

Research



Cite this article: Lv J, Zhu J, Yang G. 2021 Which GAN? A comparative study of generative adversarial network-based fast MRI reconstruction. *Phil. Trans. R. Soc. A* **379**: 20200203.
<https://doi.org/10.1098/rsta.2020.0203>

Accepted: 14 December 2020

One contribution of 9 to a theme issue 'Synergistic tomographic image reconstruction: part 1'.

Subject Areas:

artificial intelligence, image processing, medical computing, medical physics

Keywords:

magnetic resonance imaging, generative adversarial network, reconstruction, deep learning

Authors for correspondence:

Jun Lv
e-mail: ljdream0710@pku.edu.cn
Guang Yang
e-mail: g.yang@imperial.ac.uk

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5336690>.

Which GAN? A comparative study of generative adversarial network-based fast MRI reconstruction

Jun Lv¹, Jin Zhu² and Guang Yang^{3,4}

¹School of Computer and Control Engineering, Yantai University, Yantai, People's Republic of China

²Department of Computer Science and Technology, University of Cambridge, Cambridge CB3 0FD, UK

³Cardiovascular Research Centre, Royal Brompton Hospital, SW3 6NP London, UK

⁴National Heart and Lung Institute, Imperial College London, London SW7 2AZ, UK

GY, 0000-0001-7344-7733

Fast magnetic resonance imaging (MRI) is crucial for clinical applications that can alleviate motion artefacts and increase patient throughput. *K*-space undersampling is an obvious approach to accelerate MR acquisition. However, undersampling of *k*-space data can result in blurring and aliasing artefacts for the reconstructed images. Recently, several studies have been proposed to use deep learning-based data-driven models for MRI reconstruction and have obtained promising results. However, the comparison of these methods remains limited because the models have not been trained on the same datasets and the validation strategies may be different. The purpose of this work is to conduct a comparative study to investigate the generative adversarial network (GAN)-based models for MRI reconstruction. We reimplemented and benchmarked four widely used GAN-based architectures including DAGAN, ReconGAN, RefineGAN and KIGAN. These four frameworks were trained and tested on brain, knee and liver MRI images using twofold, fourfold and sixfold accelerations, respectively, with a random undersampling mask. Both quantitative evaluations and qualitative visualization have shown that the RefineGAN method has achieved

superior performance in reconstruction with better accuracy and perceptual quality compared to other GAN-based methods.

This article is part of the theme issue 'Synergistic tomographic image reconstruction: part 1'.

1. Introduction

Magnetic resonance imaging (MRI) is among the most important and widely used imaging methods for clinical diagnosis nowadays due to its advantages of non-invasive characteristic and excellent soft-tissue contrast. However, the scan is quite slow, taking several minutes up to more than an hour to acquire clinically serviceable images. Thus, reducing MRI scan time is beneficial for both patient experience and cost considerations. Accordingly, undersampling of MRI data in the Fourier domain is used to improve scanning efficiency but leads to blurring and aliasing artefacts for the reconstructed images.

Consequently, various reconstruction methods have been developed to reduce aliasing artefacts and recover high fidelity images such as parallel imaging [1,2] and compressed sensing (CS) [3]. However, parallel imaging is usually affected by the coil geometry factor (g-factor) [4,5] under higher acceleration factors resulting in a decrease in the signal-to-noise ratio. By contrast, CS algorithms can restore original data from undersampled k -space by exploiting the sparsity of an unknown image in the total variation (TV) [6,7] or wavelet [8] transform domains, and incoherent sampling schemes such as Gaussian random or Poisson disc sampling are usually required. These CS studies focused on designing new objective functions and optimization algorithms, but they ignored the fact that the reconstruction process was computationally inefficient and required a significant amount of iterations to converge.

Recently, deep learning has been widely used in medical image processing, such as disease classification [9–13], anatomic segmentation [14–19], organ registration [20–22] and lesion detection [23–25]. Besides, emerging studies based on deep learning have reported promising results to reconstruct high-quality MRI images from the undersampled acquisition. Wang *et al.* [26] first used the convolutional neural networks (CNN) for MRI reconstruction, which mapped between the undersampled MR images to fully sampled reconstruction. Jin *et al.* [27] proposed a U-net based FBPCovNet to solve the inverse problems in biomedical imaging with residual learning. Schlemper *et al.* [28] introduced a deep network architecture, which formed a cascade of CNNs for reconstructing MR images from undersampled data. Lv *et al.* [29] used the stacked convolutional auto-encoders based strategy to remove streaking artefacts from undersampled free-breathing three-dimensional abdominal MRI. Yang *et al.* [30] reformulated the ADMM algorithm as a deep network for fast CS MRI. This method combined the advantages of a model-based method incorporating domain knowledge and the deep learning approach in parameter learning. Dar *et al.* [31] proposed a transfer learning-based approach for accelerated MRI using deep neural networks. This transfer learning paradigm was first trained on thousands of samples obtained from public datasets consisted of either natural images or MR images. The network was then fine-tuned on a few T1 and T2 MR images to reduce overfitting.

Nowadays, generative adversarial networks (GANs) have shown outstanding performance in modelling the prior distributions of images. In general, a GAN consists of a generator and a discriminator. The generator is used to learn the distribution of a dataset, while the discriminator is used to distinguish the generated data from the real ones. Since the discriminator error propagates back to the generator, the discriminator and generator errors are adversarial, leading to an adversarial loss. The use of adversarial loss improves the quality of perceptual images compared to other loss functions. The GAN can also generate data without explicitly modelling its probability density function. Many related techniques have been explored to preserve the textural information in MRI reconstruction. Shitrit *et al.* [32] proposed a framework that uses a GAN to perform reconstruction directly from the k -space. The generator was used to estimate the whole k -space grid from undersampled data, and the discriminator was adopted to distinguish

between the ‘generated’ images and the real ones. Yang *et al.* [33] developed a novel deep de-aliasing generative adversarial network (DAGAN) for fast CS MRI reconstruction. To preserve details in the reconstruction, the image loss, frequency loss, adversarial loss and perceptual loss were unified. Mardani *et al.* [34] also proposed a CS framework that used a GAN to map the undersampled zero-filled MR images to the fully sampled high-quality MR images. Quan *et al.* [35] employed a cyclic loss with a fully residual convolutional GAN. Their network consisted of two consecutive networks, one for reconstruction (ReconGAN) and one for refining the results (RefineGAN). This proposed model was a variant of the fully residual U-Net and GAN. Shaul *et al.* [36] developed a two-stage GAN framework to estimate the missing k -space data and fix aliasing artefacts in the image-space (KIGAN). Li *et al.* [37] proposed the Structure-Enhanced GAN (SEGAN) to restore structure information of MRI images at both local and global scale. The loss function consisted of MSE in the image domain, structural similarity index measure (SSIM) and patch correlation regularization terms. Murugesan *et al.* [38] developed a novel GAN-based architecture named Reconstruction Global-Local GAN (Recon-GLGAN). This model was composed of a U-Net generator and a context discriminator, which incorporated global and local contextual information from images. Deora *et al.* [39] proposed a novel GAN-based framework by leveraging a combination of patch-based discriminator and SSIM-based loss. This architecture focused on preserving high-frequency content as well as fine textural details in the reconstructed MR images. Guo *et al.* [40] used attention model with recurrent neural networks for enhancing the GAN performance, which was capable of reconstructing more accurate anatomical structures for the MRI data.

Although numerous deep learning studies have been proposed for solving fast MRI, the comparison of these methods remains limited because the models have not been trained on the same datasets and the validation strategies may be different. Ramzi *et al.* [41] have benchmarked different convolutional neural network-based methods for MRI reconstruction. However, to the best of our knowledge, there was no research has been carried out for the comparison of the recent GAN-based methods for MR reconstruction.

In this work, we conduct a comparative and benchmark study that focuses on investigating which GAN-based deep learning models can achieve superior performance in fast MRI reconstruction, including reimplementation and comparison of DAGAN, ReconGAN, RefineGAN and KIGAN. The performance of these GAN-based methods was verified using brain, knee and liver MRI data with a random undersampling factor of $2\times$, $4\times$ and $6\times$, respectively. The MRI reconstruction results were then evaluated using multiple traditional and newly developed quantitative metrics to demonstrate the effectiveness of the compared methods.

2. Method

(a) MRI reconstruction pipeline

To reconstruct undersampled data from an accelerated MR acquisition, the reconstruction process can be expressed as a system of linear equations

$$y = E_u x + n, \quad (2.1)$$

where the MR encoding operator E_u includes the Fourier transform and the undersampling operation, x is the image to be recovered and y is the acquired undersampled k -space data. The image x is thus reconstructed by solving an inverse problem that aims to recover an estimation of x . This inverse problem is ill-posed because it is affected by the limited amount of phase increment steps and the acquisition noise that corrupts the signal.

To tackle this problem, the equation can be reformulated as a regularized optimization procedure, that is

$$\min_x \frac{1}{2} \|E_u x - y\|_2^2 + \lambda R(x), \quad (2.2)$$

where $R(x)$ is a regularization term to stabilize the solution, and a data consistency term $(1/2)\|E_u x - y\|_2^2 < \epsilon$, where ϵ is the noise level. The weighting parameter λ controls the degree of regularization and needs to be chosen according to the noise level of the acquired data.

For CS MRI reconstruction, it is assumed that the k -space data are pseudo-randomly undersampled, the image admits a sparse representation in some transform domain Ψ , and a nonlinear optimization is performed to enforce the data consistency and sparsity of the MR image in the transform domain. For example, Ψ can be chosen as the Fourier transform, TV or wavelets. Thus, the problem illustrated in equation (2.2) can be rewritten as follows:

$$\min_x \frac{1}{2} \|E_u x - y\|_2^2 + \lambda \|\Psi x\|_p, \quad (2.3)$$

where p represents the L_p norm. This equation can be solved using convex optimization techniques. However, there are still two major drawbacks of CS MRI reconstruction. Firstly, it needs considerable computational time to solve the problem. Then, it depends heavily on empirical tuning of the regularization parameters and suitable transform basis Ψ . An inappropriate choice of reconstruction parameters can lead to over-smoothing or images with residual undersampling artefacts.

To solve equation (2.3) more efficiently, deep learning-based approaches are introduced to reconstruct the MRI images. In general, these deep learning-based methods can be formulated as an optimization problem

$$\min_x \frac{1}{2} \|E_u x - y\|_2^2 + \lambda \|x - F_{\text{CNN}}(x_u | \phi)\|_2^2, \quad (2.4)$$

where $x_u = F_u^H y$ is the zero-filled reconstruction, $F_{\text{CNN}}(x_u | \phi)$ is an image generated by the CNN network and ϕ represents the optimal parameters of the trained CNN.

When the CNN model is replaced by the GAN, the objective is to train a generator G that can generate an image $G_{\theta_G}(x_u)$ from the zero-filled reconstruction image x_u under the constraint that $G_{\theta_G}(x_u)$ is indistinguishable from the ground truth image x_t .

The objective function of D is used to maximize the log-likelihood for estimating the conditional probability, where $D(G_{\theta_G}(x_u)) = 0$ and $D(x_t) = 1$. Hence, this can be addressed by defining an adversarial loss L_{adv} , which can be considered as a minimax problem between the generator $G_{\theta_G}(x)$ and the discriminator $D_{\theta_D}(x)$. The training process of the GAN can be parametrized by θ_G and θ_D as follows:

$$\min_{\theta_G} \max_{\theta_D} L_{\text{adv}}(\theta_D, \theta_G) = E_{x_t \sim P_{\text{train}}(x_t)} [\log D_{\theta_D}(x_t)] + E_{x_u \sim P_G(x_u)} [\log (1 - D_{\theta_D}(G_{\theta_G}(x_u)))]. \quad (2.5)$$

Here, x_u is sampled from a fixed latent distribution $P_G(x_u)$ and real samples x_t come from a real data distribution $P_{\text{train}}(x_t)$. Once the training coverages, G_{θ_G} can generate the image $G_{\theta_G}(x_u)$, which is similar to x_t , and D_{θ_D} is unable to differentiate between them.

(b) Benchmarked GAN-based fast MRI methods

(i) DAGAN

Yang *et al.* [33] proposed a deep de-aliasing generative adversarial network, named as DAGAN, for the fast MRI reconstruction. The framework of DAGAN is shown in figure 1a. As can be observed, they used a U-Net-based framework to construct the generator G (figure 1b), in which the encoder path contained eight convolutional layers and the decoder path included corresponding eight deconvolutional layers, and each of them was followed by the batch normalization (BN) and leaky ReLU (LReLU) layers. Moreover, skip connections (purple lines) were used to connect the corresponding layers of the encoding and decoding path, which could pass features of different scales to the decoding layer in order to obtain better reconstruction details. In addition, the discriminator D (figure 1c) was applied to differentiate the de-aliased reconstruction \hat{x}_u from the fully sampled ground truth image x_t . It consisted of 11 convolutional layers, and each was also followed by the BN and LReLU layers. Finally, a fully connected (FC)

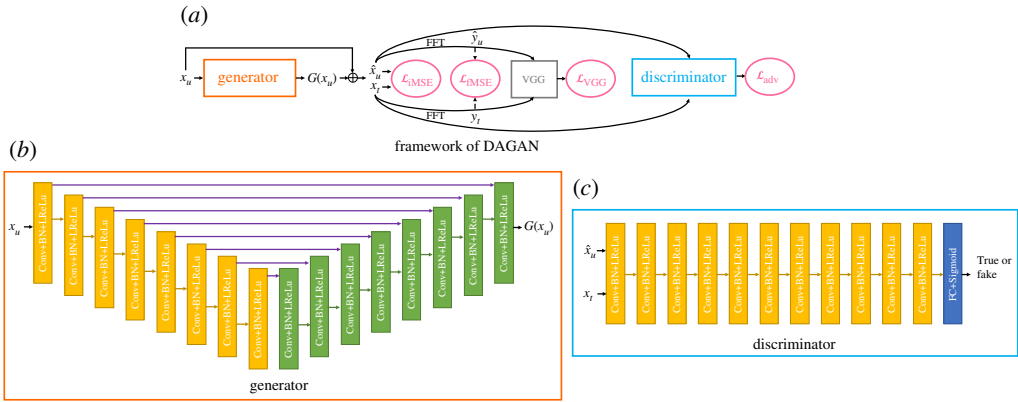


Figure 1. Schema of the DAGAN for the fast MRI reconstruction. (Online version in colour.)

layer was connected, and the classification result was output through the sigmoid activation function. To stabilize the training process, $\hat{x}_u = G_\theta(x_u) + x_u$ was proposed to replace $\hat{x}_u = G_\theta(x_u)$.

In DAGAN, a content loss was applied to improve the perceptual quality of the reconstructed image. The loss consisted of three parts, a pixel-wise image domain mean square error (MSE) loss, a frequency domain MSE loss and a perceptual VGG loss. The VGG loss was computed by the VGG network [42]. In DAGAN, the VGG network was pre-trained on ImageNet [43] and the Conv4 output of the VGG16 was used as the encoded path of the de-aliased image and the ground truth. These loss functions can be defined as following:

$$\min_{\theta_G} L_{iMSE}(\theta_G) = \frac{1}{2} \|x_t - \hat{x}_u\|_2^2, \quad (2.6)$$

$$\min_{\theta_G} L_{fMSE}(\theta_G) = \frac{1}{2} \|y_t - \hat{y}_u\|_2^2 \quad (2.7)$$

$$\text{and} \quad \min_{\theta_G} L_{VGG}(\theta_G) = \frac{1}{2} \|f_{VGG}(x_t) - f_{VGG}(\hat{x}_u)\|_2^2. \quad (2.8)$$

The y_t and \hat{y}_u is the corresponding frequency domain image of x_t and \hat{x}_u . Meanwhile, the adversarial loss of the generator can be defined as

$$\min_{\theta_G} L_{GEN}(\theta_G) = -\log(D_{\theta_D}(G_{\theta_G}(x_u))). \quad (2.9)$$

Thus, the overall loss function can be defined as

$$L_{TOTAL} = \alpha L_{iMSE} + \beta L_{fMSE} + \gamma L_{VGG} + L_{GEN}, \quad (2.10)$$

where α , β and γ are the weights, which help to control the balance between each loss terms. The hyperparameters were optimized and set to $\alpha = 15$, $\beta = 0.1$, $\gamma = 0.0025$ according to the original DAGAN paper.

(ii) ReconGAN and RefineGAN

Quan *et al.* [35] proposed an MRI reconstruction method using a GAN with a cyclic loss. Figure 2a illustrates the overall structure of ReconGAN and RefineGAN. As can be seen, the generator G included twofold networks, in which the first checkpoint was referred to as ReconGAN and the second checkpoint was represented as RefineGAN, the corresponding output was \hat{x}_u and \hat{x}_u , respectively. This entire framework was built based on the architecture of a convolutional autoencoder, which consisted of the encoder and decoder paths. In the encoder path, there were four encoder blocks with the insertion of the residual block (figure 2b). Each encoder block included a convolutional layer, a residual block and another convolutional layer. In the decoder

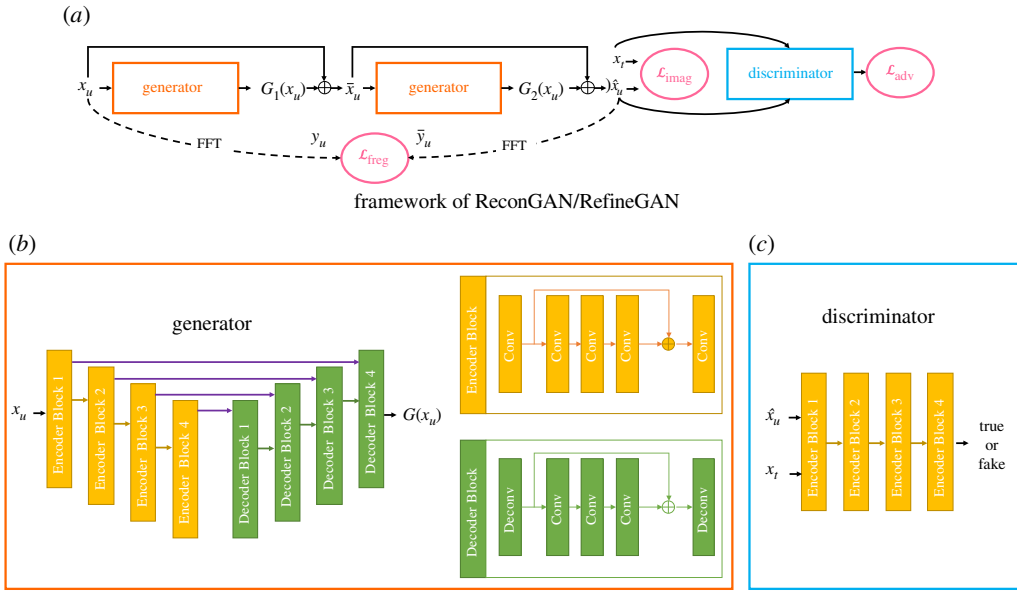


Figure 2. Schema of the ReconGAN/RefineGAN for the fast MRI reconstruction. (Online version in colour.)

path, there were four decoder blocks, and each of them consisted of a deconvolutional layer, a residual block and another deconvolutional layer. The residual block was composed of three convolutional layers, while the number of feature maps of the second convolutional layer was half of the first and the last one. The residual block was designed to increase the depth of generator G and discriminator D networks. In addition, the generator G was trained to generate the inverse amplitude of the noise results from the undersampled data instead of reconstructing the images directly. Thus, the final reconstruction was acquired by adding the zero-filled image to the output from the generator G . For the Discriminator D , as shown in figure 2c, it used a similar architecture of the encoding path of the generator G .

In RefineGAN, the cyclic data consistency loss was a combination of undersampled frequency loss and the fully reconstructed image loss. Thus, L_{cyc} can be defined as

$$L_{\text{cyc}}(G) = L_{\text{freq}}(G) + L_{\text{imag}}(G), \quad (2.11)$$

$$L_{\text{freq}}(G) = d(\hat{y}_u[i], y_t[i]) \quad (2.12)$$

and

$$L_{\text{imag}}(G) = d(\hat{x}_u[i], x_t[i]). \quad (2.13)$$

It is worth noting that L_{cyc} only affects the generator G and not the discriminator D . The adversarial loss can be defined as

$$L_{\text{adv}}(G, D) = E_{m \in M} [1 - \log D(G(x_u))] + E_{s \in S} [\log D(x_t)]. \quad (2.14)$$

These loss functions were used to map the undersampled data $x_u[n]$ to the fully-sampled image $x_t[n]$. The term L_{freq} guaranteed that the difference between $y_t[i]$ and $\hat{y}_u[i]$ was minimal when we undersampled the reconstructed image $\hat{x}_u[j]$ with the Fourier transform operation to obtain $\hat{y}_u[i]$. Besides, L_{imag} ensured that for any fully sampled image $x_t[j]$, when $x_t[j]$ had undergone a subsampling operation RF and the input to the generator G was $x_u[n]$ to obtain a reconstructed image $\hat{x}_u[j]$, then $x_t[j]$ and $\hat{x}_u[j]$ should be similar.

Thus, the overall loss function can be defined as

$$L_{\text{TOTAL}} = \alpha L_{\text{imag}}(G) + \beta L_{\text{freq}}(G) + L_{\text{adv}}(G, D), \quad (2.15)$$

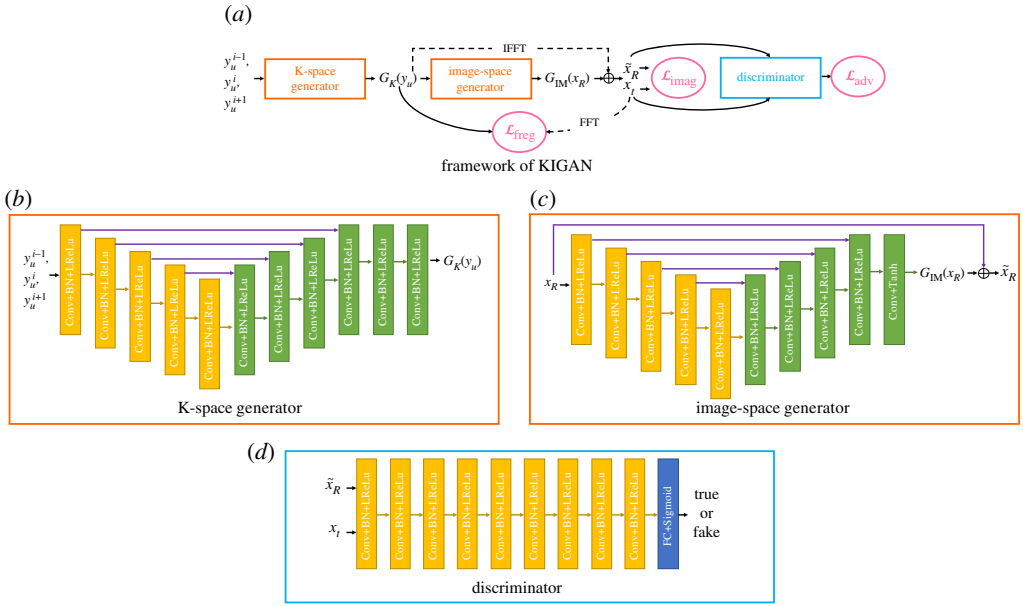


Figure 3. Schema of the KIGAN for the fast MRI reconstruction. (Online version in colour.)

where α and β are the weights that help to control the balance between each loss terms. We set $\alpha = 10.0$ and $\beta = 1.0$ according to the original paper.

(iii) KIGAN

Shaul *et al.* [36] used a cascade of a k -space (G_K) and an image-space (G_{IM}) U-Net architecture for the MRI reconstruction from undersampled data. The entire network structure is depicted in figure 3. The two generators had the same architecture. As shown in figure 3*b,c*, both G_K and G_{IM} networks were composed of five convolutional layers in the encoding path and corresponding five deconvolutional layers in the decoding path, and each layer was followed by BN and leaky-ReLU activation functions. The only difference was that the last layer of G_{IM} needed to be refined by adding the input of the G_{IM} . As indicated in figure 3*d*, the discriminator D consisted of nine convolutional layers and each layer was followed by BN and leaky ReLU activation functions and a final FC layer.

As shown in figure 3*a*, G_K was used to reconstruct the missing k -space data, then the output of G_K was subjected to an inverse Fourier transform, and finally G_{IM} was used to correct the output of G_K in the image domain. Thus, the k -space loss function can be defined as follows:

$$L_K(\theta) = \|G_K(y_u; \theta) - y_t\|_2, \quad (2.16)$$

in which $G_K(y_u; \theta)$ is the output of the generator G_K , which represents the reconstructed data in the k -space, y_t represents the entire k -space samples. It is defined that u is the under-sampling mask and $\bar{u} = \mathbf{1}_u - u$, in which $\mathbf{1}_u$ is an all-ones metric of the same size as u . As shown in figure 3*a*, the input of the G_K is three adjacent slices named as y_u^{i-1} , y_u^i and y_u^{i+1} . In addition, the noise is used as a substitute for the missing data. Then the un-masked k -space data can be obtained by multiplying the output of k -space generator G_K and \bar{u} . Next, the new reconstructed k -space y_R can be acquired by combining the input samples and un-masked k -space data

$$y_R = y_u + \bar{u} \odot G_K(y_u; \theta). \quad (2.17)$$

The input of the image-space generator is $x_R = |F^H y_R|$, the output of G_{IM} is $\tilde{x}_R = G_{IM}(|F^H y_R|; \varphi)$, where H is the Hermitian transpose operation and F^H is an orthonormal 2D IFT operator, where

$F^H F = I$. Thus, the image-space loss function can be defined as the L_p norm between the ground truth x_t and the reconstructed image \tilde{x}_R , as follows:

$$\begin{aligned} L_{\text{IM}}(\theta, \varphi) &= \|x_t - \tilde{x}_R\|_p \\ &= \|F^H y_t - G_{\text{IM}}(F^H(y_u + \tilde{u} \odot G_K(y_u; \theta)); \varphi)\|_p. \end{aligned} \quad (2.18)$$

Besides, the adversarial loss function is formulated as follows:

$$\begin{aligned} L_{\text{adv}}(\theta, \varphi) &= E_{Z \sim P_u(z)} \log[1 - D(\tilde{x}_R)] \\ &= E_{Z \sim P_u(z)} \log\left[1 - D\left(G_{\text{IM}}\left(F^H(y_u + \tilde{u} \odot G_K(y_u; \theta)); \varphi\right)\right)\right]. \end{aligned} \quad (2.19)$$

(c) MR datasets

In this study, we benchmarked and compared four GAN-based methods using two widely used open access datasets, including (1) Calgary–Campinas brain MRI dataset, (2) Stanford knee MRI dataset and (3) a private liver MRI dataset.

- (1) Calgary–Campinas brain MR raw data included 35 fully sampled T1-weighted MR datasets acquired from a clinical 3T MR scanner (Discovery MR750; General Electric Healthcare, Waukesha, WI). Acquisition parameters were repetition time = 6.3 ms, echo time = 2.6 ms, inversion time = 650 ms with 100 contiguous 1.0 mm slices. The data were acquired using a 12-channel imaging coil, and all the multi-coil k -space data were combined using the vendor-supplied tools (Orchestra Toolbox; General Electric Healthcare). The matrix size was 256×256 . Thus, the brain training dataset had 2500 slices extracted from 25 subjects. The brain test dataset had 1000 slices corresponding to the remaining 10 subjects.
- (2) Stanford knee database consisted of 20 subjects and each scan contained 8-channel signal. The images were fully sampled using a 3D FSE CUBE sequence with proton density weighting on a 3T General Electric MR scanner using the following parameters, including repetition time = 25 ms, echo time = 1550 ms, slice thickness = 0.6 mm, field of view = 160 mm. The matrix size of the data were $320 \times 320 \times 256$. For data preparation, k -space data from multi-channels were first compressed into single-channel data using the coil-compression approach [44]. Thus, the knee training dataset had 1800 slices extracted from 18 subjects. The knee test dataset had 200 slices corresponding to the remaining two subjects.
- (3) A private liver MRI datasets were obtained from 106 subjects. The images were fully sampled using T2-weighted turbo spin echo pulse sequence on Philips 3T MRI scanner with the following parameters, including repetition time = 2750 ms, echo time = 78 ms, slice thickness = 6 mm. The matrix size of the data were $430 \times 430 \times 40$. We randomly selected 80 subjects for training and 26 subjects for test that contained 3200 and 1040 two-dimensional images, separately.

For network training, the k -space data were down-sampled retrospectively along the phase-encoding direction with various acceleration factors ($R=2\times$, $4\times$ and $6\times$) using the one-dimensional Gaussian random undersampling operations.

(d) Quantitative evaluation metrics

To quantify the reconstruction error, conventionally, peak signal-to-noise ratio (PSNR), SSIM, normalized mean square error (NMSE), Visual Information Fidelity (VIF) [45] and Frechet Inception Distance (FID) score [46] were measured between the reconstructed images and the fully-sampled images (ground truth) to evaluate the reconstruction quality by each method. Lower values of NMSE and FID and higher values of SSIM, PSNR and VIF would indicate better

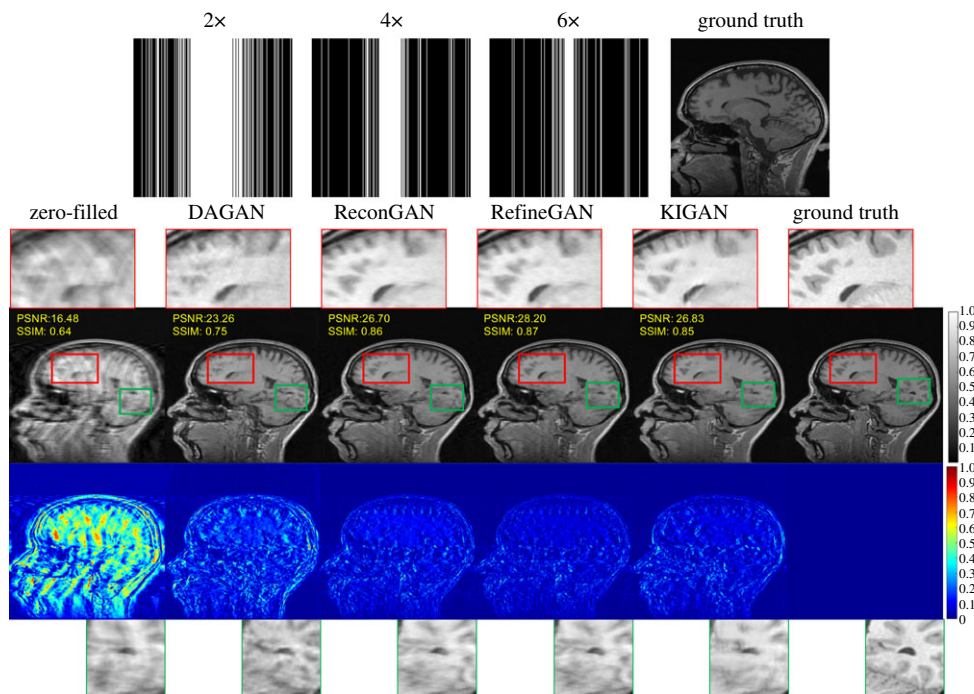


Figure 4. Reconstruction results of an example brain MRI dataset by various methods at sixfold undersampling. Red and Green boxes illustrate the zoomed-in areas. From left to right: zero-filled, DAGAN, ReconGAN, RefineGAN, KIGAN results and the ground truth. (Online version in colour.)

reconstruction performance. The statistical significance was assessed using the paired Wilcoxon signed-rank test with an alpha of 0.05.

3. Results

The visual comparisons with zoomed-in areas of the reconstructed images are shown in figure 4. The sampling masks and the ground truth images are shown in the top row. It can be seen that the results obtained by DAGAN suffered severe blurring, and KIGAN results still existed some residual artefacts. Although ReconGAN showed a significant reduction of aliasing artefacts, the red and green boxes indicated that the loss of structural details still existed in some areas. The RefineGAN reduced the blurring artefacts and provided relatively sharper edges. Moreover, electronic supplementary material, figure S1 records the visual comparison of various algorithms at 2 \times , 4 \times and 6 \times undersampling, respectively. Table 1 lists the average quantitative results of the brain MRI datasets. RefineGAN outperformed the competing algorithms with significantly higher PSNR and SSIM results and significantly lower NMSE and FID values ($p < 0.05$). On the other hand, with different acceleration factors, KIGAN always obtained the highest VIF values. When $R = 2$, Wilcoxon rank-sum test showed no significant difference between DAGAN and KIGAN ($p = 0.59$) in terms of PSNR and NMSE. Besides, there was no remarkable difference between ReconGAN and RefineGAN for the VIF values regardless of the acceleration factors.

The zoomed-in visual comparisons of an example knee dataset reconstruction are shown in figure 5. We can observe that RefineGAN produced sharper images compared to other GAN-based models, and it preserved more details. Moreover, we presented the zoomed-in areas and the corresponding error maps to reveal the blood vessel and structure details preserved by each method. As can be observed in the zoomed-in areas enclosed by the red boxes, KIGAN performed the worst with obvious residual artefacts. The reconstruction errors were nearly

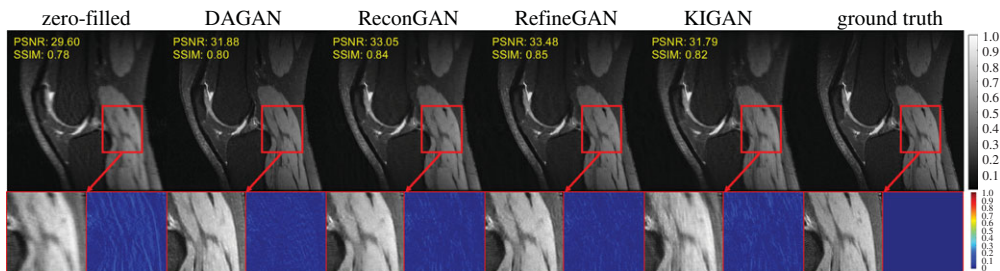


Figure 5. Reconstruction results of an example knee MRI dataset using various methods at sixfold undersampling rate. Red boxes illustrate the zoomed-in areas and the corresponding error maps. From left to right: zero-filled, DAGAN, ReconGAN, RefineGAN, KIGAN results and the ground truth. (Online version in colour.)

Table 1. Quantitative evaluation comparison in terms of PSNR (mean \pm s.d.), SSIM (mean \pm s.d.), NMSE (mean \pm s.d.), VIF (mean \pm s.d.) and FID (mean) for brain MRI reconstructions using different acceleration factors. The best results of each metric are highlighted in bold.

<i>R</i>	metric	zero-filled	DAGAN	ReconGAN	RefineGAN	KIGAN
2 \times	PSNR	30.94 \pm 2.75	33.79 \pm 1.88	39.08 \pm 1.34	39.40 \pm 1.33	33.90 \pm 2.55
	SSIM	0.92 \pm 0.02	0.93 \pm 0.01	0.97 \pm 0.00	0.97 \pm 0.00	0.96 \pm 0.01
	NMSE ($\times 10^{-2}$)	1.57 \pm 0.01	0.72 \pm 0.43	0.20 \pm 0.09	0.19 \pm 0.08	0.78 \pm 0.53
	VIF	—	0.99 \pm 0.08	1.00 \pm 0.03	1.00 \pm 0.03	1.12 \pm 0.07
	FID	—	31.42	8.41	7.44	13.09
4 \times	PSNR	23.69 \pm 3.02	28.76 \pm 1.95	32.07 \pm 1.65	32.67 \pm 1.56	28.14 \pm 1.84
	SSIM	0.79 \pm 0.03	0.86 \pm 0.02	0.92 \pm 0.01	0.93 \pm 0.01	0.88 \pm 0.02
	NMSE ($\times 10^{-2}$)	8.86 \pm 7.01	2.35 \pm 1.50	1.05 \pm 0.54	0.91 \pm 0.45	2.67 \pm 1.4
	VIF	—	0.92 \pm 0.11	0.97 \pm 0.07	0.97 \pm 0.06	1.09 \pm 0.09
	FID	—	47.02	28.05	24.98	80.79
6 \times	PSNR	19.47 \pm 2.31	25.4 \pm 1.57	29.23 \pm 1.68	29.95 \pm 1.61	27.91 \pm 1.57
	SSIM	0.66 \pm 0.04	0.77 \pm 0.03	0.88 \pm 0.02	0.89 \pm 0.02	0.86 \pm 0.02
	NMSE ($\times 10^{-2}$)	21.3 \pm 14.1	4.89 \pm 2.33	2.03 \pm 1.00	1.71 \pm 0.83	2.75 \pm 1.31
	VIF	—	0.87 \pm 0.12	0.93 \pm 0.07	0.94 \pm 0.07	0.97 \pm 0.08
	FID	—	79.91	46.86	43.97	52.91

identical in DAGAN and ReconGAN. Besides, electronic supplementary material, figure S2 shows the reconstruction of a typical knee image. Table 2 exhibits the quantitative results of our knee MRI reconstruction. It can be observed that RefineGAN achieved the best values for all metrics at the sixfold undersampling. When $R = 2$, Wilcoxon rank-sum test showed no significant differences between DAGAN and KIGAN in terms of PSNR and NMSE ($p = 0.17$), which was also consistent with the statistical results for our brain MRI reconstruction. When $R = 4$ and $R = 6$, Wilcoxon rank-sum test showed that there only existed significant differences between ReconGAN and KIGAN in terms of FID ($p < 0.05$).

Results in figure 6 illustrate detailed reconstruction results from an example of liver MRI dataset. Zero-Filled approach reconstructed images with a significant amount of aliasing artefacts. Similarly, there were obvious blurring artefacts and obscure blood vessels that existed in the results obtained by DAGAN. ReconGAN not only suppressed aliasing artefacts but also

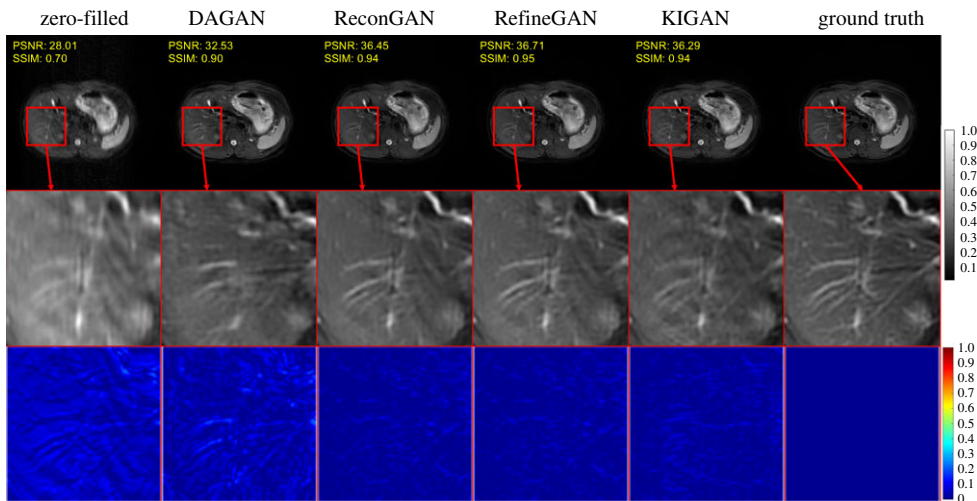


Figure 6. Reconstruction results of an example liver MRI dataset using various methods at sixfold undersampling rate. Red boxes illustrate the zoomed-in areas and the corresponding error maps. From left to right: zero-filled, DAGAN, ReconGAN, RefineGAN, KIGAN results and the ground truth. (Online version in colour.)

Table 2. Quantitative evaluation comparison in terms of PSNR (mean \pm s.d.), SSIM (mean \pm s.d.), NMSE (mean \pm s.d.), VIF (mean \pm s.d.) and FID (mean) for knee MRI reconstructions using different acceleration factors. The best results of each metric are highlighted in bold.

<i>R</i>	metric	zero-filled	DAGAN	ReconGAN	RefineGAN	KIGAN
2 \times	PSNR	34.66 \pm 2.98	38.91 \pm 1.59	42.37 \pm 1.60	42.41 \pm 1.98	38.53 \pm 2.51
	SSIM	0.95 \pm 0.01	0.94 \pm 0.01	0.97 \pm 0.00	0.98 \pm 0.00	0.96 \pm 0.01
	NMSE ($\times 10^{-2}$)	1.64 \pm 1.32	0.52 \pm 0.31	0.23 \pm 0.09	0.24 \pm 0.16	0.83 \pm 1.03
	VIF	—	0.92 \pm 0.06	0.96 \pm 0.03	0.97 \pm 0.02	1.05 \pm 0.11
	FID	—	21.22	9.96	7.95	22.63
4 \times	PSNR	27.31 \pm 3.23	34.35 \pm 1.77	34.88 \pm 1.96	35.58 \pm 1.74	34.70 \pm 1.74
	SSIM	0.84 \pm 0.02	0.86 \pm 0.02	0.90 \pm 0.02	0.91 \pm 0.02	0.89 \pm 0.02
	NMSE ($\times 10^{-2}$)	10.30 \pm 11.00	1.55 \pm 1.20	1.37 \pm 0.93	1.14 \pm 0.66	1.49 \pm 1.27
	VIF	—	0.79 \pm 0.11	0.85 \pm 0.08	0.86 \pm 0.07	0.87 \pm 0.12
	FID	—	59.6	65.25	47.81	99.5
6 \times	PSNR	25.15 \pm 3.37	32.67 \pm 1.89	32.34 \pm 2.16	33.36 \pm 1.81	30.83 \pm 2.09
	SSIM	0.79 \pm 0.03	0.82 \pm 0.03	0.86 \pm 0.02	0.87 \pm 0.02	0.84 \pm 0.02
	NMSE ($\times 10^{-2}$)	18.2 \pm 22.3	2.36 \pm 2.09	2.62 \pm 2.19	2.00 \pm 1.52	4.08 \pm 4.29
	VIF	—	0.75 \pm 0.13	0.77 \pm 0.11	0.80 \pm 0.09	0.79 \pm 0.17
	FID	—	86.07	107.66	78.83	163.2

provided sharper edges and more realistic textures. RefineGAN produced the most visually pleasing results. It provided a better reconstruction of blood vessels and finer textures compared to other competing methods. For the reconstructions produced by KIGAN, residual artefacts still existed. Moreover, electronic supplementary material, figure S3 records the reconstruction results for a specific slice of the liver MRI dataset. Table 3 shows the compared quantitative

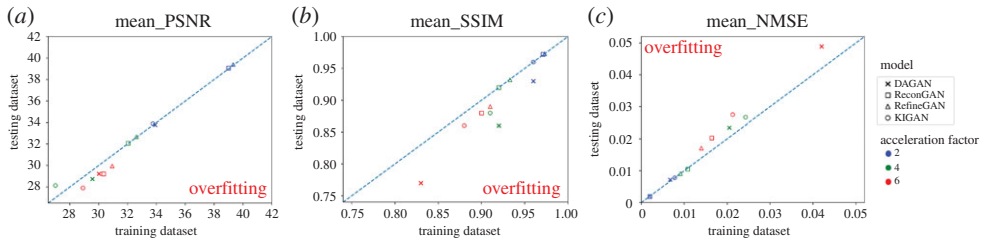


Figure 7. Each point in this figure represents the performance of a specific architecture on the training and testing brain MRI datasets. The blue dashed line indicates the points that are neither overfitting nor underfitting. (Online version in colour.)

Table 3. Quantitative evaluation comparison in terms of PSNR (mean \pm s.d.), SSIM (mean \pm s.d.), NMSE (mean \pm s.d.), VIF (mean \pm s.d.) and FID (mean) for liver MRI reconstructions using different acceleration factors. The best results of each metric are highlighted in bold.

R	metric	zero-filled	DAGAN	ReconGAN	RefineGAN	KIGAN
$2\times$	PSNR	33.49 ± 3.28	37.39 ± 2.26	44.57 ± 4.47	45.01 ± 4.59	44.38 ± 4.17
	SSIM	0.89 ± 0.04	0.95 ± 0.03	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.01
	NMSE ($\times 10^{-2}$)	1.43 ± 1.28	1.10 ± 0.39	0.68 ± 0.37	0.59 ± 0.31	0.65 ± 0.40
	VIF	—	0.74 ± 0.06	0.98 ± 0.02	0.99 ± 0.02	1.07 ± 0.07
	FID	—	23.15	10.60	7.50	8.82
$4\times$	PSNR	32.28 ± 3.27	35.99 ± 2.18	37.86 ± 3.43	38.09 ± 3.47	37.46 ± 2.87
	SSIM	0.78 ± 0.05	0.94 ± 0.03	0.95 ± 0.02	0.96 ± 0.02	0.96 ± 0.03
	NMSE ($\times 10^{-2}$)	5.75 ± 2.67	1.89 ± 1.40	2.65 ± 1.71	2.12 ± 1.54	2.50 ± 0.99
	VIF	—	0.68 ± 0.07	0.91 ± 0.06	0.93 ± 0.05	1.05 ± 0.10
	FID	—	35.09	35.96	31.94	37.42
$6\times$	PSNR	28.86 ± 3.14	34.63 ± 2.01	35.86 ± 3.37	36.25 ± 3.39	35.34 ± 2.68
	SSIM	0.70 ± 0.06	0.92 ± 0.03	0.94 ± 0.03	0.95 ± 0.02	0.94 ± 0.02
	NMSE ($\times 10^{-2}$)	11.79 ± 10.84	2.54 ± 1.7	3.15 ± 2.48	2.87 ± 2.26	3.38 ± 1.46
	VIF	—	0.64 ± 0.09	0.84 ± 0.07	0.86 ± 0.07	0.97 ± 0.14
	FID	—	67.81	49.58	48.69	68.87

performance for the liver MRI datasets using different acceleration factors. It can be observed that RefineGAN algorithm still performed better compared to other GAN-based algorithms. When $R = 4$, Wilcoxon rank-sum test showed no significant differences between ReconGAN and RefineGAN ($p > 0.37$) in terms of PSNR and NMSE.

In order to investigate the model reliability, we plot the quantified results for the training versus testing datasets for our brain MRI study. As shown in figure 7, each point indicates the performance of a particular model on the train and test datasets used in this study. The dashed line represents the points where there are neither overfitting nor underfitting. When $R = 2$, consistency between the training set and testing set was high. When $R = 4$ and $R = 6$, both DAGAN and KIGAN models tended to be overfitted.

4. Discussion

In this study, we conducted a comparative study for the GAN-based models in solving MRI reconstruction problem. Four GAN-based network architectures were reimplemented and

compared, including DAGAN, ReconGAN, RefineGAN and KIGAN. The comparisons were tested on three different but representative clinical datasets, i.e. brain, knee and liver MRI datasets, using random undersampling masks.

The reasons we have chosen these four GAN-based methods for comparison are summarized as follows: DAGAN, ReconGAN and RefineGAN are three typical single-domain networks, while KIGAN is a novel representative cross-domain network. More importantly, all four methods have open-source implementations, which have been widely validated on various MRI datasets, and therefore ensured the robustness and reproducibility.

For the brain and knee datasets, KIGAN has underperformed ReconGAN and RefineGAN. However, as for the liver dataset, the differences between KIGAN and RefineGAN has not been found significant. This might be attributed to the fact that the thickness of the liver dataset was different, i.e. thicker compared to other datasets.

As shown in electronic supplementary material, figure S4, the PSNR loss curve can represent the convergence speed. DAGAN has converged quickly, which is related to the fact that the calculation of its loss function is the simplest compared to other three GAN-based methods. RefineGAN has converged slowly but has demonstrated the best results. We analysed the size of each generator of the four GAN models to understand the potential cost to apply well-trained models in practice. Note that the discriminators and pre-trained VGG was not counted in the model size, because they were only required for training. All these four generators are U-Net based, consisting of an encoder and a decoder with several down-sample and up-sample units. Thus, the size of generators mainly depends on the depth of encoder and decoder, the implementation details of down/up-sample units, and the basic width, i.e. the number of feature maps in the first level. In particular, the parameters of each model are as follows: DAGAN has 98 589 507 parameters, ReconGAN has 15 393 313 parameters, RefineGAN has 30 786 626 parameters and KIGAN has 3 121 281 parameters. Thus, we speculated that the RefineGAN performs best is due to that it has twofold chained generators with the most convolution layers. Meanwhile, the residual bottleneck block declines the number of parameters, and avoid gradient vanishing problem of training very deep networks. Furthermore, the checkpoint in the middle of RefineGAN introduced extra restriction of training and improved the final performance. In contrast, although the architecture of KIGAN is more complicated, it is more suitable for the DCE images since it requires three adjacent images as input to the network.

For the liver MRI datasets, DAGAN has produced remarkable poor reconstruction. We believe the reason is that the image size of the public datasets (256×256 or 512×512) is more suitable for a direct application of DAGAN. This is because the VGG loss is calculated by the VGG16 network, in which the input image is specified at 224×224 . However, the image size of the liver datasets is 430×430 . Before DAGAN training, we have to firstly interpolate the image into the size of 512×512 . This results in that the whole liver occupies just a small proportion of the entire image, which cannot provide sufficient image details for the DAGAN network. This can in turn lead to unsatisfactory reconstruction results. Therefore, the prerequisite of the fixed size input is a major drawback of the DAGAN method.

Recently, there are many endeavours have been made to extend previous single-coil GAN-based MR reconstruction to a method incorporating multi-coil parallel imaging. Zhang *et al.* [47] proposed a multi-channel GAN model to learn physical data relationship in k -space instead of learning anatomy structure in the image space. Liu *et al.* [48] developed a deep learning-based reconstruction architecture called SANIST (Sampling-Augmented Neural neTwork with Incoherent Structure) for accelerated MRI reconstruction with an enhanced robustness against undersampling pattern discrepancy. It is important to see how these methods would behave using the same datasets but these are beyond the scope of our current study.

Visual inspection of the images agrees with the quantitative results. It is known that the VIF is highly correlated with the radiologist assessment of the image quality [49]. Although the VIF values of KIGAN are higher than that of RefineGAN, if we assume that an image with $VIF > 0.85$ is good enough to be introduced into the clinical settings, there would be almost no differences

between RefineGAN and KIGAN in terms of visual quality. One possible limitation of our current study is that the actual clinical usage and value of the reconstructed MRI images were not compared systematically. This could lead to our further studies by performing various GAN-based models to reconstruct pathological MRI datasets. The classification and quantification of the possible diseases could provide additional clinical validation of the compared GAN-based methods.

Interestingly, FID is now a widely used metric for performance evaluation of the GAN-based applications [46,50,51]. In this study, the value of FID is strongly correlated to other quantitative metrics, which means reconstructed images with lower FID shows more realistic structures. Thus, it is indicated that FID indeed captures the image quality of the reconstructed MRI images despite that it was originally designed for natural images.

Our study still has a couple of additional limitations. One potential criticism of our study is that the data were retrospectively undersampled to form the training and testing datasets. Training on simulated data could be suboptimal since prospectively acquired data may contain features, e.g. noise contamination and motion artefacts, which may not present in the simulated datasets. Moreover, our current analysis has been performed on normal MRI datasets. In the future work, we will explore and benchmark the GAN-based frameworks for pathological MRI datasets and assorted MRI sequences, e.g. dynamic MRI and diffusion-weighted MRI.

5. Conclusion

In this study, we have benchmarked and compared four different GAN-based network architectures for fast MRI reconstruction. Our comparison has been done on the same datasets but has covered three widely used clinical MRI scenarios. Both conventional and newly proposed quantitative metrics have been used for our comprehensive analysis and evaluation. Qualitative visualization results were also scrutinized and compared. To conclude, our benchmarking suggested that all GAN-based methods could achieve satisfactory results at lower acceleration factors compared to the information loss and aliasing artefact contamination observed in the zero-filled reconstructions. However, some GAN-based architectures, e.g. DAGAN and KIGAN, might not be able to adequate for MRI reconstruction applications when the acceleration factors are high. Besides, in comparison to other GAN-based approaches, the RefineGAN has achieved better reconstruction precision and perceptual quality improvement.

Data accessibility. This article has no additional data.

Authors' contributions. J.L. and G.Y. conceived and designed the study. J.L. and G.Y. contributed to the literature search. J.L. contributed to data collection and performed data curation. J.L. and J.Z. contributed to data analysis. J.L. and G.Y. contributed to data interpretation. J.L. contributed to the tables and figures. J.L. and G.Y. contributed to the writing of the report. All the authors have read and approved the publication of this work.

Competing interests. We declare we have no competing interests.

Funding. This work was supported in part by the National Natural Science Foundation of China No. 61902338, in part by the British Heart Foundation [PG/16/78/32402], in part by the Hangzhou Economic and Technological Development Area Strategic Grant (Imperial Institute of Advanced Technology), in part by the European Research Council Innovative Medicines Initiative on Development of Therapeutics and Diagnostics Combatting Coronavirus Infections Award 'DRAGON: rapiD and secuRe AI imaging-based diaGnosis, stratification, follow-up, and preparedness for coronavirus paNdemics' [H2020-JTI-IMI2 101005122], and in part by the AI for Health Imaging Award 'CHAIMELEON: Accelerating the Lab to Market Transition of AI Tools for Cancer Management' [H2020-SC1-FA-DTS-2019-1 952172].

References

1. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A. 2002 Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn. Reson. Med.* **47**, 1202–1210. (doi:10.1002/mrm.10171)

2. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. 1999 SENSE: sensitivity encoding for fast MRI. *Magn. Reson. Med.* **42**, 952–962. (doi:10.1002/(SICI)1522-2594(199911)42:5<952::AID-MRM16>3.0.CO;2-S)
3. Lustig M, Donoho DL, Santos JM, Pauly JM. 2008 Compressed sensing MRI. *IEEE Signal Process Mag.* **25**, 72–82. (doi:10.1109/MSP.2007.914728)
4. Robson PM, Grant AK, Madhuranthakam AJ, Lattanzi R, Sodickson DK, McKenzie CA. 2008 Comprehensive quantification of signal-to-noise ratio and g-factor for image-based and k-space-based parallel imaging reconstructions. *Magn. Reson. Med.* **60**, 895–907. (doi:10.1002/mrm.21728)
5. Deshmane A, Gulani V, Griswold MA, Seiberlich N. 2012 Parallel MR imaging. *J. Magn. Reson. Imaging* **36**, 55–72. (doi:10.1002/jmri.23639)
6. Block KT, Uecker M, Frahm J. 2007 Undersampled radial MRI with multiple coils. iterative image reconstruction using a total variation constraint. *Magn. Reson. Med.* **57**, 1086–1098. (doi:10.1002/mrm.21236)
7. Osher S, Burger M, Goldfarb D, Xu J, Yin W. 2005 An iterative regularization method for total variation-based image restoration. *Multiscale Model. Simul.* **4**, 460–489. (doi:10.1137/040605412)
8. Chaàri L, Pesquet JC, Benazza-Benyahia A, Ciuciu P. 2011 A wavelet-based regularized reconstruction algorithm for SENSE parallel MRI with applications to neuroimaging. *Med. Image Anal.* **15**, 185–201. (doi:10.1016/j.media.2010.08.001)
9. Böhle M, Eitel F, Weygandt M, Ritter K. 2019 Layer-wise relevance propagation for explaining deep neural network decisions in MRI-based Alzheimer's disease classification. *Front. Aging Neurosci.* **11**, 194. (doi:10.3389/fnagi.2019.00194)
10. Ismael SAA, Mohammed A, Hefny H. 2020 An enhanced deep learning approach for brain cancer MRI images classification using residual networks. *Artif. Intell. Med.* **102**, 101779. (doi:10.1016/j.artmed.2019.101779)
11. Wang Y, Yue W, Li X, Liu S, Guo L, Xu H, Zhang H, Yang G. 2020 Comparison study of radiomics and deep learning-based methods for thyroid nodules classification using ultrasound images. *IEEE Access* **8**, 52 010–52 017. (doi:10.1109/ACCESS.2020.2980290)
12. Hu S *et al.* 2020 Weakly supervised deep learning for COVID-19 infection detection and classification from CT images. *IEEE Access* **8**, 118 869–118 883. (doi:10.1109/ACCESS.2020.3005510)
13. Yang M, Xiao X, Liu Z, Sun L, Guo W, Cui L, Sun D, Zhang P, Yang G. 2020 Deep RetinaNet for dynamic left ventricle detection in multiview echocardiography classification. *Sci. Programming* **2020**, 7025403.
14. Thyreau B, Taki Y. 2020 Learning a cortical parcellation of the brain robust to the MRI segmentation with convolutional neural networks. *Med. Image Anal.* **61**, 101639. (doi:10.1016/j.media.2020.101639)
15. Liu Y, Yang G, Hosseiny M, Azadikhah A, Mirak SA, Miao Q, Raman SS, Sung K. 2020 Exploring uncertainty measures in Bayesian deep attentive neural networks for prostate zonal segmentation. *IEEE Access* **8**, 151817–151828. (doi:10.1109/ACCESS.2020.3017168)
16. Yang G *et al.* 2020 Simultaneous left atrium anatomy and scar segmentations via deep learning in multiview information with attention. *Fut. Gen. Comput. Syst.* **107**, 215–228. (doi:10.1016/j.future.2020.02.005)
17. Li M, Wang C, Zhang H, Yang G. 2020 MV-RAN: Multiview recurrent aggregation network for echocardiographic sequences segmentation and full cardiac cycle analysis. *Comput. Biol. Med.* **120**, 103728. (doi:10.1016/j.compbimed.2020.103728)
18. Ferreira PF, Martin RR, Scott AD, Khalique Z, Yang G, Nilles-Vallespin S, Pennell DJ, Firmin DN. 2020 Automating in vivo cardiac diffusion tensor postprocessing with deep learning-based segmentation. *Magn. Reson. Med.* **84**, 2801–2814.
19. Liu Y, Yang G, Mirak SA, Hosseiny M, Azadikhah A, Zhong X, Reiter RE, Lee Y, Raman SS, Sung K. 2019 Automatic prostate zonal segmentation using fully convolutional network with feature pyramid attention. *IEEE Access* **7**, 163 626–163 632. (doi:10.1109/ACCESS.2019.2952534)
20. Lv J, Yang M, Zhang J, Wang X. 2018 Respiratory motion correction for free-breathing 3D abdominal MRI using CNN-based image registration: a feasibility study. *Br. J. Radiol.* **91**, 20170788. (doi:10.1259/bjr.20170788)

21. Lv J, Huang W, Zhang J, Wang X. 2018 Performance of U-Net based pyramidal Lucas-Kanade registration on free-breathing multi-b-value diffusion MRI of the kidney. *Br. J. Radiol.* **91**, 20170813. (doi:10.1259/bjr.20170813)
22. de Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Išgum I. 2019 A deep learning framework for unsupervised affine and deformable image registration. *Med. Image Anal.* **52**, 128–143. (doi:10.1016/j.media.2018.11.010)
23. Maicas G, Snaauw G, Bradley AP, Reid I, Carneiro G. 2019 Model agnostic saliency for weakly supervised lesion detection from breast DCE-MRI. In *2019 IEEE 16th Int. Symp. on Biomedical Imaging (ISBI 2019)*, pp. 1057–1060. Piscataway, NJ: IEEE.
24. Nair T, Precup D, Arnold DL, Arbel T. 2020 Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. *Med. Image Anal.* **59**, 101557. (doi:10.1016/j.media.2019.101557)
25. Salem M, Valverde S, Cabezas M, Pareto D, Oliver A, Salvi J. 2020 A fully convolutional neural network for new T2-w lesion detection in multiple sclerosis. *NeuroImage: Clinical* **25**, 102149. (doi:10.1016/j.nicl.2019.102149)
26. Wang S, Su Z, Ying L, Peng X, Zhu S, Liang F, Feng D, Liang D. 2016 Accelerating magnetic resonance imaging via deep learning. In *2016 IEEE 13th Int. Symp. on Biomedical Imaging (ISBI)*, pp. 514–517. Piscataway, NJ: IEEE.
27. Jin KH, McCann MT, Froustey E, Unser M. 2017 Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Image Process.* **26**, 4509–4522. (doi:10.1109/TIP.2017.2713099)
28. Schlemper J, Caballero J, Hajnal JV, Price AN, Rueckert D. 2017 A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans. Med. Imaging* **37**, 491–503. (doi:10.1109/TMI.2017.2760978)
29. Lv J, Chen K, Yang M, Zhang J, Wang X. 2018 Reconstruction of undersampled radial free-breathing 3D abdominal MRI using stacked convolutional auto-encoders. *Med. Phys.* **45**, 2023–2032. (doi:10.1002/mp.12870)
30. Yang Y, Sun J, Li H, Xu Z. 2020 ADMM-CSNet: a deep learning approach for image compressive sensing. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 521–538. (doi:10.1109/TPAMI.2018.2883941)
31. Dar SUH, Zbey M, Atl AB, Ukur T. 2020 A transfer-learning approach for accelerated MRI using deep neural networks. *Magn. Reson. Med.* **84**, 663–685.
32. Shitrit O, Raviv TR. 2017 Accelerated magnetic resonance imaging by adversarial neural network. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pp. 30–38. Cham, Switzerland: Springer.
33. Yang G *et al.* 2017 DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans. Med. Imaging* **37**, 1310–1321. (doi:10.1109/TMI.2017.2785879)
34. Mardani M, Gong E, Cheng JY, Vasanawala SS, Zaharchuk G, Xing L, Pauly JM. 2018 Deep generative adversarial neural networks for compressive sensing MRI. *IEEE Trans. Med. Imaging* **38**, 167–179. (doi:10.1109/TMI.2018.2858752)
35. Quan TM, Nguyen-Duc T, Jeong WK. 2018 Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. *IEEE Trans. Med. Imaging* **37**, 1488–1497. (doi:10.1109/TMI.2018.2820120)
36. Shaul R, David I, Shitrit O, Raviv TR. 2020 Subsampled brain MRI reconstruction by generative adversarial neural networks. *Med. Image Anal.* **65**, 101747. (doi:10.1016/j.media.2020.101747)
37. Li Z, Zhang T, Wan P, Zhang D. 2019 SEGAN: structure-enhanced generative adversarial network for compressed sensing MRI reconstruction. In *Proc. of the AAAI Conf. on Artificial Intelligence*, vol. 33, pp. 1012–1019. Menlo Park, CA: AAAI.
38. Murugesan B, Raghavan SV, Sarveswaran K, Ram K, Sivaprakasam M. 2019 Recon-GLGAN: a global-local context based generative adversarial network for MRI reconstruction. In *Int. Workshop on Machine Learning for Medical Image Reconstruction*, pp. 3–15. Berlin, Germany: Springer.
39. Deora P, Vasudeva B, Bhattacharya S, Pradhan PM. 2020 Structure preserving compressive sensing MRI reconstruction using generative adversarial networks. In *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops*, pp. 522–523. Piscataway, NJ: IEEE.

40. Guo Y, Wang C, Zhang H, Yang G. 2020 Deep attentive Wasserstein generative adversarial networks for MRI reconstruction with recurrent context-awareness. In *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, pp. 167–177. Berlin, Germany: Springer.
41. Ramzi Z, Ciuciu P, Starck JL. 2020 Benchmarking MRI reconstruction neural networks on large public datasets. *Appl. Sci.* **10**, 1816. (doi:10.3390/app10051816)
42. Simonyan K, Zisserman A. 2014 Very deep convolutional networks for large-scale image recognition. (<http://arxiv.org/abs/1409.1556>).
43. Russakovsky O *et al.* 2015 Imagenet large scale visual recognition challenge. *Int. J. Comput. Vision* **115**, 211–252. (doi:10.1007/s11263-015-0816-y)
44. Zhang T, Pauly JM, Vasanawala SS, Lustig M. 2013 Coil compression for accelerated imaging with cartesian sampling. *Magn. Reson. Med.* **69**, 571–582. (doi:10.1002/mrm.24267)
45. Sheikh HR, Bovik AC. 2006 Image information and visual quality. *IEEE Trans. Image Process.* **15**, 430–444. (doi:10.1109/TIP.2005.859378)
46. Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S. 2017 GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *Advances in neural information processing systems*, pp. 6626–6637. San Diego, CA: NIPS.
47. Zhang P, Wang F, Xu W, Li Y. 2018 Multi-channel generative adversarial network for parallel magnetic resonance image reconstruction in k-space. In *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, pp. 180–188. Cham, Switzerland: Springer.
48. Liu F, Samsonov A, Chen L, Kijowski R, Feng L. 2019 Santis: sampling-augmented neural network with incoherent structure for MR image reconstruction. *Magn. Reson. Med.* **82**, 1890–1904. (doi:10.1002/mrm.27827)
49. Mason A, Rioux J, Clarke SE, Costa A, Beyea S. 2019 Comparison of objective image quality metrics to expert radiologists' scoring of diagnostic quality of MR images. *IEEE Trans. Med. Imaging*, pp. 1–1.
50. Lucic M, Kurach K, Michalski M, Gelly S, Bousquet O. 2018 Are GANs created equal? A large-scale study. In *Advances in neural information processing systems*, pp. 700–709. San Diego, CA: NIPS.
51. Borji A. 2019 Pros and cons of GAN evaluation measures. *Comput. Vision and Image Understanding* **179**, 41–65. (doi:10.1016/j.cviu.2018.10.009)