

DEEP LEARNING FOR MULTI-CONTRAST MRI SYNTHESIS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

By
Mahmut Yurt
July 2021

Deep learning for multi-contrast MRI synthesis

By Mahmut Yurt

July 2021

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Tolga ukur (Advisor)

Emine lk Sarıtař ukur

İbrahim Aykut Erdem

Approved for the Graduate School of Engineering and Science:

Ezhan Karařan
Director of the Graduate School

ABSTRACT

DEEP LEARNING FOR MULTI-CONTRAST MRI SYNTHESIS

Mahmut Yurt

M.S. in Electrical and Electronics Engineering

Advisor: Tolga Çukur

July 2021

Magnetic resonance imaging (MRI) possesses the unique versatility to acquire images under a diverse array of distinct tissue contrasts. Multi-contrast images, in turn, better delineate tissues, accumulate diagnostic information, and enhance radiological analyses. Yet, prolonged, costly exams native to multi-contrast protocols often impair the diversity, resulting in missing images from some contrasts. A promising remedy against this limitation arises as image synthesis that recovers missing target contrast images from available source contrast images. Learning-based models demonstrated remarkable success in this source-to-target mapping due to their prowess in solving even the most demanding inverse problems. Mainstream approaches proposed for synthetic MRI were typically subjected to a model training to perform either one-to-one or many-to-one mapping. One-to-one models manifest elevated sensitivity to detailed features of the given source, but they perform suboptimally when source-target images are poorly linked. Meanwhile, many-to-one counterparts pool information from multiple sources, yet this comes at the expense of losing detailed features uniquely present in certain sources. Furthermore, regardless of the mapping, they both innately demand large training sets of high-quality source and target images Fourier-reconstructed from Nyquist-sampled acquisitions. However, time and cost considerations put significant challenges in compiling such datasets. To address these limitations, here we first propose a novel multi-stream model that task-adaptively fuses unique and shared image features from a hybrid of multiple one-to-one streams and a single many-to-one stream. We then introduce a novel semi-supervised learning framework based on selective tensor loss functions to learn high-quality image synthesis directly from a training dataset of undersampled acquisitions, bypassing the undesirable data requirements of deep learning. Demonstrations on brain MRI images from healthy subjects and glioma patients indicate the superiority of the proposed approaches against state-of-the-art baselines.

Keywords: MRI synthesis, deep learning, multi-stream, semi-supervised.

ÖZET

ÇOKLU KONTRAST MRG İÇİN DERİN ÖĞRENME

Mahmut Yurt

Elektrik Elektronik Mühendisliği, Yüksek Lisans

Tez Danışmanı: Tolga Çukur

Temmuz 2021

Manyetik rezonans görüntüleme (MRG), farklı doku kontrastları altında görüntü elde etmek için benzersiz çok yönlülüğe sahiptir. Çoklu kontrast görüntüler, dokuları daha iyi betimler, tanı bilgisini artırır ve radyolojik analizleri geliştirir. Yine de, çoklu kontrast protokollerine özgü uzun süreli, maliyetli muayeneler çeşitliliği bozabilir ve bazı kontrastlardan eksik görüntülere neden olur. Buna karşı umut verici bir çözüm, mevcut kaynak kontrast görüntülerinden eksik hedef kontrast görüntülerini kurtaran görüntü sentezi olarak ortaya çıkmaktadır. Öğrenmeye dayalı modeller, en zorlu ters problemleri bile çözmedeki hünerlerinden dolayı bu haritalamada büyük bir başarı göstermiştir. Sentetik MRG için önerilen temel yaklaşımlar, bire bir veya çoktan bire haritalama gerçekleştirmek için model eğitime tabi tutulmaktadır. Bire bir modeller, verilen kaynağın ayrıntılı özelliklerine karşı yüksek hassasiyet gösterir, ancak kaynak-hedef görüntüler zayıf bir şekilde bağlantılı olduğunda yetersiz kalabilirler. Çoktan bire muadilleri ise birden fazla kaynaktan bilgi toplar, ancak bu belirli kontrastlarda benzersiz bir şekilde bulunan ayrıntılı özelliklerin kaybına yol açar. Ayrıca, haritalamadan bağımsız olarak, her ikisi de yüksek kaliteli Nyquist örneklenmiş edinimlerden geri çatılmış kaynak ve hedef görüntülerinden oluşan büyük eğitim setleri talep ederler. Fakat, zaman ve maliyet hususları, bu tür veri kümelerinin derlenmesinde önemli zorluklar ortaya çıkarmaktadır. Bu sınırlamaları çözmek için, burada ilk önce, bire bir akışlardan ve çoktan bire akıştan oluşan bir hibrit ile, benzersiz ve ortak görüntü özelliklerini göreve uyarlanabilir bir şekilde birleştiren yeni bir model önerilmiştir. Ardından, derin öğrenmenin istenmeyen veri gereksinimlerini aşarak, alt örneklenmiş edinimlerden oluşan eğitim veri kümesinden yüksek kaliteli görüntü sentezini öğrenmek için seçici tensör kaybı yitimlerine dayanan yeni bir yarı denetimli öğrenme yöntemi sunulmuştur. Sağlıklı deneklerden ve hastalardan alınan beyin MRG görüntüleri üzerinde yapılan deneyler, önerilen yaklaşımların en gelişmiş temel yöntemlere göre üstünlüğünü göstermektedir.

Anahtar sözcükler: MRG sentezleme, derin öğrenme, çoklu akış, yarı denetim.

Acknowledgement

First of all, I would like to thank my advisor, Prof. Tolga Çukur, for his consistent support and encouragement. He has always been very patient, empathetic, and gentle during my academic studies, and I feel truly fortunate to have worked with him. I should particularly acknowledge that his unfailing quest for perfection in research was not only motivational but also contagious, thanks to which I was able to start growing as a zealous researcher. Therefore, I could not have hoped for a better advisor than him for my Master's studies, and I will surely feel his absence in my doctorate.

I would also like to thank my jury members. I express my deepest gratitude to Prof. Emine Sarıtaş for always showing a very polite and supportive approach towards me. Her excellence in the departmental courses helped me acquire layers of knowledge in medical imaging that prove crucial in many components of my studies. I am also grateful to Prof. Aykut Erdem for always being very sincere and helpful. I should acknowledge that he has put critical time and effort into the studies in this thesis and constantly guided us about the latest trends in machine learning.

I would also like to thank Prof. Erkut Erdem and Prof. Kader Oğuz for their meaningful collaboration. Much of the work in this thesis would not have been possible without their essential input on the computational and radiological viewpoints.

As a tradition, I would like to thank all my fellow labmates at UMRAM. To name a few, Salman never failed to become the first address to new lab members needing help. Emin proved how poor it could be for a tennis player to serve yet having a solid forehand. Özgür provided great talks about the components of our personal computers. Muzaffer showed how stressless upcoming deadlines could be. I still remember Yılmaz and me starting to write a workshop paper a day before the deadline. Onat will assuredly come up with newer projects built on the work in this thesis. We had very crucial conversations with Bora about career

plans. Batu has always been very supportive, and I am reserving all rights to extend my words about him in my doctoral thesis.

I could not help but thank my fellow friends at Bilkent that I am delighted to make acquainted with. Onurcan offered his invaluable friendship for years and over countless adventures. Furkan took place among the kindest people, and we constantly recall the glorious summer. I still cannot explain how Gökçe provides conversations without needing content. Emir proved one of my biggest supporters, and he was the official winner of our challenge (see his thesis for details). Berk was the first address to discuss our bizarre research ideas that we never pursued (for now). I cannot forget Melih suggesting to perform a Monte Carlo simulation for our already creditable project (5 minutes before the deadline). Bahadır constantly looked for an excuse to start a debate to present his unique ideas yet remained one of the most gentle people. Elif made us smile with the funniest stickers in her spare time apart from reviewing literature for research. Fatih was one of my earliest collaborators in the machine learning domain, and I hope we can reconvene the lunch council together with Oğuzhan.

I would also like to give a special shout-out to my close friends from earlier years who did not only supported me throughout my journey at Bilkent but also had encouraged me beforehand in making it possible (so do I theirs). Ömer has never failed to show his invaluable friendship since we were kids. Hasan regularly visited Bilkent to see me so many times that even the security started recognizing him. Reyhan perpetually demonstrated her supportive friendship.

Most importantly, I would like to thank my family – my father Hasan, my mother Mukadder, and my brother Emre. I feel very privileged to have such a supportive, encouraging, patient, joyful, and caring family. They always believed in me, trusted my decisions, and shared my feelings. Without their unconditional love and guidance, I could not have gone this far. Therefore, every success I achieve in my life is surely theirs.

Contents

1	Introduction	1
2	Fundamentals of Multi-Contrast MRI	6
3	Multi-Stream Generative Adversarial Networks for MR Image Synthesis	9
3.1	Introduction	10
3.2	Theory	12
3.2.1	Generative Adversarial Networks	12
3.2.2	Conditional Generative Adversarial Networks	13
3.3	Methods	14
3.3.1	Multi-Stream GAN Model	14
3.3.2	Network Architecture	19
3.3.3	Datasets	20
3.3.4	Network Training	22

<i>CONTENTS</i>	x
3.3.5	Competing Methods 24
3.3.6	Experiments 25
3.4	Results 26
3.4.1	Task-Specific Fusion Across Multiple Streams 26
3.4.2	Demonstrations Against One-to-one and Many-to-one Mappings 28
3.4.3	Demonstrations Against Competing Methods 39
3.5	Discussion 49
4	Semi-Supervised Learning of Mutually Accelerated MRI Synthesis without Fully-Sampled Ground Truths 52
4.1	Introduction 53
4.2	Methods 55
4.2.1	Generative Adversarial Networks 56
4.2.2	Semi-Supervised Generative Adversarial Networks 57
4.2.3	Datasets 61
4.2.4	Implementation Details 63
4.2.5	Competing Methods 63
4.2.6	Experiments 65
4.3	Results 67

4.3.1	Robustness of semi-supervised learning against deficiencies in training data	67
4.3.2	Single-coil image synthesis in mutually accelerated multi- contrast MRI	70
4.3.3	Multi-coil image synthesis in mutually accelerated multi- contrast MRI	73
4.3.4	Ablation Studies	76
4.4	Discussion	76
5	Conclusion	80

List of Figures

2.1	Multi-contrast images from the IXI dataset (https://brain-development.org/ixi-dataset/).	8
3.1	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.	15

3.2 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) $T_1, T_2 \rightarrow FLAIR$ 27

3.3 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 8th 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 256×256 . 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 model). 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. 29

3.4 FLAIR image synthesis from T_1 - and T_2 -weighted images in the ISLES dataset. Feature maps from different network layers (1, 4, 8, 12) within the one-to-one (first and second rows) and many-to-one (third row) models are displayed along with source and ground truth target images. The two sources for the many-to-one model are displayed as red and green color channels in a combined image. Concatenation of the feature maps is also shown (fourth row), where separate feature maps from the one-to-one and many-to-one mappings are taken as RGB color channels. While one-to-one models manifest increased sensitivity to unique, detailed features that are predominantly present in the given source, many-to-one mapping pools information across sources with high sensitivity to shared features. Note that more discriminable representations among one-to-one and many-to-one streams emerge towards later layers in the network. 30

3.5 Methods were compared in terms of quality of T_1 synthesis in the IXI 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 as input T_2 -weighted images, and pGAN-B receives as input PD-weighted images. 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). . . 32

- 3.6 Methods were compared in terms of quality of T_2 synthesis in the IXI 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 as input T_1 -weighted images, and pGAN-B receives as input PD-weighted images. 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). . 33
- 3.7 Methods were compared in terms of quality of PD synthesis in the IXI 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 as input T_1 -weighted images, and pGAN-B receives as input T_2 -weighted images. 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). . 34

- 3.8 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). 35
- 3.9 Methods were compared in terms of quality of T_2 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 as input T_1 -weighted images, and pGAN-B receives as input FLAIR images. 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). 36

- 3.10 Methods were compared in terms of quality of FLAIR 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 as input T₁-weighted images, and pGAN-B receives as input T₂-weighted images. 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). 37
- 3.11 Methods were compared in terms of radiological opinion scores for three synthesis tasks in IXI: a) T₂, PD → T₁, b) T₁, PD → T₂, c) T₁, T₂ → PD, and for three synthesis tasks in ISLES: d) T₂, FLAIR → T₁, e) T₁, FLAIR → T₂, f) T₁, T₂ → 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 truth images, images of 3 subjects for T₂ 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), (T₂, PD), (T₁, PD), (T₁, T₂), and in ISLES: (1, 2), (T₂, FLAIR), (T₁, FLAIR), (T₁, T₂). 41

3.12 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. 42

3.13 The proposed method was demonstrated on healthy subjects from the IXI dataset for T_2 -weighted image synthesis from T_1 - and PD-weighted images. Synthesized images from mustGAN, pGAN, pGAN_{many}, MM-GAN, and Multimodal are shown along with the ground truth target image. Due to synergistic use of information captured by one-to-one and many-to-one streams, mustGAN improves synthesis accuracy in many regions that are recovered suboptimally in competing methods (shown in zoom-in displays). Overall, mustGAN yields less noisy depiction of tissues and sharper depiction of tissue boundaries. 43

- 3.14 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. 45
- 3.15 The proposed method was demonstrated on high/low grade glioma patients from the ISLES dataset for T_2 -weighted image synthesis from T_1 -weighted and FLAIR images. Synthesized images from mustGAN, pGAN, pGAN_{many}, MM-GAN, and Multimodal are shown along with the ground truth target image. Due to synergistic use of information captured by one-to-one and many-to-one streams, mustGAN improves synthesis accuracy in many regions that are recovered suboptimally in competing methods (shown in zoom-in displays). Overall, mustGAN yields less noisy depiction of tissues and sharper depiction of tissue boundaries. 46

4.1 Illustration of the proposed semi-supervised ssGAN model. As opposed to fully-supervised models that demand Nyquist-sampled acquisitions for training (1), ssGAN learns to synthesize high-quality images given a dataset of undersampled source and target acquisitions (2). ssGAN initially synthesizes a coil-combined target image that is backprojected onto individual coils via sensitivity maps. These multi-coil target images are subsampled in Fourier domain with the target acquisition mask in order to define the selective multi-coil tensor losses in image, k-space and adversarial domains (3). 58

4.2 Reliability of ssGAN against training data deficiencies. Evaluations were performed for $n_T = [8 : 8 : 64]$. For each n_T , pix2pix and CycleGAN were trained with $R_{target} = 1$, whereas ssGAN was trained with $R_{target} \in [2 : 1 : 10]$, ssGAN- k with $k = R_{target}$. All models were trained with $R_{source} = 1$. Performance metrics for CycleGAN at $n_T = 8$ remain outside the display windows. 69

4.3 ssGAN was demonstrated on IXI for $T_2 \rightarrow T_1$ mapping against pix2pix and CycleGAN with ($R_{source} = 1$). Synthesized images from ssGAN-2, ssGAN-3, ssGAN-4, pix2pix, and CycleGAN are displayed together with the reference and source images in the first row. The corresponding error maps for the synthesized images are displayed in the second row. 71

4.4 Synthesis quality of ssGAN, fsGAN and CasGAN was demonstrated on IXI for $T_1 \rightarrow T_2$ synthesis ($R_{source} = 2$). Synthesized images from the competing methods are displayed together with the source and reference (i.e. target) images in the first row, and the corresponding error maps for the synthesized images are displayed in the second row. 73

- 4.5 The proposed ssGAN model was demonstrated on the in vivo brain dataset for multi-coil complex PD \rightarrow T₂ synthesis task with source contrast acquisitions undersampled by $R_{target} = 2$. Synthesized coil-combined images from fsGAN, ssGAN, and CasGAN are displayed along with the reference image (i.e. target) in the first row, and the corresponding error images for the synthesized images are displayed in the second row (see colorbar). 75
- 4.6 Synthesis quality of ssGAN, fsGAN and CasGAN was demonstrated on the in vivo dataset for T₂ \rightarrow PD synthesis ($R_{source} = 2$). Representative results from two different subjects are displayed, along with the source and reference images. 75
- 4.7 Radiological evaluations for fsGAN, ssGAN-2, and CasGAN-2 are shown. T₂ \rightarrow PD and PD \rightarrow T₂ synthesis tasks at $R_{source} = 2, 4$ were assessed on the in vivo dataset. 76

List of Tables

- 3.1 Synthesis Quality Enhancement due to the Many-to-one Stream: Synthesis quality of the proposed mustGAN method is compared against a variant of itself, namely mustGAN_{one}, that recovers the target contrast image from only the unique feature maps generated in the one-to-one streams. PSNR and SSIM measurements of mustGAN and mustGAN_{one} are reported as mean±std calculated across test subjects for three distinct synthesis tasks in the IXI dataset: T₂, PD→T₁; T₁, PD→T₂; T₁, T₂→PD, and for three distinct synthesis tasks in the ISLES dataset: T₂, FLAIR→T₁; T₁, FLAIR→T₂; T₁, T₂→FLAIR. Boldface marks the model having the highest performance. 38
- 3.2 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: T₂, PD → T₁; T₁, PD → T₂; T₁, T₂ → PD. pGAN-A receives the 1st source contrast and pGAN-B receives the 2nd source contrast i.e., (1,2): (T₂, PD), (T₁, PD), (T₁, T₂). Boldface marks the model having the highest performance. 39

3.3	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 → T ₁ ; T ₁ , FLAIR → T ₂ ; T ₁ , T ₂ → 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.	44
3.4	Synthesis Quality of mustGAN Models with Varying Fusion Block Positions: PSNR and SSIM measurements between the ground truth and synthesized target images from multiple independent mustGAN models with varying fusion block positions ($i \in \{1, 2, \dots, 14\}$) are reported as mean±std for the test set. Three distinct synthesis tasks are considered in the IXI dataset: T ₂ , PD→T ₁ ; T ₁ , PD→T ₂ ; T ₁ , T ₂ →PD, and three distinct synthesis tasks are considered in the ISLES dataset: T ₂ , FLAIR→T ₁ ; T ₁ , FLAIR→T ₂ ; T ₁ , T ₂ →FLAIR.	48
4.1	Quality of Synthesis in the IXI Dataset for Nyquist-Sampled Source Acquisitions	70
4.2	Image Quality for Mutually Accelerated Single-Coil MRI Synthesis	72
4.3	Image Quality for Mutually Accelerated Multi-Coil MRI Synthesis	74
4.4	Effects of Tensor Losses on Synthesis Quality	77

Chapter 1

Introduction

Magnetic resonance imaging (MRI) offers the unique versatility to acquire bodily tissues under an array of distinct contrasts. Multi-contrast MR images, in turn, better delineate tissues, accumulate diagnostic information, and enhance subsequent radiological image analysis tasks. Yet, prolonged exams and elevated scan costs often impair the diversity of multi-contrast protocols.

A promising remedy against the impaired MRI protocols arises as image synthesis that aims to recover missing or poor-quality images of target contrasts from high-quality images of available source contrasts. This image-to-image translation task depicts an ill-posed inverse problem as distinct contrasts manifest stark differences and their images are high-dimensional. To handle this inverse problem, earlier studies introduced two fundamental approaches: atlas-based registration and voxel-intensity mapping. Atlas-based registration undertakes a geometric transformation between the anatomical space of the given subject and a multi-modal atlas, intrinsically yielding poor performance in recovering subject-specific patterns. Meanwhile, voxel-intensity mapping performs a nonlinear translation between the source and target contrast images, so it offers enhanced capture for specific relationships between the contrasts.

Guided by intensity transformation, recent studies in the medical imaging domain embraced learning-based deep neural networks for multi-contrast MR image synthesis. The proposed deep network models are subject to a training procedure using a large dataset of images from source and target contrasts. Pioneering studies in the field introduced encoder-decoder-type convolutional neural network (CNN) models for image-level processing. These CNN models processed the given source images via an encoder module to embed intermediate-level latent representations and then recovered the target images via a decoder module. Later studies demonstrated generative adversarial networks (GAN) for MRI contrast synthesis, where the previous CNN models taking a generator role were supplemented with a competing discriminator network. The game-theoretical interplay between the generator and discriminator enabled advanced capture of detailed structural information associated with high-frequency details in the recovered images. The reported CNN- and GAN-based deep networks demonstrated success in synthetic MRI, so they effectively leveraged impaired multi-contrast MRI protocols. That said, the previous models involved single- or multi-source inputs isolated from each other, and they demanded large training databases of high-quality source and target contrast images. These shortcomings restricted the utility and practicality of multi-contrast MRI synthesis.

Contributions

This thesis first addresses the shortcomings of isolated one-to-one and many-to-one recovery models that are common for multi-contrast MRI synthesis. One-to-one models receive as input images from a single source contrast, so they can manifest increased sensitivity to unique, detailed features of the given source. Meanwhile, many-to-one models pool information from images of multiple source contrasts and learn representations shared across the sources. That said, isolated one-to-one methods might suffer from poor recovery of information available within the entire set of source contrasts, whereas many-to-one models may suboptimally capture unique features available within specific sources. To handle these limitations, the proposed multi-stream GAN model complementarily captures unique and shared image features across multiple sources via a hybrid of multiple one-to-one streams and a single many-to-one stream. The learned feature maps from the streams are integrated via a novel fusion block that adaptively takes place in the network to maximize task-specific synthesis performance. The proposed model outperforms the state-of-the-art baselines based on quantitative measurements and radiological evaluations on brain images from healthy subjects and glioma patients.

Secondly, this thesis introduces a novel semi-supervised generative model for multi-contrast MRI synthesis that eliminates the need for fully-sampled ground truth acquisitions for training. Previous approaches for synthetic MRI typically involve deep models trained with high-quality source and target contrast images from fully-sampled acquisitions regardless of whether the training samples are paired or unpaired. Inherently, this leads to undesirable dependence on the

availability of fully-sampled acquisitions for training, which may prove infeasible in many cases due to increased scan durations or costs. The proposed semi-supervised model enables model training with undersampled target ground truths by introducing a novel selective loss function. This loss function is defined only on the acquired k-space coefficients of the given target ground truths, and subjectwise randomized sampling patterns are utilized during training to effectively capture the relationship among the acquired and non-acquired k-space coefficients. Demonstrations performed on multi-coil and single-coil neuroimaging datasets indicate that the proposed model maintains equivalent performance to the gold-standard fully-supervised models and alternative approaches. Hence, the proposed model offers great promise to improve the practicality and utility of MRI synthesis trained with accelerated data.

Outline

In the remainder of this thesis, the organization is as follows. Chapter 2 discusses the fundamentals of multi-contrast MRI. Chapter 3 describes in detail the multi-stream GAN model for multi-contrast MRI synthesis. Chapter 4 then introduces the semi-supervised learning approach for multi-contrast MRI synthesis without requiring fully-sampled ground truths. Lastly, Chapter 5 portrays the conclusive remarks.

Chapter 2

Fundamentals of Multi-Contrast MRI

Magnetic resonance imaging (MRI) typically delineates hydrogen nuclei that possess a single proton (^1H) due to their abundance in bodily tissues. The hydrogen nuclei spin around an axis lying along their magnetic moment direction. If no external magnetic field is applied, the axes of the spins are randomly oriented, so the net magnetization yields to zero. Contrarily, performing an external magnetic field denoted as B_0 leads MR-relevant hydrogen nuclei to precess around the longitudinal dimension of B_0 with the Larmor frequency ω_0 . In MRI scans, a radiofrequency (RF) magnetic pulse denoted as B_1 tips the orientation of the spins to the transverse plane up to a degree designated as the flip angle. This magnetization tends to relax to equilibrium as the RF magnetic pulse is switched off. During this relaxation, the longitudinal component of the magnetization increases as per a time constant T_1 , whereas the transverse component decays as per a second time constant T_2 . Here, the signal in the transverse plane is detected via a set of receive coils by sensing flux changes. Meanwhile, magnetic field gradients are performed to spatially encode the received signals, resulting in:

$$s(t) = \int_z \int_y \int_x M(x, y, z) e^{-t/T_2(r)} \exp\{-i\gamma \int G(\tau) \dot{r} d\tau\} dx dy dz \quad (2.1)$$

where $M(x, y, z)$ stands for the transverse magnetization in the Cartesian coordinates, γ denotes gyromagnetic ratio, $r = [x, y, z]$, and $G = [G_x, G_y, G_z]$. Expressing the gradients with spatial frequency variables $k_x(t) = \frac{\gamma}{2\pi} \int G_x(\tau) d\tau$, $k_y(t) = \frac{\gamma}{2\pi} \int G_y(\tau) d\tau$, $k_z(t) = \frac{\gamma}{2\pi} \int G_z(\tau) d\tau$ yields a frequency domain (k-space) formulation:

$$s(t) = \int_z \int_y \int_x M(x, y, z) e^{-t/T_2(r)} \exp\{-i2\pi(k_x(t)x + k_y(t)y + k_z(t)z)\} dx dy dz \quad (2.2)$$

Note that the signal levels in various tissues depend on tissue-specific relaxation parameters characterized by the execution of gradient and RF fields. This suggests that tailored RF pulses and timing can adjust contrast in the acquired MRI images. The spin-echo sequence commonly used in scans yields the following MR signal:

$$S_0 = K_0 \rho(r) [1 - e^{-TR/T_1(r)}] e^{-TE/T_2(r)} \quad (2.3)$$

where K_0 is a scaling constant, $\rho(r)$ stands for the density of the nuclei, TR denotes the repetition time, and TE denotes the echo time, where $T_1(r)$ and $T_2(r)$ are relaxation parameters characterizing the longitudinal and transverse magnetization. This equation indicates that the contrast in the acquired images can be adjusted by merely changing TR and TE. For example, T_1 -weighted MRI images can be acquired by selecting low TE and moderate TR, and T_2 -weighted MRI images can be acquired by selecting moderate TE and long TR. An example of multi-contrast images is provided in 2.1. Furthermore, repetition of the exams with a set of TR and TE values enables the acquisition of multi-contrast images, improving the radiological information available for diagnosis.

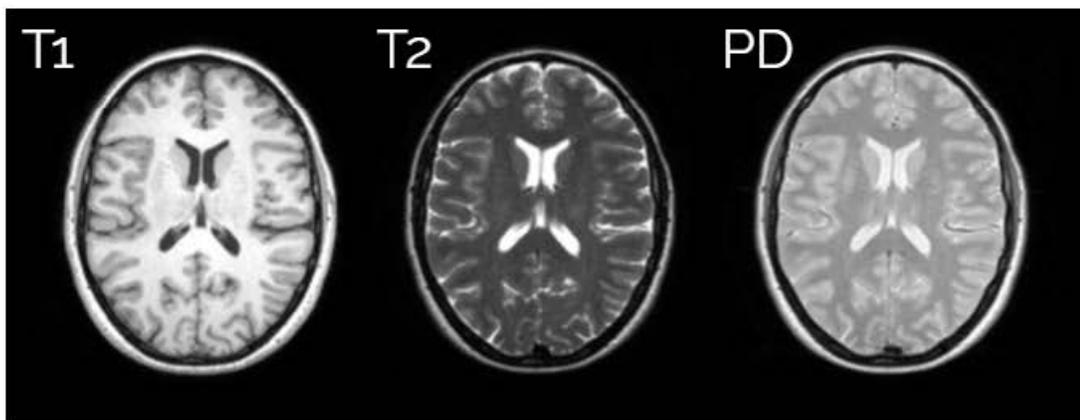


Figure 2.1: Multi-contrast images from the IXI dataset (<https://brain-development.org/ixi-dataset/>).

Chapter 3

Multi-Stream Generative Adversarial Networks for MR Image Synthesis

The content of this chapter reflects the study described in the following publications:

- M. Yurt, S. U. H. Dar, A. Erdem, E. Erdem, K. K. Oğuz, and T. Çukur, “mustGAN: multi-stream generative adversarial networks for MR image synthesis,” *Medical Image Analysis*, vol. 70, p. 101944, 2021
- M. Yurt, S. U. H. Dar, A. Erdem, E. Erdem, and T. Çukur, “A multi-stream GAN approach for multi-contrast MRI synthesis,” in *28th annual meeting of International Society for Magnetic Resonance Imaging (ISMRM)*, (Virtual Conference), 8 2020
- M. Yurt, S. U. H. Dar, A. Erdem, E. Erdem, and T. Çukur, “Adaptive fusion via dual-branch GAN for multi-contrast MRI synthesis,” in *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, (Virtual Conference), 4 2020

3.1 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 [4, 5]. 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 [6].

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 [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18] and many-to-one methods [19, 20, 10, 11, 21, 22, 23, 24, 25, 26, 27, 28, 29]. 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 [30, 31, 32, 33, 34, 35, 36, 37], 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 linear combination of dictionary atoms of b_S , and this combination is then used for synthesizing y from b_T [30, 31, 32, 33, 34, 35, 36, 37]. For improved performance, patch-based non-linear regression using random forests [33] or location-sensitive neural networks [16] 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 [12, 14]. One powerful method is based on the encoder-decoder architecture [14], 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 [14]. Recent deep network models have further incorporated an adversarial loss to better capture the high frequency details in the

target image [7, 9, 15, 17, 18, 38, 39]. An important adversarial method is pGAN [9], which additionally utilizes pixel-wise and perceptual losses [40] 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 [19, 20, 10, 11, 21, 22, 23, 24, 41, 42, 43, 25, 26, 27, 28, 29]. 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 $X = \{x_m : m = 1, 2, \dots, K\}$ in varying contrasts $C_S = \{c_{S_m} : m = 1, 2, \dots, K\}$. As in one-to-one synthesis, a common method is to perform non-linear regression using random forests [41, 43]. The random-forest method in [43] fits a non-linear regression model in feature space to estimate intensities of the target contrast given multiple source contrasts [43]. Deep neural network methods have also been proposed for many-to-one synthesis [19, 10, 11, 26, 25]. In [19], 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 [19]. Similar to one-to-one methods, recent studies have leveraged an adversarial loss to improve the quality of many-to-one synthesis [20, 21, 22, 24, 27, 28, 29]. An important example is MM-GAN [28], 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 [28].

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 contrast. 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.

3.2 Theory

3.2.1 Generative Adversarial Networks

A generative adversarial network (GAN) consists of a pair of competing networks; a generator (G) and a discriminator (D) [44]. Recently, GAN models have been successfully demonstrated for various tasks including data augmentation [45, 46, 47] and image synthesis [48, 49, 50]. 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 [44]. This can be formulated as a minimax game based on an adversarial loss function L_{GAN} .

$$\min_G \max_D L_{GAN} = \min_G \max_D \left(\mathbb{E}_y [\log D(y)] + \mathbb{E}_z [\log (1 - D(G(z)))] \right) \quad (3.1)$$

where \mathbb{E} denotes expectation. To improve stability, the negative log-likelihood in L_{GAN} is typically replaced by a squared loss function [51]:

$$L_{GAN} = -\mathbb{E}_y [(D(y) - 1)^2] - \mathbb{E}_z [D(G(z))^2] \quad (3.2)$$

3.2.2 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 [52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62], 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 [56]. The adversarial loss is then expressed as:

$$L_{\text{condGAN}} = -\mathbb{E}_{x,y} [(D(x, y) - 1)^2] - \mathbb{E}_x [D(G(x))^2] \quad (3.3)$$

When source and target images are spatially registered, a pixel-wise loss can be added between the ground truth and generated images [63]:

$$L_{\text{pixel-wise}} = \mathbb{E}_{x,y} [\|y - G(x)\|_1] \quad (3.4)$$

The joint loss function then becomes:

$$L_{\text{condGAN}} = -\mathbb{E}_{x,y} [(D(x, y) - 1)^2] - \mathbb{E}_x [D(x, G(x))^2] + \mathbb{E}_{x,y} [\|y - G(x)\|_1] \quad (3.5)$$

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 $T_2 \rightarrow T_1$ [9]. 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 $X = \{x_m : m = 1, 2, \dots, K\}$, a many-to-one condGAN model is formulated as:

$$L_{\text{condGAN}} = -\mathbb{E}_{X,y} [(D(X, y) - 1)^2] - \mathbb{E}_X [D(X, G(X))^2] + \mathbb{E}_{X,y} [\|y - G(X)\|_1] \quad (3.6)$$

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

3.3 Methods

3.3.1 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. 3.1). To synthesize the target image y , mustGAN receives as input a collection of source images denoted as $X = \{x_m : m = 1, 2, \dots, K\}$. 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.

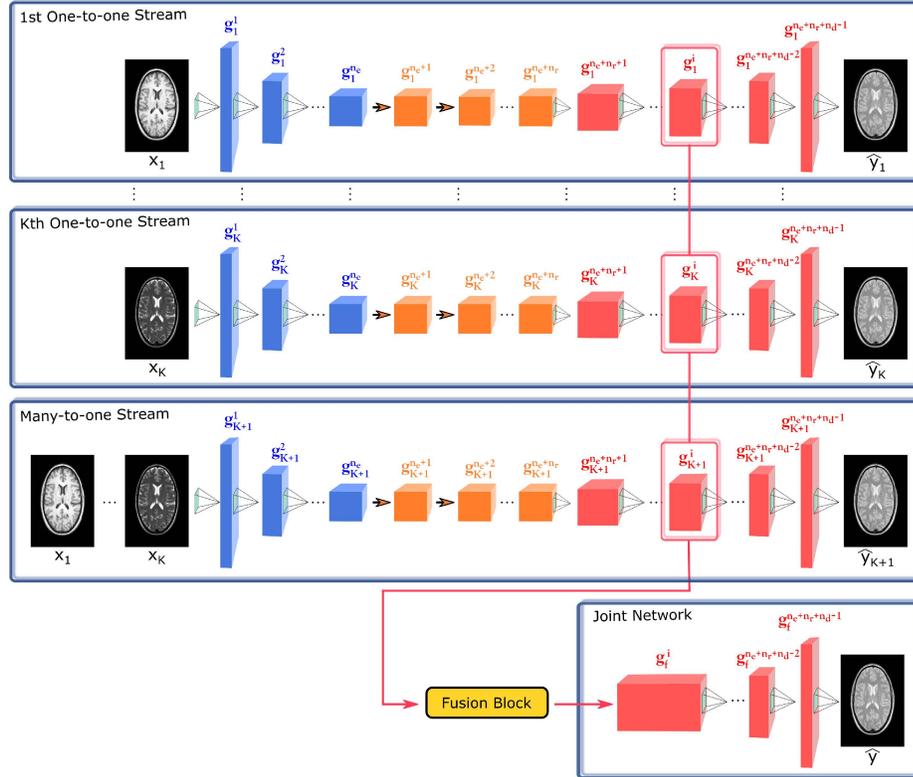


Figure 3.1: The generator (G) in mustGAN consists of K one-to-one streams and a many-to-one stream, followed by an adaptively positioned fusion block, and a joint network for final recovery. The independent one-to-one streams generate unique feature maps of individual source images, whereas the many-to-one 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.

3.3.1.1 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:

$$\hat{y}_m = G_m(x_m) = d_m(r_m(e_m(x_m))) \quad (3.7)$$

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

$$D_m(x_m, s) = c_m(x_m, s) \quad (3.8)$$

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

$$\begin{aligned} L_m = & -\mathbb{E}_{x_m, y} [(D_m(x_m, y) - 1)^2] - \mathbb{E}_{x_m} [D_m(x_m, G_m(x_m))^2] \\ & + \mathbb{E}_{x_m, y} [\|y - G_m(x_m)\|_1] \end{aligned} \quad (3.9)$$

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

3.3.1.2 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:

$$\hat{y}_{K+1} = G_{K+1}(X) = d_{K+1}(r_{K+1}(e_{K+1}(X))) \quad (3.10)$$

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

$$D_{K+1}(X, s) = c_{K+1}(X, s) \quad (3.11)$$

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

$$L_{K+1} = -\mathbb{E}_{X,y} [(D_{K+1}(X, y) - 1)^2] - \mathbb{E}_X [D_{K+1}(X, G_{K+1}(X))^2] + \mathbb{E}_{X,y} [\|y - G_{K+1}(X)\|_1] \quad (3.12)$$

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

3.3.1.3 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:

$$\begin{aligned} g_m^i &= e_m(x_m|i) \\ g_{K+1}^i &= e_{K+1}(X|i) \end{aligned} \quad (3.13)$$

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

$$g_f^i = f(g_1^i, g_2^i, \dots, g_K^i, g_{K+1}^i) \quad (3.14)$$

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

$$\hat{y} = J(g_f^i) = d_J(r_J(e_J(g_f^i|i))) \quad (3.15)$$

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:

$$\begin{aligned} g_m^i &= r_m(e_m(x_m)|i) \\ g_{K+1}^i &= r_{K+1}(e_{K+1}(X)|i) \end{aligned} \quad (3.16)$$

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

$$g_f^i = f(g_1^i, g_2^i, \dots, g_K^i, g_{K+1}^i) \quad (3.17)$$

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

$$\hat{y} = J(g_f^i) = d_J(r_J(g_f^i|i)) \quad (3.18)$$

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:

$$\begin{aligned} g_m^i &= d_m(r_m(e_m(x_m))|i) \\ g_{K+1}^i &= d_{K+1}(r_{K+1}(e_{K+1}(X))|i) \end{aligned} \quad (3.19)$$

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

$$g_f^i = f(g_1^i, g_2^i, \dots, g_K^i, g_{K+1}^i) \quad (3.20)$$

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

$$\hat{y} = J(g_f^i) = d_J(g_f^i|i) \quad (3.21)$$

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:

$$D_J(X, s) = c_J(X, s) \quad (3.22)$$

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:

$$\begin{aligned} L_J = & -\mathbb{E}_{X,y} [(D_J(X, y) - 1)^2] - \mathbb{E}_X [(D_J(X, J(g_f^i)))^2] \\ & + \mathbb{E}_{X,y} [\|y - J(g_f^i)\|_1] \end{aligned} \quad (3.23)$$

3.3.2 Network Architecture

The K one-to-one streams and the many-to-one stream had identical generator (G) and discriminator (D) architectures adopted from [40] and [63], respectively. 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 30×30 binary matrix of fake/real predictions, with each element of the matrix denoting the prediction for the corresponding 70×70 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.

3.3.3 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 [64] 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. Prior to PSNR/SSIM measurements and visualizations, the pixel intensities of the output images from the generators were linearly mapped onto the original 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 = $0.94 \times 0.94 \times 1.2 \text{ mm}^3$, matrix size = $256 \times 256 \times 150$. T_2 -weighted images: TE = 100 ms, TR = 8178.34 ms, flip angle = 90° , spatial resolution = $0.94 \times 0.94 \times 1.2 \text{ mm}^3$, matrix size = $256 \times 256 \times 150$. PD-weighted images: TE = 8 ms, TR = 8178.34 ms, flip angle = 90° , spatial resolution = $0.94 \times 0.94 \times 1.2 \text{ mm}^3$, matrix size = $256 \times 256 \times 150$. 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 [65]. 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 heterogeneous 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 [64], 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 [64], 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.

3.3.4 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 [9], 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 [66], where the learning rate was set to 2×10^{-4} in the first 50 epochs, and was linearly decayed from 2×10^{-4} to 0 in the last 50 epochs. During the training, the decay rates of the first moment β_1 and the second moment β_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 position 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 [9]. During the training, the Adam optimizer was employed, where the decay rates of the first moment β_1 and the second moment β_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 2×10^{-4} , and for models trained for more than 50 epochs, the learning rate was set to 2×10^{-4} in the first 50 epochs and decreased by 4×10^{-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 β_1 and the second moment β_2 of gradient estimates were again set to 0.5 and 0.999, respectively.

The total number of network parameters was approximately 1.1×10^7 for the one-to-one and many-to-one streams, and 2.3×10^6 for the discriminators. Meanwhile, it varied from 10^4 to 1.1×10^7 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.

3.3.5 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 [9]. 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 [28]. 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 [28] except for the curriculum learning to ensure a standard sample selection procedure among GAN models during trainings. The last competing method was Multimodal [19], 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 [19].

In addition, we also implemented a constraint 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 T₁, (10, 25) for T₂, and (6, 20) for PD synthesis in the IXI dataset, and (8, 45) for T₁, (5, 55) for T₂, and (12, 20) for FLAIR synthesis in the ISLES dataset. The remaining hyperparameters were matched to mustGAN.

3.3.6 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 (T₂ → T₁; PD → T₁; T₁ → T₂; PD → T₂; T₁ → PD; T₂ → PD) were considered for pGAN. Meanwhile, 3 distinct synthesis tasks (T₂, PD → T₁; T₁, PD → T₂; T₁, T₂ → 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 (T₂ → T₁; FLAIR → T₁; T₁ → T₂; FLAIR → T₂; T₁ → FLAIR; T₂ → FLAIR) were considered for pGAN. Meanwhile, 3 distinct synthesis tasks (T₂, FLAIR → T₁; T₁, FLAIR → T₂; T₁, T₂ → FLAIR) were considered for mustGAN, pGAN_{many}, MM-GAN, and Multimodal. T₁-weighted and FLAIR images were synthesized using axial cross-sections, whereas T₂-weighted images were synthesized using sagittal cross-sections (see Section 3.3.3 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 T₂ synthesis, and a separate MM-GAN model was trained using only axial cross-sections for T₁ and FLAIR synthesis.

For quantitative evaluations, the synthesized and ground truth target images were compared via PSNR and SSIM [67]. 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 were 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 differences between 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 mustGAN, pGAN and pGAN_{many} models will be publicly available on <http://github.com/icon-lab/mrirecon>.

3.4 Results

3.4.1 Task-Specific Fusion Across Multiple Streams

To optimize the mustGAN model for specific tasks, we performed experiments to determine the optimal position of the fusion 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. 3.2a-c for T_1 -weighted, T_2 -weighted and PD-weighted image synthesis in the IXI dataset, respectively.

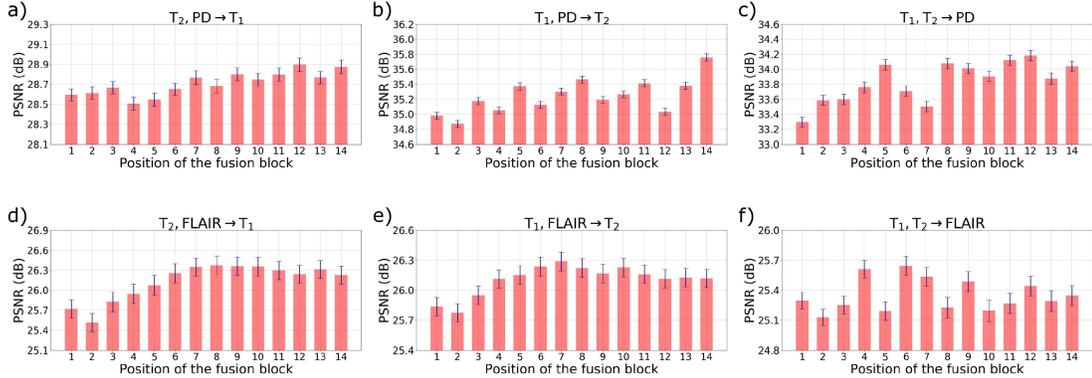


Figure 3.2: Synthesis quality of mustGAN models with varying positions of the fusion block was evaluated on the validation set in terms of PSNR measurements. The mean of the results are plotted together with standard error intervals for three distinct synthesis tasks in the IXI dataset: a) $T_2, PD \rightarrow T_1$, 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) $T_1, T_2 \rightarrow FLAIR$

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. 3.2d-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.

3.4.2 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 Figs. 3.3, 3.4. 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.

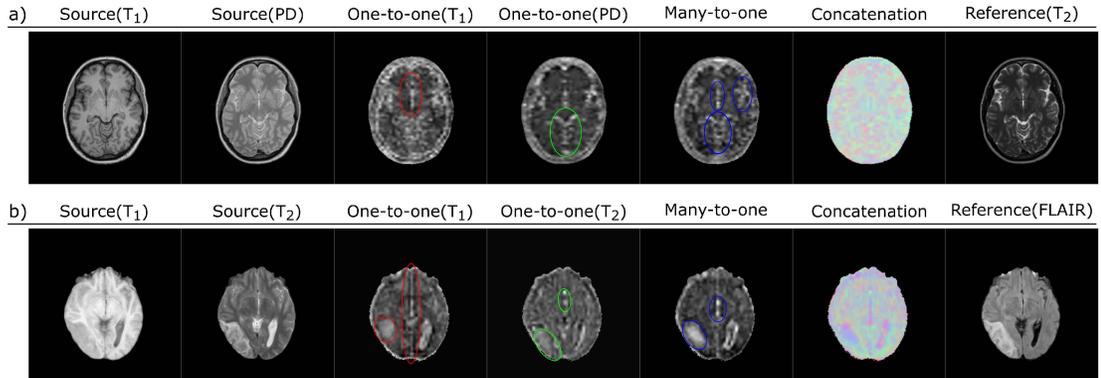


Figure 3.3: a) T₂-weighted image synthesis from T₁- and PD-weighted images in the IXI dataset, b) FLAIR image synthesis from T₁- and T₂-weighted images in the ISLES dataset. Feature maps at a central network layer (8th) in one-to-one (third and fourth columns) and many-to-one models (fifth column) are displayed along with source images and ground truth target images (reference). At the 8th 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 256×256 . 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 model). 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.

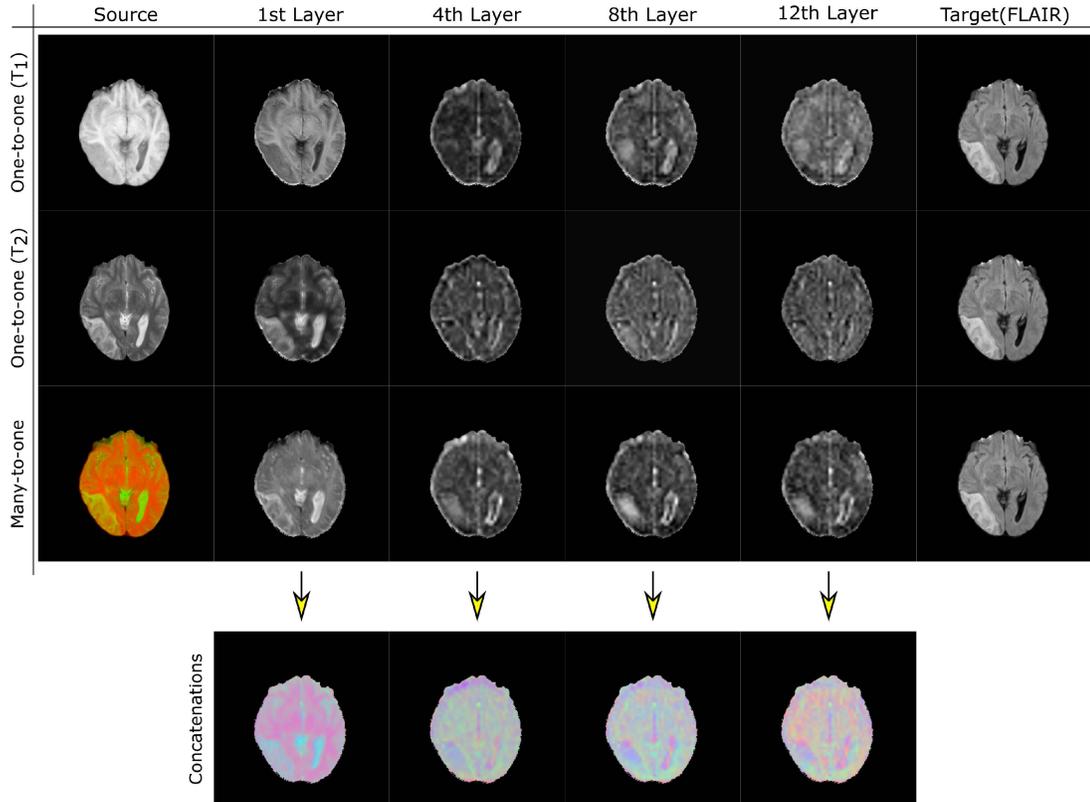


Figure 3.4: FLAIR image synthesis from T_1 - and T_2 -weighted images in the ISLES dataset. Feature maps from different network layers (1, 4, 8, 12) within the one-to-one (first and second rows) and many-to-one (third row) models are displayed along with source and ground truth target images. The two sources for the many-to-one model are displayed as red and green color channels in a combined image. Concatenation of the feature maps is also shown (fourth row), where separate feature maps from the one-to-one and many-to-one mappings are taken as RGB color channels. While one-to-one models manifest increased sensitivity to unique, detailed features that are predominantly present in the given source, many-to-one mapping pools information across sources with high sensitivity to shared features. Note that more discriminable representations among one-to-one and many-to-one streams emerge towards later layers in the network.

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 Figs. 3.5-3.7 for IXI, and in Figs. 3.8-3.10 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 proposed 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 3.3.5). Table 3.1 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.

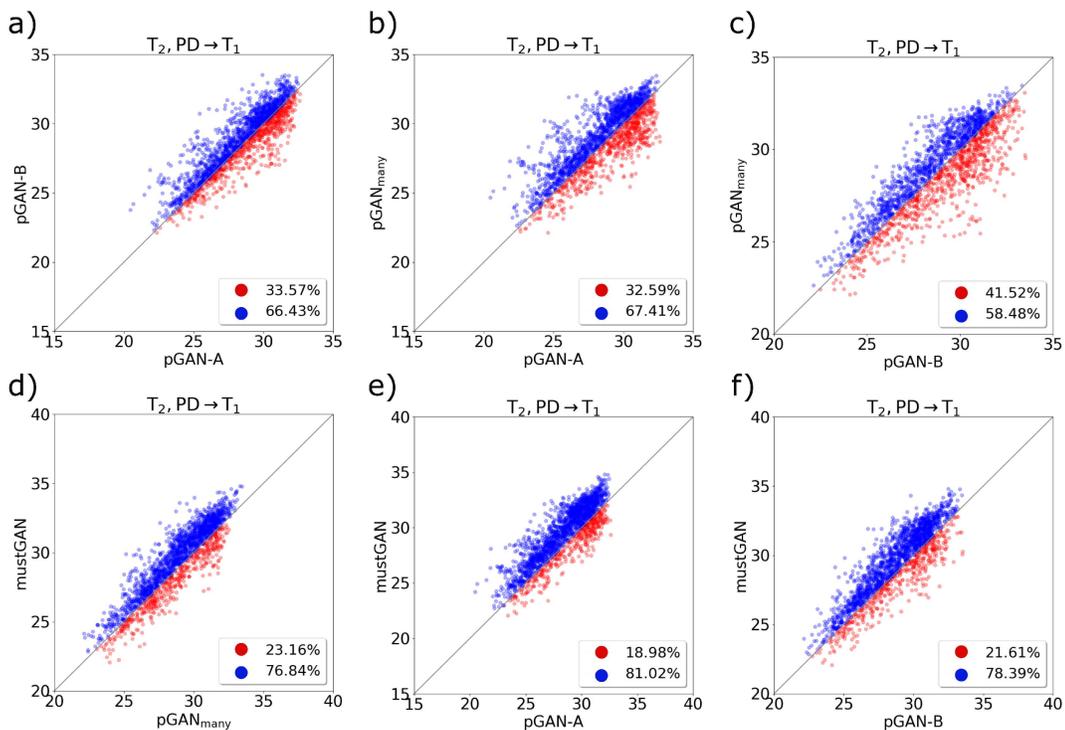


Figure 3.5: Methods were compared in terms of quality of T_1 synthesis in the IXI dataset: a) pGAN-A versus pGAN-B, b) pGAN-A versus pGAN_{many}, c) pGAN-B versus pGAN_{many}, d) pGAN_{many} versus mustGAN, e) pGAN-A versus mustGAN, f) pGAN-B versus mustGAN. Note that pGAN-A receives as input T_2 -weighted images, and pGAN-B receives as input PD-weighted images. 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).

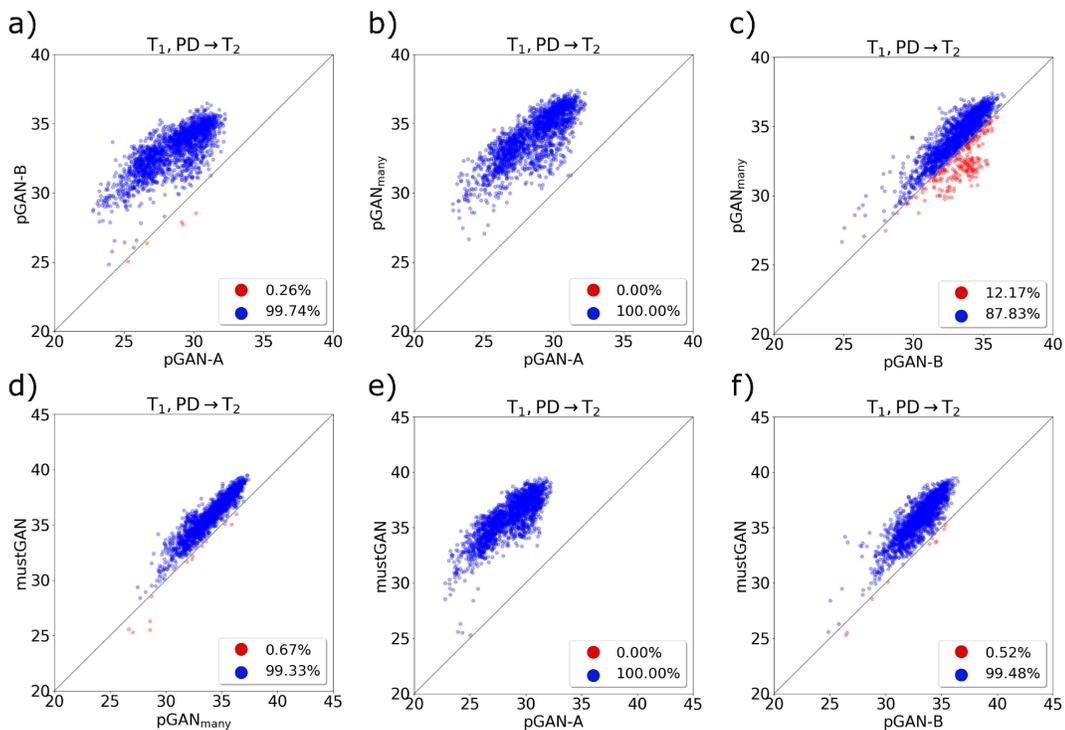


Figure 3.6: Methods were compared in terms of quality of T₂ synthesis in the IXI dataset: a) pGAN-A versus pGAN-B, b) pGAN-A versus pGAN_{many}, c) pGAN-B versus pGAN_{many}, d) pGAN_{many} versus mustGAN, e) pGAN-A versus mustGAN, f) pGAN-B versus mustGAN. Note that pGAN-A receives as input T₁-weighted images, and pGAN-B receives as input PD-weighted images. 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).

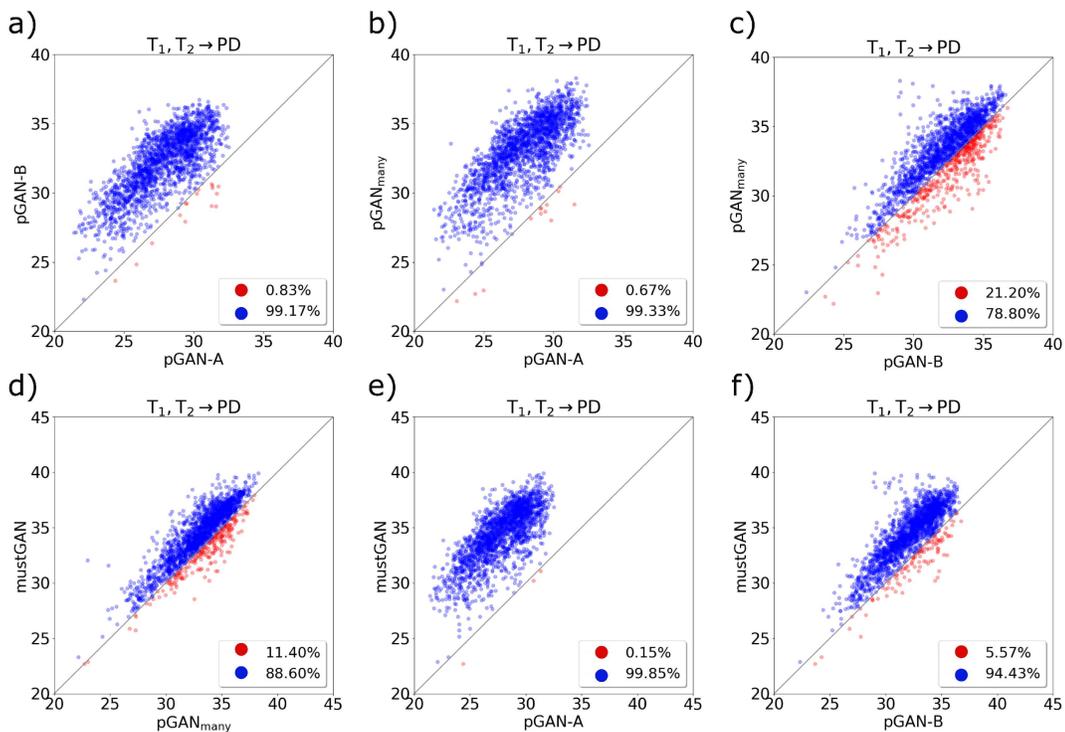


Figure 3.7: Methods were compared in terms of quality of PD synthesis in the IXI dataset: a) pGAN-A versus pGAN-B, b) pGAN-A versus pGAN_{many}, c) pGAN-B versus pGAN_{many}, d) pGAN_{many} versus mustGAN, e) pGAN-A versus mustGAN, f) pGAN-B versus mustGAN. Note that pGAN-A receives as input T_1 -weighted images, and pGAN-B receives as input T_2 -weighted images. 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).

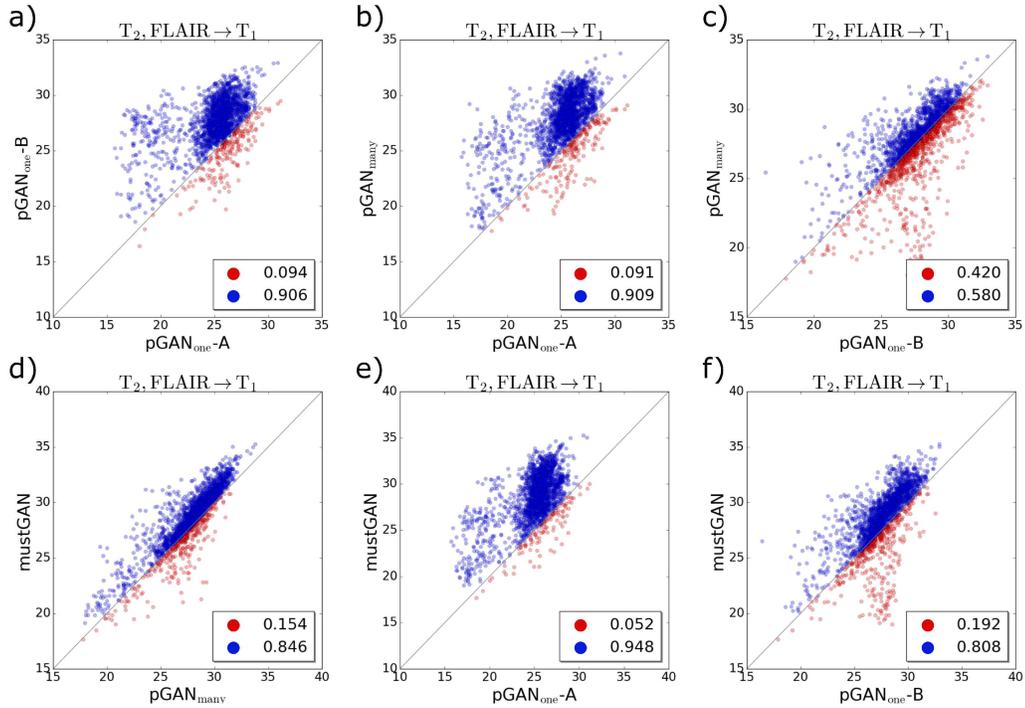


Figure 3.8: Methods were compared in terms of quality of T_1 synthesis in the ISLES dataset: a) pGAN-A versus pGAN-B, b) pGAN-A versus pGAN_{many}, c) pGAN-B versus pGAN_{many}, d) pGAN_{many} versus mustGAN, e) pGAN-A versus mustGAN, f) pGAN-B versus mustGAN. Note that pGAN-A receives T_2 -weighted 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).

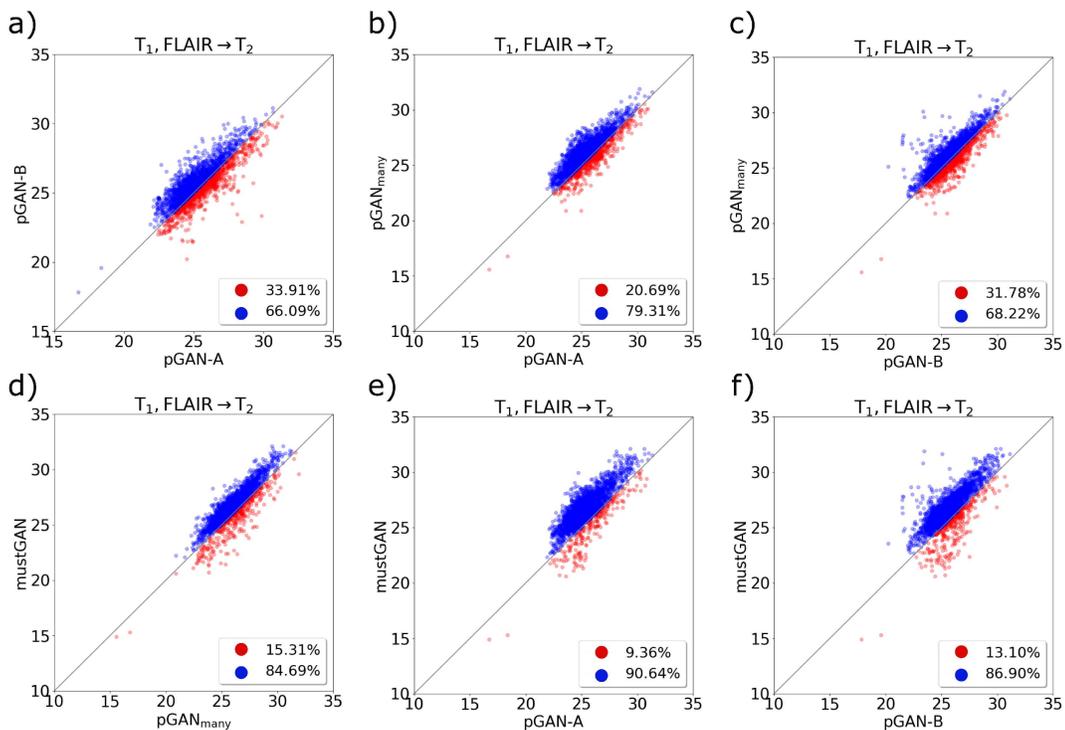


Figure 3.9: Methods were compared in terms of quality of T₂ synthesis in the ISLES dataset: a) pGAN-A versus pGAN-B, b) pGAN-A versus pGAN_{many}, c) pGAN-B versus pGAN_{many}, d) pGAN_{many} versus mustGAN, e) pGAN-A versus mustGAN, f) pGAN-B versus mustGAN. Note that pGAN-A receives as input T₁-weighted images, and pGAN-B receives as input FLAIR images. 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).

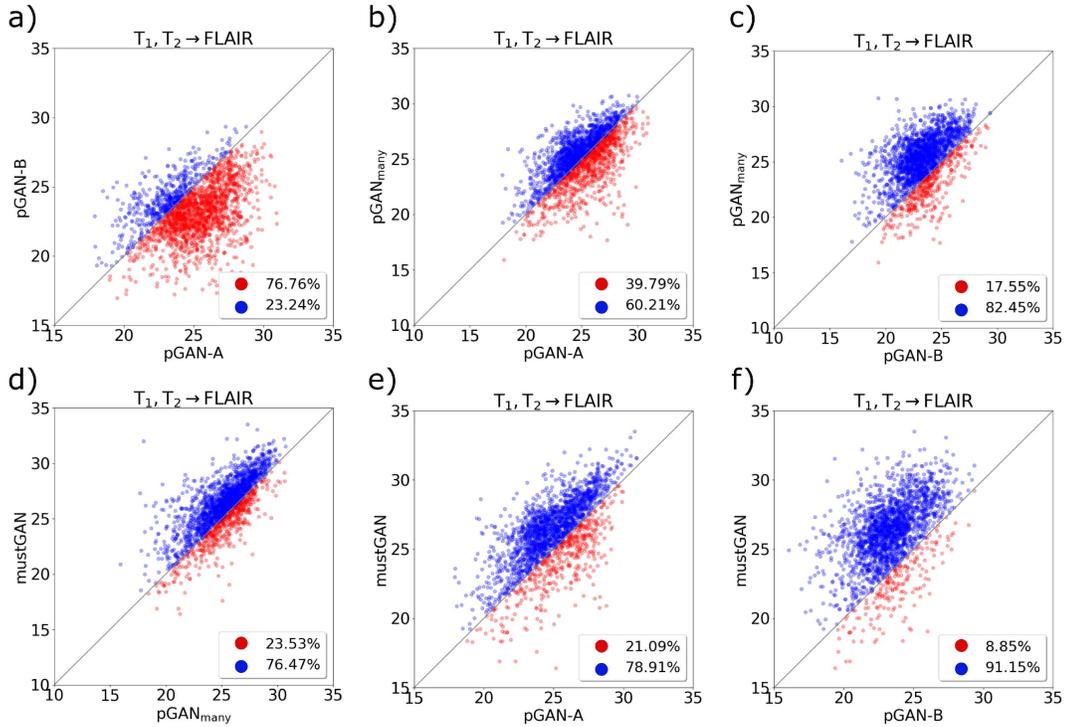


Figure 3.10: Methods were compared in terms of quality of FLAIR synthesis in the ISLES dataset: a) pGAN-A versus pGAN-B, b) pGAN-A versus pGAN_{many}, c) pGAN-B versus pGAN_{many}, d) pGAN_{many} versus mustGAN, e) pGAN-A versus mustGAN, f) pGAN-B versus mustGAN. Note that pGAN-A receives as input T_1 -weighted images, and pGAN-B receives as input T_2 -weighted images. 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).

Table 3.1: Synthesis Quality Enhancement due to the Many-to-one Stream: Synthesis quality of the proposed mustGAN method is compared against a variant of itself, namely mustGAN_{one}, that recovers the target contrast image from only the unique feature maps generated in the one-to-one streams. PSNR and SSIM measurements of mustGAN and mustGAN_{one} are reported as mean \pm std calculated across test subjects for three distinct synthesis tasks in the IXI dataset: T₂, PD \rightarrow T₁; T₁, PD \rightarrow T₂; T₁, T₂ \rightarrow PD, and for three distinct synthesis tasks in the ISLES dataset: T₂, FLAIR \rightarrow T₁; T₁, FLAIR \rightarrow T₂; T₁, T₂ \rightarrow FLAIR. Boldface marks the model having the highest performance.

		mustGAN		mustGAN _{one}	
		PSNR	SSIM	PSNR	SSIM
IXI	T2, PD \rightarrow T ₁	29.45 \pm 1.19	0.947 \pm 0.012	29.40 \pm 1.25	0.945 \pm 0.011
	T1, PD \rightarrow T2	35.89 \pm 1.20	0.977 \pm 0.005	35.70 \pm 1.24	0.976 \pm 0.005
	T1, T2 \rightarrow PD	34.40 \pm 0.97	0.974 \pm 0.005	34.16 \pm 1.11	0.972 \pm 0.005
ISLES	T2, FLAIR \rightarrow T1	28.51 \pm 2.10	0.929 \pm 0.018	28.00 \pm 2.37	0.925 \pm 0.019
	T1, FLAIR \rightarrow T2	26.63 \pm 0.74	0.904 \pm 0.013	26.62 \pm 0.77	0.902 \pm 0.012
	T1, T2 \rightarrow FLAIR	26.08 \pm 1.04	0.910 \pm 0.016	25.88 \pm 1.04	0.906 \pm 0.017

Table 3.2: 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: T₂, PD → T₁; T₁, PD → T₂; T₁, T₂ → PD. pGAN-A receives the 1st source contrast and pGAN-B receives the 2nd source contrast i.e., (1,2): (T₂, PD), (T₁, PD), (T₁, T₂). Boldface marks the model having the highest performance.

	T ₂ , PD→T ₁		T ₁ , PD→T ₂		T ₁ , T ₂ →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

3.4.3 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 methods (pGAN, pGAN_{many}, MM-GAN, and Multimodal). Three synthesis tasks were considered in the IXI dataset: T₂, PD → T₁; T₁, PD → T₂; T₁, T₂ → PD. Table 3.2 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 ($p < 0.05$). 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. 3.11a-c display radiological scores for T_1 -, T_2 - and PD-weighted image synthesis. mustGAN outperforms the competing methods in all synthesis tasks ($p < 0.05$) in terms of radiological evaluation scores, except for pGAN_{many} in T_2 synthesis, where the two methods perform similarly ($p > 0.05$). 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 Figs. 3.12, 3.13. 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: $T_2, \text{FLAIR} \rightarrow T_1$; $T_1, \text{FLAIR} \rightarrow T_2$; $T_1, T_2 \rightarrow \text{FLAIR}$. Table 3.3 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 ($p < 0.05$). 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 quantitative measurements of synthesis quality. Fig. 3.11d-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 ($p < 0.05$). 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 Figs. 3.14, 3.14. Compared to other methods, mustGAN depicts tissues with apparently lower noise levels and sharper tissue boundaries.

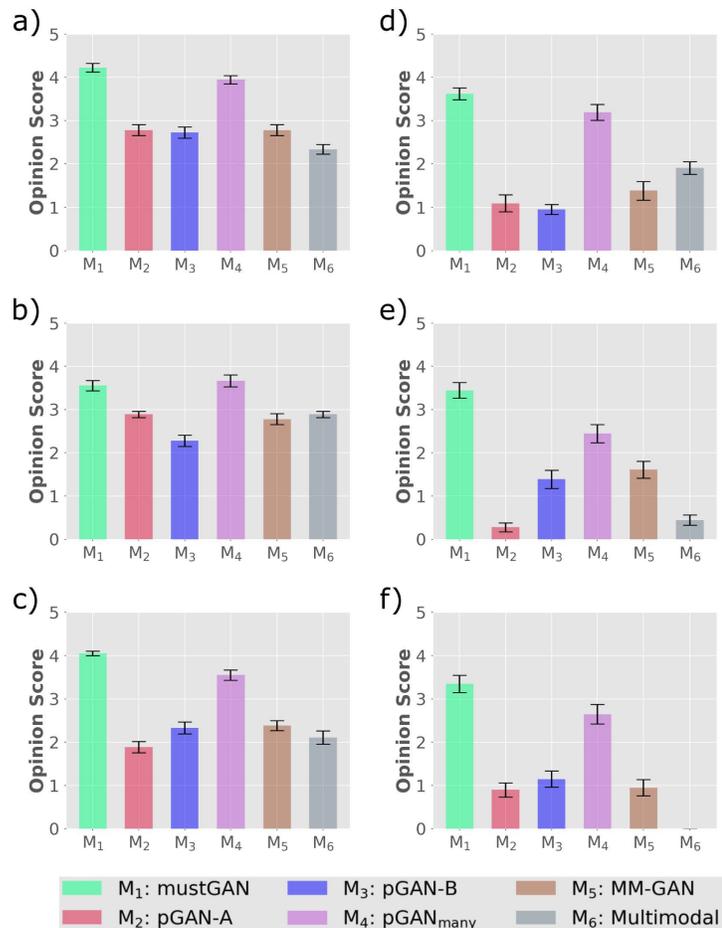


Figure 3.11: Methods were compared in terms of radiological opinion scores for three synthesis tasks in IXI: a) T_2 , PD \rightarrow T_1 , b) T_1 , PD \rightarrow T_2 , c) T_1 , $T_2 \rightarrow$ PD, and for three synthesis tasks in ISLES: d) T_2 , FLAIR \rightarrow T_1 , e) T_1 , FLAIR \rightarrow T_2 , 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 truth images, 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).

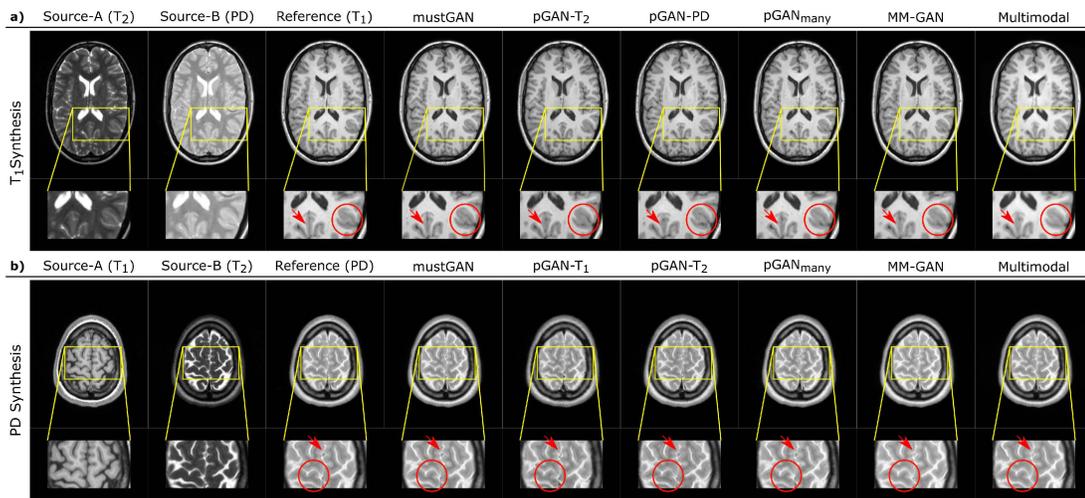


Figure 3.12: The proposed method was demonstrated on healthy subjects from the IXI dataset for two synthesis tasks: a) T_1 -weighted image synthesis from T_2 - and PD-weighted images, b) PD-weighted image synthesis from T_1 - and T_2 -weighted images. Synthesized images from mustGAN, pGAN, pGAN_{many}, 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.

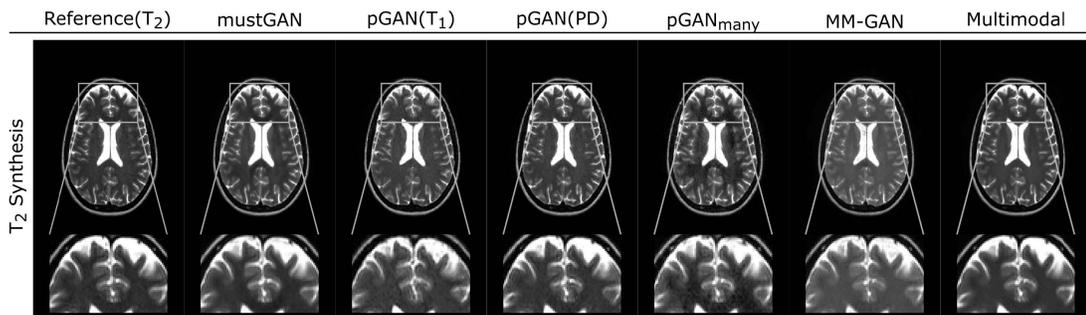


Figure 3.13: The proposed method was demonstrated on healthy subjects from the IXI dataset for T₂-weighted image synthesis from T₁- and PD-weighted images. Synthesized images from mustGAN, pGAN, pGAN_{many}, MM-GAN, and Multimodal are shown along with the ground truth target image. Due to synergistic use of information captured by one-to-one and many-to-one streams, mustGAN improves synthesis accuracy in many regions that are recovered suboptimally in competing methods (shown in zoom-in displays). Overall, mustGAN yields less noisy depiction of tissues and sharper depiction of tissue boundaries.

Table 3.3: 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 \pm std calculated across test subjects for three different synthesis tasks: T₂, FLAIR \rightarrow T₁; T₁, FLAIR \rightarrow T₂; T₁, T₂ \rightarrow FLAIR. pGAN-A receives the 1st source contrast and pGAN-B receives the 2nd source contrast i.e., (1,2): (T₂, FLAIR), (T₁, FLAIR), (T₁, T₂). 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
	\pm 2.10	\pm 0.018	\pm 0.74	\pm 0.013	\pm 1.04	\pm 0.016
pGAN-A	25.03	0.886	25.36	0.884	24.91	0.889
	\pm 1.92	\pm 0.015	\pm 0.70	\pm 0.012	\pm 0.94	\pm 0.015
pGAN-B	27.55	0.919	25.68	0.890	23.32	0.861
	\pm 1.35	\pm 0.015	\pm 0.70	\pm 0.012	\pm 0.67	\pm 0.012
pGAN _{many}	27.64	0.921	26.00	0.895	25.11	0.894
	\pm 1.88	\pm 0.017	\pm 0.68	\pm 0.012	\pm 0.81	\pm 0.013
MM-GAN	27.79	0.924	25.67	0.888	24.57	0.898
	\pm 1.65	\pm 0.017	\pm 0.54	\pm 0.010	\pm 0.88	\pm 0.015
Multimodal	27.73	0.934	25.71	0.910	19.62	0.895
	\pm 1.91	\pm 0.017	\pm 0.66	\pm 0.011	\pm 1.29	\pm 0.011

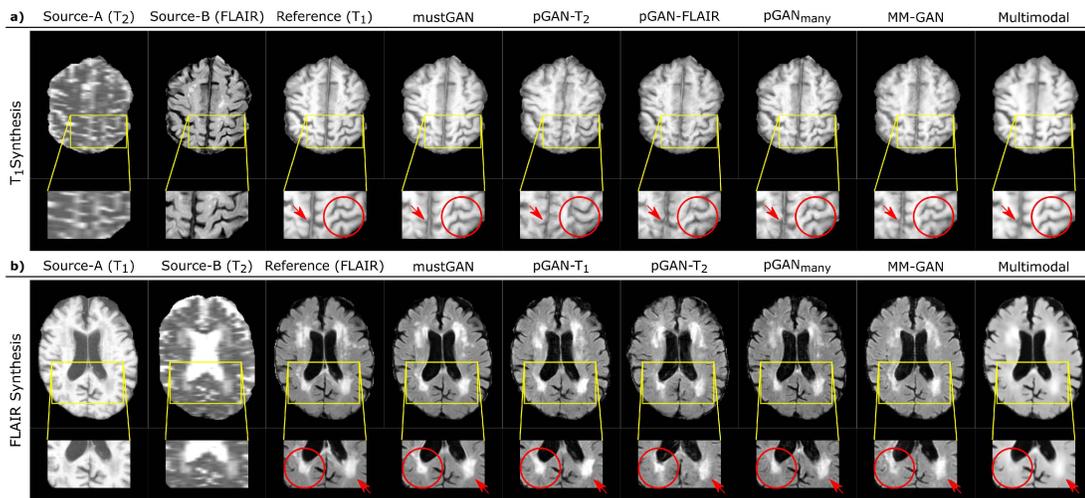


Figure 3.14: The proposed method was demonstrated on high/low grade glioma patients from the ISLES dataset for three synthesis tasks: a) T_1 -weighted image synthesis from T_2 -weighted and FLAIR images, b) FLAIR image synthesis from T_1 - and T_2 -weighted images. Synthesized images from mustGAN, pGAN, 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.

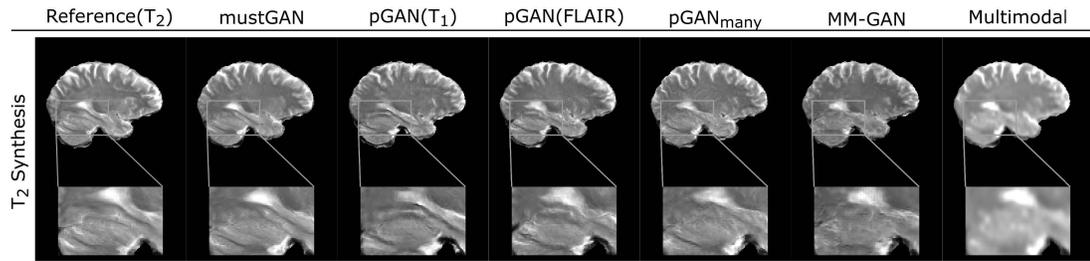


Figure 3.15: The proposed method was demonstrated on high/low grade glioma patients from the ISLES dataset for T_2 -weighted image synthesis from T_1 -weighted and FLAIR images. Synthesized images from mustGAN, pGAN, pGAN_{many}, MM-GAN, and Multimodal are shown along with the ground truth target image. Due to synergistic use of information captured by one-to-one and many-to-one streams, mustGAN improves synthesis accuracy in many regions that are recovered suboptimally in competing methods (shown in zoom-in displays). Overall, mustGAN yields less noisy depiction of tissues and sharper depiction of 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.

Table 3.4: Synthesis Quality of mustGAN Models with Varying Fusion Block Positions: PSNR and SSIM measurements between the ground truth and synthesized target images from multiple independent mustGAN models with varying fusion block positions ($i \in \{1, 2, \dots, 14\}$) are reported as mean \pm std for the test set. Three distinct synthesis tasks are considered in the IXI dataset: $T_2, PD \rightarrow T_1$; $T_1, PD \rightarrow T_2$; $T_1, T_2 \rightarrow PD$, and three distinct synthesis tasks are considered in the ISLES dataset: $T_2, FLAIR \rightarrow T_1$; $T_1, FLAIR \rightarrow T_2$; $T_1, T_2 \rightarrow FLAIR$.

	i=1	i=2	i=3	i=4	i=5	i=6	i=7	i=8	i=9	i=10	i=11	i=12	i=13	i=14
$T_2, PD \rightarrow T_1$	PSNR	29.18	29.07	29.17	29.06	29.14	29.33	29.30	29.19	29.45	29.36	29.37	29.45	29.47
	SSIM	± 1.19	± 1.09	± 1.17	± 1.24	± 1.20	± 1.13	± 1.13	± 1.10	± 1.26	± 1.19	± 1.30	± 1.19	± 1.15
$T_1, PD \rightarrow T_2$	PSNR	0.946	0.944	0.948	0.945	0.950	0.952	0.949	0.950	0.949	0.947	0.947	0.948	0.949
	SSIM	± 0.012	± 0.012	± 0.013	± 0.013	± 0.012	± 0.011	± 0.010	± 0.011	± 0.012	± 0.012	± 0.012	± 0.012	± 0.011
$T_1, T_2 \rightarrow PD$	PSNR	35.00	34.90	35.14	35.24	35.54	35.29	35.54	35.67	35.44	35.39	35.36	35.34	35.89
	SSIM	± 1.33	± 1.27	± 1.20	± 1.18	± 1.21	± 1.13	± 1.24	± 1.20	± 1.11	± 1.08	± 1.07	± 1.12	± 1.13
$T_2, FLAIR \rightarrow T_1$	PSNR	0.974	0.970	0.970	0.972	0.977	0.970	0.977	0.975	0.974	0.973	0.977	0.972	0.975
	SSIM	± 0.006	± 0.006	± 0.006	± 0.006	± 0.005	± 0.006	± 0.005	± 0.005	± 0.005	± 0.005	± 0.005	± 0.005	± 0.005
$T_1, FLAIR \rightarrow T_2$	PSNR	33.54	33.81	33.89	34.03	34.36	34.16	33.81	34.40	34.25	34.29	34.38	34.40	34.12
	SSIM	± 1.15	± 1.23	± 1.15	± 1.19	± 1.13	± 1.25	± 1.26	± 1.20	± 1.17	± 1.14	± 1.15	± 0.97	± 1.28
$T_1, T_2 \rightarrow FLAIR$	PSNR	0.969	0.970	0.973	0.972	0.974	0.976	0.970	0.978	0.974	0.972	0.976	0.974	0.977
	SSIM	± 0.005	± 0.006	± 0.004	± 0.005	± 0.005	± 0.005	± 0.006	± 0.005	± 0.005	± 0.005	± 0.004	± 0.005	± 0.004
$T_2, PD \rightarrow T_1$	PSNR	27.96	27.82	28.08	28.23	28.38	28.34	28.27	28.51	28.40	28.39	28.49	28.43	28.11
	SSIM	± 1.91	± 2.10	± 2.19	± 2.05	± 2.04	± 2.27	± 2.16	± 2.10	± 2.08	± 2.06	± 1.79	± 1.84	± 2.01
$T_1, FLAIR \rightarrow T_2$	PSNR	0.924	0.923	0.927	0.926	0.929	0.928	0.928	0.929	0.928	0.928	0.929	0.928	0.936
	SSIM	± 0.017	± 0.019	± 0.018	± 0.018	± 0.019	± 0.018	± 0.018	± 0.018	± 0.018	± 0.017	± 0.017	± 0.018	± 0.018
$T_1, T_2 \rightarrow FLAIR$	PSNR	26.22	26.35	26.32	26.37	26.47	26.61	26.63	26.51	26.54	26.55	26.48	26.48	26.43
	SSIM	± 0.76	± 0.86	± 1.15	± 0.92	± 1.02	± 0.89	± 0.88	± 0.86	± 0.85	± 0.78	± 0.73	± 0.70	± 0.68
$T_1, T_2 \rightarrow FLAIR$	PSNR	0.898	0.904	0.903	0.903	0.904	0.904	0.904	0.902	0.902	0.904	0.903	0.901	0.900
	SSIM	± 0.013	± 0.013	± 0.012	± 0.012	± 0.012	± 0.013	± 0.013	± 0.013	± 0.013	± 0.013	± 0.013	± 0.013	± 0.013
$T_1, T_2 \rightarrow FLAIR$	PSNR	25.30	25.19	25.50	25.75	25.74	26.08	25.74	25.60	25.89	25.53	25.92	25.85	25.96
	SSIM	± 0.88	± 1.06	± 0.88	± 0.92	± 1.01	± 1.04	± 1.06	± 1.03	± 1.14	± 1.04	± 1.09	± 1.06	± 1.07
$T_1, T_2 \rightarrow FLAIR$	PSNR	0.899	0.899	0.899	0.905	0.909	0.910	0.907	0.906	0.909	0.906	0.905	0.910	0.910
	SSIM	± 0.014	± 0.015	± 0.015	± 0.016	± 0.016	± 0.016	± 0.017	± 0.016	± 0.016	± 0.015	± 0.017	± 0.017	± 0.017

3.5 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 [19]. 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 [9].

Several recent studies have proposed GAN-based architectures for multi-contrast MRI synthesis. In [9], we have proposed pGAN that uses conditional GAN models for one-to-one synthesis. In [28], 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: [68, 69, 70], MRI synthesis from CT: [71, 72], PET synthesis from CT: [73, 74], CT synthesis from PET: [75], PET synthesis from MRI: [76, 77], MRI synthesis from PET: [78]. 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, 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 [9] and many-to-one [22] 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 [79] 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.

Here, the joint network in mustGAN was designed to fuse feature maps from a single network layer that is jointly optimized across all underlying streams. Alternatively, a distributed fusion strategy could be used to flexibly fuse feature maps from multiple different layers within each stream. The network layers for the feature maps can also be independently optimized across separate streams. These advances can help further increase synthesis quality, albeit at the expense of elevated computational load.

In summary, here we proposed a novel synthesis model for multi-contrast MRI that achieves state-of-the-art performance due to its hybrid architecture. A similar approach utilizing ensemble learning can be adapted to other anatomies, imaging protocols, recovery tasks, or modalities [80, 81, 82, 83, 84].

Chapter 4

Semi-Supervised Learning of Mutually Accelerated MRI Synthesis without Fully-Sampled Ground Truths

The content of this chapter reflects the study described in the following publications:

- M. Yurt, S. U. H. Dar, M. Özbey, B. Tınaz, K. K. Oğuz, and T. Çukur, “Semi-supervised learning of mutually accelerated MRI synthesis without fully-sampled ground truths,” *arXiv preprint arXiv:2011.14347*, 2021
- M. Yurt, B. Tınaz, M. Özbey, S. U. H. Dar, and T. Çukur, “Semi-supervised learning of multi-contrast MR image synthesis without fully-sampled ground-truth acquisitions,” in *Medical Imaging Meets NeurIPS*, (Virtual Conference), 12 2020
- M. Yurt, S. U. H. Dar, B. Tınaz, M. Özbey, Y. Korkmaz, and T. Çukur, “A semi-supervised learning framework for jointly accelerated multi-contrast

MRI synthesis without fully-sampled ground-truths,” in *29th annual meeting of International Society for Magnetic Resonance Imaging (ISMRM)*, (Virtual Conference), 5 2021

4.1 Introduction

MRI is a clinical powerhouse in neuroimaging due to its noninvasiveness and excellent soft-tissue contrast. Its unique ability to image the same anatomy under a diverse set of tissue contrasts empowers it to accumulate complementary diagnostic information within a single exam session [88, 89]. However, prolonged scans and increased costs associated with multi-contrast protocols often limit the diversity and quality of MRI exams [5, 4]. A promising solution against this limitation is synthesis of missing or unacceptably low-quality images within the protocol from available high-quality images [6, 90]. Multi-contrast MRI synthesis methods can enhance radiological assessments as well as image analysis tasks such as registration, segmentation, or detection [91, 92, 35, 93].

In recent years, there has been emerging interest in learning-based MRI synthesis based on deep neural networks, given their state-of-the-art performance in other computer vision [44, 56, 94, 63, 53] and medical imaging tasks [95, 96, 97, 98, 99]. An earlier group of studies proposed deep models with convolutional neural networks (CNNs) to learn nonlinear latent representations that mediate conversion from source to target images [19, 14, 25, 29, 8]. These studies typically involved encoder-decoder architectures, where the encoder embeds hierarchical image features onto a latent space that is later used by the decoder to recover the target image [19, 14, 25, 29, 8]. For improved capture of structural details, a second group has proposed deep architectures based on conditional generative adversarial networks (GAN) [9, 28, 100, 69, 22, 24, 39, 75, 48, 21, 101, 38], where the generator that performs the source-to-target mapping benefits from the game-theoretic interplay with the discriminator [44]. Pioneering studies have exploited pixel- or feature-wise correspondence between source-target images in an

adversarial setup [9, 21, 48]. Later studies have proposed unified models capable of multiple types of contrast conversion [28, 22, 24], or multi-tasking frameworks [101, 22, 28] to reduce computational complexity. These previous studies have collectively highlighted the immense potential of learning-based synthesis in multi-contrast MRI. That said, both CNN and GAN models are canonically trained in a fully-supervised setup based on pixel-wise, adversarial or perceptual losses between synthesized and ground truth target images. Supervised models require large datasets of high-quality images from Nyquist-sampled source and target acquisitions, paired within subjects [19, 28, 9, 102]. Yet, compilation of paired, high-quality datasets might prove impractical due to scan time and cost considerations [5, 4]. As such, there is a dire need for methods with lower reliance on supervision to improve practicality of learning-based MRI synthesis.

Recent efforts to lower supervision requirements in MRI synthesis have predominantly focused on model training in the absence of paired images across subjects. For unpaired training, a successful approach has been to replace pixel-wise losses in GAN models with cycle-consistency, shape-consistency or mutual information losses [27, 22, 15, 9, 103, 104]. Similar to supervised models, unpaired models that unify multiple contrast conversion tasks have also been introduced to reduce computational complexity [15, 53]. As an alternative, [105, 106] have proposed a hybrid method where the model is trained on a composite dataset with both paired and unpaired samples. These previous methods have increased the capacity of synthesis models to learn from unpaired data, but they still leverage high-quality MR images reconstructed from fully-sampled k-space acquisitions. While training of MRI reconstruction [107, 108] models from undersampled data has received recent interest [109, 110], to the best of our knowledge, no prior study has considered learning of MRI synthesis models from undersampled source or target acquisitions.

Here, we propose a novel semi-supervised deep generative model for multi-contrast MRI synthesis, namely ssGAN, to avoid reliance on fully-sampled k-space acquisitions. The proposed model is trained directly on undersampled acquisitions, and it generates high-quality target images given undersampled multi-coil source acquisitions. To do this, ssGAN introduces novel multi-coil tensor losses

in image, k-space and adversarial domains. These selective losses are based only on acquired k-space samples, and randomized sampling masks are used across subjects to capture relationships among acquired and non-acquired k-space regions. Comprehensive experiments performed on brain MRI clearly demonstrate that ssGAN achieves equivalent performance to gold-standard models based on fully-supervised training across a broad range of acceleration factors. Meanwhile, ssGAN outperforms a cascade-model that first reconstructs undersampled acquisitions using compressive sensing [111, 112, 113], and then trains a learning-based synthesis model.

Contributions

- To the best of our knowledge, this is the first semi-supervised learning method for multi-contrast MRI synthesis that performs model training based on undersampled source and target acquisitions.
- The proposed method synthesizes target images directly from undersampled multi-coil source acquisitions.
- The proposed method introduces novel multi-coil tensor losses in image, k-space and adversarial domains, selectively expressed based on acquired k-space samples in target contrast acquisitions.
- The proposed method substantially lowers data requirements in MRI synthesis by enabling model training and inference from undersampled acquisitions.

4.2 Methods

In this section, we first overview basics of generative adversarial networks, and the foundation of the proposed architecture for semi-supervised multi-contrast MRI synthesis. We then describe in detail the datasets and experiments conducted to evaluate the proposed methodology.

4.2.1 Generative Adversarial Networks

Generative adversarial networks (GANs) [44] are deep generative models comprising a pair of competing subnetworks: a generator (G) and a discriminator (D). G aims to map a random noise vector z to a sample resembling a target domain distribution, whereas D aims to distinguish between real and fake samples of the target domain [44]. These two subnetworks are alternately trained via an adversarial loss function, formulated as follows:

$$L_{GAN} = -\mathbb{E}_y[(D(y) - 1)^2] - \mathbb{E}_z[D(G(z))^2] \quad (4.1)$$

where \mathbb{E} denotes expectation, and y is an arbitrary real sample in the target domain. Upon convergence, G is expected to generate realistic target domain samples that D cannot tell apart from the real ones [44]. While the initial GAN models generated target samples from a random noise vector, later studies have demonstrated success in image-to-image translation with conditional GAN (cGAN) models that additionally receive as input a source domain image x [56]. The adversarial loss function is therefore modified by conditioning G on x :

$$L_{cGAN} = -\mathbb{E}_{x,y}[(D(y) - 1)^2] - \mathbb{E}_x[D(G(x))^2] \quad (4.2)$$

When spatially aligned source-target images are available, a pixel-wise loss can be further included [63]:

$$\begin{aligned} L_{cGAN} = & -\mathbb{E}_{x,y}[(D(y) - 1)^2] - \mathbb{E}_x[D(G(x))^2] \\ & + \mathbb{E}_{x,y}[\|y - G(x)\|_1] \end{aligned} \quad (4.3)$$

Several studies have demonstrated variants of cGAN models on multi-contrast MRI that synthesize target contrast images from source contrast images of the same underlying anatomy [9, 28, 100, 69, 22, 24, 39, 75, 48, 21, 101, 38]. These models typically learn the source-to-target mapping in a fully-supervised setup.

A comprehensive training set is needed containing high-quality source and target images reconstructed from fully-sampled k-space acquisitions $(x_{\mathbf{1}}, y_{\mathbf{1}})$, where $x_{\mathbf{1}}$ is an arbitrary source, $y_{\mathbf{1}}$ is an arbitrary target image in the training set, and $\mathbf{1}$ denotes the the sampling mask for Nyquist-sampled acquisitions. These fully-supervised models have demonstrated state-of-the-art performance for synthetic multi-contrast MRI. However, they are limited due to reliance on fully-sampled acquisitions that might prove impractical. Therefore, there is a critical need for methods that can directly learn from undersampled MRI data.

4.2.2 Semi-Supervised Generative Adversarial Networks

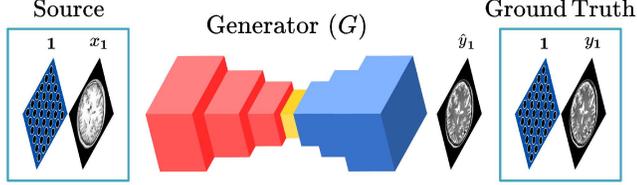
Here, we propose a novel semi-supervised GAN model, namely ssGAN, to mitigate the dependency of MRI synthesis models on supervised training with Nyquist-sampled source and target acquisitions. ssGAN is trained on undersampled acquisitions of source and target contrasts, and it synthesizes multi-coil target images directly from undersampled multi-coil acquisitions of the source contrast. To do this, ssGAN introduces novel selective loss functions expressed based on only the acquired subset of k-space samples in the target contrast (Fig. 4.1). Details regarding the optimization objectives of ssGAN are provided in the remainder of this section.

ssGAN receives as input Fourier reconstructions of either fully-sampled or undersampled acquisitions of the source contrast, and learns to synthesize high-quality images of the target contrast. The generator G in ssGAN produces target contrast images via a forward mapping:

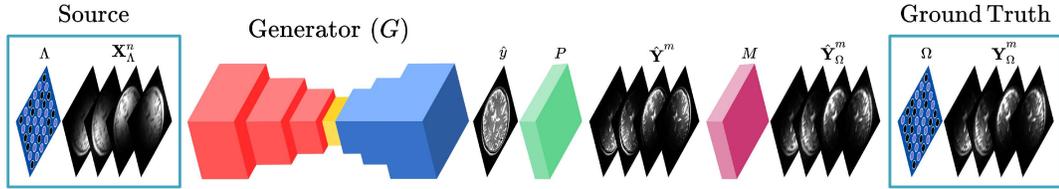
$$G(\mathbf{X}_{\Lambda}^n) = \hat{y}, \quad \text{with } \mathbf{X}_{\Lambda}^n = \{x_{\Lambda}^1, \dots, x_{\Lambda}^n\} \quad (4.4)$$

where \mathbf{X}_{Λ}^n denotes multi-coil source contrast images acquired with a k-space sampling mask Λ , n denotes the number of receive coils with sensitivity maps $\hat{\mathbf{C}}_{\mathbf{X}}^n$ computed via ESPIRiT [114], and \hat{y} denotes the synthesized coil-combined target contrast image. Note that ssGAN considers that only undersampled acquisitions

1) Fully-Supervised Generative Adversarial Networks



2) Semi-Supervised Generative Adversarial Networks



3) Selective Loss Function

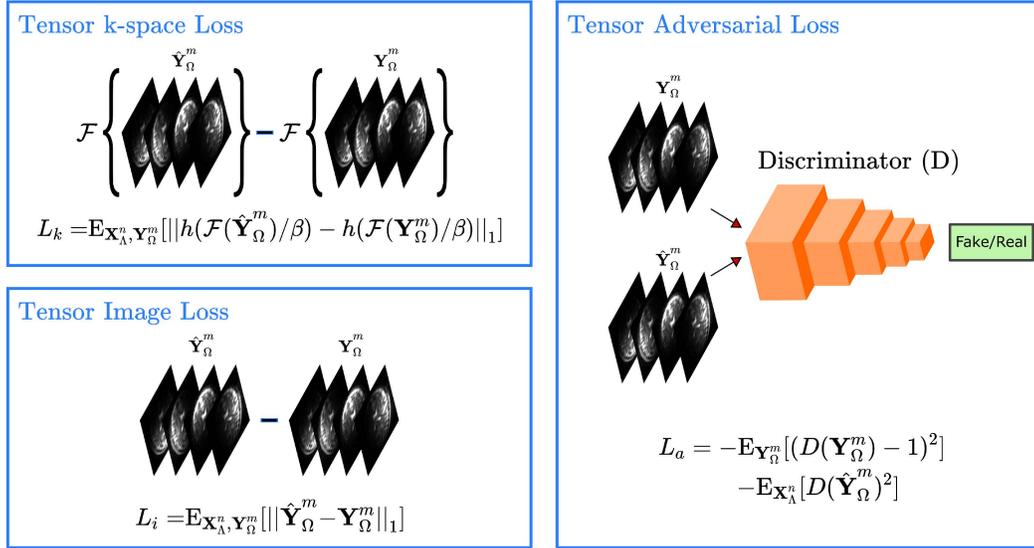


Figure 4.1: Illustration of the proposed semi-supervised ssGAN model. As opposed to fully-supervised models that demand Nyquist-sampled acquisitions for training (1), ssGAN learns to synthesize high-quality images given a dataset of undersampled source and target acquisitions (2). ssGAN initially synthesizes a coil-combined target image that is backprojected onto individual coils via sensitivity maps. These multi-coil target images are subsampled in Fourier domain with the target acquisition mask in order to define the selective multi-coil tensor losses in image, k-space and adversarial domains (3).

of the target contrast are available, where $\mathbf{Y}_\Omega^m = \{y_\Omega^1, \dots, y_\Omega^m\}$ denotes Fourier reconstructions of multi-coil target acquisitions collected with a sampling mask Ω and m receive coils of true coil sensitivities \mathbf{C}_Y^m . As no high-quality reference for the target contrast image is assumed, ssGAN expresses novel selective loss functions based on only the acquired subset of k-space samples. To do this, the synthesized coil-combined image is first projected onto individual coils as follows:

$$\hat{\mathbf{Y}}^m = P(\hat{y}, \hat{\mathbf{C}}_Y^m) = \hat{y} \cdot \hat{\mathbf{C}}_Y^m \quad (4.5)$$

where $\hat{\mathbf{Y}}^m$ denotes the synthesized multi-coil target contrast images, $\hat{\mathbf{C}}_Y^m$ denotes estimated coil sensitivity maps computed via ESPIRiT [114], and P is the operator that performs the coil projection in the image domain as dot product takes vectors and outputs a scalar, element-wise multiplication between the input image and coil sensitivity maps. The multi-coil target image projections are then subjected to the binary sampling mask in Fourier domain:

$$\begin{aligned} \hat{k}_{Y_\Omega^m} &= M(\mathcal{F}(\hat{\mathbf{Y}}^m), \Omega) = \mathcal{F}(\hat{\mathbf{Y}}^m) \cdot \Omega \\ \hat{\mathbf{Y}}_\Omega^m &= \mathcal{F}^{-1}(\hat{k}_{Y_\Omega^m}) \end{aligned} \quad (4.6)$$

where \mathcal{F} denotes the forward and \mathcal{F}^{-1} denotes the inverse Fourier transform, M is the operator that performs binary masking in k-space to with a given sampling mask. In Eq. (6) $\hat{k}_{Y_\Omega^m}$ and $\hat{\mathbf{Y}}_\Omega^m$ denote undersampled multi-coil data respectively in k-space and image domain for the synthesized target contrast image. The selective loss function in ssGAN is then defined between undersampled synthesized and undersampled ground truth data for the target contrast, based on three loss components: multi-coil tensor image loss, multi-coil tensor k-space loss, and multi-coil tensor adversarial loss. Each loss term is described below.

4.2.2.1 Multi-Coil Tensor Image Loss

The first component of the selective loss function is a multi-coil tensor image loss defined based on undersampled multi-coil data in image domain, between

synthesized and ground truth target images:

$$L_i = E_{\mathbf{X}_\Lambda^n, \mathbf{Y}_\Omega^m} [\|\hat{\mathbf{Y}}_\Omega^m - \mathbf{Y}_\Omega^m\|_1] \quad (4.7)$$

where \mathbf{Y}_Ω^m denotes the multi-coil ground truth target images from accelerated acquisitions, and $\hat{\mathbf{Y}}_\Omega^m$ denotes the undersampled target images generated by ssGAN.

4.2.2.2 Multi-Coil Tensor k-space Loss

The quality of the synthesized images in ssGAN is further enhanced via a multi-coil tensor k-space loss expressed between the Fourier-domain data of the synthesized and ground truth images.

$$L_k = E_{\mathbf{X}_\Lambda^n, \mathbf{Y}_\Omega^m} [\|h(\mathcal{F}(\hat{\mathbf{Y}}_\Omega^m)/\beta) - h(\mathcal{F}(\mathbf{Y}_\Omega^m)/\beta)\|_1] \quad (4.8)$$

where h is a *tanh* function with a normalization constant β to provide a comparable signal intensities across k-space, and $\mathcal{F}(\mathbf{Y}_\Omega^m)$ - $\mathcal{F}(\hat{\mathbf{Y}}_\Omega^m)$ stand for k-space data of the ground truth and synthesized multi-coil images, respectively.

4.2.2.3 Multi-Coil Tensor Adversarial Loss

The level of realism in the synthesized images is advanced via a multi-coil adversarial loss function evaluated between image-domain data of the synthesized and ground truth multi-coil images:

$$L_a = -E_{\mathbf{Y}_\Omega^m} [(D(\mathbf{Y}_\Omega^m) - 1)^2] - E_{\mathbf{X}_\Lambda^n} [D(\hat{\mathbf{Y}}_\Omega^m)^2] \quad (4.9)$$

where D denotes the discriminator that distinguishes between undersampled ground truth and synthesized images.

The final selective loss function for ssGAN is constructed as a weighted combination of the three multi-coil tensor loss terms described as $L_{ssGAN} = \lambda_k L_k + \lambda_i L_i + \lambda_a L_a$, where λ_k , λ_i , and λ_a denote the relative weighting of the tensor k-space, image, and adversarial losses. Note that the selective loss function along with randomization of the k-space sampling masks across training subjects enables ssGAN to effectively capture complex relationships between acquired and non-acquired k-space coefficients. In turn, ssGAN can successfully recover high-quality target images without requiring Nyquist-sampled acquisitions of the target contrast.

4.2.3 Datasets

The proposed ssGAN model was demonstrated on the public IXI dataset (<https://brain-development.org/ixi-dataset/>) containing multi-contrast single-coil magnitude brain images and an in-house dataset of multi-contrast multi-coil complex brain images.

4.2.3.1 The IXI Dataset

T_1 - and T_2 -weighted single-coil magnitude brain MR images of 94 subjects were used, where 64 were reserved for training, 10 for validation, and 20 for testing. The acquisitions parameters were as follows:

- **T_1 -weighted images:** TR = 9.81 ms, TE = 4.603 ms, flip angle = 8° , matrix size = $256 \times 256 \times 150$, spatial resolution = $0.94 \times 0.94 \times 1.2 \text{ mm}^3$, acquisition time = 4 : 42.
- **T_2 -weighted images:** TR = 8178.34 ms, TE = 100 ms, flip angle = 90° , matrix size = $256 \times 256 \times 150$, spatial resolution = $0.94 \times 0.94 \times 1.2 \text{ mm}^3$, acquisition time = 3 : 11.

Since multi-contrast images within subjects were spatially unaligned, T_2 -weighted images were registered onto T_1 -weighted images via FSL [65] prior to experiments. Registration was performed using an affine transformation based on mutual information. For demonstrations, brain images for individual cross-sections were retrospectively undersampled in two-dimensions to yield acceleration ratios $R = [2 : 1 : 10]$, via uniform random sampling with a 10×10 central fully-sampled k-space region.

4.2.3.2 In vivo Brain Dataset

T_2 - and PD-weighted multi-coil complex images of 10 subjects were used, where 7 were reserved for training, 1 for validation, and 2 for testing. The acquisitions parameters were as follows:

- **T_2 -weighted images:** 3D spin-echo sequence, $TR = 1000\text{ ms}$, $TE = 118\text{ ms}$, flip angle = 90° , imaging matrix = $256 \times 192 \times 88$, spatial resolution = $1 \times 1 \times 2\text{ mm}^3$, acquisition time = 17 : 39.
- **PD-weighted images:** 3D spin-echo sequence, $TR = 750\text{ ms}$, $TE = 12\text{ ms}$, flip angle = 90° , imaging matrix = $256 \times 192 \times 88$, spatial resolution = $1 \times 1 \times 2\text{ mm}^3$, acquisition time = 13 : 14.

Because there was negligible interscan motion, no spatial registration was performed. Data were collected on a 3T Siemens Magnetom scanner using a 32-channel receive-only head coil at Bilkent University, Ankara, Turkey. Imaging protocols were approved by the local ethics committee at Bilkent University, and all participants provided written informed consent. To lower computational complexity, geometric-decomposition coil compression was performed to reduce the number of coils from 32 to 5 [115]. For demonstrations, brain images for individual cross-sections were retrospectively undersampled in two-dimensions to yield acceleration ratios $R = [2 : 1 : 4]$, via uniform random sampling with a 16×16 central fully-sampled k-space region.

4.2.4 Implementation Details

The architecture of the generator and discriminator in ssGAN were adopted from a previous state-of-the-art study that demonstrated success in multi-contrast MRI synthesis [9]. The generator contained an encoder of 3 convolutional layers, a residual network of 9 ResNet blocks, and a decoder of 3 convolutional layers in series. The discriminator contained a convolutional neural network of 5 convolutional layers in series. An unlearned coil-combination block was placed at the input of the generator, so the generator recovered real-imaginary parts of the target image given real-imaginary parts of the coil-combined source image. The coil-combined target image was backprojected onto individual coils, and the complex target images from each coil were sequentially fed to an unconditional patch discriminator. The generator and the discriminator were alternately trained for 100 epochs with a batch size of 1 using the ADAM optimizer with first and second gradient moments of $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The learning rate of the optimizer was set to 0.0002 in the first 50 epochs and was linearly decayed to 0 in the last 50 epochs. Cross-validation was used to select the relative weighting of the selective loss function components $(\lambda_i, \lambda_k, \lambda_a, \beta)$ by maximizing synthesis performance based on network loss in the validation set. The set of parameters $(\lambda_i = 100, \lambda_k = 3000, \lambda_a = 1, \beta = 5000)$ that yielded near-optimal performance in both datasets were used in all experiments. Implementations were run on nVidia 1080 Ti and 2080 Ti GPUs in Python2.7 using PyTorch. Code will be available at <http://github.com/icon-lab/mrirecon>.

4.2.5 Competing Methods

The proposed semi-supervised ssGAN model was comparatively demonstrated for multi-contrast MRI synthesis against several state-of-the-art methods.

4.2.5.1 pix2pix [63] (fully-sampled source, fully-sampled target)

The fully-supervised pix2pix model based on paired, Nyquist-sampled source-target acquisitions sets a gold-standard for synthesis performance. pix2pix learns a mapping between coil-combined, magnitude source-target images, so it is geared for single-coil MRI synthesis. Here, pix2pix was trained with single-coil nonselective variants of tensor losses in ssGAN. The generator-discriminator architectures were taken from [9], and hyperparameters were optimized via cross-validation.

4.2.5.2 CycleGAN [94] (fully-sampled source, fully-sampled target)

The CycleGAN model based on unpaired albeit fully-sampled source-target acquisitions is another gold-standard reference. CycleGAN also learns the contrast mapping between coil-combined magnitude MR images, making it suitable for single-coil synthesis. The network architecture, hyperparameters, and loss functions (single-coil nonselective adversarial and cycle-consistency) in [9] were adopted.

4.2.5.3 fsGAN (undersampled source, fully-sampled target)

The fsGAN model was constructed as a gold-standard supervised baseline in cases where target acquisitions were Nyquist sampled but source acquisitions were undersampled. fsGAN learns to map Fourier reconstructions of undersampled, multi-coil source acquisitions onto target images. Here it was trained using nonselective variants of tensor image, k-space and adversarial loss functions in ssGAN. The network architecture was matched to ssGAN. Hyperparameters were selected via cross-validation and identical to ssGAN.

4.2.5.4 CasGAN (undersampled source, undersampled target)

CasGAN is a cascaded method that sequentially performs reconstruction and synthesis to cope with mutually accelerated source-target acquisitions. Here compressed-sensing reconstructions were first performed to recover source-target images from undersampled acquisitions. A fully-supervised synthesis model was then learned based on these reconstructions. CS reconstructions were implemented using SparseMRI [111] for single-coil data, and L_1 -SPIRiT [113] for multi-coil data (<https://people.eecs.berkeley.edu/mlustig/Software.html>). Hyperparameters were selected via cross-validation. In SparseMRI, the number of iterations was 4, weight for total variation regularization was 0.0001 and weight for wavelet-domain L_1 regularization was 0.0001. In SPIRiT, the kernel size was 5×5 , weight for wavelet-domain L_1 -regularization was 0.1, weight for Tikhonov regularization during kernel estimation was 0.001, number of iterations was 10 for PD-weighted images and 20 for T_2 -weighted images. The synthesis model had identical architecture, loss functions and hyperparameters to fsGAN.

4.2.6 Experiments

4.2.6.1 Reliability against deficiencies in training data

Robustness against deficiencies in the quality and amount of training data was examined on the IXI dataset. Multiple independent ssGAN models were trained for $T_1 \rightarrow T_2$ synthesis while varying the training dataset. Variations were introduced by altering the acceleration ratio of target contrast acquisitions across $R_{target} = [2 : 1 : 10]$, and by altering the number of training subjects across $n_T = [8 : 8 : 64]$. As gold-standard baselines, independent pix2pix and CycleGAN models were trained for the same variations in n_T (albeit with $R_{target} = 1$). Additional experiments were conducted on $T_2 \rightarrow T_1$ synthesis, where ssGAN models with $R_{target} = \{2, 3, 4\}$ were compared against pix2pix and CycleGAN. Fully-sampled source acquisitions $R_{source} = 1$ were assumed for all experiments.

4.2.6.2 Single-coil synthesis

Experiments were conducted on brain images from the IXI dataset to demonstrate synthesis performance on single-coil data. Demonstrations were performed on the $T_1 \rightarrow T_2$ and $T_2 \rightarrow T_1$ synthesis tasks with ssGAN, fsGAN and CasGAN methods. Independent ssGAN and CasGAN models were trained for different target accelerations: ssGAN- k and CasGAN- k trained with $R_{target} = k$, where $k \in \{2, 3, 4\}$. ssGAN and all competing methods were separately trained for $R_{source} = \{2, 3, 4\}$.

4.2.6.3 Multi-coil synthesis

Experiments were conducted on brain images from the in vivo dataset to demonstrate synthesis performance on multi-coil data. Demonstrations were performed on the $T_2 \rightarrow PD$ and $PD \rightarrow T_2$ synthesis tasks with ssGAN, fsGAN and CasGAN. Independent ssGAN and CasGAN models were trained for different target accelerations: ssGAN- k and CasGAN- k trained with $R_{target} = k$, where $k \in \{2, 3, 4\}$. All competing methods were separately trained for $R_{source} = \{2, 3, 4\}$. A radiological evaluation was conducted on $T_2 \rightarrow PD$ and $PD \rightarrow T_2$ synthesis tasks with $R_{source} = 2, 4$. Opinion scores of an expert radiologist with more than 25 years of experience were considered. The quality of the synthesized images was rated based on similarity to reference images from fully-sampled acquisitions, on a five-point scale (0: unacceptable, 1: poor, 2: limited, 3: moderate, 4: good, 5: perfect match). For each synthesis task, radiological evaluations were performed on 5 different cross-sections randomly taken from each subject.

4.2.6.4 Ablation studies

Experiments were conducted to individually examine the effects of the tensor image, k-space and adversarial loss functions on synthesis quality. Demonstrations

were performed on IXI for $T_1 \rightarrow T_2$ and $T_2 \rightarrow T_1$ synthesis tasks. Four independent ssGAN models were trained: ssGAN with all loss functions, ssGAN(w/o image) without the image loss, ssGAN(w/o k-space) without the k-space loss, and ssGAN(w/o adv) without the adversarial loss. Acceleration rates of $R_{source} = 4$ and $R_{target} = 4$ were assumed.

Synthesis performance was evaluated using peak signal-to-noise ratio (PSNR), structural similarity (SSIM) and mean-squared error (MSE) metrics (all MSE reports reflect measured $MSE \times 100$). Metrics were measured on coil-combined magnitude images derived from synthesized and reference target contrasts. The reference image was based on Fourier reconstructions of fully-sampled target acquisitions. In Tables, summary statistics of quantitative metrics were provided as mean \pm std across test subjects. Significance of PSNR, SSIM, MSE, and radiological opinions scores was assessed via Kruskal Wallis H-test ($p < 0.05$) to collectively compare ssGAN models vs pix2pix, ssGAN models vs CycleGAN, and ssGAN models vs fsGAN, and via Wilcoxon signed-rank test ($p < 0.05$) to individually compare ssGAN-2 vs CasGAN-2, ssGAN-3 vs CasGAN-3, and ssGAN-4 vs CasGAN-4.

4.3 Results

4.3.1 Robustness of semi-supervised learning against deficiencies in training data

We first performed comprehensive experiments to examine the reliability of ssGAN against deficiencies in training data. Models were learned for $T_1 \rightarrow T_2$ mapping in the IXI dataset, while acceleration ratio of target acquisitions ranged in $R_{target} = [2 : 1 : 10]$ and number of training subjects ranged in $n_T = [8 : 8 : 64]$. As baselines, gold-standard pix2pix and CycleGAN models were also trained for matching n_T but with $R_{target} = 1$. All models were input high-quality source images reconstructed from Nyquist-sampled acquisitions ($R_{source} = 1$). Synthesis

performance in terms of PSNR, SSIM, and MSE is displayed in Fig. 4.2 as a function of R_{target} and n_T . The reported measurements indicate that synthesis quality of ssGAN is on par with the gold-standard pix2pix model ($p > 0.05$), where ssGAN performance is within $[-0.51, 0.13]$ dB PSNR, $[-0.70, 0.02]$ % SSIM, and $[0.019, -0.004]$ MSE windows of pix2pix. Meanwhile, ssGAN outperforms CycleGAN with 1.18 dB higher PSNR, 1.16 % SSIM, and 0.081 lower MSE ($p < 0.05$). Importantly, ssGAN models trained with varying R_{target} yield highly similar performance, where ssGAN achieves near-optimal synthesis quality while undersampling target acquisitions up to 10-fold.

Measurements reported in Fig. 4.2 also demonstrate that increasing n_T improves performance of all competing methods. Comparing $n_T = 16$ against $n_T = 32$, average improvements in (PSNR, SSIM, MSE) are (0.68 dB, 0.57 %, -0.025) for ssGAN, (0.58 dB, 0.47 %, -0.020) for pix2pix, and (0.32 dB, 0.24 %, -0.012) for CycleGAN. Note that ssGAN improves scan efficiency by accelerating target acquisitions, so in principle training data from a larger group of subjects can be collected at high acceleration rates for training ssGAN, compared to fully-supervised or unpaired models. For instance, given a total, active scan time of 126 min, Nyquist-sampled k-space data for T_1 - and T_2 -weighted images can be collected in 16 subjects for pix2pix and CycleGAN. In the same duration, a protocol with undersampled target acquisitions ($R_{target} = 10$) can be performed in 32 subjects for ssGAN, resulting in performance benefits of (0.34 dB, 0.04 %, -0.012) over pix2pix and (0.74 dB, 0.48 %, -0.031) over CycleGAN. Therefore, ssGAN enables elevated diversity in the training set to improve accuracy and practicality of learning-based MRI synthesis.

We then extended the demonstrations of ssGAN against pix2pix and CycleGAN by comparisons on the $T_2 \rightarrow T_1$ synthesis task in the IXI dataset with a fixed number of training subjects $n_T = 32$ (used hereafter in all evaluations in IXI). Measurements of synthesis quality are reported in Table 4.1. The reported measurements reveal that ssGAN models maintain near-optimal synthesis quality on par with pix2pix, and on average they outperform CycleGAN with (0.41 dB, 0.34 %, -0.015) improvement in (PSNR, SSIM, MSE). Representative results displayed in Fig. 4.3 corroborate the quantitative findings by showing

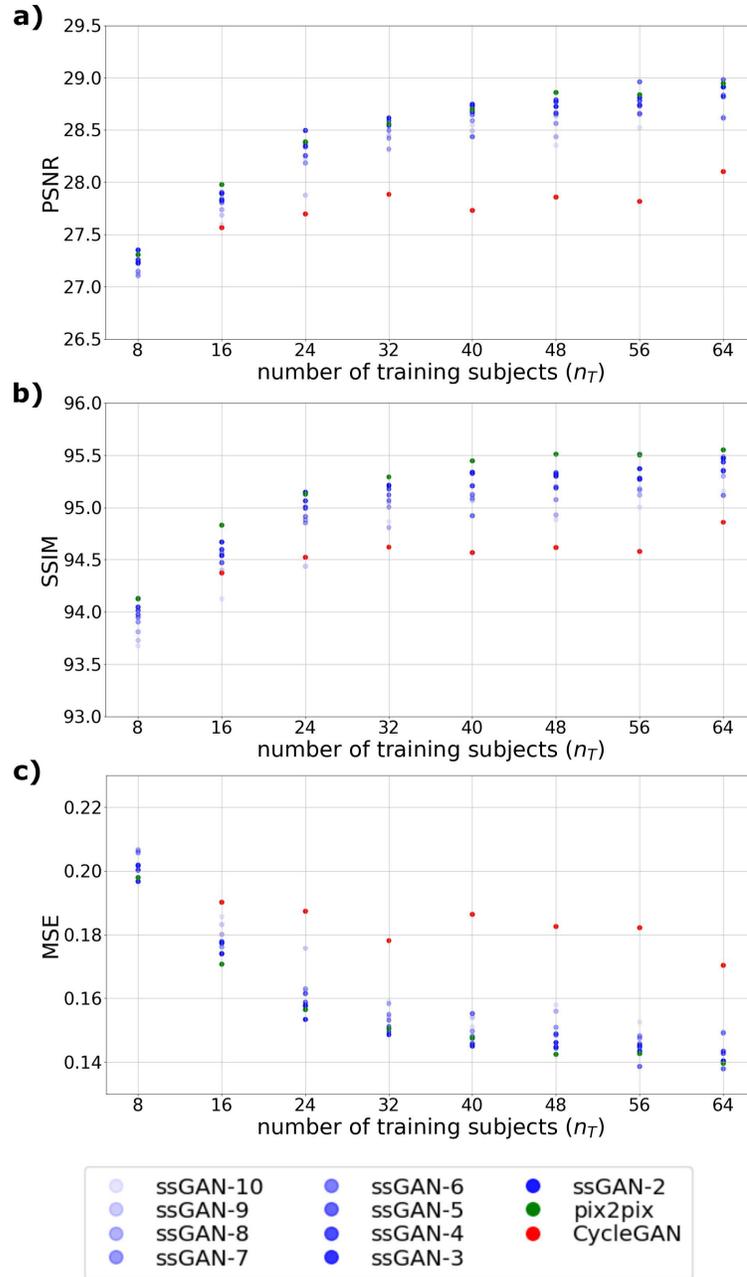


Figure 4.2: Reliability of ssGAN against training data deficiencies. Evaluations were performed for $n_T = [8 : 8 : 64]$. For each n_T , pix2pix and CycleGAN were trained with $R_{target} = 1$, whereas ssGAN was trained with $R_{target} \in [2 : 1 : 10]$, ssGAN- k with $k = R_{target}$. All models were trained with $R_{source} = 1$. Performance metrics for CycleGAN at $n_T = 8$ remain outside the display windows.

Table 4.1: Quality of Synthesis in the IXI Dataset for Nyquist-Sampled Source Acquisitions

	$T_1 \rightarrow T_2$			$T_2 \rightarrow T_1$		
	PSNR	SSIM	MSE	PSNR	SSIM	MSE
pix2pix	28.57	95.33	0.15	28.62	95.95	0.154
	± 1.39	± 1.35	± 0.048	± 1.45	± 1.38	± 0.057
CycleGAN	27.91	94.66	0.178	28.05	95.4	0.175
	± 1.52	± 1.51	± 0.062	± 1.41	± 1.52	± 0.062
ssGAN-2	28.63	95.25	0.148	28.52	95.79	0.158
	± 1.42	± 1.33	± 0.049	± 1.61	± 1.46	± 0.061
ssGAN-3	28.56	95.24	0.15	28.44	95.72	0.16
	± 1.39	± 1.32	± 0.048	± 1.55	± 1.41	± 0.059
ssGAN-4	28.6	95.21	0.149	28.42	95.7	0.162
	± 1.43	± 1.34	± 0.05	± 1.56	± 1.42	± 0.061

that ssGAN offers a similar level of accuracy in tissue depiction to pix2pix, while it synthesizes higher quality images compared to CycleGAN that suffers from elevated errors.

4.3.2 Single-coil image synthesis in mutually accelerated multi-contrast MRI

To examine the synthesis performance of ssGAN in mutually accelerated MRI, we conducted experiments on IXI where both source and target acquisitions were undersampled. Single-coil image synthesis was considered with $T_1 \rightarrow T_2$ and $T_2 \rightarrow T_1$ recovery tasks. ssGAN was compared against a gold-standard supervised model (fsGAN) that was trained on undersampled source acquisitions but Nyquist-sampled target acquisitions, and against a sequential model (CasGAN) that first reconstructed undersampled acquisitions, and then trained a synthesis model on the reconstructed source and target images. The target acceleration ratio varied in $R_{target} = \{2, 3, 4\}$ for ssGAN and CasGAN resulting in three separate models for each method: ssGAN- k and CasGAN- k with $k = R_{target}$. Meanwhile,

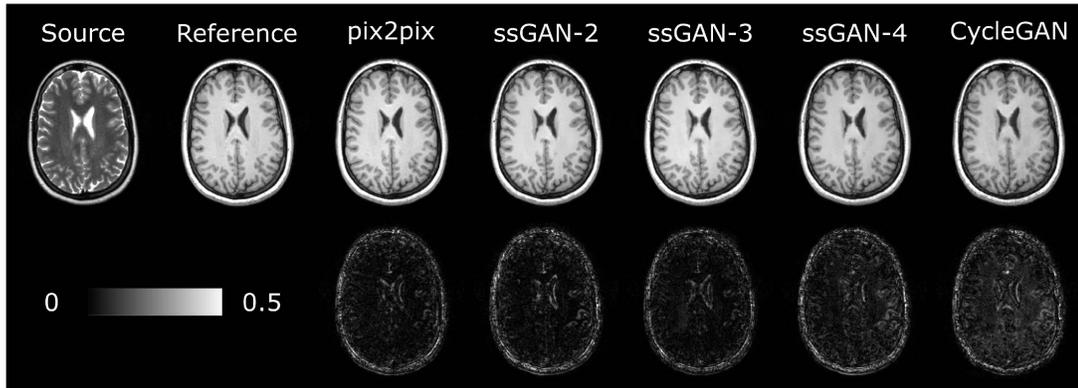


Figure 4.3: ssGAN was demonstrated on IXI for $T_2 \rightarrow T_1$ mapping against pix2pix and CycleGAN with ($R_{source} = 1$). Synthesized images from ssGAN-2, ssGAN-3, ssGAN-4, pix2pix, and CycleGAN are displayed together with the reference and source images in the first row. The corresponding error maps for the synthesized images are displayed in the second row.

the acceleration ratio for the source acquisitions varied in $R_{source} = \{2, 3, 4\}$ for all methods.

Quantitative metrics for synthesis performance are listed in Table 4.2 for varying R_{source} and R_{target} values. Overall, ssGAN models at moderate acceleration factors for the target acquisition yield near-optimal performance on par with the reference fsGAN model ($p > 0.05$), while mitigating the demands for Nyquist-sampled target acquisitions. Furthermore, ssGAN outperforms CasGAN by an average of 6.32 dB in PSNR, 11.26 % in SSIM, and -0.914 in MSE ($p < 0.05$). On average, incremental steps from $R_{target} = 1$ to $R_{target} = 4$ result in modest performance losses of 0.10 dB PSNR, 0.19 % SSIM and 0.005 MSE for ssGAN. In contrast, CasGAN suffers from elevated losses of 2.04 dB PSNR, 3.05 % SSIM and 0.805 MSE. This finding demonstrates that the selective loss function in ssGAN effectively copes with the reduction in quality of target acquisitions.

Representative synthesis results from the methods under comparison are shown in Fig. 4.4. Quality of synthetic images from ssGAN are virtually identical to those from the supervised fsGAN model, and they are superior to the CasGAN model that suffers from residual artifacts and noise that carry over from the initial reconstruction stage.

Table 4.2: Image Quality for Mutually Accelerated Single-Coil MRI Synthesis

(a) $R_{source} = 2$

	$T_1 \rightarrow T_2$			$T_2 \rightarrow T_1$		
	PSNR	SSIM	MSE	PSNR	SSIM	MSE
fsGAN	27.01	93.63	0.215	27.54	94.34	0.194
	± 1.41	± 1.53	± 0.072	± 1.45	± 1.51	± 0.067
ssGAN-2	26.9	93.36	0.219	27.46	94.09	0.196
	± 1.37	± 1.52	± 0.072	± 1.35	± 1.53	± 0.061
ssGAN-3	26.82	93.21	0.223	27.3	93.93	0.206
	± 1.38	± 1.53	± 0.074	± 1.47	± 1.6	± 0.072
ssGAN-4	26.78	93.13	0.224	27.29	93.86	0.204
	± 1.34	± 1.5	± 0.072	± 1.38	± 1.58	± 0.066
CasGAN-2	24.04	85.22	0.409	21.25	85.56	0.84
	± 0.81	± 2.12	± 0.079	± 1.05	± 2.31	± 0.176
CasGAN-3	21.62	81.46	0.714	18.07	81.45	1.672
	± 0.72	± 2.43	± 0.122	± 0.8	± 2.58	± 0.287
CasGAN-4	20.06	79.03	1.026	16.81	79.29	2.167
	± 0.6	± 2.35	± 0.141	± 0.69	± 2.68	± 0.345

(b) $R_{source} = 3$

	$T_1 \rightarrow T_2$			$T_2 \rightarrow T_1$		
	PSNR	SSIM	MSE	PSNR	SSIM	MSE
fsGAN	26.18	92.58	0.259	27.01	93.54	0.215
	± 1.4	± 1.63	± 0.089	± 1.27	± 1.54	± 0.065
ssGAN-2	26.14	92.22	0.261	27.05	93.38	0.214
	± 1.35	± 1.61	± 0.087	± 1.31	± 1.58	± 0.065
ssGAN-3	25.99	92.06	0.269	26.91	93.14	0.22
	± 1.32	± 1.57	± 0.087	± 1.3	± 1.59	± 0.066
ssGAN-4	25.84	91.69	0.279	26.83	93.06	0.225
	± 1.34	± 1.65	± 0.093	± 1.3	± 1.61	± 0.067
CasGAN-2	23.71	84.35	0.442	21.16	84.95	0.853
	± 0.84	± 2.23	± 0.089	± 1.05	± 2.3	± 0.19
CasGAN-3	21.47	80.82	0.743	17.97	81.03	1.705
	± 0.74	± 2.49	± 0.131	± 0.86	± 2.62	± 0.312
CasGAN-4	19.73	78.08	1.112	16.91	78.94	2.126
	± 0.67	± 2.48	± 0.172	± 0.68	± 2.72	± 0.338

(c) $R_{source} = 4$

	$T_1 \rightarrow T_2$			$T_2 \rightarrow T_1$		
	PSNR	SSIM	MSE	PSNR	SSIM	MSE
fsGAN	25.51	91.61	0.302	26.81	93.1	0.227
	± 1.39	± 1.71	± 0.103	± 1.29	± 1.64	± 0.068
ssGAN-2	25.46	91.27	0.305	26.63	92.75	0.236
	± 1.38	± 1.73	± 0.104	± 1.29	± 1.64	± 0.069
ssGAN-3	25.35	91.13	0.313	26.54	92.59	0.241
	± 1.37	± 1.75	± 0.109	± 1.32	± 1.68	± 0.072
ssGAN-4	25.26	90.79	0.321	26.5	92.32	0.243
	± 1.43	± 1.83	± 0.114	± 1.28	± 1.73	± 0.072
CasGAN-2	23.49	83.7	0.465	21.04	84.42	0.88
	± 0.9	± 2.31	± 0.104	± 1.19	± 2.39	± 0.213
CasGAN-3	21.3	80.07	0.773	18.01	80.66	1.707
	± 0.75	± 2.58	± 0.134	± 0.89	± 2.69	± 0.329
CasGAN-4	19.85	77.72	1.078	16.87	78.55	2.142
	± 0.66	± 2.52	± 0.169	± 0.68	± 2.73	± 0.347

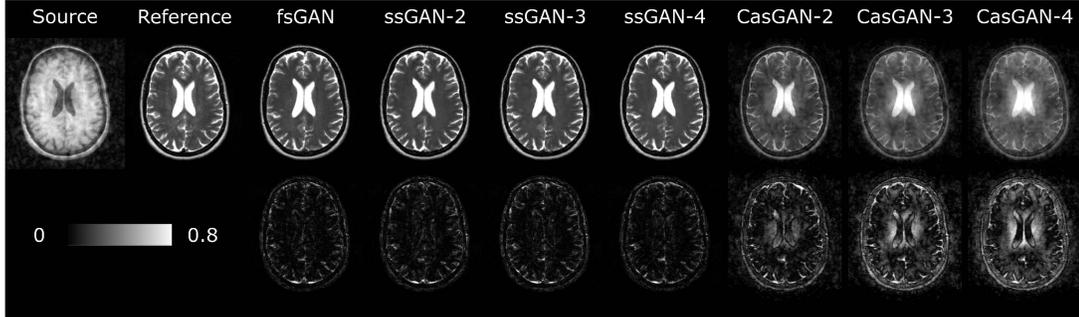


Figure 4.4: Synthesis quality of ssGAN, fsGAN and CasGAN was demonstrated on IXI for $T_1 \rightarrow T_2$ synthesis ($R_{source} = 2$). Synthesized images from the competing methods are displayed together with the source and reference (i.e. target) images in the first row, and the corresponding error maps for the synthesized images are displayed in the second row.

4.3.3 Multi-coil image synthesis in mutually accelerated multi-contrast MRI

Next, we conducted experiments on the in vivo brain dataset to demonstrate multi-coil MRI synthesis with the proposed ssGAN model. Multi-coil image synthesis was considered for $T_2 \rightarrow PD$ and $PD \rightarrow T_2$ recovery tasks. As in single-coil synthesis, ssGAN was compared against fsGAN and CasGAN. $R_{target} = \{2, 3, 4\}$ and $R_{source} = \{2, 3, 4\}$ were considered.

Quantitative measurements for synthesis quality are reported in Table 4.3 for various R_{source} and R_{target} values. Overall, ssGAN models at distinct acceleration factors for the target acquisitions yield near-optimal performance on par with the gold-standard fsGAN model ($p > 0.05$). On average across R_{target} , ssGAN outperforms CasGAN by 0.92 dB in PSNR, 3.18 % in SSIM, and -0.130 in MSE ($p < 0.05$, except for $PD \rightarrow T_2$ with $R_{source} = 4$). Incremental steps from $R_{target} = 1$ to $R_{target} = 4$ result in an average performance loss of 0.12 dB PSNR, 0.30 % SSIM, and 0.010 MSE for ssGAN, and 0.82 dB PSNR, 2.25 % SSIM, and 0.07 MSE for CasGAN. Similar to single-coil results, this finding demonstrates the utility of the selective loss function in ssGAN to cope with moderately undersampled target acquisitions.

Table 4.3: Image Quality for Mutually Accelerated Multi-Coil MRI Synthesis

(a) $R_{source} = 2$

	$T_2 \rightarrow PD$			$PD \rightarrow T_2$		
	PSNR	SSIM	MSE	PSNR	SSIM	MSE
fsGAN	25.48	87.8	0.295	25.35	87.78	0.306
	± 0.13	± 1.46	± 0.015	± 0.46	± 1.63	± 0.032
ssGAN-2	25.75	88.14	0.279	24.59	86.56	0.361
	± 0.01	± 1.83	± 0.005	± 0.21	± 0.52	± 0.016
ssGAN-3	25.16	87.42	0.315	24.59	86.25	0.359
	± 0.26	± 0.97	± 0.022	± 0.13	± 0.42	± 0.01
ssGAN-4	25.47	87.72	0.296	24.46	85.97	0.37
	± 0.15	± 1.92	± 0.006	± 0.2	± 0.3	± 0.017
CasGAN-2	25.01	86.96	0.347	24.44	85.02	0.466
	± 0.03	± 1.61	± 0.021	± 0.43	± 2.45	± 0.102
CasGAN-3	23.91	84.72	0.443	24.23	83.54	0.465
	± 0.15	± 1.76	± 0.042	± 0.81	± 3.74	± 0.137
CasGAN-4	22.95	82.22	0.562	23.55	81.53	0.487
	± 0.59	± 2.32	± 0.112	± 0.24	± 3.03	± 0.061

(b) $R_{source} = 3$

	$T_2 \rightarrow PD$			$PD \rightarrow T_2$		
	PSNR	SSIM	MSE	PSNR	SSIM	MSE
fsGAN	25.04	86.5	0.338	24.77	86.73	0.345
	± 0.42	± 0.56	± 0.048	± 0.07	± 0.79	± 0.003
ssGAN-2	25.14	87.08	0.315	24.65	86.37	0.356
	± 0.01	± 1.61	± 0.003	± 0.02	± 0.56	± 0.006
ssGAN-3	25.2	87.15	0.311	24.61	86.09	0.358
	± 0.05	± 1.37	± 0.005	± 0.06	± 0.67	± 0.003
ssGAN-4	25.01	86.78	0.325	24.16	85.25	0.398
	± 0.19	± 1.25	± 0.018	± 0.12	± 0.19	± 0.016
CasGAN-2	24.77	86.41	0.374	24.57	82.86	0.428
	± 0.32	± 1.81	± 0.051	± 1.44	± 6.67	± 0.169
CasGAN-3	23.44	83.9	0.496	24.2	82.11	0.444
	± 0.36	± 2.05	± 0.073	± 1.24	± 6.18	± 0.153
CasGAN-4	22.68	81.76	0.579	22.83	77.46	0.59
	± 0.04	± 1.8	± 0.037	± 1.0	± 5.75	± 0.164

(c) $R_{source} = 4$

	$T_2 \rightarrow PD$			$PD \rightarrow T_2$		
	PSNR	SSIM	MSE	PSNR	SSIM	MSE
fsGAN	24.9	86.51	0.337	24.51	86.16	0.365
	± 0.06	± 1.9	± 0.003	± 0.33	± 0.07	± 0.031
ssGAN-2	24.85	86.64	0.338	24.35	85.66	0.379
	± 0.0	± 1.47	± 0.002	± 0.35	± 0.2	± 0.034
ssGAN-3	24.6	86.12	0.359	24.3	85.5	0.383
	± 0.08	± 1.74	± 0.006	± 0.37	± 0.32	± 0.036
ssGAN-4	24.73	86.22	0.35	24.05	84.84	0.406
	± 0.12	± 1.83	± 0.008	± 0.38	± 0.39	± 0.039
CasGAN-2	23.72	84.78	0.479	25.12	85.91	0.353
	± 0.48	± 2.24	± 0.085	± 1.33	± 3.81	± 0.124
CasGAN-3	23.09	82.59	0.553	24.84	84.83	0.374
	± 0.41	± 2.37	± 0.099	± 1.29	± 3.87	± 0.127
CasGAN-4	22.31	80.32	0.663	23.51	81.68	0.491
	± 0.23	± 2.5	± 0.095	± 1.17	± 3.79	± 0.145

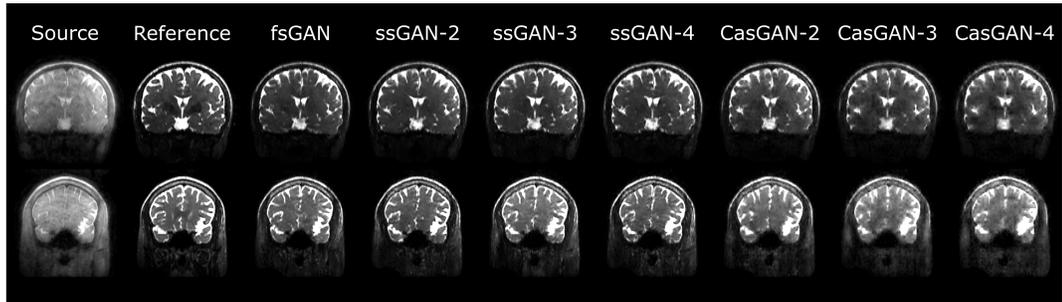


Figure 4.5: The proposed ssGAN model was demonstrated on the in vivo brain dataset for multi-coil complex PD \rightarrow T₂ synthesis task with source contrast acquisitions undersampled by $R_{target} = 2$. Synthesized coil-combined images from fsGAN, ssGAN, and CasGAN are displayed along with the reference image (i.e. target) in the first row, and the corresponding error images for the synthesized images are displayed in the second row (see colorbar).

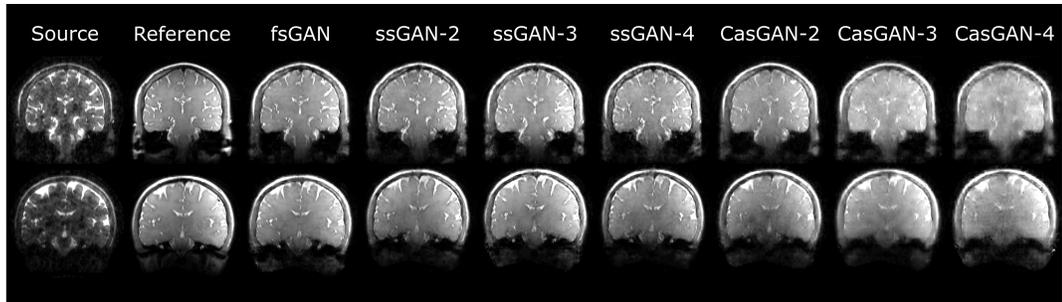


Figure 4.6: Synthesis quality of ssGAN, fsGAN and CasGAN was demonstrated on the in vivo dataset for T₂ \rightarrow PD synthesis ($R_{source} = 2$). Representative results from two different subjects are displayed, along with the source and reference images.

For further validation of the quantitative assessments, radiological evaluations were performed for ssGAN-2, fsGAN and CasGAN-2 on T₂ \rightarrow PD and PD \rightarrow T₂ synthesis tasks. Representative synthetic images are displayed in Figs. 4.5, 4.6, whereas results of radiological evaluation are shown in Fig. 4.7. ssGAN images are visually similar to fsGAN, whereas they manifest superior synthesis quality compared to CasGAN. In terms of opinion score, ssGAN maintains a high-level of synthesis quality on par with fsGAN ($p > 0.05$, except for $R_{source} = 2$), and on average a modest score difference of 0.325 is observed. In contrast, ssGAN yields superior performance to CasGAN with an average improvement of 1.075 in opinion score across tasks ($p < 0.05$).

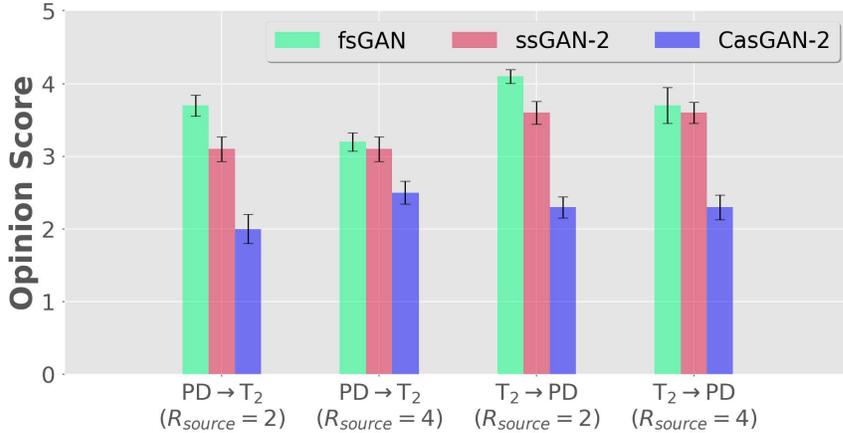


Figure 4.7: Radiological evaluations for fsGAN, ssGAN-2, and CasGAN-2 are shown. T₂ \rightarrow PD and PD \rightarrow T₂ synthesis tasks at $R_{source} = 2, 4$ were assessed on the in vivo dataset.

4.3.4 Ablation Studies

Ablation experiments were conducted to demonstrate the contribution of individual loss components in ssGAN. Independent ssGAN models were trained while the loss components were selectively ablated. Models were learned for T₁ \rightarrow T₂ and T₂ \rightarrow T₁ mappings in the IXI dataset. The effects of image and k-space losses were evaluated using PSNR, SSIM and MSE metrics, whereas the effect of adversarial loss was assessed using Frechlet Inception Distance (FID) scores and visual inspection as common in literature [116]. Quantitative metrics listed in Table 4.4 indicate that the selective image and k-space losses serve to improve synthesis quality in both T₁ and T₂ recovery tasks. Meanwhile, the selective adversarial loss component increases the realism of synthetic images with decreased FID scores.

4.4 Discussion

Here we introduced a novel semi-supervised deep generative model for image synthesis in multi-contrast MRI that is mutually accelerated across both contrast sets and k-space. Please note that ssGAN does not attempt at any reconstruction

Table 4.4: Effects of Tensor Losses on Synthesis Quality

	T1 to T2			T2 to T1		
	PSNR	SSIM	MSE	PSNR	SSIM	MSE
ssGAN	25.25	90.75	0.32	26.47	92.27	0.24
ssGAN (w/o image)	25.12	90.36	0.328	26.12	91.78	0.264
ssGAN (w/o k-space)	24.96	90.30	0.341	26.41	92.19	0.246
	T1 to T2			T2 to T1		
	FID			FID		
ssGAN	22.89			16.80		
ssGAN (w/o adv)	23.13			24.35		

[108, 117, 118, 119], but it only performs synthesis. As opposed to supervised models [19, 14, 25, 29, 8, 9, 28], ssGAN learns to synthesize high-quality target-contrast images in the absence of training sets composed of costly acquisitions of Nyquist-sampled source and target contrasts. ssGAN achieves synthesis quality on par with gold-standard supervised models for a broad range of acceleration ratios. This performance leap is mediated by selective loss functions in image, k-space, and adversarial domains. Unlike prior synthesis methods, ssGAN processes multi-coil complex MRI data and learns to synthesize directly from undersampled source acquisitions. Therefore, ssGAN holds great promise in advancing the practicality and utility of multi-contrast MRI synthesis.

Comprehensive experiments were conducted on single- and multi-coil neuroimaging datasets to demonstrate the proposed approach. Our experiments indicate that ssGAN achieves equivalent performance to gold-standard fully-supervised models, pix2pix [94] with fully-sampled source acquisitions, and fsGAN with undersampled source acquisitions. Importantly, ssGAN maintains near-optimal synthesis for acceleration factors up to 10 for target acquisitions given the same amount of training subjects, significantly reducing the data requirements for model training. Furthermore, ssGAN outperforms an alternative weakly-supervised synthesis model CasGAN based on a sequential hybrid

of compressed-sensing reconstructions followed by supervised synthesis. Compared to CasGAN, ssGAN enables end-to-end learning of target contrast recovery, alleviating potential propagation of errors across the pipeline and reducing computational complexity.

Here, we demonstrated ssGAN with uniform-density random undersampling in k-space. An alternative would be to collect low-resolution source-target images by Nyquist-sampling within a central k-space region to achieve similar acceleration. The problem would then be transformed into a superresolution task on coil-combined images [120, 121]. However, this superresolution task is challenging when both source and target images are low resolution, so external priors might be required to enable recovery of high-spatial-frequency information. Another alternative would be to perform variable-density sampling where central k-space is more densely sampled [111, 122]. Variable-density schemes trade-off high-frequency coverage in return for increased signal-to-noise ratio in acquired data. As such, they might improve aggregate performance metrics (e.g., PSNR, MSE) that are dominated by low-spatial-frequency errors, particularly for higher acceleration factors used in ssGAN and CasGAN. Note, however, that uniform-density sampling improves peripheral k-space coverage to expand high-spatial-frequency information, and in turn contributes to recovery of detailed tissue structure.

In the current study, we demonstrated the proposed method on one-to-one synthesis tasks with a single source and a single target contrast. In multi-contrast protocols, multiple source and/or target contrasts might be available, one might additionally be interested in many-to-one and many-to-many synthesis tasks [28, 22]. In such cases, a many-to-many variant of ssGAN can be constructed by concatenating the multitude of source and target contrasts as separate input and output channels, respectively [28]. The selective loss function along with the k-space masking and coil projection operators can then be defined separately for each target contrast.

The ssGAN implementation considered here leverages a multi-coil tensor loss between undersampled versions of the synthesized and reference target images. This image-domain loss implicitly assumes that the source and target-contrast

acquisitions are spatially registered. The datasets examined here were either aligned or a registration step was performed during preprocessing. If an end-to-end alternative is desired that can cope with misaligned source-target acquisitions, deep network-based registration models can be cascaded to the input of ssGAN for spatial registration [79, 123]. It remains important future work to explore the extent of improvements in synthesis performance with integrated registration and synthesis.

The semi-supervised learning framework that ssGAN leverages undersampled albeit paired acquisitions of source and target contrasts from the same set of subjects. Our results suggest that successful ssGAN models can be trained even with relatively modest size datasets. However, more complex models including order of magnitude higher number of parameters such as 3D architectures might require substantial datasets for reliable training. In such cases, a variant of ssGAN that permits training on a hybrid of paired and unpaired images or directly on unpaired images would be valuable. To do this, the cycle-consistent counterpart of the selective loss function in ssGAN can be devised [9, 22, 27, 105, 106].

In summary, here we proposed a semi-supervised learning framework based on generative adversarial networks that can recover high-quality target images without demanding Nyquist-sampled ground truths. While the superior data-efficiency of ssGAN was primarily demonstrated for within-modality contrast conversion in the brain, it can also be adopted to other anatomies, other recovery tasks including multi-parametric MRI synthesis, or cross-modality mappings between MRI and other imaging modalities [105, 103, 124, 11, 125, 126, 127, 128].

Chapter 5

Conclusion

Here, in this thesis, we first introduced a novel multi-stream GAN model to address the limitations of isolated one-to-one and many-to-one mappings that are common in multi-contrast MRI synthesis. The proposed model leverages a hybrid of multiple one-to-one streams and a many-to-one stream. The unique and shared features generated in the streams are task-adaptively combined with a fusion block, and then the fused feature maps are fed to a joint stream to recover target contrast images. Experiments performed on neuroimaging datasets clearly demonstrate that complementary use of the one-to-one and many-to-one streams yield superiority against the state-of-the-art baselines.

Secondly, we proposed a novel semi-supervised generative model for data-efficient MRI synthesis. The proposed model bypasses the need for large training datasets of high-quality source and target contrast images from fully-sampled k-space acquisitions. This performance leap is achieved via a group of selective tensor loss functions defined only on the acquired k-space coefficients of the under-sampled ground truths and via randomization of sampling masks across training subjects. Demonstrations conducted on single-coil and multi-coil datasets reveal that the proposed semi-supervised model maintains equivalent performance to the gold-standard baselines, and it achieves enhanced performance against cascaded models.

Bibliography

- [1] M. Yurt, S. U. H. Dar, A. Erdem, E. Erdem, K. K. Oğuz, and T. Çukur, “mustGAN: multi-stream generative adversarial networks for MR image synthesis,” *Medical Image Analysis*, vol. 70, p. 101944, 2021.
- [2] M. Yurt, S. U. H. Dar, A. Erdem, E. Erdem, and T. Çukur, “A multi-stream GAN approach for multi-contrast MRI synthesis,” in *28th annual meeting of International Society for Magnetic Resonance Imaging (ISMRM)*, (Virtual Conference), 8 2020.
- [3] M. Yurt, S. U. H. Dar, A. Erdem, E. Erdem, and T. Çukur, “Adaptive fusion via dual-branch GAN for multi-contrast MRI synthesis,” in *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, (Virtual Conference), 4 2020.
- [4] K. Krupa and M. Bekiesińska-Figatowska, “Artifacts in magnetic resonance imaging,” *Polish Journal of Radiology*, vol. 80, pp. 93–106, 02 2015.
- [5] B. Thukral, “Problems and preferences in pediatric imaging,” *Indian Journal of Radiology and Imaging*, vol. 25, pp. 359–364, 11 2015.
- [6] J. E. Iglesias, E. Konukoglu, D. Zikic, B. Glocker, K. Van Leemput, and B. Fischl, “Is synthesizing MRI contrast useful for inter-modality analysis?,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, pp. 631–638, Springer, 2013.
- [7] D. Abramian and A. Eklund, “Refacing: Reconstructing anonymized facial features using GANs,” in *IEEE 16th International Symposium on Biomedical Imaging (ISBI)*, pp. 1104–1108, April 2019.

- [8] C. Bowles, C. Qin, C. Ledig, R. Guerrero, R. Gunn, A. Hammers, E. Sakka, D. A. Dickie, M. V. Hernández, N. Royle, J. Wardlaw, H. Rhodius-Meester, B. Tijms, A. W. Lemstra, W. van der Flier, F. Barkhof, P. Scheltens, and D. Rueckert, “Pseudo-healthy image synthesis for white matter lesion segmentation,” in *Simulation and Synthesis in Medical Imaging*, (Cham), pp. 87–96, Springer International Publishing, 2016.
- [9] S. U. H. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Cukur, “Image synthesis in multi-contrast MRI with conditional generative adversarial networks,” *IEEE Transactions on Medical Imaging*, vol. 38, pp. 2375–2388, 2019.
- [10] B. E. Dewey, C. Zhao, A. Carass, J. Oh, P. A. Calabresi, P. C. M. van Zijl, and J. L. Prince, “Deep harmonization of inconsistent MR data for consistent volume segmentation,” in *Simulation and Synthesis in Medical Imaging* (A. Gooya, O. Goksel, I. Oguz, and N. Burgos, eds.), (Cham), pp. 20–30, Springer International Publishing, 2018.
- [11] B. E. Dewey, C. Zhao, J. C. Reinhold, A. Carass, K. C. Fitzgerald, E. S. Sotirchos, S. Saidha, J. Oh, D. L. Pham, P. A. Calabresi, P. C. van Zijl, and J. L. Prince, “DeepHarmony: A deep learning approach to contrast harmonization across scanner changes,” *Magnetic Resonance Imaging*, vol. 64, pp. 160 – 170, 2019.
- [12] R. Li, W. Zhang, H.-I. Suk, L. Wang, J. Li, D. Shen, and S. Ji, “Deep learning based imaging data completion for improved brain disease diagnosis,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 17, pp. 305–12, 09 2014.
- [13] F. Liu, “SUSAN: Segment unannotated image structure using adversarial network,” *Magnetic Resonance in Medicine*, vol. 81, no. 5, pp. 3330–3345, 2019.
- [14] V. Sevetlidis, M. V. Giuffrida, and S. A. Tsaftaris, “Whole image synthesis using a deep encoder-decoder network,” in *Simulation and Synthesis in Medical Imaging* (S. A. Tsaftaris, A. Gooya, A. F. Frangi, and J. L. Prince, eds.), (Cham), pp. 127–137, Springer International Publishing, 2016.

- [15] M. Sohail, M. N. Riaz, J. Wu, C. Long, and S. Li, “Unpaired multi-contrast MR image synthesis using generative adversarial networks,” in *Simulation and Synthesis in Medical Imaging* (N. Burgos, A. Gooya, and D. Svoboda, eds.), (Cham), pp. 22–31, Springer International Publishing, 2019.
- [16] H. Van Nguyen, K. Zhou, and R. Vemulapalli, “Cross-domain synthesis of medical images using efficient location-sensitive deep network,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, (Cham), pp. 677–684, Springer International Publishing, 2015.
- [17] P. Welander, S. Karlsson, and A. Eklund, “Generative adversarial networks for image-to-image translation on multi-contrast MR images - A comparison of CycleGAN and UNIT,” *arXiv preprint arXiv:1806.07777*, 06 2018.
- [18] Q. Yang, N. Li, Z. Zhao, X. Fan, E. I-Chao Chang, and Y. Xu, “MRI image-to-image translation for cross-modality image registration and segmentation,” *arXiv preprint arXiv:1801.06940v1*, 2018.
- [19] A. Chatsias, T. Joyce, M. Valerio Giuffrida, and S. Tsiftaris, “Multimodal MR synthesis via modality-invariant latent representation,” *IEEE Transactions on Medical Imaging*, vol. 37, pp. 803–814, 10 2017.
- [20] S. U. H. Dar, M. Yurt, M. E. Ildiz, M. Shahdloo, B. Tinaz, and T. Cukur, “Prior-guided image reconstruction for accelerated multi-contrast MRI via generative adversarial networks,” *IEEE Journal of Selected Topics in Signal Processing*, pp. 1–1, 2020.
- [21] A. Hagiwara, Y. Otsuka, M. Hori, Y. Tachibana, K. Yokoyama, S. Fujita, C. Andica, K. Kamagata, R. Irie, S. Koshino, T. Maekawa, L. Chougar, A. Wada, and S. Aoki, “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*, vol. 40, pp. 224–230, 01 2019.
- [22] D. Lee, J. Kim, W. Moon, and J. C. Ye, “Collagan: Collaborative GAN for missing image data imputation,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2482–2491, June 2019.

- [23] Y. Lei, Y. Fu, H. Mao, W. J. Curran, T. Liu, and X. Yang, “Multi-modality MRI arbitrary transformation using unified generative adversarial networks,” in *Medical Imaging 2020: Image Processing* (I. Išgum and B. A. Landman, eds.), vol. 11313, pp. 1 – 6, International Society for Optics and Photonics, SPIE, 2020.
- [24] H. Li, J. C. Paetzold, A. Sekuboyina, F. Kofler, J. Zhang, J. S. Kirschke, B. Wiestler, and B. Menze, “Diamondgan: Unified multi-modal generative adversarial networks for MRI sequences synthesis,” in *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, (Cham), pp. 795–803, Springer International Publishing, 2019.
- [25] T. Joyce, A. Chatsias, and S. A. Tsaftaris, “Robust multi-modal MR image synthesis,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 347–355, 2017.
- [26] R. Mehta and T. Arbel, “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*, (Cham), pp. 119–129, Springer International Publishing, 2018.
- [27] S. Olut, Y. H. Sahin, U. Demir, and G. Unal, “Generative adversarial training for MRA image synthesis using multi-contrast MRI,” in *PRedictive Intelligence in MEDicine (PRIME)*, (Cham), pp. 147–154, Springer International Publishing, 2018.
- [28] A. Sharma and G. Hamarneh, “Missing MRI pulse sequence synthesis using multi-modal generative adversarial network,” *IEEE Transactions on Medical Imaging*, vol. 39, pp. 1170–1183, 2019.
- [29] W. Wei, E. Poirion, B. Boudini, S. Durrleman, O. Colliot, B. Stankoff, and N. Ayache, “FLAIR MR image synthesis by using 3D fully convolutional networks for multiple sclerosis,” in *ISMRM-ESMRMB - Joint Annual Meeting*, pp. 1–6, 06 2018.
- [30] Y. Huang, L. Beltrachini, L. Shao, and A. F. Frangi, “Geometry regularized joint dictionary learning for cross-modality image synthesis in magnetic

- resonance imaging,” in *Simulation and Synthesis in Medical Imaging* (S. A. Tsiftaris, A. Gooya, A. F. Frangi, and J. L. Prince, eds.), (Cham), pp. 118–126, Springer International Publishing, 2016.
- [31] Y. Huang, L. Shao, and A. F. Frangi, “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, 2017.
- [32] Y. Huang, L. Shao, and A. F. Frangi, “Cross-modality image synthesis via weakly-coupled and geometry co-regularized joint dictionary learning,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 3, pp. 815 – 827, 2018.
- [33] A. Jog, A. Carass, S. Roy, D. L. Pham, and J. Prince, “MR image synthesis by contrast learning on neighborhood ensembles,” *Medical Image Analysis*, vol. 24, pp. 63–76, 05 2015.
- [34] S. Roy, A. Carass, and J. Prince, “A compressed sensing approach for MR tissue contrast synthesis,” in *Information Processing in Medical Imaging*, pp. 371–383, 2011.
- [35] S. Roy, A. Carass, and J. L. Prince, “Magnetic resonance image example-based contrast synthesis,” *IEEE Transactions on Medical Imaging*, vol. 32, no. 12, pp. 2348–2363, 2013.
- [36] S. Roy, A. Jog, A. Carass, and J. L. Prince, “Atlas based intensity transformation of brain MR images,” in *Multimodal Brain Image Analysis (MBIA)*, pp. 51–62, 2013.
- [37] S. Roy, Y.-Y. Chou, A. Jog, J. Butman, and D. L. Pham, “Patch based synthesis of whole head MR images: application to EPI distortion correction,” in *Simulation and Synthesis in Medical Imaging* (S. A. Tsiftaris, A. Gooya, A. F. Frangi, and J. L. Prince, eds.), vol. 9968, pp. 146–156, 10 2016.
- [38] B. Yu, L. Zhou, L. Wang, J. Fripp, and P. Bourgeat, “3D cGAN based cross-modality MR image synthesis for brain tumor segmentation,” in *IEEE 15th International Symposium on Biomedical Imaging*, pp. 626–630, IEEE, apr 2018.

- [39] B. Yu, L. Zhou, L. Wang, Y. Shi, J. Fripp, and P. Bourgeat, “Ea-GANs: edge-aware generative adversarial networks for cross-modality MR image synthesis,” *IEEE Transactions on Medical Imaging*, vol. 38, pp. 1750–1762, 2019.
- [40] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *European Conference on Computer Vision (ECCV)*, pp. 694–711, 2016.
- [41] A. Jog, A. Carass, D. L Pham, and J. Prince, “Random forest FLAIR reconstruction from T1, T2, and PD-weighted MRI,” in *IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, vol. 2014, pp. 1079–1082, 05 2014.
- [42] A. Jog, A. Carass, D. L Pham, and J. Prince, “Tree-encoded conditional random fields for image synthesis,” *Information Processing in Medical Imaging*, vol. 24, pp. 733–45, 07 2015.
- [43] A. Jog, A. Carass, S. Roy, D. L Pham, and J. Prince, “Random forest regression for magnetic resonance image synthesis,” *Medical Image Analysis*, vol. 35, pp. 475–488, 08 2016.
- [44] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *Advances in Neural Information Processing Systems (NIPS)*, vol. 27, pp. 2672–2680, 06 2014.
- [45] C. Bermudez, A. J. Plassard, L. T. Davis, A. T. Newton, S. M. Resnick, and B. A. Landman, “Learning implicit brain MRI manifolds with deep learning,” in *Medical Imaging 2018: Image Processing*, vol. 10574, p. 105741L, International Society for Optics and Photonics, 2018.
- [46] C. Bowles, L. Chen, R. Guerrero, P. Bentley, R. Gunn, A. Hammers, D. A. Dickie, M. V. Hernández, J. Wardlaw, and D. Rueckert, “GAN augmentation: Augmenting training data using generative adversarial networks,” *arXiv preprint arXiv:1810.10863*, 2018.

- [47] F. Calimeri, A. Marzullo, C. Stamile, and G. Terracina, “Biomedical data augmentation using generative adversarial neural networks,” in *International Conference on Artificial Neural Networks*, pp. 626–634, Springer, 2017.
- [48] A. Beers, J. Brown, K. Chang, J. P. Campbell, S. Ostmo, M. F. Chiang, and J. Kalpathy-Cramer, “High-resolution medical image synthesis using progressively grown generative adversarial networks,” *arXiv preprint arXiv:1805.03144*, 2018.
- [49] Z. Han, B. Wei, A. Mercado, S. Leung, and S. Li, “Spine-GAN: Semantic segmentation of multiple spinal structures,” *Medical Image Analysis*, vol. 50, pp. 23–35, 2018.
- [50] H. Zhao, H. Li, S. Maurer-Stroh, and L. Cheng, “Synthesizing retinal and neuronal images with generative adversarial nets,” *Medical Image Analysis*, vol. 49, pp. 14–26, 2018.
- [51] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, “Least squares generative adversarial networks,” in *IEEE International Conference on Computer Vision (ICCV)*, pp. 2813–2821, Oct 2017.
- [52] J. Cai, Z. Zhang, L. Cui, Y. Zheng, and L. Yang, “Towards cross-modal organ translation and segmentation: A cycle-and shape-consistent generative adversarial network,” *Medical Image Analysis*, vol. 52, pp. 174–184, 2019.
- [53] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, “StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
- [54] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, “StarGAN v2: Diverse image synthesis for multiple domains,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020.
- [55] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, “A survey on

- deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [56] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” *arXiv preprint arXiv:1411.1784*, 11 2014.
- [57] M. Ran, J. Hu, Y. Chen, H. Chen, H. Sun, J. Zhou, and Y. Zhang, “Denoising of 3D magnetic resonance images using a residual encoder–decoder Wasserstein generative adversarial network,” *Medical Image Analysis*, vol. 55, pp. 165–180, 2019.
- [58] D. Ravì, A. B. Szczotka, S. P. Pereira, and T. Vercauteren, “Adversarial training with cycle consistency for unsupervised super-resolution in endomicroscopy,” *Medical Image Analysis*, vol. 53, pp. 123–131, 2019.
- [59] M. Rubin, O. Stein, N. A. Turko, Y. Nygate, D. Roitshtain, L. Karako, I. Barnea, R. Giryes, and N. T. Shaked, “TOP-GAN: Stain-free cancer cell classification using deep learning with a small training set,” *Medical Image Analysis*, vol. 57, pp. 176–185, 2019.
- [60] J. Wang, J. H. Noble, and B. M. Dawant, “Metal artifact reduction for the segmentation of the intra cochlear anatomy in CT images of the ear with 3D-conditional GANs,” *Medical Image Analysis*, vol. 58, p. 101553, 2019.
- [61] C. Xu, J. Howey, P. Ohorodnyk, M. Roth, H. Zhang, and S. Li, “Segmentation and quantification of infarction without contrast agents via spatiotemporal generative adversarial learning,” *Medical Image Analysis*, vol. 59, p. 101568, 2020.
- [62] X. Yi, E. Walia, and P. Babyn, “Generative adversarial network in medical imaging: A review,” *Medical Image Analysis*, p. 101552, 2019.
- [63] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967–5976, July 2017.

- [64] O. Maier, B. H. Menze, J. von der Gablentz, L. Häni, M. P. Heinrich, M. Liebrand, S. Winzeck, A. Basit, P. Bentley, L. Chen, *et al.*, “ISLES 2015- A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI,” *Medical Image Analysis*, vol. 35, pp. 250–269, 2017.
- [65] M. Jenkinson and S. Smith, “A global optimisation method for robust affine registration of brain images,” *Medical Image Analysis*, vol. 5, pp. 143–156, 2001.
- [66] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 2015.
- [67] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, apr 2004.
- [68] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, “Medical image synthesis with context-aware generative adversarial networks,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 417–425, Springer, 2017.
- [69] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, “Medical image synthesis with deep convolutional adversarial networks,” *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 12, pp. 2720–2730, 2018.
- [70] H. Emami, M. Dong, S. P. Nejad-Davarani, and C. K. Glide-Hurst, “Generating synthetic CTs from magnetic resonance images using generative adversarial networks,” *Medical Physics*, vol. 45, no. 8, pp. 3627–3636, 2018.
- [71] J. Jiang, Y.-C. Hu, N. Tyagi, P. Zhang, A. Rimner, G. S. Mageras, J. O. Deasy, and H. Veeraraghavan, “Tumor-aware, adversarial domain adaptation from CT to MRI for lung cancer segmentation,” in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2019*, pp. 777–785, Springer, 2018.

- [72] C.-B. Jin, H. Kim, M. Liu, W. Jung, S. Joo, E. Park, Y. S. Ahn, I. H. Han, J. I. Lee, and X. Cui, “Deep CT to MR synthesis using paired and unpaired data,” *Sensors*, vol. 19, no. 10, p. 2361, 2019.
- [73] A. Ben-Cohen, E. Klang, S. P. Raskin, S. Soffer, S. Ben-Haim, E. Konen, M. M. Amitai, and H. Greenspan, “Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection,” *Engineering Applications of Artificial Intelligence*, vol. 78, pp. 186–194, 2019.
- [74] L. Bi, J. Kim, A. Kumar, D. Feng, and M. Fulham, “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*, pp. 43–51, Springer, 2017.
- [75] K. Armanious, C. Jiang, M. Fischer, T. Küstner, T. Hepp, K. Nikolaou, S. Gatidis, and B. Yang, “MedGAN: Medical image translation using GANs,” *Computerized Medical Imaging and Graphics*, vol. 79, p. 101684, 2020.
- [76] Y. Pan, M. Liu, C. Lian, T. Zhou, Y. Xia, and D. Shen, “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)*, pp. 455–463, Springer, 2018.
- [77] W. Wei, E. Poirion, B. Boudini, S. Durrleman, N. Ayache, B. Stankoff, and O. Colliot, “Learning myelin content in multiple sclerosis from multimodal MRI through adversarial training,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 514–522, Springer, 2018.
- [78] H. Choi and D. S. Lee, “Generation of structural MR images from amyloid PET: Application to MR-less quantification,” *Journal of Nuclear Medicine*, vol. 59, no. 7, pp. 1111–1117, 2018.

- [79] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, “Quicksilver: Fast predictive image registration - a deep learning approach,” *NeuroImage*, vol. 158, pp. 378–396, 03 2017.
- [80] M. Shahdloo, E. Çelik, and T. Çukur, “Biased competition in semantic representation during natural visual search,” *Neuroimage*, vol. 216, p. 116383, 2020.
- [81] E. Bıyık, K. Keskin, S. UH Dar, A. Koç, and T. Çukur, “Factorized sensitivity estimation for artifact suppression in phase-cycled bssfp mri,” *NMR in Biomedicine*, vol. 33, no. 4, p. e4228, 2020.
- [82] E. Kopanoglu, A. Güngör, T. Kilic, E. U. Saritas, K. K. Oguz, T. Çukur, and H. E. Güven, “Simultaneous use of individual and joint regularization terms in compressive sensing: Joint reconstruction of multi-channel multi-contrast mri acquisitions,” *NMR in Biomedicine*, vol. 33, no. 4, p. e4247, 2020.
- [83] I. Kiremitçi, Ö. Yilmaz, E. Çelik, M. Shahdloo, A. G. Huth, and T. Çukur, “Attentional modulation of hierarchical speech representations in a multi-talker environment,” *bioRxiv*, 2020.
- [84] M. Shahdloo, E. Ilicak, M. Tofghi, E. U. Saritas, A. E. Çetin, and T. Çukur, “Projection onto epigraph sets for rapid self-tuning compressed sensing mri,” *IEEE transactions on medical imaging*, vol. 38, no. 7, pp. 1677–1689, 2018.
- [85] M. Yurt, S. U. H. Dar, M. Özbey, B. Tınaz, K. K. Oğuz, and T. Çukur, “Semi-supervised learning of mutually accelerated MRI synthesis without fully-sampled ground truths,” *arXiv preprint arXiv:2011.14347*, 2021.
- [86] M. Yurt, B. Tınaz, M. Özbey, S. U. H. Dar, and T. Çukur, “Semi-supervised learning of multi-contrast MR image synthesis without fully-sampled ground-truth acquisitions,” in *Medical Imaging Meets NeurIPS*, (Virtual Conference), 12 2020.

- [87] M. Yurt, S. U. H. Dar, B. Tinaz, M. Özbey, Y. Korkmaz, and T. Çukur, “A semi-supervised learning framework for jointly accelerated multi-contrast MRI synthesis without fully-sampled ground-truths,” in *29th annual meeting of International Society for Magnetic Resonance Imaging (ISMRM)*, (Virtual Conference), 5 2021.
- [88] B. Moraal, S. D. Roosendaal, P. J. W. Pouwels, H. Vrenken, R. A. van Schijndel, D. S. Meier, C. R. G. Guttman, J. J. G. Geurts, and F. Barkhof, “Multi-contrast, isotropic, single-slab 3D MR imaging in multiple sclerosis,” *The Neuroradiology Journal*, vol. 22, pp. 33–42, 2009.
- [89] S. Bakas, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, “Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features,” *Scientific Data*, vol. 4, 2017.
- [90] S. Ilbey, C. B. Top, T. Çukur, E. U. Saritas, and H. E. Güven, “Image reconstruction for magnetic particle imaging using an augmented lagrangian method,” in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 64–67, IEEE, 2017.
- [91] D. Lee, W.-J. Moon, and J. C. Ye, “Assessing the importance of magnetic resonance contrasts using collaborative generative adversarial networks,” *Nature Machine Intelligence*, vol. 2, p. 34–42, 2020.
- [92] Y. Huo, Z. Xu, S. Bao, A. Assad, R. G. Abramson, and B. A. Landman, “Adversarial synthesis learning enables segmentation without target modality ground truth,” in *2018 IEEE 15th International Symposium on Biomedical Imaging*, pp. 1217–1220, 2018.
- [93] L. K. Senel, T. Kilic, A. Gungor, E. Kopanoglu, H. E. Guven, E. U. Saritas, A. Koc, and T. Cukur, “Statistically segregated k-space sampling for accelerating multiple-acquisition mri,” *IEEE transactions on medical imaging*, vol. 38, no. 7, pp. 1701–1714, 2019.

- [94] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *IEEE International Conference on Computer Vision*, pp. 2223–2232, 2017.
- [95] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, “Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [96] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, “A deep cascade of convolutional neural networks for dynamic MR image reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 491–503, 2018.
- [97] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser, “CNN-based projected gradient descent for consistent CT image reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1440–1453, 2018.
- [98] O. Dalmaz, M. Yurt, and T. Çukur, “Resvit: Residual vision transformers for multi-modal medical image synthesis,” *arXiv preprint arXiv:2106.16031*, 2021.
- [99] Y. Korkmaz, S. U. Dar, M. Yurt, M. Özbey, and T. Çukur, “Unsupervised mri reconstruction via zero-shot learned adversarial transformers,” *arXiv preprint arXiv:2105.08059*, 2021.
- [100] T. Zhou, H. Fu, G. Chen, J. Shen, and L. Shao, “Hi-Net: hybrid-fusion network for multi-modal MR image synthesis,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 9, pp. 2772–2781, 2020.
- [101] G. Wang, E. Gong, S. Banerjee, D. Martin, E. Tong, J. Choi, H. Chen, M. Wintermark, J. M. Pauly, and G. Zaharchuk, “Synthesize high-quality multi-contrast magnetic resonance imaging from multi-echo acquisition using multi-task deep generative model,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 10, pp. 3089–3099, 2020.

- [102] M. Özbey, M. Yurt, S. U. H. Dar, and T. Çukur, “Three dimensional mr image synthesis with progressive generative adversarial networks,” *arXiv preprint arXiv:2101.05218*, 2020.
- [103] J. M. Wolterink, A. M. Dinkla, M. H. F. Savenije, P. R. Seevinck, C. A. T. van den Berg, and I. Išgum, “Deep MR to CT synthesis using unpaired data,” in *Simulation and Synthesis in Medical Imaging*, (Cham), pp. 14–23, 2017.
- [104] Y. Ge, D. Wei, Z. Xue, Q. Wang, X. Zhou, Y. Zhan, and S. Liao, “Unpaired MR to CT synthesis with explicit structural constrained adversarial learning,” in *IEEE 16th International Symposium on Biomedical Imaging*, pp. 1096–1099, 2019.
- [105] C.-B. Jin, H. Kim, M. Liu, W. Jung, S. Joo, E. Park, Y. S. Ahn, I. H. Han, J. I. Lee, and X. Cui, “Deep CT to MR synthesis using paired and unpaired data,” *Sensors*, vol. 19, no. 10, 2019.
- [106] Z. Wang, Y. Lin, K.-T. T. Cheng, and X. Yang, “Semi-supervised mp-MRI data synthesis with StitchLayer and auxiliary distance maximization,” *Medical Image Analysis*, vol. 59, p. 101565, 2020.
- [107] E. Ilicak, S. Cetin, E. Bulut, K. K. Oguz, E. U. Saritas, G. Unal, and T. Çukur, “Targeted vessel reconstruction in non-contrast-enhanced steady-state free precession angiography,” *NMR in Biomedicine*, vol. 29, no. 5, pp. 532–544, 2016.
- [108] T. Çukur, “Accelerated phase-cycled ssfp imaging with compressed sensing,” *IEEE transactions on medical imaging*, vol. 34, no. 1, pp. 107–115, 2014.
- [109] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya, “Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data,” *Magnetic Resonance in Medicine*, vol. 84, no. 6, pp. 3172–3191, 2020.

- [110] E. K. Cole, J. M. Pauly, S. S. Vasanawala, and F. Ong, “Unsupervised MRI reconstruction with generative adversarial networks,” *arXiv:2008.13065*, 2020.
- [111] M. Lustig, D. Donoho, and J. M. Pauly, “Sparse MRI: The application of compressed sensing for rapid MR imaging,” *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [112] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, “Compressed sensing MRI,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 72–82, 2008.
- [113] M. Lustig and J. M. Pauly, “SPIRiT: iterative self-consistent parallel imaging reconstruction from arbitrary k-space,” *Magnetic Resonance in Medicine*, vol. 64, no. 2, pp. 457–471, 2010.
- [114] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig, “ESPIRiT—an eigenvalue approach to auto-calibrating parallel MRI: Where SENSE meets GRAPPA,” *Magnetic Resonance in Medicine*, vol. 71, no. 3, pp. 990–1001, 2014.
- [115] T. Zhang, J. M. Pauly, S. S. Vasanawala, and M. Lustig, “Coil compression for accelerated imaging with Cartesian sampling,” *Magnetic Resonance in Medicine*, vol. 69, no. 2, pp. 571–582, 2013.
- [116] T. DeVries, A. Romero, L. Pineda, G. W. Taylor, and M. Drozdal, “On the evaluation of conditional GANs,” *arXiv:1907.08175*, 2019.
- [117] B. Quist, B. A. Hargreaves, T. Cukur, G. R. Morrell, G. E. Gold, and N. K. Bangerter, “Simultaneous fat suppression and band reduction with large-angle multiple-acquisition balanced steady-state free precession,” *Magnetic resonance in medicine*, vol. 67, no. 4, pp. 1004–1012, 2012.
- [118] T. Çukur, J. M. Santos, J. M. Pauly, and D. G. Nishimura, “Variable-density parallel imaging with partially localized coil sensitivities,” *IEEE transactions on medical imaging*, vol. 29, no. 5, pp. 1173–1181, 2010.

- [119] T. Cukur, M. Lustig, and D. G. Nishimura, “Improving non-contrast-enhanced steady-state free precession angiography with compressed sensing,” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 61, no. 5, pp. 1122–1131, 2009.
- [120] C. Zhao, B. E. Dewey, D. L. Pham, P. A. Calabresi, D. S. Reich, and J. L. Prince, “Smore: A self-supervised anti-aliasing and super-resolution algorithm for mri using deep learning,” *IEEE Transactions on Medical Imaging*, vol. 40, no. 3, pp. 805–817, 2021.
- [121] Q. Lyu, H. Shan, C. Steber, C. Helis, C. Whitlow, M. Chan, and G. Wang, “Multi-contrast super-resolution MRI through a progressive network,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 9, pp. 2738–2749, 2020.
- [122] B. Gözcü, R. K. Mahabadi, Y.-H. Li, E. Ilıcak, T. Cukur, J. Scarlett, and V. Cevher, “Learning-based compressive mri,” *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1394–1406, 2018.
- [123] M. Aksoy, C. Forman, M. Straka, T. Çukur, J. Hornegger, and R. Bammer, “Hybrid prospective and retrospective head motion correction to mitigate cross-calibration errors,” *Magnetic resonance in medicine*, vol. 67, no. 5, pp. 1237–1251, 2012.
- [124] A. Ozaslan, A. Alacaoglu, O. Demirel, T. Çukur, and E. Saritas, “Fully automated gridding reconstruction for non-cartesian x-space magnetic particle imaging,” *Physics in Medicine & Biology*, vol. 64, no. 16, p. 165018, 2019.
- [125] S. U. H. Dar, M. Yurt, and T. Çukur, “A few-shot learning approach for accelerated mri via fusion of data-driven and subject-driven priors,” *arXiv preprint arXiv:2103.07790*, 2021.
- [126] K. Keskin, U. Yılmaz, and T. Çukur, “Constrained ellipse fitting for efficient parameter mapping with phase-cycled bssfp mri,” *arXiv preprint arXiv:2106.03239*, 2021.

- [127] Ö. Yılmaz, E. Çelik, and T. Çukur, “Informed feature regularization in voxelwise modeling for naturalistic fmri experiments,” *European Journal of Neuroscience*, vol. 52, no. 5, pp. 3394–3410, 2020.
- [128] E. Biyik, E. Ilicak, and T. Cukur, “Reconstruction by calibration over tensors for multi-coil multi-acquisition balanced ssfp imaging,” *Magnetic resonance in medicine*, vol. 79, no. 5, pp. 2542–2554, 2018.