Challenge Report

# mustGAN: multi-stream Generative Adversarial Networks for MR Image Synthesis

Mahmut Yurt [a,b], Salman UH Dar [a,b], Aykut Erdem [d], Erkut Erdem [e], Kader K Oguz [b,f], Tolga Çukur [a,b,c,*]

[a] Department of Electrical and Electronics Engineering, Bilkent University, Ankara, TR-06800, Turkey
[b] National Magnetic Resonance Research Center, Bilkent University, Ankara, TR-06800, Turkey
[c] Neuroscience Program, Aysel Sabuncu Brain Research Center, Bilkent, Ankara, TR-06800, Turkey
[d] Department of Computer Engineering, Koç University, İstanbul, TR-34450, Turkey
[e] Department of Computer Engineering, Hacettepe University, Ankara, TR-06800, Turkey
[f] Department of Radiology, Hacettepe University, Ankara, TR-06100, Turkey

## ARTICLE INFO

## ABSTRACT

Multi-contrast MRI protocols increase the level of morphological information available for diagnosis. Yet, the number and quality of contrasts are limited in practice by various factors including scan time and patient motion. Synthesis of missing or corrupted contrasts from other high-quality ones can alleviate this limitation. When a single target contrast is of interest, common approaches for multi-contrast MRI involve either one-to-one or many-to-one synthesis methods depending on their input. One-to-one methods take as input a single source contrast, and they learn a latent representation sensitive to unique features of the source. Meanwhile, many-to-one methods receive multiple distinct sources, and they learn a shared latent representation more sensitive to common features across sources. For enhanced image synthesis, we propose a multi-stream approach that aggregates information across multiple source images via a mixture of multiple one-to-one streams and a joint many-to-one stream. The complementary feature maps generated in the one-to-one streams and the shared feature maps generated in the many-to-one stream are combined with a fusion block. The location of the fusion block is adaptively modified to maximize task-specific performance. Quantitative and radiological assessments on $T_1$-, $T_2$-, PD-weighted, and FLAIR images clearly demonstrate the superior performance of the proposed method compared to previous state-of-the-art one-to-one and many-to-one methods.

© 2020 Elsevier B.V. All rights reserved.

## 1. Introduction

Magnetic resonance imaging (MRI) enables a given anatomy to be imaged under different tissue contrasts by simply manipulating pulse sequences. In turn, images acquired in several distinct contrasts help better distinguish tissues and increase diagnostic information. For instance, gray and white matter can be better delineated in $T_1$-weighted brain images, whereas fluids and cortical tissues can be better delineated in PD-weighted images. Yet, multi-contrast acquisitions often prove impractical due to scan time limitations or excessive artifacts related to patient motion (Krupa and Bekiesińska-Figatowska, 2015; Thukral, 2015). Therefore, within-domain synthesis of missing or corrupted contrasts from other high-quality contrasts is a promising tool to improve the clinical feasibility and utility of multi-contrast MRI (Iglesias et al., 2013).

Prior methods proposed for synthesis of a single target contrast within a multi-contrast MRI protocol can be grouped under two main titles depending on their input: one-to-one methods (Abramian and Eklund, 2019; Bowles et al., 2016; Dar et al., 2019; Dewey et al., 2018; 2019; Li et al., 2014; Liu, 2019; Sevetlidis et al., 2016; Sohail et al., 2019; Van Nguyen et al., 2015; Welander et al., 2018; Yang et al., 2018) and many-to-one methods (Chartsias et al., 2017; Dar et al., 2020; Dewey et al., 2018; 2019; Hagiwara et al., 2019; Joyce et al., 2017; Lee et al., 2019; Lei et al., 2020; Li et al., 2019; Mehta and Arbel, 2018; Olut et al., 2018; Sharma and Hamarneh, 2019; Wei et al., 2018b). One-to-one synthesis aims to generate a subject's image $y$ in a target contrast $c_T$ from the same subject's image $x$ in a source contrast $c_S$. Earlier studies have formulated one-to-one synthesis as a sparse dictionary reconstruction problem (Huang et al., 2016; 2017; 2018; Jog et al., 2015b; Roy et al., 2011; 2013a; 2013b; 2016), where patch-based dictionaries are formed from a set of co-registered atlas image $b_S$ of $c_S$ and atlas image $b_T$ of $c_T$. Each patch in $x$ is expressed as a sparse

* Corresponding author.
  *E-mail address:* cukur@ee.bilkent.edu.tr (T. Çukur).

linear combination of dictionary atoms of $b_S$, and this combination is then used for synthesizing $y$ from $b_T$ (Huang et al., 2016; 2017; 2018; Jog et al., 2015b; Roy et al., 2011; 2013a; 2013b; 2016). For improved performance, patch-based non-linear regression using random forests (Jog et al., 2015b) or location-sensitive neural networks (Van Nguyen et al., 2015) has been proposed for source to target mapping. To overcome limitations due to patch-based processing, later studies have proposed deep network models that process the entire source image with convolutional layers (Li et al., 2014; Sevetlidis et al., 2016). One powerful method is based on the encoder-decoder architecture (Sevetlidis et al., 2016), where latent representations of the source image are embedded via an encoder, and the target image is then recovered via a decoder from these representations (Sevetlidis et al., 2016). Recent deep network models have further incorporated an adversarial loss to better capture the high frequency details in the target image (Abramian and Eklund, 2019; Dar et al., 2019; Sohail et al., 2019; Welander et al., 2018; Yang et al., 2018; Yu et al., 2018; 2019). An important adversarial method is pGAN (Dar et al., 2019), which additionally utilizes pixel-wise and perceptual losses (Johnson et al., 2016) to enhance the synthesis performance.

When several source contrasts are available in a multi-contrast MRI protocol, a natural alternative is to perform many-to-one synthesis (Chartsias et al., 2017; Dar et al., 2020; Dewey et al., 2018; 2019; Hagiwara et al., 2019; Jog et al., 2014; 2015; 2016; Joyce et al., 2017; Lee et al., 2019; Lei et al., 2020; Li et al., 2019; Mehta and Arbel, 2018; Olut et al., 2018; Sharma and Hamarneh, 2019; Wei et al., 2018b). In this approach, the goal is to generate the subject's image $y$ in the target contrast $c_T$ from a collection of source images $X = \{x_m : m = 1, 2, \ldots, K\}$ in varying contrasts $C_S = \{c_{S_m} : m = 1, 2, \ldots, K\}$. As in one-to-one synthesis, a common method is to perform non-linear regression using random forests (Jog et al., 2014; 2016). The random-forest method in Jog et al. (2016) fits a non-linear regression model in feature space to estimate intensities of the target contrast given multiple source contrasts (Jog et al., 2016). Deep neural network methods have also been proposed for many-to-one synthesis (Chartsias et al., 2017; Dewey et al., 2018; 2019; Mehta and Arbel, 2018; Joyce et al., 2017). In Chartsias et al. (2017), latent representations of multiple source contrast images are encoded through separate encoder architectures. These latent representations are then used to synthesize the target image through a joint decoder architecture (Chartsias et al., 2017). Similar to one-to-one methods, recent studies have leveraged an adversarial loss to improve the quality of many-to-one synthesis (Dar et al., 2020; Hagiwara et al., 2019; Lee et al., 2019; Li et al., 2019; Olut et al., 2018; Sharma and Hamarneh, 2019; Wei et al., 2018b). An important example is MM-GAN (Sharma and Hamarneh, 2019), which learns recovery of missing (target) contrasts from a collection of available source contrasts. MM-GAN receives as input the concatenation of the sources, and treats them as separate information channels (Sharma and Hamarneh, 2019).

In general, one-to-one synthesis methods attempt to recover the target image from the latent representation of a given source image. Since these methods are optimized for a single input channel, they can sensitively learn the unique, detailed features of the given source contrast. While this sensitivity can be preferable when the images of the source and target contrast are highly correlated, it might limit synthesis performance when the two contrasts are weakly linked. On the other hand, many-to-one synthesis methods aim to recover the target image from a shared latent representation of multiple source images. These methods naturally manifest increased sensitivity for capturing features that are shared across distinct source contrasts, even when these features are weakly present in individual contrasts. Yet, a shared latent representation might also be less sensitive to complementary features that are uniquely present in a specific source con-

trast. When such unique information is predominantly predictive of the target image, many-to-one synthesis might perform suboptimally.

Here, we propose a novel method, multi-stream generative adversarial network (mustGAN), for enhanced image synthesis in multi-contrast MRI. To alleviate limitations of one-to-one and many-to-one synthesis, mustGAN leverages both shared and complementary features of multiple source images via a mixture of multiple one-to-one streams and a joint many-to-one stream. The complementary feature maps generated in the one-to-one streams and the shared feature maps generated in the many-to-one stream are later combined with a fusion block. The optimal position of the fusion block is searched over network layers to maximize task-specific performance. A joint network is then trained to recover the target image from the fused feature maps. Comprehensive quantitative assessments and radiological evaluations are performed on multi-contrast MR images (T$_1$-, T$_2$-, PD-weighted, and FLAIR images) from healthy subjects and high/low grade glioma patients. The proposed method yields both quantitatively and qualitatively higher performance in multi-contrast MRI synthesis compared to state-of-the-art one-to-one and many-to-one methods.

## 2. Theory

### 2.1. Generative adversarial networks

A generative adversarial network (GAN) consists of a pair of competing networks; a generator ($G$) and a discriminator ($D$) (Goodfellow et al., 2014). Recently, GAN models have been successfully demonstrated for various tasks including data augmentation (Bermudez et al., 2018; Bowles et al., 2018; Calimeri et al., 2017) and image synthesis (Beers et al., 2018; Han et al., 2018; Zhao et al., 2018). In an image synthesis task, $G$ maps a random noise vector $z$ to an output image $y$ from a target distribution $p(y)$, $G : z \rightarrow y$, and $D$ estimates the probability that a sample $s$ is drawn from $p(y)$, $D : s$. While $G$ is trained to synthesize fake images that are indistinguishable from real images, $D$ is trained to discriminate between real and generated images (Goodfellow et al., 2014). This can be formulated as a minimax game based on an adversarial loss function $L_{GAN}$.

$$\min_G \max_D L_{GAN} = \min_G \max_D \left( E_y[logD(y)] + E_z[log(1 - D(G(z)))] \right) \tag{1}$$

where E denotes expectation. To improve stability, the negative log-likelihood in $L_{GAN}$ is typically replaced by a squared loss function (Mao et al., 2017):

$$L_{GAN} = -E_y\left[(D(y) - 1)^2\right] - E_z\left[D(G(z))^2\right] \tag{2}$$

### 2.2. Conditional generative adversarial networks

Recent studies on image-to-image translation have demonstrated that conditional GANs (condGANs) are highly effective in mapping between statistically-dependent source and target images (Cai et al., 2019; Choi et al., 2018; 2020; Litjens et al., 2017; Mirza and Osindero, 2014; Ran et al., 2019; Ravì et al., 2019; Rubin et al., 2019; Wang et al., 2019; Xu et al., 2020; Yi et al., 2019), i.e., when these images manifest the same underlying scene in distinct domains. To capture this dependency, condGANs take as input the source image $x$ as prior information (Mirza and Osindero, 2014). The adversarial loss is then expressed as:

$$L_{condGAN} = -E_{x,y}\left[(D(x, y) - 1)^2\right] - E_x\left[D(G(x))^2\right] \tag{3}$$

When source and target images are spatially registered, a pixel-wise loss can be added between the ground truth and generated

images (Isola et al., 2017):

$$L_{\text{pixel-wise}} = E_{x,y}\left[\left|\left|y - G(x)\right|\right|_1\right] \tag{4}$$

The joint loss function then becomes:

$$L_{\text{condGAN}} = -E_{x,y}\left[(D(x,y) - 1)^2\right] - E_x\left[D(x, G(x))^2\right] + E_{x,y}\left[\left|\left|y - G(x)\right|\right|_1\right] \tag{5}$$

In a previous study, we have demonstrated that condGAN-based architectures yield state-of-the-art performance for one-to-one MR image synthesis, e.g., $T_1 \rightarrow T_2$ and $T_2 \rightarrow T_1$ (Dar et al., 2019). Yet, numerous different contrasts are often collected in an MR exam. When multiple source images are available, many-to-one condGAN models may offer improved performance. Given $K$ source contrast images denoted as $X = \{x_m : m = 1, 2, \ldots, K\}$, a many-to-one condGAN model is formulated as:

$$L_{\text{condGAN}} = -E_{X,y}\left[(D(X,y) - 1)^2\right] - E_X\left[D(X, G(X))^2\right] + E_{X,y}\left[\left|\left|y - G(X)\right|\right|_1\right] \tag{6}$$

Note that this formulation corresponds to a single-stream many-to-one network that concatenates multiple source images at the input level.

## 3. Methods

### 3.1. Multi-Stream GAN model

Here, we propose a multi-stream GAN architecture (mustGAN) that leverages information from multiple source contrasts by adaptively combining one-to-one and many-to-one streams (Fig. 1). To synthesize the target image $y$, mustGAN receives as input a collection of source images denoted as $X = \{x_m : m = 1, 2, \ldots, K\}$. First, mustGAN learns $K$ independent one-to-one streams, where each stream is a condGAN model trained to generate the target image from a distinct source image. Second, mustGAN learns a single many-to-one stream -again a condGAN model- that is trained to generate the target image from all source images concatenated at the input level. mustGAN then fuses the unique feature maps generated in the one-to-one streams and the shared feature maps generated in the many-to-one stream. The position of the fusion block is searched over pre-defined network levels to maximize task-specific performance. Finally, mustGAN trains a joint network that synthesizes the target image given the fused feature maps. The architecture of this joint network varies depending on the position of the fusion block.

### 3.1.1. One-to-One streams

The proposed architecture contains $K$ separate one-to-one streams, where the $m$th stream learns to synthesize $y$ from the source contrast image $x_m$ via a generator $G_m$ and a discriminator $D_m$. $G_m$ consists of three sub-networks: an encoder ($e_m$) with $n_e$ convolutional layers, a residual network ($r_m$) with $n_r$ ResNet blocks, and a decoder ($d_m$) with $n_d$ convolutional layers. $G_m$ is expressed as:

$$\hat{y}_m = G_m(x_m) = d_m(r_m(e_m(x_m))) \tag{7}$$

where $\hat{y}_m$ denotes the predicted target image. Meanwhile, $D_m$ is a convolutional network ($c_m$) with $n_c$ layers:

$$D_m(x_m, s) = c_m(x_m, s) \tag{8}$$

where $s$ is either $G_m(x_m)$ or $y$. To train $G_m$ and $D_m$, adversarial and pixel-wise losses are used:

$$L_m = -E_{x_m,y}\left[(D_m(x_m, y) - 1)^2\right] - E_{x_m}\left[D_m(x_m, G_m(x_m))^2\right] + E_{x_m,y}\left[\left|\left|y - G_m(x_m)\right|\right|_1\right] \tag{9}$$

$G_m$ aims to map $x_m$ to $y$, and $D_m$ aims to discriminate between $\hat{y}_m$ and $y$.

### 3.1.2. Many-to-One stream

mustGAN contains a $(K + 1)$th stream that performs many-to-one synthesis given all source images. This generator $G_{K+1}$ again consists of an encoder ($e_{K+1}$) with $n_e$ convolutional layers, a residual network ($r_{K+1}$) with $n_r$ ResNet blocks, and a decoder ($d_{K+1}$) with $n_d$ convolutional layers:

$$\hat{y}_{K+1} = G_{K+1}(X) = d_{K+1}(r_{K+1}(e_{K+1}(X))) \tag{10}$$

The discriminator $D_{K+1}$ containing $n_c$ convolutional layers is formulated as:

$$D_{K+1}(X, s) = c_{K+1}(X, s) \tag{11}$$

where $s$ is either $G_{K+1}(X)$ or $y$. The loss function for the $(K + 1)$th stream is given as:

$$L_{K+1} = -E_{X,y}\left[(D_{K+1}(X, y) - 1)^2\right] - E_X\left[D_{K+1}(X, G_{K+1}(X))^2\right] + E_{X,y}\left[\left|\left|y - G_{K+1}(X)\right|\right|_1\right] \tag{12}$$

$G_{K+1}$ learns to predict $y$ given $x_1, x_2, \ldots, x_K$ concatenated at the input level, and $D_{K+1}$ learns to discriminate between $\hat{y}_{K+1}$ and $y$.

### 3.1.3. Joint network

Once the $K + 1$ streams are trained, source images are propagated separately through the streams up to the fusion block $f$ positioned at the $i$th network layer. $f$ combines the feature maps generated at the $i$th layer of the one-to-one and many-to-one streams by performing concatenation. A joint network ($J$) is then trained to recover the target contrast image from the fused feature maps. The position of the fusion block, $i$, is searched over $n_e + n_r + n_d - 1$ possible positions in the network to maximize task-specific performance, where $n_e$ denotes the number of network layers in the encoders, $n_r$ denotes the number of network layers in the residual blocks, and $n_d$ denotes the number of network layers in the decoders. Therefore, there are $n_e + n_r + n_d - 1$ possible architectures for $J$ depending on the fusion block position. mustGAN chooses the task-optimal fusion strategy among these options, so the embodied joint network receives the fused feature maps from a single network layer. To simplify analytical descriptions, we separated the precise architectures of $J$ under three distinct fusion titles: early, intermediate, and late.

**Early Fusion:** Early fusion occurs when $f$ is within the encoder (i.e., $0 < i < n_e$). The feature maps generated by the $m$th one-to-one stream ($g_m^i$) and by the many-to-one stream ($g_{K+1}^i$) at the $i$th layer are formulated as:

$$g_m^i = e_m(x_m|i)$$
$$g_{K+1}^i = e_{K+1}(X|i) \tag{13}$$

These feature maps are concatenated by $f$ yielding the fused feature maps ($g_f^i$):

$$g_f^i = f(g_1^i, g_2^i, \ldots, g_K^i, g_{K+1}^i) \tag{14}$$

$J$ receives as input these fused maps to recover the target image. Thus, architecture of $J$ for early fusion is as follows:

$$\hat{y} = J(g_F^i) = d_J(r_J(e_J(g_f^i|i))) \tag{15}$$

**Intermediate Fusion:** Intermediate fusion occurs when $f$ is within the residual block (i.e., $n_e \leq i < n_e + n_r$). In this case, the feature maps generated by the $m$th one-to-one stream ($g_m^i$) and the many-to-one stream ($g_{K+1}^i$) are formulated as:

$$g_m^i = r_m(e_m(x_m)|i)$$
$$g_{K+1}^i = r_{K+1}(e_{K+1}(X)|i) \tag{16}$$

The fused feature maps ($g_f^i$) are then:

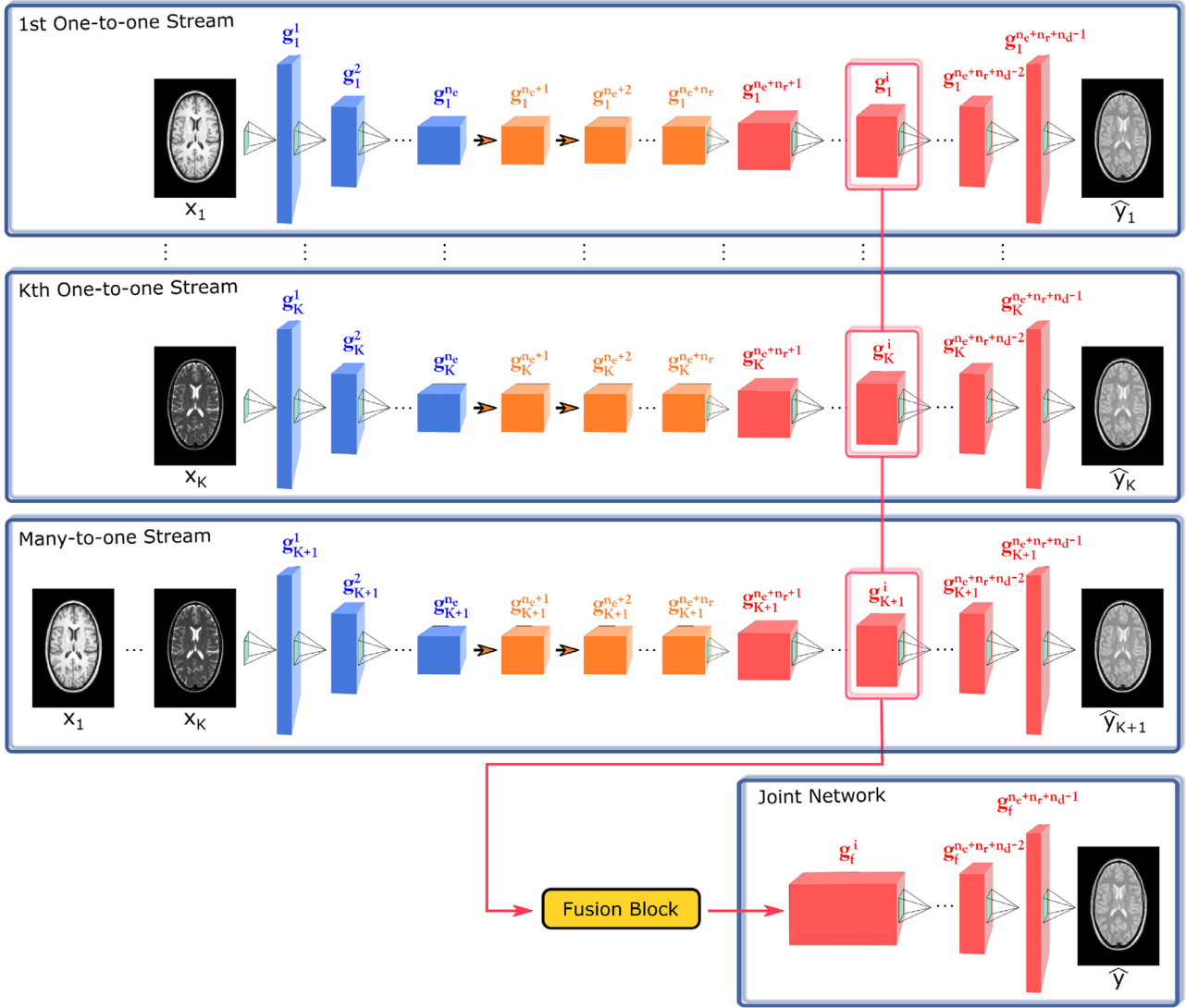$$g_f^i = f(g_1^i, g_2^i, \ldots, g_K^i, g_{K+1}^i) \tag{17}$$

**Fig. 1.** The generator (G) in mustGAN consists of $K$ one-to-one streams and a many-to-one stream, followed by an adaptively positioned fusion block, and a joint network for final recovery. The independent one-to-one streams generate unique feature maps of individual source images, whereas the many-to-one stream generates the shared feature maps across all source images. The fusion block then fuses the feature maps generated at the fusion layer by concatenation. Lastly, the joint network synthesizes the target contrast image from the fused feature maps. The architecture of the one-to-one and many-to-one streams consists of an encoder with $n_e$ convolutional layers (shown with color blue), a residual block with $n_r$ ResNet blocks (shown with color orange), and a decoder with $n_d$ convolutional layers (shown with color red). The unique position of the fusion block is optimized over $n_e + n_r + n_d - 1$ possible network layers for task-specific positioning. The precise architecture of the joint network is therefore adaptively modified depending on the fusion block position, and it consists of the remaining network layers after the fusion level. Note that a single fusion level candidate out of $n_e + n_r + n_d - 1$ possible positions is displayed here for illustration. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$J$ again receives as input the fused maps to recover the target image. Architecture of $J$ for intermediate fusion is as follows:

$$\hat{y} = J(g_f^i) = d_J(r_J(g_f^i|i)) \tag{18}$$

**Late Fusion:** Late fusion occurs when $f$ is within the decoder (i.e., $n_e + n_r \leq i < n_e + n_r + n_d$). The feature maps by the $m$th one-to-one stream ($g_m^i$) and by the many-to-one stream at the $i$th layer ($g_{K+1}^i$) are given as:

$$g_m^i = d_m(r_m(e_m(x_m))|i)$$
$$g_{K+1}^i = d_{K+1}(r_{K+1}(e_{K+1}(X))|i) \tag{19}$$

In turn, the fused feature maps ($g_f^i$) are:

$$g_f^i = f(g_1^i, g_2^i, \ldots, g_K^i, g_{K+1}^i) \tag{20}$$

$J$ receives as input the fused maps to recover the target image, yielding the following architecture for late fusion:

$$\hat{y} = J(g_f^i) = d_J(g_f^i|i) \tag{21}$$

$J$ is also trained in an adversarial setup with a conditional discriminator $D_J$ that has a fixed architecture independent of $i$. Therefore, $D_J$ receives as input the source and target contrast images, and consists of a convolutional network ($c_J$) with $n_c$ layers:

$$D_J(X, s) = c_J(X, s) \tag{22}$$

where s is either $J(g_f^i)$ or $y$. To train $J$ and $D_J$, a loss function consisting of an adversarial loss and pixel-wise $L1$ loss is used:

$$L_J = -\mathrm{E}_{X,y}\big[(D_J(X, y) - 1)^2\big] - \mathrm{E}_X\big[(D_J(X, J(g_f^i)))^2\big]$$
$$+ \mathrm{E}_{X,y}\big[||y - J(g_f^i)||_1\big] \tag{23}$$

### 3.2. Network architecture

The $K$ one-to-one streams and the many-to-one stream had identical generator ($G$) and discriminator ($D$) architectures adopted from (Johnson et al., 2016) and (Isola et al., 2017), respectively (see

Supp. Methods for further details). $G$ consisted of an encoder ($e$) of 3 convolutional layers, a residual network ($r$) of 9 ResNet blocks, and a decoder ($d$) of 3 convolutional layers. Meanwhile, $D$ is a conditional patch discriminator that consisted of a convolutional network ($c$) of 5 layers. It receives as input the concatenation of the source and target contrast images, and returns a $30 \times 30$ binary matrix of fake/real predictions, with each element of the matrix denoting the prediction for the corresponding $70 \times 70$ patch within the images. The number of input channels was 1 for $G_m$, 2 for $D_m$, $K$ for $G_{K+1}$, and $K+1$ for $D_{K+1}$.

The architecture of the joint network ($J$) was adaptively modified based on the position of the fusion block ($i$). When $0 < i < 3$ (early fusion), $J$ consisted of $3 - i$ convolutional layers, 9 ResNet blocks, and 3 convolutional layers connected in series. When $3 \leq i < 12$ (intermediate fusion), $J$ consisted of $12 - i$ ResNet blocks and 3 convolutional layers. When $12 \leq i < 15$ (late fusion), $J$ consisted of $15 - i$ convolutional layers. $J$ receives as input the fused feature maps from the streams that have variable tensor dimensions across the network layers. Therefore, the number of input channels for $J$ depended on $i$. The corresponding conditional discriminator $D_J$ had identical architecture with that in the many-to-one stream. The number of input channels was $K+1$ for $D_J$.

All generator and discriminator architectures used in this study were two-dimensional (2D), so mustGAN mapped cross-sectional images of the source and target contrasts. We opted for 2D models since they offer reduced model complexity compared to volumetric models. Furthermore, individual cross-sections in a volume are taken as separate data samples in 2D models, so they also have lower demand on training data.

The tensor dimensions of the source-target images and the feature maps are denoted as $[t_1, t_2, t_3, t_4]$, where $t_1$ indicates the batch size, $t_2$ indicates the number of channels, $t_3$ indicates the height, and $t_4$ indicates the width. The tensor dimensions of the source images were [1,1,256,256] for the one-to-one streams, and $[1, K, 256, 256]$ for the many-to-one stream, where $K$ denotes the total number of source contrasts. The ordering of the source contrasts in the many-to-one stream was arbitrarily chosen for a given synthesis task, and the chosen ordering was used thereafter in all experiments. The tensor dimensions of the target images were [1,1,256,256] for all streams and the joint network. The tensor dimensions of the feature maps varied across network levels but remained identical across streams for a fixed layer. The feature fusion implemented in this study was performed as a concatenation along the second tensor dimension $t_2$. Therefore, the fused feature maps had tensor dimensions of $[t_1, (K+1)t_2, t_3, t_4]$. The tensor dimensions of the subsequent feature maps within the joint network returned back to their original values (those in the one-to-one and many-to-one streams) immediately after the next convolutional layer.

### 3.3. Datasets

Demonstrations were performed on two separate neuroimaging datasets: the IXI dataset (http://brain-development.org/ixi-dataset/) that contained data from healthy subjects and the ISLES dataset (Maier et al., 2017) that contained data from high/low grade glioma patients. Data normalization was performed to provide comparable voxel intensities across subjects. To do this, the maximum intensity of each brain volume was normalized to 1. This normalization was performed separately for each subject and each MR contrast. The normalized images were then linearly mapped onto a range of $[-1, 1]$ as the last network layer of the generators included a tanh function (see Supp. Methods for details). Prior to PSNR/SSIM measurements and visualizations, the pixel intensities of the output images from the generators were linearly mapped onto the orig-

inal range of [0,1]. No data augmentation was performed during the experiments.

*IXI Dataset:* $T_1$-, $T_2$- and PD-weighted images from 53 subjects were used, where 25 were reserved for training, 10 were reserved for validation, and 18 were reserved for testing. Subject selection was performed sequentially. Approximately 100 axial cross-sections that contained artifact-free brain tissue were manually selected from each subject. The images were acquired with the following parameters. $T_1$-weighted images: TE = 4.603 $ms$, TR = 9.813 $ms$, flip angle = 8°, spatial resolution = 0.94 × 0.94 × 1.2 $mm^3$, matrix size = 256 × 256 × 150. $T_2$-weighted images: TE = 100 $ms$, TR = 8178.34 $ms$, flip angle = 90°, spatial resolution = 0.94 × 0.94 × 1.2 $mm^3$, matrix size = 256 × 256 × 150. PD-weighted images: TE = 8 $ms$, TR = 8178.34 $ms$, flip angle = 90°, spatial resolution = 0.94 × 0.94 × 1.2 $mm^3$, matrix size = 256 × 256 × 150. Note that images of separate contrasts were unregistered in this dataset. Therefore, $T_2$- and PD-weighted images were registered onto $T_1$-weighted images by rigid transformation based on mutual-information. Registration was performed via FSL (Jenkinson and Smith, 2001). The axial cross-sections used in the experiments were zero-padded to achieve a consistent input-output image size of 256 × 256.

*ISLES Dataset:* $T_1$-, $T_2$-weighted and FLAIR images from 56 subjects were used, where 25 were reserved for training, 10 were reserved for validation, and 21 were reserved for testing. Subject selection was performed sequentially. Note that the ISLES dataset comprised images acquired under a heterogenous set of scanning parameters, where separate contrasts were collected in different orientations (i.e., $T_1$ and FLAIR were collected axially, $T_2$ was collected sagittally). Although all images were resampled to an isotropic resolution of 1 $mm^3$ (Maier et al., 2017), synthesis models were built to recover cross-sections in the original orientation of each target contrast. Approximately 100 axial cross-sections containing artifact-free brain tissues from all contrasts were manually selected for $T_1$-weighted and FLAIR image synthesis. Meanwhile, approximately 110 sagittal cross-sections containing artifact-free brain tissues from all contrasts were manually selected for $T_2$-weighted image synthesis. Since $T_1$- and $T_2$-weighted images were already aligned to FLAIR images (Maier et al., 2017), no other registration was performed. The image matrix size in this dataset was variable, so for consistency axial and sagittal cross-sections were zero-padded to 256 × 256 image size.

### 3.4. Network training

The network training procedure for mustGAN comprises two sequential phases: the individual training of the one-to-one and many-to-one streams, and the training of the joint network following fusion. For the first phase, we adopted hyperparameter selection from a previous study (Dar et al., 2019), where successful one-to-one image synthesis was demonstrated in multi-contrast MRI via conditional GAN models. The streams were trained for 100 epochs via the Adam optimizer (Kingma and Ba, 2015), where the learning rate was set to $2 \times 10^{-4}$ in the first 50 epochs, and was linearly decayed from $2 \times 10^{-4}$ to 0 in the last 50 epochs. During the training, the decay rates of the first moment $\beta_1$ and the second moment $\beta_2$ of gradient estimates were set to 0.5 and 0.999, respectively. Relative weighting of the pixel-wise loss to adversarial loss was selected as 100. The one-to-one pGAN and many-to-one pGAN$_{many}$ models were also trained using the same hyperparameter set as mustGAN. Training pGAN and pGAN$_{many}$ beyond 100 epochs yielded lower or similar synthesis performance, so these models were also reported for 100 epochs.

For the second phase, we performed analyses to determine the optimal position of the fusion block for each synthesis task. Since the complexity of the joint network also depends on the posi-

tion of the fusion block, we reasoned that the required number of epochs for convergence should also be optimized. Therefore, we performed grid-search for the fusion block position and number of epochs using the validation set. To do this, multiple joint network architectures were trained for varying number of epoch values (5 : 5 : 100) and fusion block positions (1 : 1 : 14). The maximum limit for the number of training epochs was determined to be 100 since training the joint network architectures beyond 100 epochs yielded lower synthesis performance. To mitigate excessive demand for model storage space, the search step for number of epochs was selected as 5. The multiple joint networks were evaluated in terms of PSNR, SSIM and network loss measurements in the validation set. While these metrics all gave correlated results (Supp. Figs. 1-6), we opted for considering PSNR measurements in optimizations since it is a direct and robust metric based solely on image quality, unlike network loss that includes an adversarial component suggested to introduce instabilities. Based on the PSNR evaluations, task-specific values for the position of the fusion block and the number of epochs denoted as (fusion block position, number of epochs) were determined to be $(12, 40)$ for $T_1$-weighted, $(14, 15)$ for $T_2$-weighted and $(12, 20)$ for PD-weighted image synthesis in the IXI dataset, and $(8, 50)$ for $T_1$-weighted, $(7, 55)$ for $T_2$-weighted and $(6, 10)$ for FLAIR image synthesis in the ISLES dataset. Remaining hyperparameters were again adopted from Dar et al. (2019). During the training, the Adam optimizer was employed, where the decay rates of the first moment $\beta_1$ and the second moment $\beta_2$ of gradient estimates were set to 0.5 and 0.999, respectively. Relative weighting of the pixel-wise loss to adversarial loss was selected as 100. For models trained for fewer than 50 epochs, the learning rate was set to $2 \times 10^{-4}$, and for models trained for more than 50 epochs, the learning rate was set to $2 \times 10^{-4}$ in the first 50 epochs and decreased by $4 \times 10^{-6}$ in each remaining epoch. While training the joint network, the neural network layers in the one-to-one and many-to-one streams prior to the fusion block were also fine-tuned. To do this, the Adam optimizer was employed with half the learning rate of the joint network. The decay rates of the first moment $\beta_1$ and the second moment $\beta_2$ of gradient estimates were again set to 0.5 and 0.999, respectively.

The total number of network parameters was approximately $1.1 \times 10^7$ for the one-to-one and many-to-one streams, and $2.3 \times 10^6$ for the discriminators. Meanwhile, it varied from $10^4$ to $1.1 \times 10^7$ for the joint network depending on the fusion level. On a single NVIDIA 1080 Ti GPU, training each individual stream required 8–10 hours, and training a single joint network with fine-tuning of one-to-one and many-to-one streams required 14–16 hours (with 100 epochs and 2500 training instances). For synthesis tasks with two source contrasts, one-to-one and many-to-one streams were first trained in parallel on 3 GPUs. The candidate joint networks at 14 different fusion layers were trained on 7 GPUs in two batches. This resulted in a total training time of nearly 40 hours per synthesis task.

### 3.5. Competing methods

Four state-of-the-art multi-contrast MRI synthesis methods were implemented to comparatively evaluate the performance of the proposed method. The first competing method was pGAN that performs one-to-one mapping with adversarial and pixel-wise losses (Dar et al., 2019). Since this method receives as input a single source contrast, multiple pGAN models were trained for target contrast recovery based on available source contrasts. The second competing method was pGAN$_{many}$ that is a many-to-one variant of pGAN, where the number of input channels in the architecture was set to the number of available source contrasts. Note that pGAN and pGAN$_{many}$ correspond to the independently trained one-to-

one and many-to-one streams in mustGAN, respectively. Therefore, these methods were trained with the same hyperparameter set.

The third competing method was MM-GAN that recovers the target contrast from all available source contrasts by concatenation at the input level (Sharma and Hamarneh, 2019). MM-GAN was also trained in an adversarial setup with pixel-wise and adversarial losses. MM-GAN was implemented based on the architecture/hyperparameters described in Sharma and Hamarneh (2019) except for the curriculum learning to ensure a standard sample selection procedure among GAN models during trainings. The last competing method was Multimodal (Chartsias et al., 2017), a standard convolutional encoder-decoder network. With multiple encoders, Multimodal generates contrast-invariant latent representations that are later fused with a maximum function. The target contrast is then recovered by either the individual latent representations or the fused latent representations. Official code posted by the authors of Multimodal was used with the procedure/hyperparameters specified in Chartsias et al. (2017).

In addition, we also implemented a constrained variant of mustGAN that only performed fusion across one-to-one streams, named mustGAN$_{one}$. Similar to mustGAN, the one-to-one streams were trained independently, and then the position of the fusion block and number of epochs were optimized for each specific task. The optimal values denoted as (fusion block position, number of epochs) were determined to be $(14, 35)$ for $T_1$, $(10, 25)$ for $T_2$, and $(6, 20)$ for PD synthesis in the IXI dataset, and $(8, 45)$ for $T_1$, $(5, 55)$ for $T_2$, and $(12, 20)$ for FLAIR synthesis in the ISLES dataset. The remaining hyperparameters were matched to mustGAN.

### 3.6. Experiments

Two public multi-contrast MRI datasets (IXI and ISLES) were used to evaluate the performance of the proposed method against the competing methods. In the IXI dataset, 6 distinct synthesis tasks ($T_2 \rightarrow T_1$; PD $\rightarrow T_1$; $T_1 \rightarrow T_2$; PD $\rightarrow T_2$; $T_1 \rightarrow$ PD; $T_2 \rightarrow$ PD) were considered for pGAN. Meanwhile, 3 distinct synthesis tasks ($T_2$, PD $\rightarrow T_1$; $T_1$, PD $\rightarrow T_2$; $T_1$, $T_2 \rightarrow$ PD) were considered for mustGAN, pGAN$_{many}$, MM-GAN, and Multimodal. All synthesis tasks in the IXI dataset were performed using axial cross-sections. Overall, 6 pGAN, 3 mustGAN, 3 pGAN$_{many}$, 1 MM-GAN, and 3 Multimodal models were trained.

In the ISLES dataset, 6 distinct synthesis tasks ($T_2 \rightarrow T_1$; FLAIR $\rightarrow T_1$; $T_1 \rightarrow T_2$; FLAIR $\rightarrow T_2$; $T_1 \rightarrow$ FLAIR; $T_2 \rightarrow$ FLAIR) were considered for pGAN. Meanwhile, 3 distinct synthesis tasks ($T_2$, FLAIR $\rightarrow T_1$; $T_1$, FLAIR $\rightarrow T_2$; $T_1$, $T_2 \rightarrow$ FLAIR) were considered for mustGAN, pGAN$_{many}$, MM-GAN, and Multimodal. $T_1$-weighted and FLAIR images were synthesized using axial cross-sections, whereas $T_2$-weighted images were synthesized using sagittal cross-sections (see Section 3.3 for details). Overall, 6 pGAN, 3 mustGAN, 3 pGAN$_{many}$, and 3 Multimodal models were trained. For fair comparison, an MM-GAN model was trained using only sagittal cross-sections for $T_2$ synthesis, and a separate MM-GAN model was trained using only axial cross-sections for $T_1$ and FLAIR synthesis.

For quantitative evaluations, the synthesized and ground truth target images were compared via PSNR and SSIM (Wang et al., 2004). Prior to measurements, maximum pixel intensity of the synthesized and ground truth cross-sectional images was normalized to 1. For qualitative evaluations, opinion scores of an expert radiologist with 25 years of experience were considered. To do this, an intermediate cross-section from each subject was randomly selected. The quality of the synthesized images was then rated by the radiologist by evaluating their similarity to the ground truth images using a five-point scale (0: unacceptable, 1: very poor, 2: limited, 3: moderate, 4: good, 5: perfect match). To assess significance of PSNR, SSIM and radiological evaluation score differ-
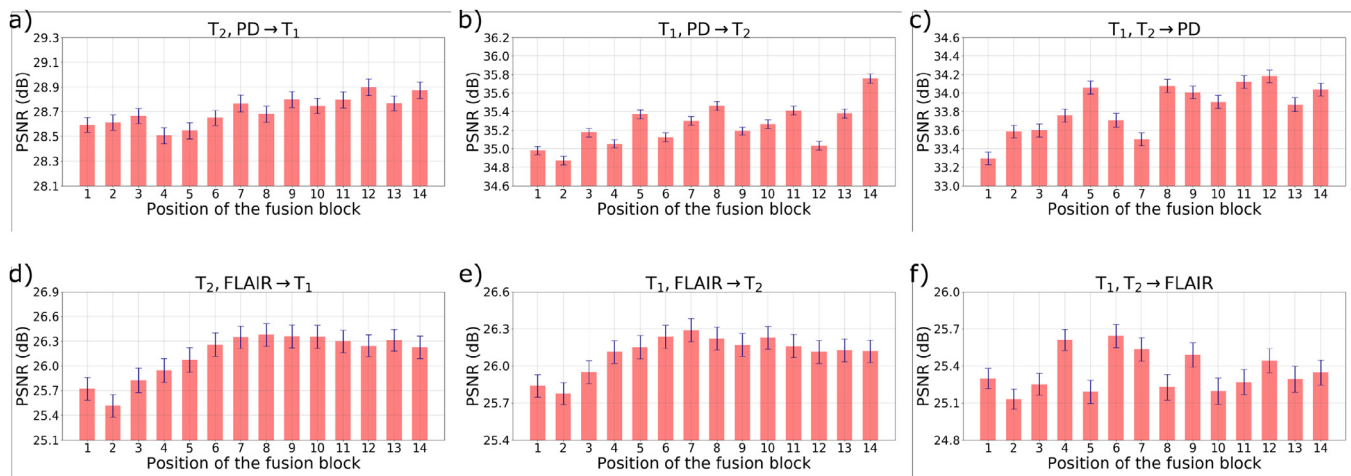
**Fig. 2.** Synthesis quality of mustGAN models with varying positions of the fusion block was evaluated on the validation set in terms of PSNR measurements. The mean of the results are plotted together with standard error intervals for three distinct synthesis tasks in the IXI dataset: a) $T_2$, PD $\rightarrow$ $T_1$, b) $T_1$, PD $\rightarrow$ $T_2$, c) $T_1$, $T_2$ $\rightarrow$ PD, and for three distinct synthesis tasks in the ISLES dataset: d) $T_2$, FLAIR $\rightarrow$ $T_1$, e) $T_1$, FLAIR $\rightarrow$ $T_2$, f) $T_1$, $T_2$ $\rightarrow$ FLAIR.
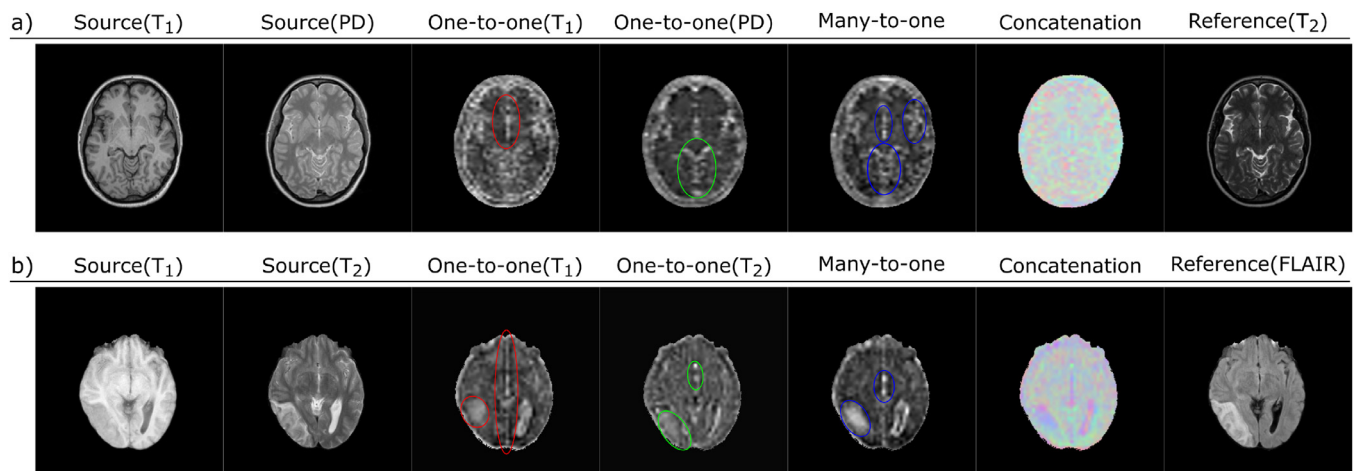


**Fig. 3.** a) $T_2$-weighted image synthesis from $T_1$- and PD-weighted images in the IXI dataset, b) FLAIR image synthesis from $T_1$- and $T_2$-weighted images in the ISLES dataset. Feature maps at a central network layer (8th) in one-to-one (third and fourth columns) and many-to-one models (fifth column) are displayed along with source images and ground truth target images (reference). At the $8^{th}$ network layer, the feature maps have tensor dimensions of $[t_1 = 1, t_2 = 256, t_3 = 64, t_4 = 64]$, where $t_1$ denotes the batch size, $t_2$ denotes the number of channels, $t_3$ denotes the height, and $t_4$ denotes the width. Lastly, the feature maps were averaged along the second tensor dimension, and were resampled to achieve a final image size of $256 \times 256$. Representative features that are captured with increased sensitivity for each model are marked with ellipses (red and green color for the one-to-one models, blue color for the many-to-one models). A concatenation of the feature maps is also shown (sixth column), where separate feature maps from the one-to-one and many-to-one mappings are taken as the red-green-blue color channels. Overall, one-to-one mappings manifest increased sensitivity to unique, detailed features that are predominantly present in the given source. Contrarily, many-to-one mapping pools information from different sources and manifests high sensitivity to shared features, especially when these features are jointly present in multiple sources. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ences between the competing methods, non-parametric Wilcoxon signed-rank tests were employed. All methods were trained and tested on the same set of data samples. Training and evaluation procedures were run on NVIDIA Titan X Pascal, Xp and 1080 Ti GPUs. Implementations of mustGAN, pGAN, pGAN$_{many}$, and MM-GAN were performed via the PyTorch framework in Python, and implementation of Multimodal was performed via Keras using the Theano backend in Python. Code and data for replicating the must-GAN, pGAN and pGAN$_{many}$ models will be publicly available on http://github.com/icon-lab/mrirecon.

## 4. Results

### 4.1. Task-Specific fusion across multiple streams

To optimize the mustGAN model for specific tasks, we performed experiments to determine the optimal position of the fu-

sion block in the architecture. Multiple mustGAN models were trained while varying the layer of fusion in [1 : 1 : 14]. Experiments were conducted separately on the IXI and ISLES datasets. 3 synthesis tasks were considered in the IXI dataset: $T_2$, PD $\rightarrow$ $T_1$; $T_1$, PD $\rightarrow$ $T_2$; $T_1$, $T_2$ $\rightarrow$ PD. Performance as a function of fusion layer is plotted in terms of PSNR measurements on the validation set in Fig. 2a-c for $T_1$-weighted, $T_2$-weighted and PD-weighted image synthesis in the IXI dataset, respectively. Across all synthesis tasks in the IXI dataset, mustGAN models performing late fusion mostly yield enhanced synthesis performance. Particularly, the optimal position of the fusion block is determined to be the 12th layer for $T_1$ synthesis, the 14th layer for $T_2$ synthesis, and the 12th layer for PD synthesis. Furthermore, optimization of the fusion layer noticeably improves model performance, where PSNR difference between highest-lowest performing models is 0.390 dB for $T_1$ synthesis, 0.885 dB for $T_2$ synthesis, and 0.885 dB for PD synthesis.

In the ISLES dataset, 3 distinct synthesis tasks were considered: $T_2$, FLAIR $\rightarrow T_1$; $T_1$, FLAIR $\rightarrow T_2$; $T_1$, $T_2 \rightarrow$ FLAIR. Performance as a function of fusion layer is plotted in terms of PSNR measurements on the validation set in Fig. 2d-f for $T_1$-weighted, $T_2$-weighted and FLAIR image synthesis, respectively. Across all synthesis tasks in the ISLES dataset, mustGAN models performing intermediate fusion mostly yield enhanced synthesis performance. Particularly, the optimal position of the fusion block is determined to be the 8th layer for $T_1$ synthesis, the 7th layer for $T_2$ synthesis, and the 6th layer for FLAIR synthesis. Again, optimization of the fusion layer noticeably improves model performance, where PSNR difference between highest-lowest performing models is 0.860 dB for $T_1$ synthesis, 0.525 dB for $T_2$ synthesis, and 0.511 dB for FLAIR synthesis. These task-specific fusion layers identified on the validation sets in the IXI and ISLES datasets were utilized in all evaluations thereafter unless otherwise stated.

Here, we observed that the optimal fusion block position varies as a function of both synthesis task and MRI dataset. Yet, separate synthesis tasks performed on the same dataset yield limited changes in block position, whereas separate datasets with distinct MRI protocols result in more substantial differences in block position. In IXI, synthesis quality is enhanced by performing the fusion within the decoder, where the fused feature maps have larger width and height and so they reflect a high-resolution representation. On the other hand, in ISLES, synthesis quality is enhanced by performing the fusion within the residual block, where the fused feature maps have smaller size, reflecting a relatively lower-resolution representation. Note that the IXI dataset contains high-quality, high-SNR images, so fusion at the decoder might help better recover fine structural details. In contrast, the ISLES dataset mostly contains images of relatively moderate quality, so fusing at the residual block might help better recover global structural information.

### 4.2. Demonstrations against one-to-one and many-to-one mappings

We then performed experiments to demonstrate potential differences in feature maps learned in one-to-one versus many-to-one mappings. Three synthesis tasks were considered in the IXI dataset ($T_2$, PD $\rightarrow T_1$; $T_1$, PD $\rightarrow T_2$; $T_1$, $T_2 \rightarrow$ PD) and in the ISLES dataset ($T_2$, FLAIR $\rightarrow T_1$; $T_1$, FLAIR $\rightarrow T_2$; $T_1$, $T_2 \rightarrow$ FLAIR). Representative feature maps generated in the one-to-one and many-to-one mappings are displayed along with the source and ground truth target images in Fig. 3 and in Supp. Fig. 7. The feature maps indicate that one-to-one mappings sensitively capture detailed features that are uniquely present in the given source, whereas many-to-one mapping pools information across shared features that are jointly present in multiple sources.

To assess benefits of pooling complementary information from unique and shared feature maps, we compared pGAN, pGAN$_{many}$ and mustGAN models. Comparisons in terms of PSNR measured across cross-sections in the test sets are displayed in Supp. Figs. 8–10 for IXI, and in Fig. 4 and Supp. Figs. 11,12 for ISLES. On average, pGAN$_{many}$ outperforms pGAN for 81.98% of test samples in IXI and for 63.14% in ISLES; whereas pGAN outperforms pGAN$_{many}$ for 18.02% in IXI and for 36.86% in ISLES. This finding demonstrates that not only shared but also unique features can be critical for successful synthesis of the target contrast. In comparison, mustGAN outperforms both competing methods, with higher PSNR than pGAN for 92.20% of test samples in IXI and for 87.19% in ISLES, and with higher PSNR than pGAN$_{many}$ for 88.26% in IXI and for 81.94% in ISLES. Taken together, these results indicate that aggregation of information from unique and shared feature maps helps significantly improve model performance.

We further demonstrated the improvements in synthesis quality due to inclusion of the many-to-one stream within the pro-
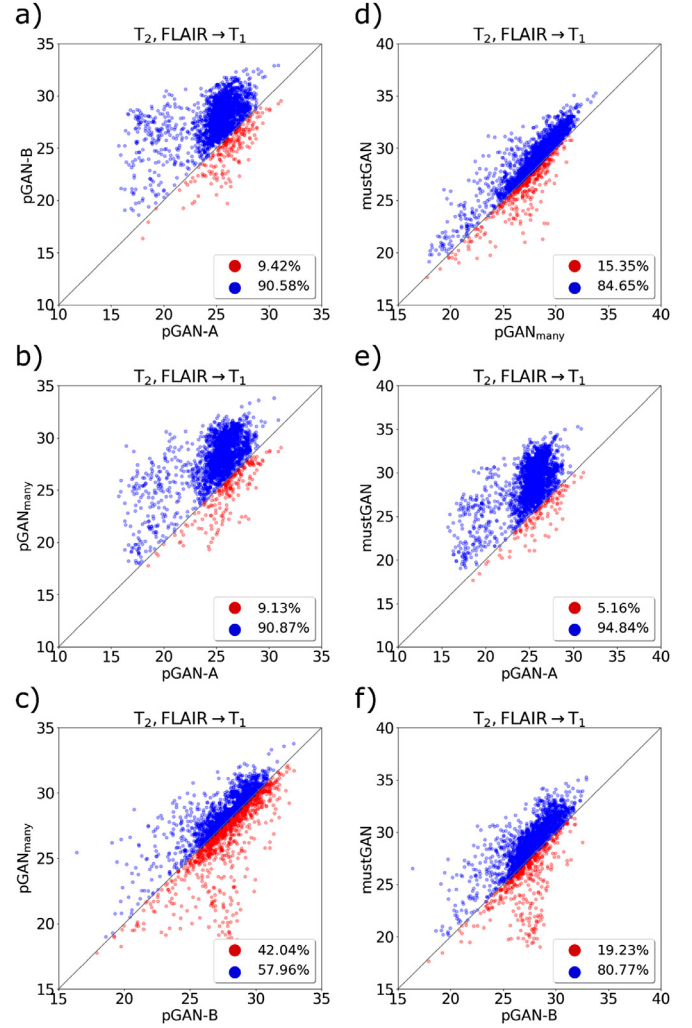


**Fig. 4.** Methods were compared in terms of quality of $T_1$ synthesis in the ISLES dataset: a) pGAN-A versus pGAN-B, b) pGAN-A versus pGAN$_{many}$, c) pGAN-B versus pGAN$_{many}$, d) pGAN$_{many}$ versus mustGAN, e) pGAN-A versus mustGAN, f) pGAN-B versus mustGAN. Note that pGAN-A receives $T_2$-weighted images as input and pGAN-B receives FLAIR images as input. Scatter plots show PSNR measurements for methods under comparison, and each point denotes a cross-section in the test set. The proportion of test samples in which either method yields superior performance is also noted in figure legends (blue font for the method on the vertical axis, red font for the method on the horizontal axis). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

posed method. To do this, we introduced a mustGAN variant, namely mustGAN$_{one}$, that recovers the target contrast image from fusion of only the unique feature maps generated in the one-to-one streams. We then compared mustGAN with mustGAN$_{one}$ on the IXI and ISLES datasets. Note that, we also comprehensively optimized the position of the fusion block and the number of epochs for this variant on the validation set for specific tasks (see Section 3.5 for details). Supp. Table 1 reports the PSNR and SSIM measurements of mustGAN and mustGAN$_{one}$ for all possible synthesis tasks in the datasets. The results indicate that the aggregation of the shared feature maps in the many-to-one stream enhances synthesis quality, where the average increase is 0.16 dB PSNR and 0.2% SSIM in the IXI dataset, and 0.24 dB PSNR and 0.4% SSIM in the ISLES dataset.

### 4.3. Demonstrations against competing methods

Next, we comparatively evaluated the performance of mustGAN against several state-of-the-art one-to-one and many-to-one

**Table 1**

Quality of Synthesis in the IXI Dataset: PSNR and SSIM measurements between the ground truth and synthesized target images from mustGAN, pGAN, pGAN$_{many}$, MM-GAN, and Multimodal are given as mean±std calculated across test subjects for three different synthesis tasks: $T_2$, PD $\rightarrow T_1$; $T_1$, PD $\rightarrow T_2$; $T_1$, $T_2 \rightarrow$ PD. pGAN-A receives the 1st source contrast and pGAN-B receives the 2nd source contrast i.e., (1,2): $(T_2, PD)$, $(T_1, PD)$, $(T_1, T_2)$. Boldface marks the model having the highest performance.

| | $T_2$, PD $\rightarrow T_1$ | | $T_1$, PD $\rightarrow T_2$ | | $T_1$, $T_2 \rightarrow$ PD | |
|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| mustGAN | **29.45** | **0.947** | **35.89** | **0.977** | **34.40** | **0.974** |
| | **±1.19** | **±0.012** | **±1.20** | **±0.005** | **±0.97** | **±0.005** |
| pGAN-A | 28.39 | 0.934 | 28.52 | 0.925 | 27.80 | 0.929 |
| | ±1.17 | ±0.013 | ±1.18 | ±0.015 | ±1.16 | ±0.012 |
| pGAN-B | 28.73 | 0.936 | 33.08 | 0.962 | 32.17 | 0.962 |
| | ±1.18 | ±0.013 | ±0.99 | ±0.007 | ±1.01 | ±0.005 |
| pGAN$_{many}$ | 28.80 | 0.940 | 34.04 | 0.964 | 33.09 | 0.967 |
| | ±1.09 | ±0.013 | ±1.18 | ±0.006 | ±1.09 | ±0.005 |
| MM-GAN | 28.27 | 0.943 | 32.67 | 0.972 | 33.35 | 0.971 |
| | ±1.51 | ±0.013 | ±1.53 | ±0.005 | ±1.65 | ±0.007 |
| Multimodal | 27.35 | 0.934 | 34.56 | 0.965 | 32.31 | 0.958 |
| | ±1.69 | ±0.013 | ±1.04 | ±0.006 | ±1.16 | ±0.005 |

methods (pGAN, pGAN$_{many}$, MM-GAN, and Multimodal). Three synthesis tasks were considered in the IXI dataset: $T_2$, PD $\rightarrow T_1$; $T_1$, PD $\rightarrow T_2$; $T_1$, $T_2 \rightarrow$ PD. Table 1 lists the average PSNR and SSIM measurements across test subjects of mustGAN, pGAN, pGAN$_{many}$, MM-GAN, and Multimodal. In all synthesis tasks, mustGAN outperforms the competing methods in terms of PSNR and SSIM measurements ($p < .05$). On average, mustGAN achieves 1.01 *dB* higher PSNR and 0.416% higher SSIM compared to the second-best performing method in each synthesis task.

Confirming quantitative results, radiological evaluations in the IXI dataset also reveal superiority of the proposed method. Fig. 5a-c display radiological scores for $T_1$-, $T_2$- and PD-weighted image synthesis. mustGAN outperforms the competing methods in all synthesis tasks ($p < .05$) in terms of radiological evaluation scores, except for pGAN$_{many}$ in $T_2$ synthesis, where the two methods perform similarly ($p > .05$). On average across tasks, mustGAN achieves 3.981 radiological opinion score, whereas the second-best competing method (pGAN$_{many}$) achieves 3.685. Superior performance of mustGAN on the IXI dataset is also clearly visible in representative results shown in Fig. 6 and Supp. Fig. 13. Fig. 6a,b display results for $T_1$-weighted and PD-weighted image synthesis. Compared to other methods, mustGAN depicts tissues with apparently lower noise levels and sharper tissue boundaries.

Having demonstrated mustGAN on healthy subjects, we next evaluated mustGAN on the ISLES dataset containing images of high/low grade glioma patients. Three synthesis tasks were considered: $T_2$, FLAIR $\rightarrow T_1$; $T_1$, FLAIR $\rightarrow T_2$; $T_1$, $T_2 \rightarrow$ FLAIR. Table 2 lists the PSNR and SSIM measurements of mustGAN, pGAN, pGAN$_{many}$, MM-GAN, and Multimodal on the test set. mustGAN again outperforms the competing methods in all synthesis tasks in terms of PSNR and SSIM measurements ($p < .05$). The only exception is Multimodal for SSIM in $T_1$ and $T_2$ synthesis. On average, mustGAN achieves 0.77 *dB* higher PSNR compared to the second-best performing method in each synthesis task. Radiological evaluations in the ISLES dataset concur with the quantative measurements of synthesis quality. Fig. 5d-f display radiological scores for $T_1$-weighted, $T_2$-weighted, and FLAIR image synthesis. mustGAN outperforms the competing methods in all synthesis tasks in terms of radiological evaluation scores ($p < .05$). On average across tasks, mustGAN achieves 3.475 radiological opinion score, whereas the second-best method (pGAN$_{many}$) achieves 2.780. Superior performance of mustGAN on the ISLES dataset is also clearly visible in the representative results shown in Fig. 7 and Supp. Fig. 14. Fig. 7a,b display results for $T_1$-weighted and FLAIR image synthesis.
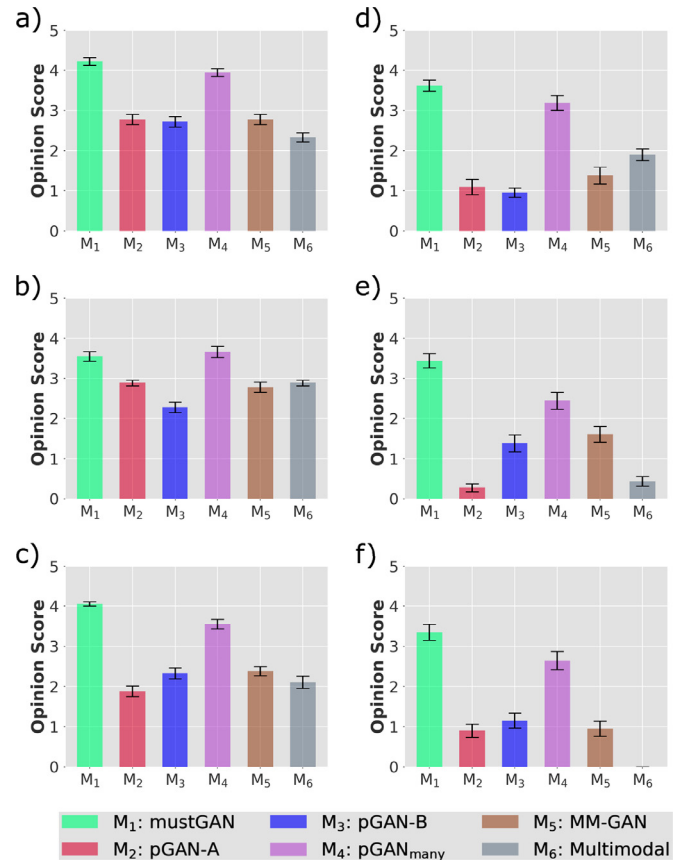


**Fig. 5.** Methods were compared in terms of radiological opinion scores for three synthesis tasks in IXI: a) $T_2$, PD $\rightarrow T_1$, b) $T_1$, PD $\rightarrow T_2$, c) $T_1$, $T_2 \rightarrow$ PD, and for three synthesis tasks in ISLES: d) $T_2$, FLAIR $\rightarrow T_1$, e) $T_1$, FLAIR $\rightarrow T_2$, f) $T_1$, $T_2 \rightarrow$ FLAIR. The quality of the synthesized cross-sections were rated by an expert radiologist by evaluating their similarity to the ground truth images using a five-point scale (0: unacceptable, 1: very poor, 2: limited, 3: moderate, 4: good, 5: perfect match). For each synthesis task, an intermediate cross-section from each of the 18 test subjects in IXI and 21 test subjects in ISLES were evaluated. Due to poor quality ground truths, images of 3 subjects for $T_2$ synthesis and image of a subject for FLAIR synthesis were removed by the radiologist. The resulting radiological opinion scores of the methods are displayed along with standard error intervals. Figure legend denotes the abbreviations and colors used for the methods under comparison. pGAN-A receives the 1st source contrast and pGAN-B receives the 2nd source contrast i.e., in IXI: (1, 2), $(T_2, PD)$, $(T_1, PD)$, $(T_1, T_2)$, and in ISLES: (1, 2), $(T_2, FLAIR)$, $(T_1, FLAIR)$, $(T_1, T_2)$.

**Table 2**

Quality of Synthesis in the ISLES Dataset: PSNR and SSIM measurements between the ground truth and synthesized target images from mustGAN, pGAN, pGAN$_{many}$, MM-GAN, and Multimodal are given as mean±std calculated across test subjects for three different synthesis tasks: $T_2$, FLAIR $\rightarrow T_1$; $T_1$, FLAIR $\rightarrow T_2$; $T_1$, $T_2 \rightarrow$ FLAIR. pGAN-A receives the 1st source contrast and pGAN-B receives the 2nd source contrast i.e., (1,2): $(T_2, FLAIR)$, $(T_1, FLAIR)$, $(T_1, T_2)$. Boldface marks the model having the highest performance.

| | $T_2$, FLAIR$\rightarrow T_1$ | | $T_1$, FLAIR$\rightarrow T_2$ | | $T_1$, $T_2 \rightarrow$FLAIR | |
|---|---|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| mustGAN | **28.51** | 0.929 | **26.63** | 0.904 | **26.08** | **0.910** |
| | **±2.10** | ±0.018 | **±0.74** | ±0.013 | **±1.04** | **±0.016** |
| pGAN-A | 25.03 | 0.886 | 25.36 | 0.884 | 24.91 | 0.889 |
| | ±1.92 | ±0.015 | ±0.70 | ±0.012 | ±0.94 | ±0.015 |
| pGAN-B | 27.55 | 0.919 | 25.68 | 0.890 | 23.32 | 0.861 |
| | ±1.35 | ±0.015 | ±0.70 | ±0.012 | ±0.67 | ±0.012 |
| pGAN$_{many}$ | 27.64 | 0.921 | 26.00 | 0.895 | 25.11 | 0.894 |
| | ±1.88 | ±0.017 | ±0.68 | ±0.012 | ±0.81 | ±0.013 |
| MM-GAN | 27.79 | 0.924 | 25.67 | 0.888 | 24.57 | 0.898 |
| | ±1.65 | ±0.017 | ±0.54 | ±0.010 | ±0.88 | ±0.015 |
| Multimodal | 27.73 | **0.934** | 25.71 | **0.910** | 19.62 | 0.895 |
| | ±1.91 | **±0.017** | ±0.66 | **±0.011** | ±1.29 | ±0.011 |
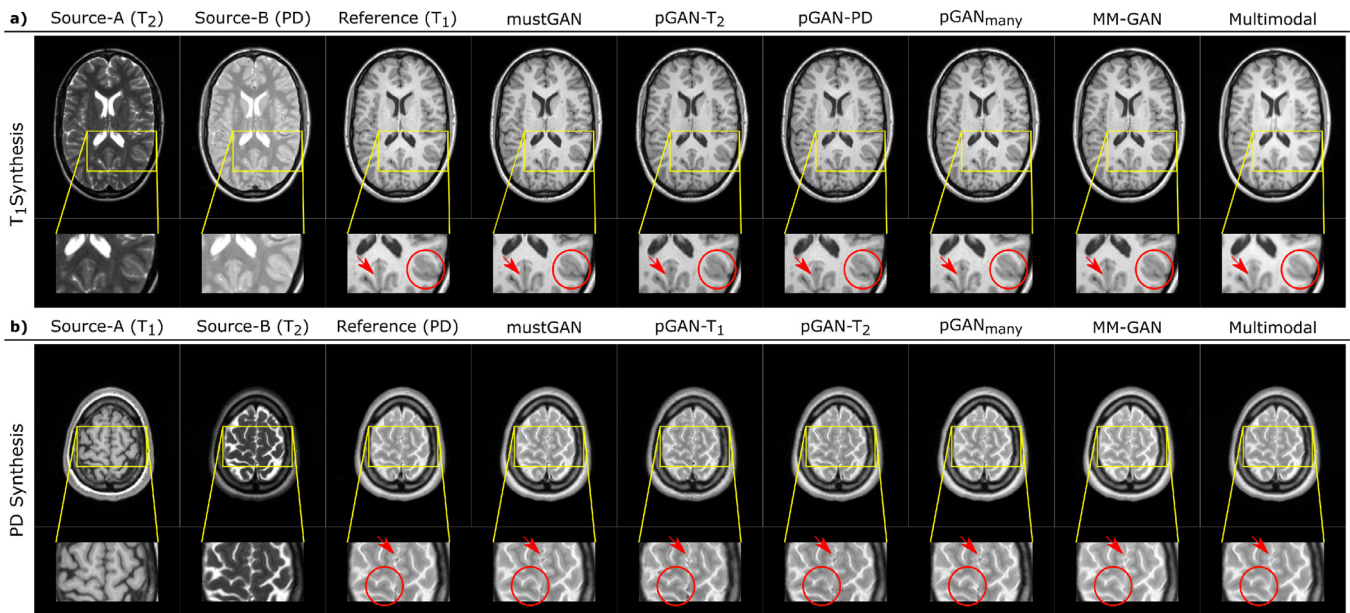
**Fig. 6.** The proposed method was demonstrated on healthy subjects from the IXI dataset for two synthesis tasks: a) $T_1$-weighted image synthesis from $T_2$- and PD-weighted images, b) PD-weighted image synthesis from $T_1$- and $T_2$-weighted images. Synthesized images from mustGAN, pGAN, pGAN$_{many}$, MM-GAN, and Multimodal are shown along with the source images and the ground truth target image. Due to synergistic use of information captured by the one-to-one and many-to-one streams, mustGAN improves synthesis accuracy in many regions that are recovered suboptimally in competing methods (marked with arrows or circles in zoom-in displays). Overall, mustGAN yields less noisy depiction of tissues and sharper depiction of tissue boundaries.
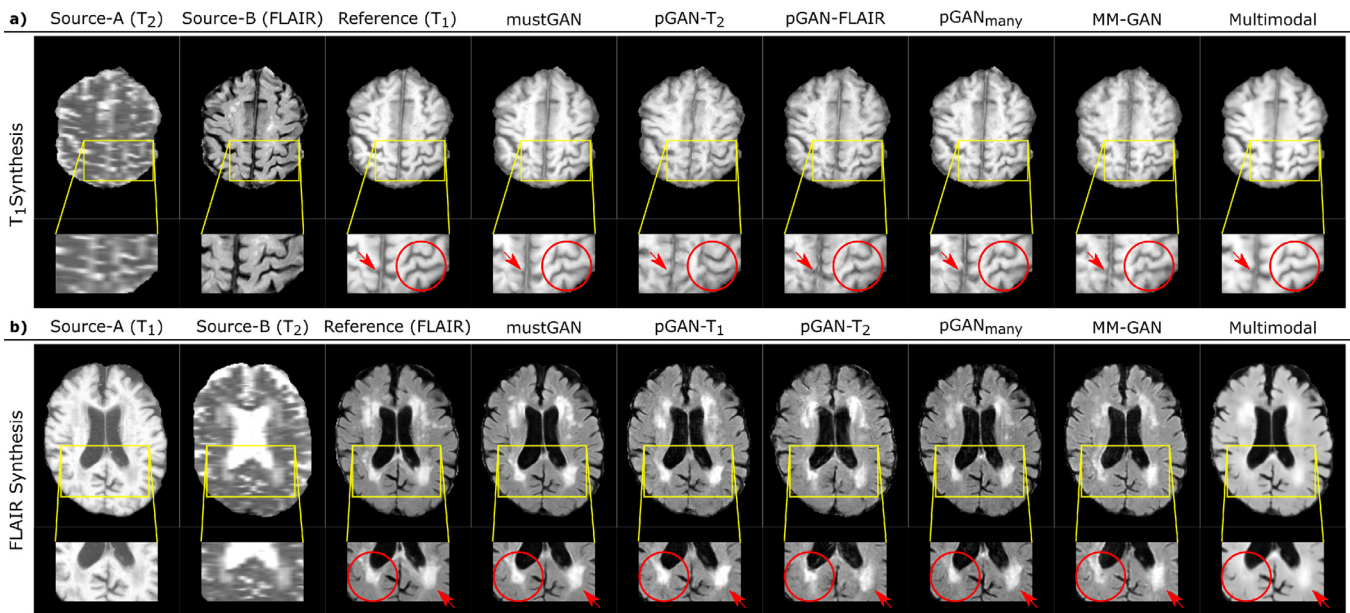


**Fig. 7.** The proposed method was demonstrated on high/low grade glioma patients from the ISLES dataset for three synthesis tasks: a) $T_1$-weighted image synthesis from $T_2$-weighted and FLAIR images, b) FLAIR image synthesis from $T_1$- and $T_2$-weighted images. Synthesized images from mustGAN, pGAN, pGAN$_{many}$, MM-GAN, and Multimodal are shown along with the source images and the ground truth target image. Due to synergistic use of information captured by the one-to-one and many-to-one streams, mustGAN improves synthesis accuracy in many regions that are recovered suboptimally in competing methods (marked with arrows or circles in zoom-in displays). Overall, mustGAN yields less noisy depiction of tissues and sharper depiction of tissue boundaries.

Compared to other methods, mustGAN depicts tissues with apparently lower noise levels and sharper tissue boundaries.

For further evaluations, the performance of the mustGAN models with varying fusion block positions are additionally reported in Supp. Table 2 in terms of PSNR and SSIM measurements on the test set for all synthesis tasks in the IXI and ISLES datasets. These results indicate that even non-optimized mustGAN variants outperform state-of-the-art methods for almost all fusion levels. Therefore, the computational cost of the fusion block optimization in mustGAN (see Section 4.1 for details) can be effectively alleviated

by performing a coarser search over fewer positions, without significantly compromising synthesis quality.

## 5. Discussion

A within-modality synthesis method was introduced for multi-contrast MRI based on conditional generative adversarial networks. The proposed method aggregates information across one-to-one streams that are sensitive to unique information in individual source contrasts and a many-to-one stream that is sensitive

to shared information across multiple source contrasts. Enhanced synthesis performance was demonstrated in a number of synthesis tasks on brain MRI datasets from normals and glioma patients. Compared to isolated one-to-one or many-to-one methods, mustGAN recovered higher quality images with reduced noise and improved sharpness.

A prior state-of-the-art method for multi-contrast MRI synthesis, Multimodal, is based on an encoder-decoder architecture with standard convolutional layers (Chartsias et al., 2017). Given multiple source contrasts, Multimodal learns contrast-invariant latent representations for source images by enforcing latent representations from separate encoders to be similar. These individual latent representations are then fused across source contrasts via a maximum function, and the decoder recovers target images based on fused representations. For improved sensitivity to unique features of individual sources, mustGAN does not explicitly seek similarity across latent representations in one-to-one streams and instead uses a separate many-to-one stream to capture shared representations across source contrasts. While the position of the fusion block is fixed to the initial layer of the decoder in Multimodal, the proposed method adaptively modifies the position of the fusion block to optimize the task-specific performance. Moreover, unlike Multimodal that uses a mean absolute error metric, mustGAN uses adversarial loss that has been demonstrated to better capture high-spatial-frequency information (Dar et al., 2019).

Several recent studies have proposed GAN-based architectures for multi-contrast MRI synthesis. In Dar et al. (2019), we have proposed pGAN that uses conditional GAN models for one-to-one synthesis. In Sharma and Hamarneh (2019), a multi-input generalization of pGAN was proposed, MM-GAN, that receives as input multiple source contrasts for enhanced recovery of missing contrasts. MM-GAN fuses multiple source contrast at the input level by treating them as separate information channels, and so it is similar in nature to $pGAN_{many}$ implemented here. Our results indicate that, compared to both $pGAN_{many}$ and MM-GAN, mustGAN achieves enhanced sensitivity to unique features of individual source contrasts due to the adaptive fusion and presence of additional one-to-one streams.

Different medical imaging modalities can provide complementary information of the underlying anatomy and thereby can enhance the diagnostic utility. Yet, acquiring an entire multi-modal set may not be feasible due to scanning cost and time limitations. In such cases, cross-modality medical image synthesis can be critical. Several studies based on GAN architectures have been recently proposed for cross-modality synthesis, CT synthesis from MRI: (Nie et al., 2017; 2018; Emami et al., 2018), MRI synthesis from CT: (Jiang et al., 2018; Jin et al., 2019), PET synthesis from CT: (Ben-Cohen et al., 2019; Bi et al., 2017), CT synthesis from PET: (Armanious et al., 2020), PET synthesis from MRI: (Pan et al., 2018; Wei et al., 2018a), MRI synthesis from PET: (Choi and Lee, 2018). Although the proposed multi-stream GAN method was primarily demonstrated for within-modality multi-contrast MRI synthesis here, it can also be adapted to cross-modality synthesis. In such cases, with synergistic use of multiple one-to-one streams and a many-to-one stream, mustGAN can offer improved performance due to better capture of unique and complementary features in different source modalities.

An important requirement for successful training of deep network architectures is the availability of large datasets. The current implementation of mustGAN assumes availability of paired source-target images from the same group of subjects. However, size of paired datasets might be limited especially when relatively less common contrasts are involved. In such cases, several lines of improvement can be considered. (1) When the source images are all paired but the target images are unpaired, the pixel-wise loss used in one-to-one and many-to-one streams can be replaced with a

cycle-consistency loss. Training procedures for the cycle-consistent models can be adopted from prior studies for both one-to-one (Dar et al., 2019) and many-to-one (Lee et al., 2019) GAN models. (2) When the source images are also unpaired, the many-to-one stream can be removed. The one-to-one streams can again be trained with a cycle-consistency loss and then fused for enhanced performance.

The proposed network model takes as input spatially registered source and target images. The datasets analyzed in this study were either pre-registered, or registration was implemented as a pre-processing step (see Methods for procedures on the IXI dataset). When an end-to-end network alternative is desired, deep-network-based registration models (Yang et al., 2017) can instead be cascaded to the input of mustGAN to spatially align source-target images. It remains important future work to investigate potential benefits of an end-to-end registration approach over pre-processing.

A central aim of the experiments in this study was to demonstrate improvements in synthesis quality over prior state-of-the-art with the novel mustGAN architecture. These demonstrations involved comprehensive quantitative (i.e., PSNR and SSIM measurement) and radiological evaluations for numerous multi-contrast MRI synthesis tasks in multiple datasets. It remains important future work to examine to what extent improvements in image quality translate to downstream tasks such as segmentation, registration or detection.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

**Mahmut Yurt:** Conceptualization, Methodology, Data curation, Software, Formal analysis, Visualization, Investigation, Validation, Writing - original draft. **Salman UH Dar:** Data curation, Software, Formal analysis, Visualization, Investigation, Validation, Writing - original draft. **Aykut Erdem:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Erkut Erdem:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Kader K Oguz:** Formal analysis, Investigation, Validation, Writing - review & editing. **Tolga Çukur:** Conceptualization, Methodology, Supervision, Funding acquisition, Writing - original draft, Writing - review & editing.

## Acknowledgments

## Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.media.2020.101944.

## References

Abramian, D., Eklund, A., 2019. Refacing: Reconstructing anonymized facial features using GANs. In: IEEE 16th International Symposium on Biomedical Imaging (ISBI), pp. 1104–1108.

Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S., Yang, B., 2020. MedGAN: medical image translation using GANs. Computerized Medical Imaging and Graphics 79, 101684.

Beers, A., Brown, J., Chang, K., Campbell, J.P., Ostmo, S., Chiang, M.F., Kalpathy-Cramer, J., 2018. High-resolution medical image synthesis using progressively grown generative adversarial networks. arXiv preprint arXiv:1805.03144.

Ben-Cohen, A., Klang, E., Raskin, S.P., Soffer, S., Ben-Haim, S., Konen, E., Amitai, M.M., Greenspan, H., 2019. Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection. Eng. Appl. Artif. Intell. 78, 186–194.

Bermudez, C., Plassard, A.J., Davis, L.T., Newton, A.T., Resnick, S.M., Landman, B.A., 2018. Learning implicit brain MRI manifolds with deep learning. In: Medical Imaging 2018: Image Processing, 10574. International Society for Optics and Photonics, p. 105741L.

Bi, L., Kim, J., Kumar, A., Feng, D., Fulham, M., 2017. Synthesis of Positron Emission Tomography (PET) Images via Multi-channel Generative Adversarial Networks (GANs). In: Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment. Springer, pp. 43–51.

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

Bowles, C., Qin, C., Ledig, C., Guerrero, R., Gunn, R., Hammers, A., Sakka, E., Dickie, D.A., Hernández, M.V., Royle, N., Wardlaw, J., Rhodius-Meester, H., Tijms, B., Lemstra, A.W., van der Flier, W., Barkhof, F., Scheltens, P., Rueckert, D., 2016. Pseudo-healthy image synthesis for white matter lesion segmentation. In: Simulation and Synthesis in Medical Imaging. Springer International Publishing, Cham, pp. 87–96.

Cai, J., Zhang, Z., Cui, L., Zheng, Y., Yang, L., 2019. Towards cross-modal organ translation and segmentation: a cycle-and shape-consistent generative adversarial network. Med. Image Anal. 52, 174–184.

Calimeri, F., Marzullo, A., Stamile, C., Terracina, G., 2017. Biomedical data augmentation using generative adversarial neural networks. In: International Conference on Artificial Neural Networks. Springer, pp. 626–634.

Chartsias, A., Joyce, T., Valerio Giuffrida, M., Tsaftaris, S., 2017. Multimodal MR synthesis via modality-invariant latent representation. IEEE Trans. Med. Imaging 37, 803–814.

Choi, H., Lee, D.S., 2018. Generation of structural MR images from amyloid PET: application to MR-less quantification. J. Nucl. Med. 59 (7), 1111–1117.

Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., Choo, J., 2018. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Choi, Y., Uh, Y., Yoo, J., Ha, J.-W., 2020. StarGAN v2: Diverse image synthesis for multiple domains. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Dar, S.U.H., Yurt, M., Ildiz, M.E., Shahdloo, M., Tinaz, B., Cukur, T., 2020. Prior-guided image reconstruction for accelerated multi-contrast MRI via generative adversarial networks. IEEE J Sel Top Signal Process 14 (6), 1072–1087.

Dar, S.U.H., Yurt, M., Karacan, L., Erdem, A., Erdem, E., Cukur, T., 2019. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. IEEE Trans. Med. Imaging 38 (10), 2375–2388.

Dewey, B.E., Zhao, C., Carass, A., Oh, J., Calabresi, P.A., van Zijl, P.C.M., Prince, J.L., 2018. Deep harmonization of inconsistent MR data for consistent volume segmentation. In: Gooya, A., Goksel, O., Oguz, I., Burgos, N. (Eds.), Simulation and Synthesis in Medical Imaging. Springer International Publishing, Cham, pp. 20–30.

Dewey, B.E., Zhao, C., Reinhold, J.C., Carass, A., Fitzgerald, K.C., Sotirchos, E.S., Saidha, S., Oh, J., Pham, D.L., Calabresi, P.A., van Zijl, P.C., Prince, J.L., 2019. Deepharmony: a deep learning approach to contrast harmonization across scanner changes. Magn. Reson. Imaging 64, 160–170.

Emami, H., Dong, M., Nejad-Davarani, S.P., Glide-Hurst, C.K., 2018. Generating synthetic CTs from magnetic resonance images using generative adversarial networks. Med. Phys. 45 (8), 3627–3636.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. Advances in Neural Information Processing Systems (NIPS) 27, 2672–2680.

Hagiwara, A., Otsuka, Y., Hori, M., Tachibana, Y., Yokoyama, K., Fujita, S., Andica, C., Kamagata, K., Irie, R., Koshino, S., Maekawa, T., Chougar, L., Wada, A., Aoki, S., 2019. Improving the quality of synthetic FLAIR images with deep learning using a conditional generative adversarial network for pixel-by-pixel image translation. American Journal of Neuroradiology 40, 224–230.

Han, Z., Wei, B., Mercado, A., Leung, S., Li, S., 2018. Spine-GAN: semantic segmentation of multiple spinal structures. Med. Image Anal. 50, 23–35.

Huang, Y., Beltrachini, L., Shao, L., Frangi, A.F., 2016. Geometry regularized joint dictionary learning for cross-modality image synthesis in magnetic resonance imaging. In: Tsaftaris, S.A., Gooya, A., Frangi, A.F., Prince, J.L. (Eds.), Simulation and Synthesis in Medical Imaging. Springer International Publishing, Cham, pp. 118–126.

Huang, Y., Shao, L., Frangi, A.F., 2017. Simultaneous super-resolution and cross-modality synthesis of 3D medical images using weakly-supervised joint convolutional sparse coding. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5787–5796.

Huang, Y., Shao, L., Frangi, A.F., 2018. Cross-modality image synthesis via weakly–coupled and geometry co-regularized joint dictionary learning. IEEE Trans. Med. Imaging 37 (3), 815–827.

Iglesias, J.E., Konukoglu, E., Zikic, D., Glocker, B., Van Leemput, K., Fischl, B., 2013. Is synthesizing MRI contrast useful for inter-modality analysis? In: Medical

Image Computing and Computer-Assisted Intervention–MICCAI 2013. Springer, pp. 631–638.

Isola, P., Zhu, J., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967–5976.

Jenkinson, M., Smith, S., 2001. A global optimisation methof for robust affine registration of brain images. Med. Image Anal. 5, 143–156.

Jiang, J., Hu, Y.-C., Tyagi, N., Zhang, P., Rimner, A., Mageras, G.S., Deasy, J.O., Veeraraghavan, H., 2018. Tumor-aware, adversarial domain adaptation from CT to MRI for lung cancer segmentation. In: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2019. Springer, pp. 777–785.

Jin, C.-B., Kim, H., Liu, M., Jung, W., Joo, S., Park, E., Ahn, Y.S., Han, I.H., Lee, J.I., Cui, X., 2019. Deep CT to MR synthesis using paired and unpaired data. Sensors 19 (10), 2361.

Jog, A., Carass, A., L Pham, D., Prince, J., 2014. Random forest FLAIR reconstruction from T1, T2, and PD-weighted MRI. In: IEEE 11th International Symposium on Biomedical Imaging (ISBI), 2014, pp. 1079–1082.

Jog, A., Carass, A., L Pham, D., Prince, J., 2015. Tree-encoded conditional random fields for image synthesis. Information Processing in Medical Imaging 24, 733–745.

Jog, A., Carass, A., Roy, S., L Pham, D., Prince, J., 2015. MR Image synthesis by contrast learning on neighborhood ensembles. Med. Image Anal. 24, 63–76.

Jog, A., Carass, A., Roy, S., L Pham, D., Prince, J., 2016. Random forest regression for magnetic resonance image synthesis. Med. Image Anal. 35, 475–488.

Johnson, J., Alahi, A., Fei-Fei, L., 2016. Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision (ECCV), pp. 694–711.

Joyce, T., Chartsias, A., Tsaftaris, S.A., 2017. Robust multi-modal MR image synthesis. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 347–355.

Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR).

Krupa, K., Bekiesińska-Figatowska, M., 2015. Artifacts in magnetic resonance imaging. Polish Journal of Radiology 80, 93–106.

Lee, D., Kim, J., Moon, W., Ye, J.C., 2019. Collagan: Collaborative GAN for missing image data imputation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2482–2491.

Lei, Y., Fu, Y., Mao, H., Curran, W.J., Liu, T., Yang, X., 2020. Multi-modality MRI Arbitrary Transformation Using Unified Generative Adversarial Networks. In: Išgum, I., Landman, B.A. (Eds.), Medical Imaging 2020: Image Processing. International Society for Optics and Photonics, SPIE, pp. 1–6. doi:10.1117/12.2549794.

Li, H., Paetzold, J.C., Sekuboyina, A., Kofler, F., Zhang, J., Kirschke, J.S., Wiestler, B., Menze, B., 2019. Diamondgan: Unified multi-modal generative adversarial networks for MRI sequences synthesis. In: Medical Image Computing and Computer Assisted Intervention (MICCAI). Springer International Publishing, Cham, pp. 795–803.

Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D., Ji, S., 2014. Deep learning based imaging data completion for improved brain disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 17, pp. 305–312.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88.

Liu, F., 2019. SUSAN: Segment unannotated image structure using adversarial network. Magn. Reson. Med. 81 (5), 3330–3345.

Maier, O., Menze, B.H., von der Gablentz, J., Häni, L., Heinrich, M.P., Liebrand, M., Winzeck, S., Basit, A., Bentley, P., Chen, L., et al., 2017. ISLES 2015-A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Med. Image Anal. 35, 250–269.

Mao, X., Li, Q., Xie, H., Lau, R.Y.K., Wang, Z., Smolley, S.P., 2017. Least squares generative adversarial networks. In: IEEE International Conference on Computer Vision (ICCV), pp. 2813–2821.

Mehta, R., Arbel, T., 2018. RS-Net: Regression-segmentation 3D CNN for synthesis of full resolution missing brain MRI in the presence of tumours. In: Simulation and Synthesis in Medical Imaging. Springer International Publishing, Cham, pp. 119–129.

Mirza, M., Osindero, S., 2014. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.

Nie, D., Trullo, R., Lian, J., Petitjean, C., Ruan, S., Wang, Q., Shen, D., 2017. Medical image synthesis with context-aware generative adversarial networks. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, pp. 417–425.

Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., Wang, Q., Shen, D., 2018. Medical image synthesis with deep convolutional adversarial networks. IEEE Trans. Biomed. Eng. 65 (12), 2720–2730.

Olut, S., Sahin, Y.H., Demir, U., Unal, G., 2018. Generative adversarial training for MRA image synthesis using multi-contrast MRI. In: PRedictive Intelligence in MEdicine (PRIME). Springer International Publishing, Cham, pp. 147–154.

Pan, Y., Liu, M., Lian, C., Zhou, T., Xia, Y., Shen, D., 2018. Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, pp. 455–463.

Ran, M., Hu, J., Chen, Y., Chen, H., Sun, H., Zhou, J., Zhang, Y., 2019. Denoising of 3d magnetic resonance images using a residual encoder–decoder wasserstein generative adversarial network. Med. Image Anal. 55, 165–180.

Ravì, D., Szczotka, A.B., Pereira, S.P., Vercauteren, T., 2019. Adversarial training with cycle consistency for unsupervised super-resolution in endomicroscopy. Med Image Anal 53, 123–131.

Roy, S., Carass, A., Prince, J., 2011. A compressed sensing approach for MR tissue contrast synthesis. In: Information Processing in Medical Imaging, pp. 371–383.

Roy, S., Carass, A., Prince, J.L., 2013. Magnetic resonance image example-based contrast synthesis. IEEE Trans. Med. Imaging 32 (12), 2348–2363.

Roy, S., Chou, Y.-Y., Jog, A., Butman, J., L. Pham, D., 2016. Patch based synthesis of whole head MR images: application to EPI distortion correction. In: Tsaftaris, S.A., Gooya, A., Frangi, A.F., Prince, J.L. (Eds.), Simulation and Synthesis in Medical Imaging, 9968, pp. 146–156.

Roy, S., Jog, A., Carass, A., Prince, J.L., 2013. Atlas Based Intensity Transformation of Brain MR Images. In: Multimodal Brain Image Analysis (MBIA), pp. 51–62.

Rubin, M., Stein, O., Turko, N.A., Nygate, Y., Roitshtain, D., Karako, L., Barnea, I., Giryes, R., Shaked, N.T., 2019. TOP-GAN: Stain-free cancer cell classification using deep learning with a small training set. Med. Image Anal. 57, 176–185.

Sevetlidis, V., Giuffrida, M.V., Tsaftaris, S.A., 2016. Whole image synthesis using a deep encoder-decoder network. In: Tsaftaris, S.A., Gooya, A., Frangi, A.F., Prince, J.L. (Eds.), Simulation and Synthesis in Medical Imaging. Springer International Publishing, Cham, pp. 127–137.

Sharma, A., Hamarneh, G., 2019. Missing MRI pulse sequence synthesis using multi–modal generative adversarial network. IEEE Trans. Med. Imaging 39, 1170–1183.

Sohail, M., Riaz, M.N., Wu, J., Long, C., Li, S., 2019. Unpaired multi-contrast MR image synthesis using generative adversarial networks. In: Burgos, N., Gooya, A., Svoboda, D. (Eds.), Simulation and Synthesis in Medical Imaging. Springer International Publishing, Cham, pp. 22–31.

Thukral, B., 2015. Problems and preferences in pediatric imaging. Indian Journal of Radiology and Imaging 25, 359–364.

Van Nguyen, H., Zhou, K., Vemulapalli, R., 2015. Cross-domain synthesis of medical images using efficient location-sensitive deep network. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer International Publishing, Cham, pp. 677–684.

Wang, J., Noble, J.H., Dawant, B.M., 2019. Metal artifact reduction for the segmentation of the intra cochlear anatomy in CT images of the ear with 3d-conditional GANs. Med Image Anal 58, 101553.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612.

Wei, W., Poirion, E., Bodini, B., Durrleman, S., Ayache, N., Stankoff, B., Colliot, O., 2018. Learning myelin content in multiple sclerosis from multimodal MRI through adversarial training. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, pp. 514–522.

Wei, W., Poirion, E., Bodini, B., Durrleman, S., Colliot, O., Stankoff, B., Ayache, N., 2018. FLAIR MR image synthesis by using 3D fully convolutional networks for multiple sclerosis. In: ISMRM-ESMRMB - Joint Annual Meeting, pp. 1–6.

Welander, P., Karlsson, S., Eklund, A., 2018. Generative adversarial networks for image-to-image translation on multi-contrast MR images - a comparison of cycleGAN and UNIT. arxiv preprint arXiv:1806.07777.

Xu, C., Howey, J., Ohorodnyk, P., Roth, M., Zhang, H., Li, S., 2020. Segmentation and quantification of infarction without contrast agents via spatiotemporal generative adversarial learning. Med. Image Anal. 59, 101568.

Yang, Q., Li, N., Zhao, Z., Fan, X., I-Chao Chang, E., Xu, Y., 2018. MRI Image-to-image translation for cross-modality image registration and segmentation. arXiv preprint arXiv:1801.06940v1.

Yang, X., Kwitt, R., Styner, M., Niethammer, M., 2017. Quicksilver: fast predictive image registration - a deep learning approach. Neuroimage 158, 378–396.

Yi, X., Walia, E., Babyn, P., 2019. Generative adversarial network in medical imaging: a review. Med. Image Anal. 101552.

Yu, B., Zhou, L., Wang, L., Fripp, J., Bourgeat, P., 2018. 3D cGAN based cross-modality MR image synthesis for brain tumor segmentation. In: IEEE 15th International Symposium on Biomedical Imaging. IEEE, pp. 626–630.

Yu, B., Zhou, L., Wang, L., Shi, Y., Fripp, J., Bourgeat, P., 2019. Ea-GANs: edge-aware generative adversarial networks for cross-modality MR image synthesis. IEEE Trans. Med. Imaging 38, 1750–1762.

Zhao, H., Li, H., Maurer-Stroh, S., Cheng, L., 2018. Synthesizing retinal and neuronal images with generative adversarial nets. Med. Image Anal. 49, 14–26.