to fibrous muscle tissue, granular texture in fat tissue, and fine structures in bones of real MR knee images. Additionally, due to the downsampling of several of the real images to a resolution of 256 × 256 pixels using bilinear interpolation to obtain identical image resolution, the image quality of the real images was described as inferior in certain cases, making it hard for the experts to classify. An example of acquisition parameter interpolation is shown in Figure 3, and additional examples of varying anatomy demonstrating the variability of the generated samples are shown in Figure 4. The effect of TE on the tissue contrast can be seen in signal changes of the muscle tissue (Figure 3). Varying TR results in signal differences in the fluid-cartilage contrast [42] and mainly affects contrast on T1-weighted images [43], which are not available within the dataset (see Section 5. Discussion). Therefore, the signal changes through varying TR are less prominent.

4.2. Quantitative Evaluation

The performance of a GAN to generate conditional samples depends on proper guidance during training by a well-trained auxiliary classifier. Therefore, we evaluated different conditional GAN architectures with reference to the classification and regression performance to determine the acquisition parameters on the test dataset (Table 3). Although the mean squared error was used as regression loss for TR and TE, we reported the mean absolute error (MAE), as the MAE is a more descriptive metric for these targets.
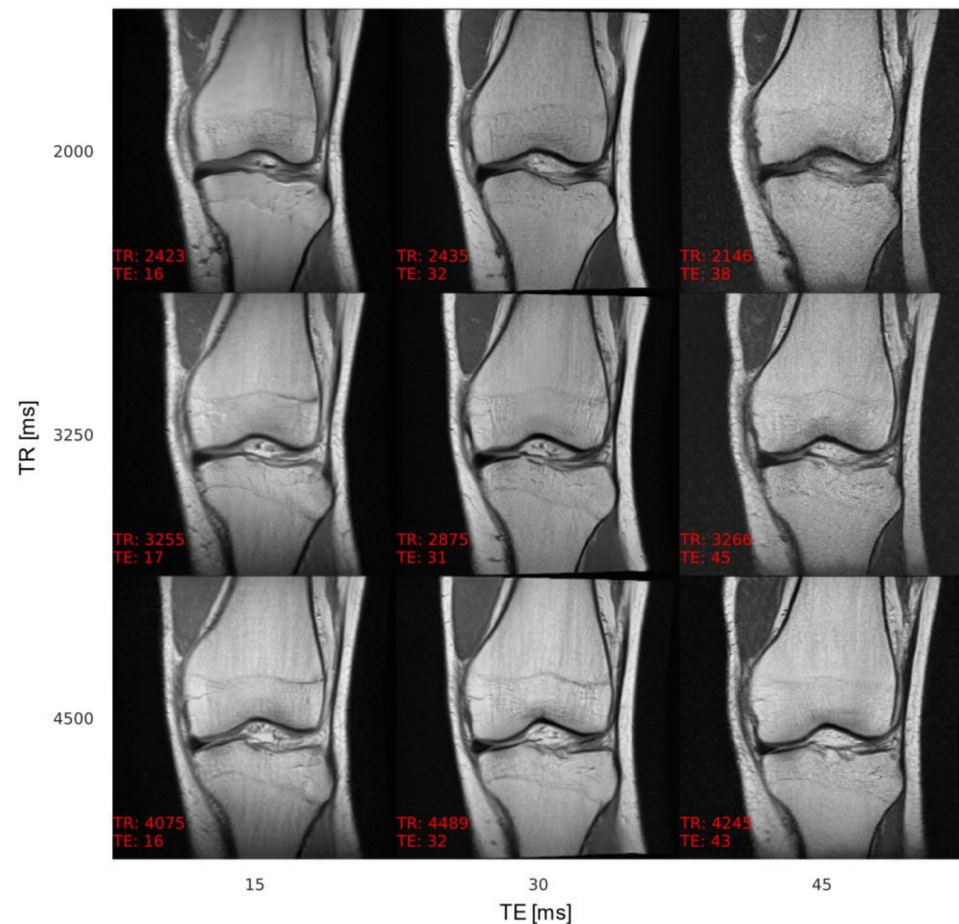
Figure 3. Acquisition parameter interpolation of TR and TE for a single latent vector. A single latent vector was reconstructed with different TR and TE values, showing the capability of the generator to synthesize MR images with adaptable image contrast. The axes describe the intended acquisition parameter values and the values at the bottom left of each image the output of the AC. The images are annotated (in red) with acquisition parameter values as determined by the AC, showing a low overall conditioning error. The contrast adapts properly within images along the axes; however, the anatomy also slightly changes, which is a sign of feature entanglement of the latent vector with the conditions.
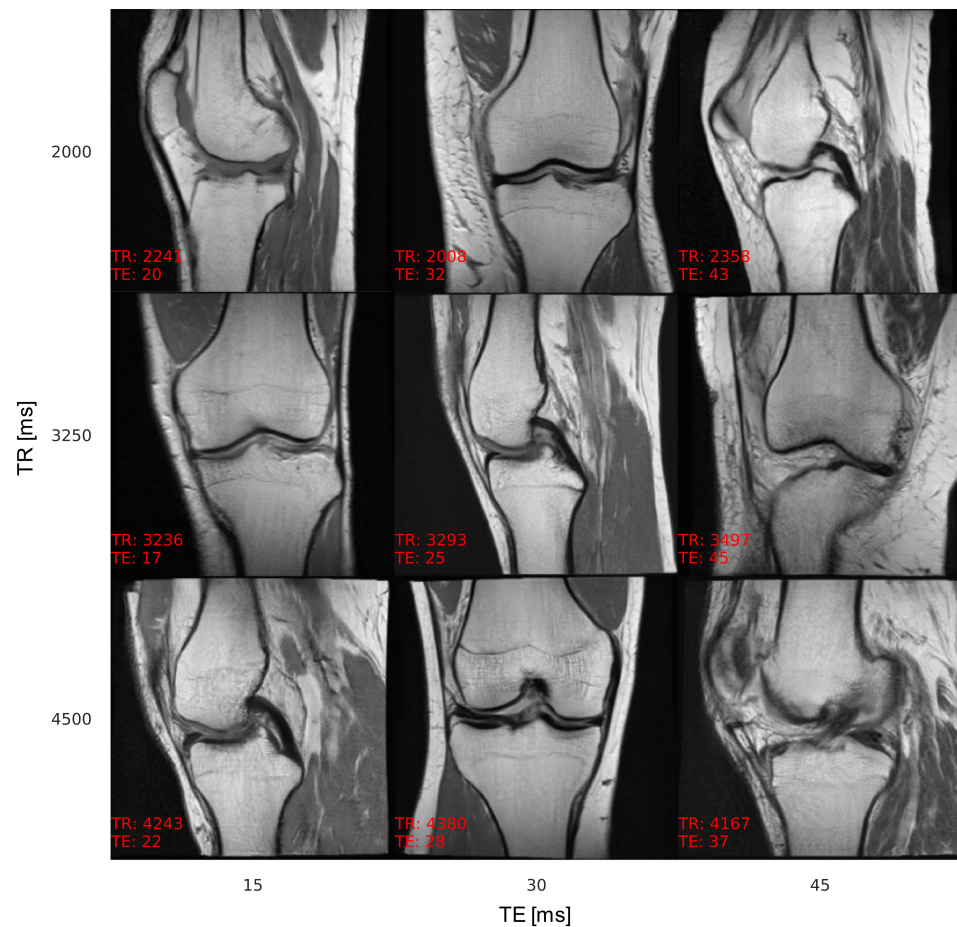
Figure 4. Additional examples of synthetic MR images with varying TR and TE to show the variety of the generated image samples. The imaging orientation alternates between sagittal and coronal. The images are annotated (in red) with acquisition parameter values as determined by the AC, showing a low overall conditioning error.

Table 3. Auxiliary classification performance optimizations.

Model Architecture	Image Orientation	TR	TE
	Accuracy [%]	MAE [ms]	MAE [ms]
ACGAN	63.8	640.0	6.4
Separate AC: DenseNet-121 and HP Tuning	100	239.6	1.6
Synthetic	100	219.4	2.8

With the original ACGAN architecture, the discriminator is trained to differentiate between real and fake samples and the conditions [6]. However, training the same model on the GAN loss and the conditioning loss leads to training instabilities and, therefore, poor performance in generating realistic images and conditioning the image synthesis. Introducing a separate AC model with the DenseNet-121 architecture as the discriminator, which is only trained on the conditions, significantly improves performance, leading to an optimized auxiliary classifier performance. Decoupling the development of the auxiliary classifier from the actual generator and discriminator with a separate network breaks down the complexity of the overall architecture. This facilitates additional data augmentation during training and hyperparameter tuning on the AC, significantly improving the GAN's conditioning performance (Table 3).

The adaptive weighting scheme for the conditioning loss based on ControlGAN yields different benefits. It removes the necessity to manually tune the loss weights for the various conditions. Furthermore, it guides the training process of the GAN properly by focusing the training on the conditioning terms that still need improvement (TR, TE) and decreasing the loss weight for conditions that were already learned sufficiently (imaging orientation). It also balances the overall conditioning learning with the GAN loss (WGAN-GP). Therefore, it prevents overfitting the generator on the conditions and produces realistic MR images. The AC's performance on synthetic data (identical size and label distribution as the test set) and the test set are similar (Table 3), which shows that the generator is neither over- nor underfitting on the conditions.

5. Discussion

Our method generates realistic MR images with high variability in the displayed anatomy that are adjustable in their image contrast through the main contrast acquisition parameter TR and TE. The images are hard to distinguish from real images for medical experts.

Since no method is currently available to retrieve the TR and TE values from an MR image alone, the quantitative evaluation of the correct contrast is difficult. However, the auxiliary classifier performs well to determine the acquisition parameters on unseen test data and can serve as a reliable method to determine the contrast settings. The GAN's capability to adjust to the contrast settings properly (i.e., TR and TE values) mainly depends on the auxiliary classifier's performance. Consequently, the AC's training must be improved with additional training data and hyperparameter tuning in future works to enhance the GAN performance further.

The generator's conditioning on the acquisition parameters was limited to values between 1800 ms to 5000 ms for TR and 12 ms to 50 ms for TE due to training data availability. All sequences within the training data were denoted as PD-weighted (by the DICOM series description). A wider range of TR and TE values within additional training data is anticipated to enable T1- and T2-weighted MR image synthesis. However, this dataset has decisive advantages over multiple publicly available datasets (e.g., [31]) despite these limitations. The dataset comprises DICOM files that includes MR acquisition parameters and offers a clinically realistic distribution of the acquisition parameters values, as shown in Figure 1.

Another issue to address in future work is the latent vector's existing feature entanglement, representing the anatomical features, with the conditioning term representing image contrast (and imaging orientation). When adapting the conditions, the image contrast adjusts properly, but the anatomy also slightly changes (see Figure 3). Feature (dis)-entanglement of generative adversarial networks is a known problem and subject of current research [44].

Applications and Future Work

GANs have proven to be a useful data generation and augmentation tool in the medical data domain with the additional benefit of anonymizing patient identifiable information [45]. Therefore, our method can be used as an advanced data augmentation technique for AI training with MR images that can generate MR images with adaptable image contrast. It enables training data generation tailored to an MRI application's contrast requirements and is also anticipated to increase AI applications' robustness against contrast changes. While AI training with synthetic data quality is not necessarily anticipated to yield the same performance as training on real data, synthetic data has several advantages. Generally, it avoids data privacy issues and can therefore be more easily shared than a proprietary dataset with patient-identifiable information. Moreover, it allows customizable training data generation in terms of dataset size and parameter distribution. Medical image datasets are often imbalanced in terms of acquisition parameter distribution or pathologies. This proof-of-concept is able to solve the acquisition imbalance and is extendable to addi-

tional conditioning (e.g., pathologies). Therefore, it may enhance the use of GAN-based, synthetic data for AI training in MRI.

Furthermore, our work can be used as a training tool for radiologists and technologists to visualize the influence of the acquisition parameters on the image contrast.

Additionally, it can serve as a tool to support protocol configuration by allowing to preview the yielded contrast for a parameterized sequence. Although medical guidelines exist, MR protocol configuration and sequence parameterization are still a matter of radiologists' and technologists' preferences and vary significantly (see Figure 1). A tool to visually support sequence parameterization is anticipated to enhance and simplify the protocol configuration workflow.

However, there are also limitations to point out. The AC's performance mainly limits the generator's ability to accurately produce an MR image with an arbitrary contrast. The use of more MR images with a wider acquisition parameter range for training of the AC is anticipated to boost the AC's performance further and, therefore, of the generator. Moreover, the incorporation of additional acquisition parameters (e.g., flip angle, inversion time, and scan options) and sequence techniques (e.g., gradient echo) can enhance the model's capability for protocoling support, training data generation, and as a training tool for radiology education.

This proof-of-concept has demonstrated the capabilities of GANs to generate customizable, synthetic MR images that are difficult to distinguish from real data. However, the performed visual Turing test must be extended to a higher number of experts in order to estimate the true "indistinguishability" of the synthetic MR images in future work. Moreover, the extension of the proposed approach to 3D is anticipated to enhance the synthesis performance as 3D image stacks offer additional information that can be utilized by a generative network to produce accurate synthetic samples. The extension to other body regions is only limited by the availability of a suitable dataset with a wide range of acquisition parameter settings as well as varying anatomy.

Additionally, more advanced GAN architectures and training procedures have been proposed recently [46] and their application to MRI will be subject to future work.

6. Conclusions

We have proposed a generative adversarial model based on the architecture presented in [13] that uses a separate auxiliary classifier and the adaptive conditioning loss from ControlGAN. The network is trained to generate synthetic MR knee images and is conditioned on the MR acquisition parameters TR, TE, and the image orientation. Our approach allows us to generate synthetic MR images that are difficult to distinguish from real images and adapt the MR image contrast based on the input acquisition parameters (TR and TE). Our MR image synthesis approach can support radiologists' and technologists' training and can adapt to AI-based MR applications' specific requirements by providing fine-tuned, custom MR images with the required contrast.

Author Contributions: Conceptualization, J.D. and J.G.; methodology, J.D.; software, J.D.; validation, J.D. and J.G.; data curation, J.D.; writing—original draft preparation, J.D.; writing—review and editing, J.G., A.M., E.R.; visualization, J.D.; supervision, J.G., A.M., E.R. All authors have read and agreed to the published version of the manuscript.

References

1. American College of Radiology. ACR–SPR–SSR Practice Parameter for the Performance and Interpretation of Magnetic Resonance Imaging (MRI) of the Knee. Available online: https://www.acr.org/-/media/ACR/Files/Practice-Parameters/MR-Knee.pdf?la=en (accessed on 23 November 2020).
2. Glazer, D.I.; DiPiro, P.J.; Shinagare, A.B.; Huang, R.Y.; Wang, A.; Boland, G.W.; Khorasani, R. CT and MRI Protocol Variation and Optimization at an Academic Medical Center. *J. Am. Coll. Radiol.* **2018**, *15*, 1254–1258. [CrossRef]
3. Sachs, P.B.; Hunt, K.; Mansoubi, F.; Borgstede, J. CT and MR Protocol Standardization across a Large Health System: Providing a Consistent Radiologist, Patient, and Referring Provider Experience. *J. Digit. Imaging* **2016**, *30*, 11–16. [CrossRef]
4. Lesjak, Ž.; Galimzianova, A.; Koren, A.; Lukin, M.; Pernuš, F.; Likar, B.; Špiclin, Ž. A Novel Public MR Image Dataset of Multiple Sclerosis Patients with Lesion Segmentations Based on Multi-rater Consensus. *Neuroinformatics* **2017**, *16*, 51–63. [CrossRef]
5. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems, Montréal, QC, Canada, 8–14 December 2014; pp. 2672–2680.
6. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
7. Odena, A.; Olah, C.; Shlens, J. Conditional image synthesis with auxiliary classifier GANs. In Proceedings of the International Conference on Machine Learning (ICML), Sydney, Australia, 6–11 August 2017; pp. 2642–2651.
8. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
9. Reed, S.; Akata, Z.; Yan, X.; Logeswaran, L.; Schiele, B.; Lee, H. Generative adversarial text to image synthesis. *arXiv* **2016**, arXiv:1605.05396.
10. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
11. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved training of wasserstein GANs. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5767–5777.
12. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. *arXiv* **2017**, arXiv:1701.07875.
13. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of GANs for improved quality, stability, and variation. *arXiv* **2018**, arXiv:1710.10196.
14. Hong, Y.; Hwang, U.; Yoo, J.; Yoon, S. How Generative Adversarial Networks and Their Variants Work. *ACM Comput. Surv.* **2019**, *52*, 1–43. [CrossRef]
15. Han, C.; Rundo, L.; Murao, K.; Noguchi, T.; Shimahara, Y.; Milacski, Z.; Koshino, S.; Sala, E.; Nakayama, H.; Satoh, S. MADGAN: Unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction. *BMC Bioinform.* **2021**, *22*, 31. [CrossRef] [PubMed]
16. Nakao, T.; Hanaoka, S.; Nomura, Y.; Murata, M.; Takenaga, T.; Miki, S.; Watadani, T.; Yoshikawa, T.; Hayashi, N.; Abe, O. Unsupervised Deep Anomaly Detection in Chest Radiographs. *J. Digit. Imaging* **2021**, *34*, 418–427. [CrossRef]
17. Kazeminia, S.; Baur, C.; Kuijper, A.; van Ginneken, B.; Navab, N.; Albarqouni, S.; Mukhopadhyay, A. GANs for medical image analysis. *Artif. Intell. Med.* **2020**, *109*, 101938. [CrossRef]
18. Yi, X.; Walia, E.; Babyn, P. Generative adversarial network in medical imaging: A review. *Med. Image Anal.* **2019**, *58*, 101552. [CrossRef]
19. Calimeri, F.; Marzullo, A.; Stamile, C.; Terracina, G. Biomedical data augmentation using generative adversarial neural networks. In Proceedings of the International Conference on Artificial Neural Networks, Alghero, Italy, 11–14 September 2017; pp. 626–634.
20. Denton, E.L.; Chintala, S.; Fergus, R. Deep generative image models using a laplacian pyramid of adversarial networks. In Proceedings of the Advances in Neural Information Processing Systems, Montréal, QC, Canada, 7–12 December 2015; pp. 1486–1494.
21. Zhang, L.; Gooya, A.; Frangi, A.F. Semi-supervised assessment of incomplete LV coverage in cardiac MRI using generative adversarial nets. In *International Workshop on Simulation and Synthesis in Medical Imaging*; Springer: Cham, Switzerland, 2017; pp. 61–68.
22. Han, C.; Hayashi, H.; Rundo, L.; Araki, R.; Shimoda, W.; Muramatsu, S.; Furukawa, Y.; Mauri, G.; Nakayama, H. GAN-based synthetic brain MR image generation. In Proceedings of the IEEE 15th International Symposium on Biomedical Imaging (ISBI), Washington, DC, USA, 4–7 April 2018; pp. 734–738.
23. Han, C.; Murao, K.; Noguchi, T.; Kawata, Y.; Uchiyama, F.; Rundo, L.; Nakayama, H.; Satoh, S. Learning more with less: Conditional PGGAN-based data augmentation for brain metastases detection using highly-rough annotation on MR images. In Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM), Beijing, China, 3–7 November 2019; pp. 119–127.
24. Han, C.; Rundo, L.; Araki, R.; Nagano, Y.; Furukawa, Y.; Mauri, G.; Nakayama, H.; Hayashi, H. Combining Noise-to-Image and Image-to-Image GANs: Brain MR Image Augmentation for Tumor Detection. *IEEE Access* **2019**, *7*, 156966–156977. [CrossRef]
25. Beers, A.; Brown, J.; Chang, K.; Campbell, J.P.; Ostmo, S.; Chiang, M.F.; Kalpathy-Cramer, J. High-resolution medical image synthesis using progressively grown generative adversarial networks. *arXiv* **2018**, arXiv:1805.03144.

26. Bermudez, C.; Plassard, A.J.; Davis, L.T.; Newton, A.T.; Resnick, S.M.; Landman, B.A. Learning implicit brain MRI manifolds with deep learning. In Proceedings of the Medical Imaging 2018: Image Processing, Houston, TX, USA, 10–15 February 2018; p. 105741L.

27. Bowles, C.; Chen, L.; Guerrero, R.; Bentley, P.; Gunn, R.; Hammers, A.; Dickie, D.A.; Hernández, M.V.; Wardlaw, J.; Rueckert, D. GAN augmentation: Augmenting training data using generative adversarial networks. *arXiv* **2018**, arXiv:1810.10863.

28. Frid-Adar, M.; Klang, E.; Amitai, M.; Goldberger, J.; Greenspan, H. Synthetic data augmentation using GAN for improved liver lesion classification. In Proceedings of the IEEE 15th International Symposium on Biomedical Imaging (ISBI), Washington, DC, USA, 4–7 April 2018; pp. 289–293.

29. Madani, A.; Moradi, M.; Karargyris, A.; Syeda-Mahmood, T. Chest X-ray generation and data augmentation for cardiovascular abnormality classification. In Proceedings of the Medical Imaging 2018: Image Processing, Houston, TX, USA, 10–15 February 2018; p. 105741M.

30. Huang, X.; Liu, M.-Y.; Belongie, S.; Kautz, J. Multimodal unsupervised image-to-image translation. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; Lecture Notes in Computer Science; pp. 172–189.

31. Menze, B.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans. Med. Imaging* **2015**, *34*, 1993–2024. [CrossRef]

32. de Coene, B.; Hajnal, J.V.; Gatehouse, P.; Longmore, D.B.; White, S.J.; Oatridge, A.; Pennock, J.M.; Young, I.R.; Bydder, G.M. MR of the brain using fluid-attenuated inversion recovery (FLAIR) pulse sequences. *AJR Am. J. Neuroradiol.* **1992**, *13*, 1555–1564.

33. Shrivastava, A.; Pfister, T.; Tuzel, O.; Susskind, J.; Wang, W.; Webb, R. Learning from simulated and unsupervised images through adversarial training. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2107–2116.

34. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–8 December 2012; pp. 1097–1105.

35. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]

36. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely connected cnvolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.

37. Lee, M.; Seok, J. Controllable Generative Adversarial Network. *IEEE Access* **2019**, *7*, 28158–28169. [CrossRef]

38. Zbontar, J.; Knoll, F.; Sriram, A.; Murrell, T.; Huang, Z.; Muckley, M.J.; Defazio, A.; Stern, R.; Johnson, P.; Bruno, M.; et al. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv* **2018**, arXiv:1811.08839.

39. Kinga, D.; Adam, B.J. A method for stochastic optimization. In Proceedings of the International Conference Learn Represent (ICLR), San Diego, CA, USA, 7–9 May 2015.

40. Kudo, A.; Kitamura, Y.; Li, Y.; Iizuka, S.; Simo-Serra, E. Virtual thin slice: 3D conditional GAN-based super-resolution for CT slice interval. In *Machine Learning for Medical Image Reconstruction*; Springer International Publishing: Cham, Switzerland, 2019; pp. 91–100.

41. Chuquicusma, M.J.M.; Hussein, S.; Burt, J.; Bagci, U. How to fool radiologists with generative adversarial networks? a visual turing test for lung cancer diagnosis. In Proceedings of the IEEE 15th International Symposium on Biomedical Imaging, Washington, DC, USA, 4–7 April 2018.

42. Gold, G.E.; Han, E.; Stainsby, J.; Wright, G.; Brittain, J.; Beaulieu, C. Musculoskeletal MRI at 3.0 T: Relaxation Times and Image Contrast. *Am. J. Roentgenol.* **2004**, *183*, 343–351. [CrossRef] [PubMed]

43. Bitar, R.; Leung, G.; Perng, R.; Tadros, S.; Moody, A.R.; Sarrazin, J.; McGregor, C.; Christakis, M.; Symons, S.; Nelson, A.; et al. MR Pulse Sequences: What Every Radiologist Wants to Know but Is Afraid to Ask. *Radiographics* **2006**, *26*, 513–537. [CrossRef] [PubMed]

44. Karras, T.; Laine, S.; Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–19 June 2019.

45. Yoon, J.; Drumright, L.N.; Van Der Schaar, M. Anonymization Through Data Synthesis Using Generative Adversarial Networks (ADS-GAN). *IEEE J. Biomed. Health Inform.* **2020**, *24*, 2378–2388. [CrossRef]

46. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and Improving the Image Quality of StyleGAN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Online, 14–19 June 2020.