Washington University in St. Louis
Washington University Open Scholarship

McKelvey School of Engineering Theses & Dissertations | McKelvey School of Engineering

5-2-2022

Optimizations in the MRI-Guided Adaptive Radiation Therapy Workflow: Towards MR-Only RT

Sven Olberg
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/eng_etds

Recommended Citation

Olberg, Sven, "Optimizations in the MRI-Guided Adaptive Radiation Therapy Workflow: Towards MR-Only RT" (2022). *McKelvey School of Engineering Theses & Dissertations*. 1088.
https://openscholarship.wustl.edu/eng_etds/1088

This Dissertation is brought to you for free and open access by the McKelvey School of Engineering at Washington University Open Scholarship. It has been accepted for inclusion in McKelvey School of Engineering Theses & Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

McKelvey School of Engineering
Department of Biomedical Engineering

Dissertation Examination Committee:
Hong Chen, Chair
Geoff Hugo
Taeho Kim
Justin Park
Tianyu Zhao

Optimizations in the MRI-Guided Adaptive Radiation Therapy Workflow: Towards MR-Only RT

by

Sven Olberg

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2022
St. Louis, Missouri

© 2022, Sven Olberg

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am incredibly thankful for the training and professional opportunities that have been afforded to me by my mentor, Dr. Justin Park. My achievements during my time at Washington University have been made under your guidance and with your help. I feel fortunate to have met the other students of my cohort who are too numerous to name here, the frequent collaborators with whom I was able to make worthwhile research contributions, and the lab mates who taught me about many things inside and outside of our work. This dissertation is the product of the person that I have become through meeting each of you. Finally, I would not be where I am today if it were not for the support of my mother, late father, brother, and fiancée—thank you for the growth that you have fostered.

The text of Chapter 2, in part or in full, has been reproduced with permission under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/) from the following publication:

Olberg S, Green O, Cai B, et al. Optimization of treatment planning workflow and tumor coverage during daily adaptive magnetic resonance image guided radiation therapy (MR-IGRT) of pancreatic cancer. *Radiat Oncol.* 2018;13(51):1–8. doi:10.1186/s13014-018-1000-7.

The dissertation author was the primary researcher and author of the work that forms the basis of this chapter.

The text of Chapter 3, in part or in full, has been reproduced with permission from the

following publication:

Chun J, Zhang H, Gach HM, et al. MRI super-resolution reconstruction for MRI-guided adaptive radiotherapy using cascaded deep learning: In the presence of limited training data and unknown translation model. *Med Phys*. 2019;46(9):4148–4164. doi: 10.1002/mp.13717. © 2019 American Association of Physicists in Medicine, John Wiley & Sons, Inc. All rights reserved.

The dissertation author made significant research contributions in model construction and substantial authorship contributions to the work that forms the basis of this chapter.

The text of Chapter 4, in part or in full, has been reproduced with permission from the following publication:

Olberg S, Zhang H, Kennedy WR, et al. Synthetic CT reconstruction using a deep spatial pyramid convolutional framework for MR-only breast radiotherapy. *Med Phys*. 2019;46(9): 4135–4147. doi: 10.1002/mp.13716. © 2019 American Association of Physicists in Medicine, John Wiley & Sons, Inc. All rights reserved.

The dissertation author was the primary researcher and author of the work that forms the basis of this chapter.

The text of Chapter 5, in part or in full, has been reproduced with permission from the following publication:

Olberg S, Chun J, Choi BS, et al. Abdominal synthetic CT reconstruction with intensity projection prior for MRI-only adaptive radiotherapy. *Phys Med Biol*. 2021;66:204001. doi: 10.1088/1361-6560/ac279e. © 2021 Institute of Physics and Engineering in Medicine. All rights reserved.

The dissertation author was the primary researcher and author of the work that forms the basis of this chapter.

The text of Chapter 6, in part or in full, has been reproduced from the material as it has been prepared for submission to *Medical Physics*:

Olberg S, Choi, BS, Park I, et al. Ensemble learning and personalized training for the improvement of unsupervised deep learning-based synthetic CT reconstruction. 2022. (Under review)

The dissertation author was the primary researcher and author of the work that forms the basis of this chapter.

Sven Olberg

Washington University in St. Louis

May 2022

Dedicated to my love.

ABSTRACT OF THE DISSERTATION

Optimizations in the MRI-Guided Adaptive Radiation Therapy Workflow:

Towards MR-Only RT

by

Sven Olberg

Doctor of Philosophy in Biomedical Engineering

Washington University in St. Louis, 2022

Professor Hong Chen, Chair

Technological developments in external beam radiation therapy (RT) made in the past decades have been largely focused on the tight integration of imaging systems and therapy delivery systems in order to enable the precise delivery of therapeutic doses of radiation to internal, complexly shaped targets. Along this line, the adoption of x-ray computed tomography (CT) in the RT workflow and the combination of volumetric cone-beam CT with linacs in the clinical setting brought the practice of image-guided RT (IGRT) to the forefront. Practical limitations related to the poor soft tissue contrast of these x-ray imaging modalities prompted the adoption of magnetic resonance imaging (MRI) as a secondary modality in the IGRT workflow, allowing for the delineation of targets and critical structures at sites throughout the body. More recently, the development and growing use of combination MR-linac platforms has placed a focus on MR-only RT workflows in which MRI is the sole imaging modality used for treatment guidance and planning. In addition to the improved soft tissue contrast, the functional imaging capabilities and capacity for real-time image guidance of MRI have opened the door to adaptive RT (ART) applications enabled by MRI guidance in which treatments are adapted based on observed changes in anatomy or biology.

These nascent MRI-guided ART workflows face several limitations, however. First, treatment planning in the adaptive setting relies on the representation of anatomy as shown in simulation scans often acquired weeks prior to treatment delivery. When the anatomy

observed at treatment differs drastically from that shown at simulation, a time-intensive re-planning process must be undertaken. Second, the trade-off between spatial resolution and acquisition time means achieving high-resolution MRI scans requires scanning periods or breath-holds that may be infeasible, introducing detrimental motion artifacts. Finally, MRI-guided workflows face a technical hurdle stemming from the requirement for electron density information for dose calculations, which may be derived directly from CT images. The inclusion of CT simulation scans in the MRI-guided RT setting brings concerns related not only to extra imaging dose but also challenges with multi-modality image registration that can give rise to geometric errors that persist throughout treatment. This dissertation presents optimizations in the MRI-guided ART workflow with a focus on these challenges.

First, a treatment planning strategy for pancreatic cancer cases that is robust to inter-fraction variations of primary critical structures is presented as a means of simplifying the daily adaptive planning workflow and improving target coverage in adapted fractions. Following this, a deep learning (DL)-based approach to MRI super-resolution reconstruction that enables the use of fast, low-resolution scans for guidance through fourfold upscaling is presented. Finally, the challenge of achieving electron density information through DL-based synthetic CT (sCT) reconstruction is explored in three contexts: 1) a parameter-efficient network architecture is explored in the context of sCT reconstruction in the breast for use in the low-data setting; 2) an approach to paired-data sCT reconstruction in the abdomen that handles the challenge posed by the variable presence of intestinal gas in corresponding MRI and CT scans is presented; and 3) two distinct approaches to improving the sCT outputs of an unpaired-data framework—a cascade ensemble approach and a personalized training strategy originally designed for use in the paired-data setting—are explored in the context of the sCT reconstruction task in the male pelvis. Ultimately, the methods presented in this dissertation represent improvements at three vital stages of the MRI-guided RT workflow that could potentially enable an effective and practical approach to MR-only ART.

Chapter 1

Introduction

1.1 Radiation Therapy

At the most fundamental level, the goal in the practice of radiation therapy (RT) is to kill tumor cells through DNA damage caused by the absorption of radiation within the body.[1] The medical uses of ionizing radiation—whether for producing images of internal tissues and structures or with a therapeutic intent for the treatment of cancer—have been explored and developed since the early years following the discovery of x-rays at the end of the 19th century.[1,2] With the expanding therapeutic use of radiation also came observations of radiation-related complications including most notably severe skin reactions.[2] These early experiences highlighted the fact that the direct and indirect action of radiation—energetic electromagnetic waves like x-rays and gamma rays or charged or uncharged particles like electrons, protons, and neutrons—does not discriminate between healthy tissues and the diseased tissues that are the target of treatment. Developing an understanding of the biological factors that influence the radiosensitivity of different types of cells and structures prompted the adoption of fractionated treatments in which the total dose is spread over multiple treatment sessions to allow for repair in healthy tissues while still achieving the ultimate goal of tumor control.[2] This highlights the fundamental challenge in the practice of RT, which is performing the balancing act of delivering a sufficient amount of radiation that will destroy a tumor while simultaneously sparing the healthy tissues that surround the target.[2–4]

1.2 Modernizing Developments

To this end, the technological developments of the past decades in the field of radiation oncology have placed the focus on improving the ability of clinicians to localize and delineate the target of treatment and shape a dose distribution that is highly conformal to that target.[2] Central to this goal are the concepts of image-guided RT (IGRT), which relies on anatomical

imaging techniques to visualize a patient's internal anatomy prior to treatment delivery, and intensity-modulated RT (IMRT), which enables the delivery of a dose distribution that is highly conformal to the boundaries of a complexly shaped target.[2–4]

1.2.1 IGRT

The primary motivation for the earliest adoption of IGRT was to enable repeatable patient setup and target localization through the use of treatment simulators that mimicked the geometry of therapy delivery machines.[4] In this setting, the emphasis remained primarily on external markers and large internal margins were still used as a means of ensuring dose delivery to a mobile, uncertain target in a large irradiated volume.[4] The development of x-ray computed tomography (CT) and its integration into the RT workflow allowed for the widespread adoption of the concept of IGRT in the early 1970s, placing the emphasis now on reducing internal margins and systematic uncertainties in treatment delivery.[2,4]

1.2.2 IMRT

Visualizing the geometric shape of the target and its position relative to the healthy tissues surrounding it grants critical insight that aids in the task of maximizing the dose delivered to the tumor while minimizing the dose received by these surrounding organs at risk (OARs).[3,5] Along this line, the development and popularization of the IMRT delivery technique followed closely behind this improvement in target visualization and is tightly intertwined with the continued development of IGRT approaches.[4,6] IMRT represents a departure from the prior approaches to planning and delivery—termed forward planning approaches—in which simple beam arrangements with large margins were selected to deliver dose to designated targets and avoid critical structures while considering setup variations and motion with relatively broad strokes.[6] IMRT instead relies on an inverse planning approach in which modulated

beam profiles are optimized to satisfy dose constraints input by the planner.[6] The conceptual basis of IMRT is that the composite dose distribution achieved through the use of multiple beams with varying intensities across the beam profile at various angles or in a continuous arc can be sculpted to conform to the contours of targets with highly complex shapes.[7] Such conformal dose distributions that maximize the sparing of healthy tissues surrounding the target are characterized by high dose gradients, meaning that geometric misses arising from inaccuracies in patient positioning are a significant clinical concern.[4–6,8]

1.2.3 Imaging and therapy systems integration

It is these geometric uncertainties in treatment delivery that highlighted the need for a tight integration of image guidance and radiation delivery systems. Of the various solutions to this problem proposed at the time, the integration of a kilovoltage x-ray source and flat-panel detector on an existing linac used for radiation delivery pursued by Jaffray et al.[9] has grown to be widely adopted as the clinical standard for IGRT.[10,11] In this setting, volumetric cone-beam CT (CBCT) data obtained in a single rotation of the linac gantry represents the patient's anatomy at the time of treatment that may be used for position verification.[9] CBCT-based IGRT is not without its drawbacks, however. Compared to magnetic resonance imaging (MRI), another popular volumetric imaging modality adopted into the field of radiation oncology, both x-ray CT and CBCT exhibit poor soft tissue contrast that can be limiting in certain anatomical sites.[12,13] Additional considerations including the imaging dose imparted in CT and CBCT examinations and the functional imaging capabilities of MRI prompted investigations into the integration of an MRI unit with radiation delivery systems at around the same time the CBCT-based IGRT solutions were being explored.[14–16]

1.3 MRI-guided RT Platforms

Since that time, several distinct MRI-guided RT solutions have been proposed and developed to various stages ranging from early research and design to commercial availability (Table 1.1).[17–21] The first generation ViewRay MRI-guided RT system, which consists of a 0.35 T magnet paired with three ^{60}Co heads, began treating patients in 2014.[17,22] More recently, linac-based systems including the second generation ViewRay MRIdian system with the same 0.35 T magnet and the Elekta Unity system with a 1.5 T magnet have been successfully adopted in the clinical setting as an alternative to the conventional CBCT-based IGRT systems.[23,24] The superior soft tissue contrast of MRI compared to that of these CBCT in-room imaging devices offers significant margin reductions and enables online adaptive radiation therapy (ART) treatments using these MRI-guided RT platforms.[25] By re-optimizing treatment plans according to anatomical or positional changes in the target and surrounding structures observed using MRI on a daily basis, the accuracy of radiation dose delivery may be significantly improved while also achieving a substantial reduction in the irradiated volume of normal tissue or a beneficial escalation of the dose delivered to the target.[26–29]

1.4 Motivation of the Thesis

Despite these attractive features of MRI-guided ART, the relatively nascent approach faces several limitations to implementing a practical and effective workflow. First is an issue regarding treatment planning itself in an adaptive setting. In the conventional IGRT case, daily variations in target positioning are handled through the registration of the planning image and the daily setup image before the original plan is delivered.[22] Optimization parameters assigned at the pre-treatment simulation stage based on the relative positioning of healthy tissues to the target may not reflect the position of these structures at the time of treatment, leading to a time-intensive re-planning process in the adaptive setting that must

Table 1.1. MRI-guided RT platforms and associated details.

Platform (Mfr./Institution)	Field	Magnet Geometry	Linac	Availability
MRIdian (ViewRay)[17,30]	0.35 T	Vertically gapped split-bore whole-body magnet, 70 cm bore	6 MV	Commercially available. Installed at clinics in the USA, Europe, the Middle East, and Asia
Unity (Elekta)[19,23]	1.5 T	Closed-bore whole-body magnet with central homogenous region absent of any coils for transverse beam, 70 cm bore	7 MV	Commercially available. Installed at clinics in the USA, Canada, Europe, Asia, and Australia
Australian MRI-Linac (Ingham Institute for Applied Medical Research)[21]	1 T	82 cm diameter open-bore whole-body magnet, 50 cm gap; Allows for inline and perpendicular beam orientations	6 MV	Pre-clinical research & development
Rotating Biplanar Linac-MR (Cross Cancer Institute)[20]	0.6 T	Open-bore whole-body high-temperature superconducting magnet, 60 cm gap; Allows for inline and perpendicular beam orientations	6 MV	Pre-clinical research & development
Facility for MRgRT (Princess Margaret Cancer Centre)[18]	1.5 T	Closed-bore whole-body mobile simulator on rails, 70 cm bore; Interfaces with conventional linac in dedicated suite	6 MV	Single clinic deployment

occur while the patient remains in treatment position.[31,32] Second, there is a persistent problem of limited spatial resolution and long acquisition time. In the MRI-guided setting, the spatial resolution fundamentally determines the setup uncertainty and the accuracy in delineating the tumor and surrounding healthy tissues.[33–36] Furthermore, motion artifacts that arise due to long acquisition times render motion management problematic, particularly when imaging organs within the abdomen.[37,38] A final hurdle in the MRI-guided RT setting is the requirement for electron density information for dose calculations during treatment planning. This information may be derived directly from CT images, but pixel intensities in MR images do not correspond to the electron densities of tissues.[39] Considering this, many existing implementations of the MRI-guided RT workflow rely on a secondary CT simulation scan in addition to the primary MRI simulation scan.[22,23,25,28] In this setting, the CT simulation scan must be registered to the corresponding MRI setup scans from each treatment fraction, which is particularly challenging in abdominal cases when the anatomy represented in each scan is incompatible due to changes in the size or position of organs of interest on an interfraction basis.[40,41] Considering these limitations, achieving an effective MRI-guided RT workflow relies on optimizing the robustness of the treatment planning process, acquiring fast and high-quality MR images, and producing electron density data without requiring an additional CT simulation scan.

1.5 Specific Aims

With a focus on these limitations, the specific aims of the dissertation are as follows:

1. Develop a treatment planning approach for the pancreas that is robust to daily anatomical variations observed in MRI-guided ART, leading to improved target coverage in adapted fractions compared to the conventional approach.

2. Construct a framework for deep learning (DL)-based MRI super-resolution reconstruc-

tion in the low-data setting to enable the use of fast, low-resolution acquisitions for treatment guidance through fourfold upsampling.

3. Explore technical challenges in the DL-based synthetic CT (sCT) reconstruction task to achieve accurate electron density information from MRI simulation and setup scans, eliminating the requirement for CT simulation scans.

    3.1. Demonstrate the training and performance benefits of a parameter-efficient network architecture in the low-data setting using the sCT reconstruction task in the breast and illustrate the feasibility of sCT-based dose calculations in the breast.

    3.2. Develop a preprocessing approach that handles the issue of the variable presence of intestinal gas to enable a supervised, paired-data DL approach to sCT reconstruction in the abdomen and demonstrate the dosimetric impact of intestinal gas in this setting.

    3.3. Demonstrate improvements in image quality for unsupervised, unpaired-data sCT results through the use of a cascade ensemble approach and a personalized training strategy initially designed for the supervised, paired-data setting.

1.6 Organization of the Dissertation

The body of the dissertation is organized as follows:

Chapter 2 describes the OAR grouping method, an approach to treatment planning for pancreatic cancer cases that treats the primarily involved OARs as a single structure during optimization. This approach is compared against the conventional approach in which these structures are handled separately with a focus on improvements in target coverage in adapted fractions.

Chapter 3 presents a framework developed for MRI super-resolution reconstruction that

is designed to operate in a setting in which there is little training data that may be utilized. The components of the framework including a denoising autoencoder as a front-end preprocessing unit, a down-sampling network that produces low-resolution images for training, and the super-resolution generative model that produces the ultimate outputs are described. Evaluations of image quality demonstrate the importance of each of these constituent networks in the overall framework.

Chapter 4 describes a parameter-efficient network architecture for sCT reconstruction in the context of the breast that utilizes the compact atrous spatial pyramid pooling module and demonstrates the improved performance of this architecture in terms of training time and image quality compared to a widely used and comparably parameter-rich architecture. Outputs of the proposed architecture are also used to demonstrate the feasibility of sCT-based dose calculations in the breast through comparisons to CT-based clinical plans.

Chapter 5 introduces a preprocessing approach for the supervised sCT reconstruction task in the abdomen specifically, where the variable presence of intestinal gas in corresponding MRI and CT scans presents a hurdle to training paired-data architectures. The method for creating a well-matched, clinically unavailable training data set through the propagation of air from MR images to the corresponding CT images is described and the dosimetric impact of these unhandled discrepancies in the presence of intestinal gas is illustrated.

Chapter 6 discusses the inherent gap between supervised, paired-data sCT results and unsupervised, unpaired-data sCT results observed in the context of sCT reconstruction in the male pelvis. Two distinct methods for improving the image quality and reconstruction accuracy of the unpaired-data sCT results are presented: a cascade ensemble approach in which two networks are applied in sequence to produce the ultimate output and a personalized training approach that produces patient-specific models, an approach that was originally designed for use in the paired-data setting.

Finally, Chapter 7 summarizes the research contributions made in this work and high-

lights future directions of interest.

References

[1] Hall E, Giaccia A. *Radiobiology for the Radiologist*. Philadelphia: Lippincott Williams & Wilkins. 2012

[2] Bernier J, Hall E, Giaccia A. Radiation oncology: a century of achievements. *Nat Rev Cancer*. 2004;4(9):737–747. doi:10.1038/nrc1451

[3] Verellen D, De Ridder M, Storme G. A (short) history of image-guided radiotherapy. *Radiother Oncol*. 2008;86(1):4–13. doi:10.1016/j.radonc.2007.11.023

[4] Verellen D, De Ridder M, Linthout N, Tournel K, Soete G, Storme G. Innovations in image-guided radiotherapy. *Nat Rev Cancer*. 2007;7(12):949–960. doi:10.1038/nrc2288

[5] van Herk M. Different styles of image-guided radiotherapy. *Semin Radiat Oncol*. 2007; 17(4):258–267. doi:10.1016/j.semradonc.2007.07.003

[6] Hong T, Ritter M, Tomé W, Harari P. Intensity-modulated radiation therapy: emerging cancer treatment technology. *Br J Cancer*. 2005;92(10):1819–1824. doi: 10.1038/sj.bjc.6602577

[7] Bortfeld T. IMRT: a review and preview. *Phys Med Biol*. 2006;51(13):R636–R379. doi:10.1088/0031-9155/51/13/R21

[8] Khan F, Gibbons J. *The Physics of Radiation Therapy*. Philadelphia: Lippincott Williams & Wilkins. 2014

[9] Jaffray D, Siewerdsen J, Wong J, Martinez A. Flat-panel cone-beam computed tomography for image-guided radiation therapy. *Int J Radiat Oncol Biol Phys*. 2002; 53(5):1337–1349. doi:10.1016/s0360-3016(02)02884-5

[10] Verellen D, De Ridder M, Tournel K, et al. An overview of volumetric imaging technologies and their quality assurance for IGRT. *Acta Oncol*. 2008;47(7):1271–1278. doi: 10.1080/02841860802244182

[11] Ibbott G. The need for, and implementation of, image guidance in radiation therapy. *Ann ICRP*. 2018;47(3–4):160–176. doi:10.1177/0146645318764092

[12] Kupelian P, Sonke JJ. Magnetic resonance–guided adaptive radiotherapy: A solution to the future. *Semin Radiat Oncol*. 2014;24(3):227–232. doi: 10.1016/j.semradonc.2014.02.013

[13] Pollard JM, Wen Z, Sadagopan R, Wang J, Ibbott GS. The future of image-guided radiotherapy will be MR guided. *Br J Radiol*. 2017;90(1073):20160667. doi: 10.1259/bjr.20160667

[14] Lagendijk JJW, Raaymakers BW, Van Der Heide UA, et al. MRI guided radiotherapy: MRI as position verification system for IMRT. *Radiother Oncol*. 2002;64

[15] Raaymakers B, Lagendijk J, Van Der Heide U, et al. Integrating an MRI scanner with a radiotherapy accelerator: a new concept of precise on line radiotherapy guidance and treatment monitoring. In *Proceedings of the 14th International Conference on the Use of the Computers in Radiation Therapy*

[16] Raaymakers B, Raaijmakers A, Kotte A, Jette D, Lagendijk J. Integrating a mri scanner with a 6 MV radiotherapy accelerator: dose deposition in a transverse magnetic field. *Phys Med Biol*. 2004;49(17):4109–4118. doi:10.1088/0031-9155/49/17/019

[17] Mutic S, Dempsey JF. The ViewRay system: Magnetic resonance–guided and controlled radiotherapy. *Semin Radiat Oncol*. 2014;24(3):196–199. doi: 10.1016/j.semradonc.2014.02.008

[18] Jaffray D, Carlone M, Milosevic M, et al. A facility for magnetic resonance–guided radiation therapy. *Semin Radiat Oncol*. 2014;24(3):193–195. doi: 10.1016/j.semradonc.2014.02.012

[19] Raaymakers B, Lagendijk J, Overweg J, et al. Integrating a 1.5 T MRI scanner with a 6 MV accelerator: proof of concept. *Phys Med Biol*. 2009;54(12):N229–N237. doi: 10.1088/0031-9155/54/12/N01

[20] Fallone BG. The rotating biplanar linac-magnetic resonance imaging system. *Semin Radiat Oncol*. 2014;24:200–202. doi:10.1016/j.semradonc.2014.02.011

[21] Keall PJ, Barton M, Crozier S. The Australian magnetic resonance imaging-linac program. *Semin Radiat Oncol*. 2014;24:203–206. doi:10.1016/j.semradonc.2014.02.015

[22] Fischer-Valuck BW, Henke L, Green O, et al. Two-and-a-half-year clinical experience with the world's first magnetic resonance image guided radiation therapy system. *Adv Radiat Oncol*. 2017;2(3):485–493. doi:10.1016/j.adro.2017.05.006

[23] Raaymakers BW, Jürgenliemk-Schulz IM, Bol GH, et al. First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. *Phys Med Biol*. 2017;62(23):L41–L50. doi:10.1088/1361-6560/aa9517

[24] Liney G, Whelan B, Oborn B, Barton M, Keall P. MRI-linear accelerator radiotherapy systems. *Clin Oncol*. 2018;30(11):686–691. doi:10.1016/j.clon.2018.08.003

[25] Acharya S, Fischer-Valuck B, Kashani R, et al. Online magnetic resonance image guided adaptive radiation therapy: First clinical applications. *Int J Radiat Oncol Biol Phys*. 2016;94(2):394–403. doi:10.1016/j.ijrobp.2015.10.015

[26] Henke L, Olsen J, Contreras J, et al. Stereotactic MR-guided online adaptive radiation therapy (SMART) for ultracentral thorax malignancies: Results of a phase 1 trial. *Adv Radiat Oncol*. 2018;4(1):201–209. doi:10.1016/j.adro.2018.10.003

[27] Chen H, van Sornsen de Koste J, Bohoudi O, et al. Reduction in normal tissue complication probability by stereotactic magnetic resonance-guided ablative radiotherapy for adrenal lesions. *Int J Radiat Oncol Biol Phys*. 2020;108(3):S65–S66. doi: 10.1016/j.ijrobp.2020.07.2200

[28] Mayinger M, Ludwig R, Christ S, et al. Benefit of replanning in MR-guided online adaptive radiation therapy in the treatment of liver metastasis. *Radiat Oncol*. 2021; 16(84):1–8. doi:10.1186/s13014-021-01813-6

[29] Glide-Hurst C, Lee P, Yock A, et al. Adaptive radiation therapy (ART) strategies and technical considerations: A state of the ART review from NRG oncology. *Int J Radiat Oncol Biol Phys*. 2021;109(4):1054–1075. doi:10.1016/j.ijrobp.2020.10.021

[30] Klüter S. Technical design and concept of a 0.35 T MR-linac. *Clin Transl Radiat Oncol*. 2019;18:98–101. doi:10.1016/j.ctro.2019.04.007

[31] Henke L, Kashani R, Robinson C, et al. Phase I trial of stereotactic MR-guided online adaptive radiation therapy (SMART) for the treatment of oligometastatic or unresectable primary malignancies of the abdomen. *Radiother Oncol*. 2018;126(3):519–526. doi:10.1016/j.radonc.2017.11.032

[32] Lamb J, Cao M, Kishan A, et al. Online adaptive radiation therapy: implementation of a new process of care. *Cureus*. 2017;9(8):e1618. doi:10.7759/cureus.1618

[33] Bourqe A, Bedwani S, Carrier JF, et al. Particle filter-based target tracking algorithm for magnetic resonance-guided respiratory compensation: Robustness and accuracy assessment. *Int J Radiat Oncol Biol Phys*. 2018;100(2):325–334. doi: 10.1016/j.ijrobp.2017.10.004

[34] Glitzner M, Denis de Senneville B, Lagendijk J, Raaymakers B, Crijns S. On-line 3D motion estimation using low resolution MRI. *Phys Med Biol*. 2015;60(16):N301–N310. doi:10.1088/0031-9155/60/16/N301

[35] Menten M, Fast M, Wetscherek A, et al. The impact of 2D cine MR imaging parameters on automated tumor and organ localization for MR-guided real-time adaptive radiotherapy. *Phys Med Biol*. 2018;63(23):235005. doi:10.1088/1361-6560/aae74d

[36] Kurz C, Buizza G, Landry G, et al. Medical physics challenges in clinical MR-guided radiotherapy. *Radiat Oncol*. 2020;15(93):1–16. doi:10.1186/s13014-020-01524-4

[37] Zaitsev M, Maclaren J, Herbst M. Motion artifacts in MRI: A complex problem with many partial solutions. *J Magn Reson Imaging*. 2015;42(4):887–901. doi: 10.1002/jmri.24850

[38] Khot R, McGettigan M, Patrie J, Feuerlein S. Quantification of gas exchange-related upward motion of the liver during prolonged breathholding—potential reduction of motion artifacts in abdominal MRI. *Br J Radiol*. 2020;93(1106):20190549. doi: 10.1259/bjr.20190549

[39] Schmidt MA, Payne GS. Radiotherapy planning using MRI. *Phys Med Biol*. 2015; 60(22):R323–R361. doi:10.1088/0031-9155/60/22/R323

[40] Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. *Radiat Oncol*. 2017;12(1):28. doi:10.1186/s13014-016-0747-y

[41] Johnstone E, Wyatt JJ, Henry AM, et al. Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging–only radiation therapy. *Int J Radiat Oncol Biol Phys*. 2018;100(1):199–217. doi: 10.1016/j.ijrobp.2017.08.043

Chapter 2

Optimization of treatment planning workflow and tumor coverage during daily adaptive magnetic resonance image guided radiation therapy of pancreatic cancer

2.1 Introduction

Magnetic resonance image guided radiation therapy (MR-IGRT) systems (MRIdian System; ViewRay Inc., Oakwood Village, OH) have been successfully implemented at a number of institutions in recent years and used to treat a growing group of patients that benefit from the advantages offered by Magnetic Resonance Imaging (MRI) compared to other conventional image guiding modalities (e.g. CBCT, x-ray radiography).[1,2] Anatomical variations from simulation to treatment as well as fraction to fraction represent a challenge in the delivery of radiation therapy, as changes in the size and position of target and critical structures can affect dose delivery in a clinically significant way.[3–12] The improved visualization of soft tissues gained from an MRI system compared to traditional cone-beam CT (CBCT) systems allows for the daily management of these inter-fraction anatomical variations, especially in areas of extensive soft tissue like the abdomen.[1,13] It is this improved soft tissue visualization that makes adaptive radiation therapy (ART) an attractive application of the MR-IGRT system.[1,14,15] The implementation of ART does come at the cost of an increased time investment for plan re-optimization, however, which reflects the time consuming nature of inverse treatment planning in radiation therapy.[1,16]

Inverse planning involves the assignment of weighting parameters to target and critical structures that control the balance between delivering the prescribed dose to the target and protecting healthy tissues.[17–23] The selection of these parameters is recognized as a challenging undertaking that involves a "guessing game" of repeated trial and error in a "human iteration loop."[22–26] This process becomes more complex and time consuming as the number of important structures and associated parameters involved in a plan grows, highlighting the need for a simpler treatment planning workflow.[25,27]

In pancreatic cancer cases specifically, a group of four structures—the stomach, duodenum, small bowel, and large bowel—represent organs at risk (OARs) of particular impor-

tance.[28,29] The present study proposes the use of a single OAR structure that combines these four primary OARs, restricting the number of associated weighting parameters by a factor of four and thereby simplifying the treatment planning process. The aim of the study is to simplify the daily adaptive treatment planning workflow in the treatment of pancreatic cancer using the ViewRay System while maintaining tumor coverage that is robust to inter-fraction anatomical variations. The conventional daily adaptive treatment planning workflow is described along with the OAR grouping method, and comparative dosimetric data for 16 pancreatic cancer patients treated with daily adaptive MR-IGRT is presented.

2.2 Materials and Methods

Sixteen pancreatic cancer patients previously treated with daily adaptive MR-IGRT were used as test cases, representing 208 adapted fractions. The volumes of the planning target volume (PTV), the contours of which are held constant throughout treatment for each patient, ranged from 57.7 cc to 356.3 cc with an average of 160.5 cc. Clinically delivered treatment plans were used as a baseline for comparison at each fraction. The OAR grouping method was compared to the baseline in three metrics: 1) Percentage of the PTV covered by 95% of the prescribed dose (D_{95}), 2) D_{95} coverage of the PTV OPT, and 3) Percentage of the PTV covered by 100% of the prescribed dose (D_{100}).

2.2.1 OAR grouping method

In the conventional treatment plan, each OAR is handled separately. Weighting parameters are assigned to each of these critical structures as well as the target as inputs to the objective function, which is the aggregate of the cost functions for individual structures involved in planning. Figure 2.1 illustrates the simple formulation of the cost function $f(D|\theta)$ used in this study for both OARs and the target, where D is the delivered voxel dose and θ is the set

of weighting parameters assigned to a structure. For OARs (Figure 2.1a), the cost increases for a given delivered dose D once a selected threshold T is exceeded. The shape of the curve is controlled using an importance factor ω and power u. Similarly, the cost for a dose D delivered to the target (Figure 2.1b) increases as the dose deviates from the prescribed dose D_0 plus a selected offset. The curves for doses above and below this threshold are shaped using importance and power parameters as discussed for the OAR case.

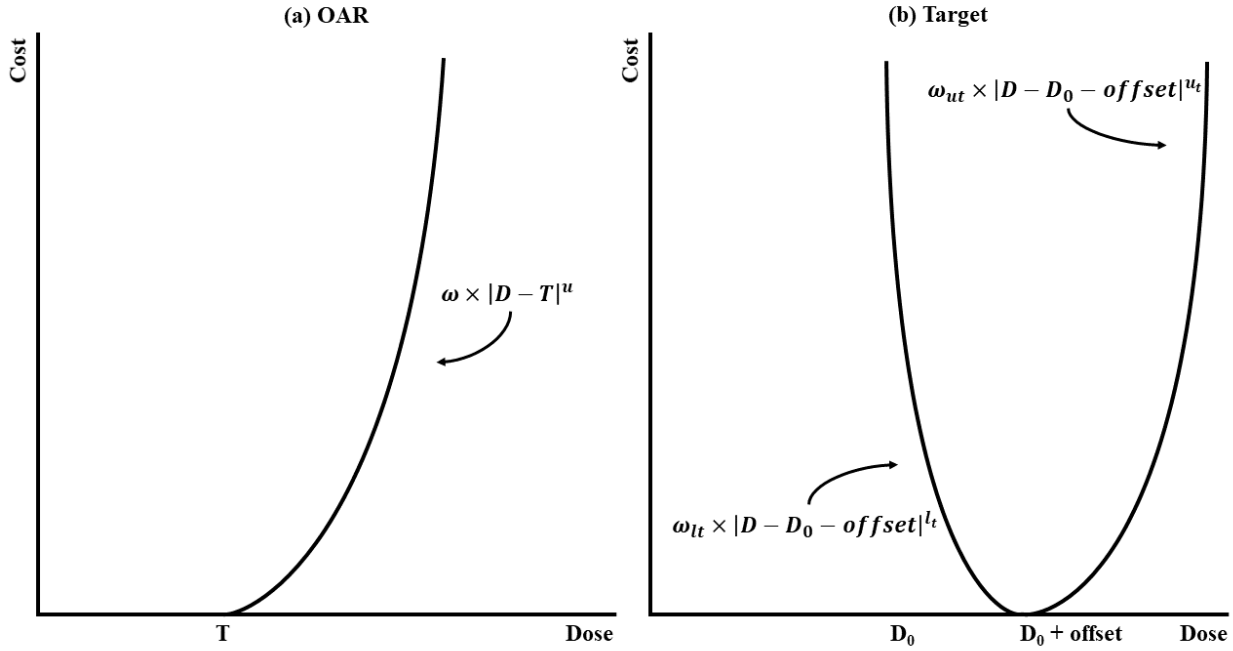

Figure 2.1. Representative cost functions plotted for an OAR (a) and the target structure (b), along with the associated weighting parameters.

The creation of the initial objective function involves tuning each one of these parameters in order to achieve a plan of acceptable quality. In a pancreatic cancer treatment plan, each of the four primary OARs is subject to the same volumetric dose constraint (Table 2.1), so any variation in the assignment of weighting parameters is likely a reflection of the position of an OAR relative to the PTV. The OAR grouping method creates a single OAR structure by combining the portions of the stomach, duodenum, small bowel, and large bowel within

3 cm of the PTV. Using OAR grouping, the shared dose constraint is maintained, but the objective function is simplified by restricting the number of parameters used as inputs.

Table 2.1. Volumetric dose constraints observed in pancreatic cancer treatment plans. Each of the four primary OARs is subject to the same volumetric constraint of less than 0.5 cc receiving a maximum dose of 45 Gy.

Structure	Volume Measure	Max Dose (Gy)
Stomach	< 0.5 cc	45
Duodenum	< 0.5 cc	45
Small Bowel	< 0.5 cc	45
Large Bowel	< 0.5 cc	45
Spinal Cord	$= 0$ cc	40
Kidneys	$< 50\%$	15

2.2.2 Simplification of objective function

The value of the proposed method is derived from a simplification of the dose optimization function:

$$f_{total}\left(D|\theta\right) = \sum_{i=1}^{k} f_{ci}\left(D|\theta_i\right) + f_t\left(D|\theta_t\right), \tag{2.1}$$

where the cost f is a function of the delivered dose D and weighting parameters $\theta_i = \{\omega_i, u_i, T_i\}$ for all critical structures and $\theta_t = \{\omega_{lt}, \omega_{ut}, l_t, u_t, D_0, offset\}$ for the target structure (Figure 2.1). In the conventional pancreas treatment plan, the stomach, duodenum, small bowel, and large bowel each carry individual weighting parameters including upper importance ω, upper power u, and threshold T. Additionally, the target carries lower parameters for importance and power, ω_l and l, respectively, as well as a prescribed dose D_0 and corresponding offset. In this conventional case, the set of all weighting parameters for critical structures and the target $\theta = \{\omega_1, u_1, T_1, ..., \omega_k, u_k, T_k, \omega_{lt}, \omega_{ut}, l_t, u_t, D_0, offset\}$ is quite large. The OAR grouping method combines the four primary critical structures, and

as a result Eq. 2.1 becomes

$$f_{total}\left(D|\theta\right) = f_c\left(D|\theta_c\right) + f_t\left(D|\theta_t\right) + \sum_{j=1}^{l} f_{r-OAR}\left(D|\theta_j\right), \tag{2.2}$$

where $\theta_{c+t} = \{\omega_c, u_c, T_c, \omega_{lt}, \omega_{ut}, l_t, u_t, D_0, offset\}$ represents the reduced set of weighting parameters for the four combined primary critical structures and the target and θ_j represents the weighting parameters of any remaining OARs (r-OAR) that may be included in the plan, such as the spinal cord or kidneys.

2.2.3 Simulation

The conventional workflow for simulation has been previously described in detail.[1,15] Briefly, patients undergo CT and MRI scans on the simulation day using the same setup device. Then, the MR images are sent to either the ViewRay treatment planning system or third-party software (e.g. Eclipse) for the structure delineation. The corresponding structure sets, MR images, and CT scans are then combined and fused into the ViewRay treatment planning system for plan creation. Treatment isocenter, number of beam entries, and beam angles are defined by the planners such that the plan is physically deliverable with respect to the couch position. Finally, the prescribed dose and dose constraints on critical organs are used to guide the selection of weighting parameters input to the objective function for the IMRT plan optimization.

In this study, the conventional simulation plan was copied for each patient to keep the same beam entry and physical setup (treatment isocenter, couch position, etc.) in order to maintain as fair a comparison as possible. Using the OAR grouping method, portions of the stomach, duodenum, small bowel, and large bowel within 3 cm of the PTV were combined into a single OAR structure. This combined structure was then used in place of the individual critical structures in constructing the simulation objective function. The

OAR grouping simulation plan was created and optimized such that the target coverage, OAR doses, and beam-on time were comparable to the conventional plan.

2.2.4 Daily adaptive MR-IGRT

In daily adaptive MR-IGRT, the volumetric MRI of the patient is scanned before each treatment fraction. After critical structures are re-contoured, the plan is re-optimized based on the patient's daily anatomy using the same objective function constructed in the pretreatment simulation plan, all while the patient remains on the couch. In the present study, the conventional plan at each treatment fraction was copied and modified using the OAR grouping method. For the purposes of comparison, all plans—both conventional and OAR grouping—were normalized to satisfy one of two scenarios, whichever came first: 1) the primary OAR receiving the greatest volumetric dose received a dose of 45 Gy to 0.5 cc, or 2) the dose to the spinal cord or kidneys met the dose constraints as outlined in Table 2.1.

2.3 Results

The percentage of fractions improved in each metric out of 208 total fractions for all patients is presented in Table 2.2. Generally, the coverage is improved across a majority of fractions when OAR grouping is utilized over the conventional method. Greater than 70% of fractions showed improvement in PTV OPT coverage, while approximately 80% of all fractions demonstrated improved PTV coverage using OAR grouping.

Table 2.2. Percentage of fractions in which coverage was improved using OAR grouping. Total = 208 fractions.

	PTV OPT D_{95}	PTV D_{95}	PTV D_{100}
Fractions Improved (%)	73	78	84

In Figure 2.2, PTV and PTV OPT D_{95} coverage relative to the prescription of 95%

target volume coverage by D_{95} is plotted as a cumulative histogram. Figure 2.2 illustrates that PTV coverage fails to meet the prescription in nearly 100% of the adapted fractions due to the close proximity of surrounding OARs. The benefits of the proposed method can be more clearly understood when examining coverage of the PTV OPT, which is comprised of portions of the PTV not overlapped by OARs. In the conventional case, only 22% of all fractions exhibited PTV OPT coverage that met the prescription. When the OAR grouping method is utilized, that ratio increases to 42% of all fractions. It should be noted that the high ratio of under-covered fractions demonstrates the challenge of treating pancreatic cancer. In a majority of cases, the PTV is overlapped to some extent by surrounding OARs. In these cases, target coverage is often compromised in order to satisfy the dose constraints assigned to these critical structures.

Average coverage across a patient's total adapted fractions is presented for each patient in Table 2.3 for both the conventional and OAR grouping plans. Similarly, Table 2.4 contains the average improvement of OAR grouping plans compared to conventional plans along with the minimum and maximum observed improvements across all patients. D_{95} coverage of the PTV and PTV OPT was improved by an average of approximately 4%, while PTV D_{100} coverage demonstrated an average improvement of greater than 6%. The relatively large standard deviations in each case are due to data points well above the mean, which can be observed for each metric in Figure 2.3.

2.4 Discussion

In the present study, the OAR grouping method was applied to pancreatic cancer cases with the aim of simplifying the daily adaptive MR-IGRT treatment planning workflow while maintaining target coverage that is robust to inter-fraction anatomical variations. In this way, the value of the proposed method is twofold. First, the simplification of the initial

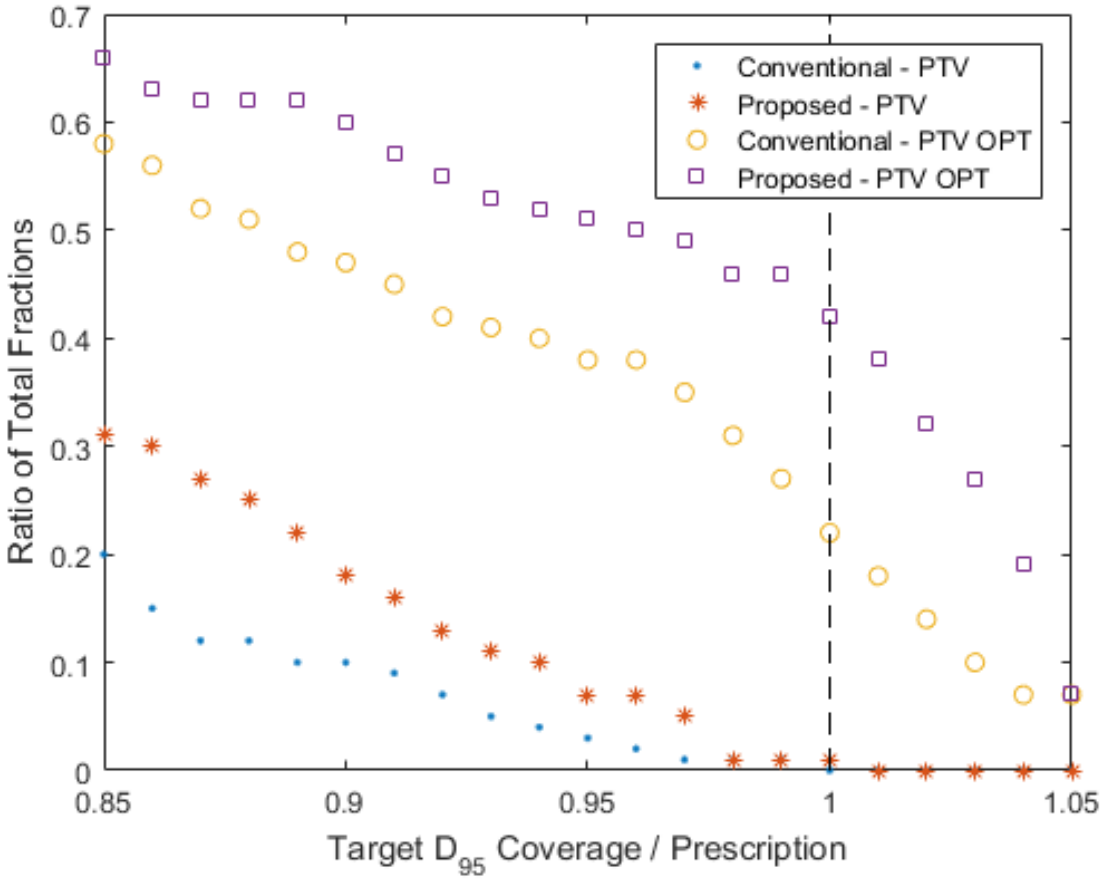

Figure 2.2. Cumulative histogram that demonstrates the ratio of total fractions receiving target coverage relative to the prescription of 95% target volume coverage by D_{95}.

Table 2.3. Average coverage in each metric using the conventional and OAR grouping method over all adapted fractions for each patient.

Adapted Fractions	Method	Average Coverage (%)		
		PTV OPT	PTV	
		D_{95}	D_{95}	D_{100}
Patient 1				
14	Conventional	94.00	77.36	66.47
	Proposed	95.78	79.64	69.68
Patient 2				
13	Conventional	99.91	82.56	77.97
	Proposed	99.65	83.26	79.39
Patient 3				
13	Conventional	94.88	84.36	78.89
	Proposed	96.85	86.40	80.13
Patient 4				
15	Conventional	94.59	81.38	72.19
	Proposed	99.48	87.88	84.24
Patient 5				
14	Conventional	76.44	48.50	40.42
	Proposed	78.49	49.99	43.29
Patient 6				
15	Conventional	67.46	49.16	39.64
	Proposed	72.62	53.08	43.93
Patient 7				
14	Conventional	78.52	76.54	63.09
	Proposed	82.32	80.37	69.21
Patient 8				
10	Conventional	61.22	52.98	42.57
	Proposed	59.21	51.30	41.78
Patient 9				
13	Conventional	81.40	67.92	51.71
	Proposed	91.22	76.20	63.37
Patient 10				
14	Conventional	86.66	70.10	53.97
	Proposed	95.60	79.10	67.78
Patient 11				
14	Conventional	93.75	70.75	60.51
	Proposed	95.66	72.79	65.45
Patient 12				
11	Conventional	90.12	80.60	61.67
	Proposed	97.30	87.89	80.22
Patient 13				
6	Conventional	84.32	61.43	45.89
	Proposed	97.99	72.68	62.57
Patient 14				
13	Conventional	75.07	51.83	44.00
	Proposed	72.37	50.12	41.84
Patient 15				
14	Conventional	75.01	59.78	49.55
	Proposed	74.62	59.49	50.00
Patient 16				
15	Conventional	79.60	62.28	52.99
	Proposed	82.27	64.54	54.92

Table 2.4. Average, minimum, and maximum coverage differences between conventional and OAR grouping plans observed over all patients ($n = 16$).

		Average	Minimum	Maximum
PTV OPT	D_{95} (%)	3.98 ± 4.97	-2.78	15.87
PTV	D_{95} (%)	3.87 ± 4.29	-1.78	13.07
	D_{100} (%)	6.47 ± 7.16	-2.29	20.19

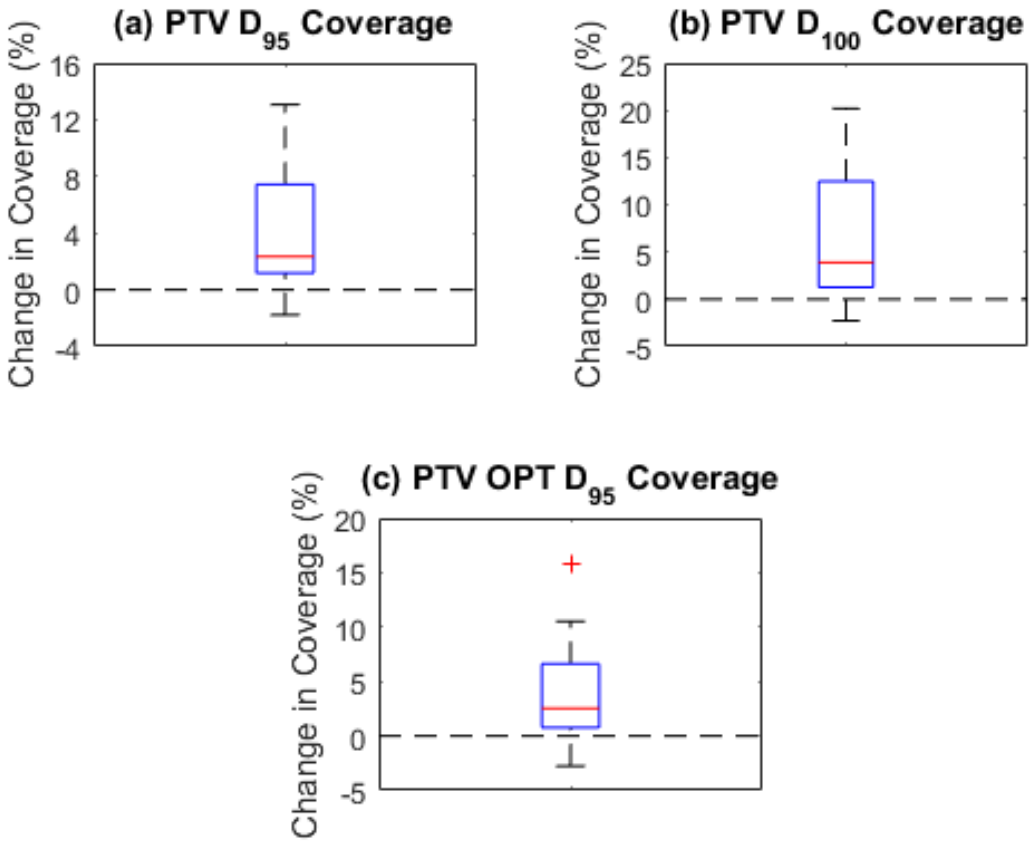

Figure 2.3. Average change from conventional to OAR grouping plans for each patient plotted for (a) PTV D_{95} coverage, (b) PTV D_{100} coverage, and (c) PTV OPT D_{95} coverage.

treatment planning process is easily understood. Reducing the number of OARs involved in planning restricts the set of weighting parameters that must be tuned iteratively to create an acceptable plan. Second and more important, the observed improvements in target coverage over a majority of fractions demonstrate the benefits of the method. The creation of the combined OAR structure makes the objective function created at simulation somewhat insensitive to significant changes in a patient's anatomy from fraction to fraction. In the conventional plan, the weighting parameters selected at simulation may not accurately reflect the patient's anatomy at a later treatment fraction. As a result, surrounding OARs may be overdosed and target coverage will suffer. In the proposed plan, the combined OAR structure lessens the impact of large differences in anatomy between simulation and treatment. Although the individual OARs may change position relative to the PTV, the position of the composite structure relative to the PTV changes less dramatically. As a result, the weighting parameters assigned at simulation more accurately reflect the present anatomy and target coverage is improved compared to the conventional case.

A number of general trends were observed for individual patients as well as the cohort as a whole. Regarding the weighting selections made in conventional plans, two general situations are relevant. For plans in which an OAR was favored or disregarded by the conventional objective function through higher or lower weighting respectively, OAR grouping plans generally performed better in terms of coverage. Included as an illustrative example, Figure 2.4 includes one slice from simulation (Figure 2.4a) and the corresponding slice at treatment fraction 9 in the conventional and proposed plans for one patient (Figure 2.4b-c). At simulation, the primary OARs in the conventional case were assigned weighting parameters based largely upon proximity to the PTV. Of the four primary OARs, the large bowel received the second lowest weighting due to the relatively small fraction of the structure located near the PTV, which is observable in Figure 2.4a. The situation at fraction 9 is considerably different, as the large bowel now represents a significant volume in close prox-

imity to the PTV. In the conventional case, the weighting assigned at simulation does not reflect the actual anatomy, and the large bowel is overdosed upon plan re-optimization as a result. Normalizing the delivered dose such that the large bowel receives the prescribed limit of 45 Gy to 0.5 cc, D_{100} coverage of the PTV is only 14.29% in the conventional case (Figure 2.4b). In contrast, by applying the OAR grouping method, the optimization function is made somewhat insensitive to these inter-fractional changes in anatomy and the resulting coverage demonstrates considerable improvement. As seen in Figure 2.4c, the isodose lines in the OAR grouping plan are moderately more conformal to the PTV compared to those in the conventional plan, resulting in PTV D_{100} coverage of 74.93% without violating any OAR dose constraints. The dose-volume histogram (DVH) presented in Figure 2.5 demonstrates improved PTV coverage and OAR doses that are generally comparable between the OAR grouping and conventional plans, save for the small bowel. The dose to the small bowel in this case, despite being higher in the OAR grouping plan, is still well below the volumetric dose limit assigned to the small bowel.

Now, the second scenario of note: for plans in which the OAR grouping method was approximated by the conventional objective function through the assignment of equal or similar weighting to each of the primary OARs, coverage for the conventional and proposed plans was generally comparable. It should be noted that the present study was limited in scope to cases in which in the volumetric dose constraints assigned to each of the four primary OARs were the same. Use of the OAR grouping method in cases where this condition is not maintained should be investigated further. However, it is anticipated that the OAR grouping method is valid as long as there does not exist any drastic difference between these critical structures.

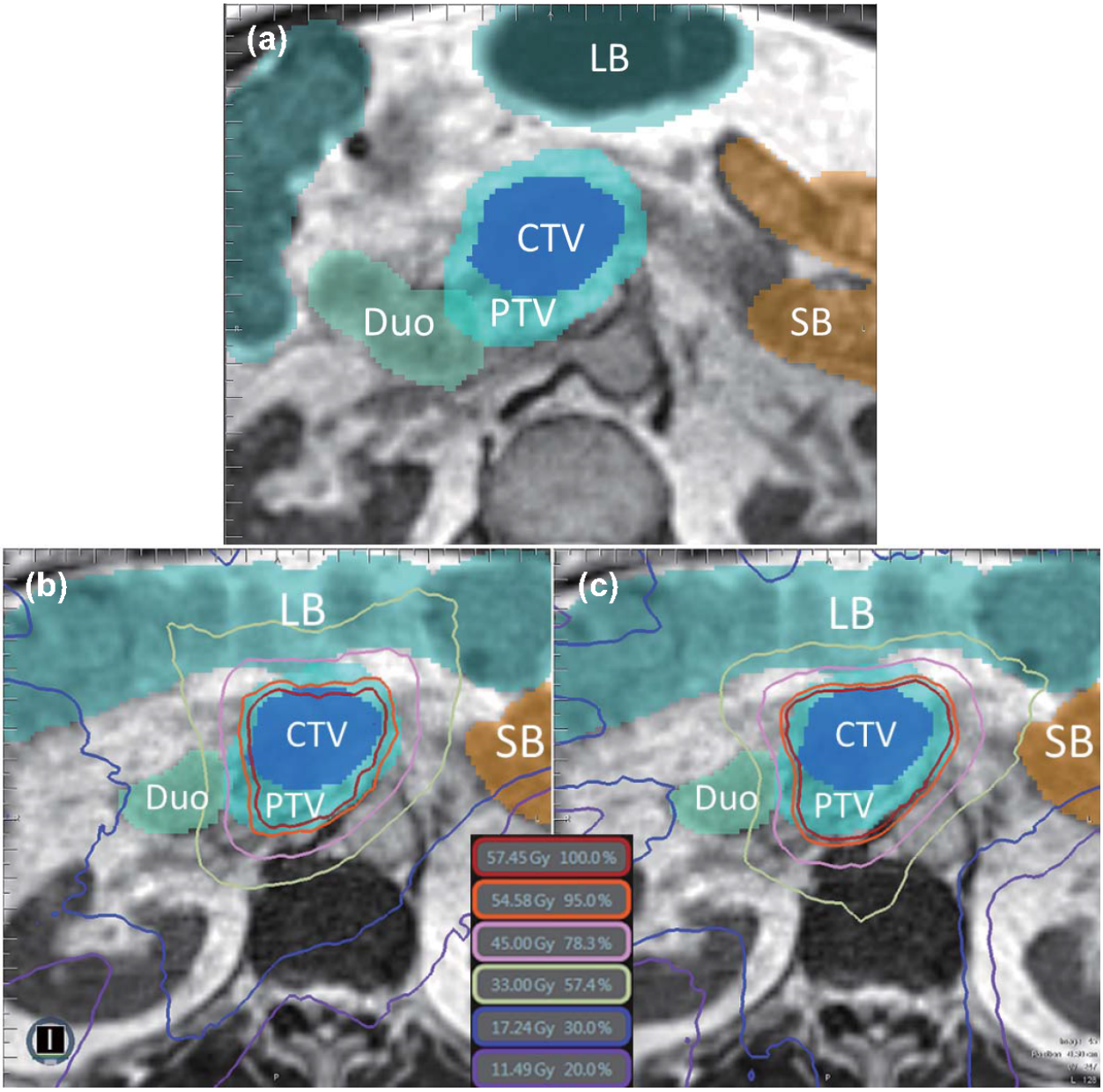

Figure 2.4. OARs and target structures in one slice at simulation (a) and treatment fraction 9 (b-c) for one patient. The large bowel demonstrates a large change in volume and proximity to the PTV from simulation to treatment. Isodose lines are displayed for the conventional plan (b) and the OAR grouping plan (c). The OAR grouping plan demonstrates improved D_{100} and D_{95} coverage over the PTV, as well as isodose lines that are moderately more conformal to the PTV compared to those in the conventional plan.

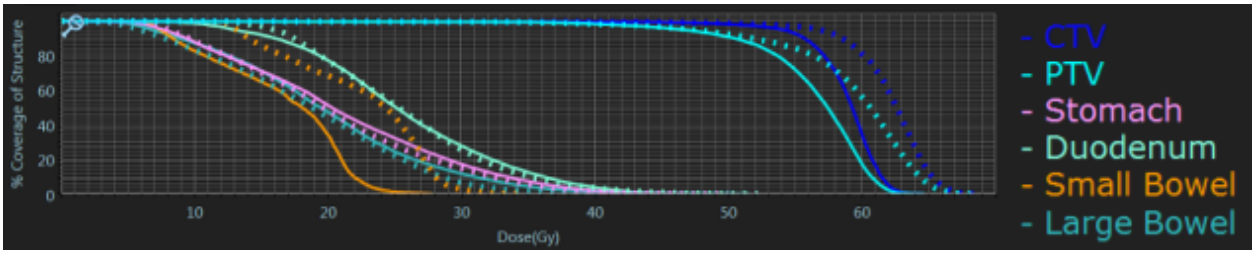

Figure 2.5. DVH for one patient comparing the conventional (solid lines) and OAR grouping (dashed lines) plans. Improved coverage of the PTV can be observed and OAR doses are generally comparable.

2.5 Conclusions

In this study, the OAR grouping method has been proposed as a means to simplify the daily adaptive treatment planning workflow and improve target coverage in adapted fractions. Characterizing the performance of OAR grouping plans reveals two scenarios of note. When the conventional simulation plan favored or disregarded an OAR through the assignment of distinct weighting parameters, OAR grouping plans generally demonstrated improved coverage compared to the conventional plan due to the decreased sensitivity of the OAR grouping objective function to inter-fraction anatomical variations. When the OAR grouping method was approximated by equal weighting in the conventional plan, coverage was generally comparable. In any case, the construction of the initial objective function at simulation is simplified by combining the four primary OARs in a pancreatic cancer case. This simplification comes along with an improvement in target coverage over a majority of fractions when comparing OAR grouping plans to conventional, clinically delivered plans.

References

[1] Fischer-Valuck BW, Henke L, Green O, et al. Two-and-a-half-year clinical experience with the world's first magnetic resonance image guided radiation therapy system. *Adv Radiat Oncol*. 2017;2(3):485–493. doi:10.1016/j.adro.2017.05.006

[2] Chen AM, Cao M, Hsu S, et al. Magnetic resonance imaging guided reirradiation of recurrent and second primary head and neck cancer. *Adv Radiat Oncol*. 2017;2(2):167–175. doi:10.1016/j.adro.2017.02.002

[3] Evans P. Anatomical imaging for radiotherapy. *Phys Med Biol*. 2008;53(12):R151–R191. doi:10.1088/0031-9155/53/12/R01

[4] McDermott L, Wendling M, Sonke J, van Herk M, Mijnheer B. Anatomy changes in radiotherapy detected using portal imaging. *Radiother Oncol*. 2006;79(2):211–217. doi:10.1016/j.radonc.2006.04.003

[5] van Herk M. Errors and margins in radiotherapy. *Semin Radiat Oncol*. 2004;14(1):52–64. doi:10.1053/j.semradonc.2003.10.003

[6] Hector C, Webb S, Evans P. The dosimetric consequences of inter-fractional patient movement on conventional and intensity-modulated breast radiotherapy treatments. *Radiother Oncol*. 2000;54(1):57–64. doi:10.1016/S0167-8140(99)00167-X

[7] Schwartz DL, Garden AS, Thomas J, et al. Adaptive radiotherapy for head-and-neck cancer: initial clinical outcomes from a prospective trial. *Int J Radiat Oncol Biol Phys*. 2012;83(3):986–993. doi:10.1016/j.ijrobp.2011.08.017

[8] Ahunbay EE, Peng C, Chen GP, et al. An on-line replanning scheme for interfractional variations. *Med Phys*. 2008;35(8):3607–3615. doi:10.1118/1.2952443

[9] Barker JL, Garden AS, Ang KK, et al. Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system. *Int J Radiat Oncol Biol Phys*. 2004;59(4):960–970. doi:10.1016/j.ijrobp.2003.12.024

[10] Spoelstra F, Pantarotto J, van Sönsen de Koste J, Slotman B, Senan S. Role of adaptive radiotherapy during concomitant chemoradiotherapy for lung cancer: Analysis of data from a prospective clinical trial. *Int J Radiat Oncol Biol Phys*. 2009;75(4):1092–1097. doi:10.1016/j.ijrobp.2008.12.027

[11] Surucu M, Shah K, Roeske J, Choi M, Small W, Emami B. Adaptive radiotherapy for head and neck cancer. *Technol Cancer Res Treat*. 2017;16(2):218–223. doi:10.1177/1533034616662165

[12] Nakamura M, Shibuya K, Nakamura A, et al. Interfractional dose variations in intensity-modulated radiotherapy with breath-hold for pancreatic cancer. *Int J Radiat Oncol Biol Phys*. 2012;82(5):1619–1626. doi:10.1016/j.ijrobp.2011.01.050

[13] Song WY, Kamath S, Ozawa S, et al. A dose comparison study between xvi and obi cbct systems. *Med Phys*. 2008;35(2):480–486. doi:10.1118/1.2825619

[14] Mutic S, Dempsey JF. The ViewRay system: Magnetic resonance–guided and controlled radiotherapy. *Semin Radiat Oncol*. 2014;24(3):196–199. doi: 10.1016/j.semradonc.2014.02.008

[15] Wooten HO, Green O, Yang M, et al. Quality of intensity modulated radiation therapy treatment plans using a ^{60} magnetic resonance image guidance radiation therapy system. *Int J Radiat Oncol Biol Phys*. 2015;92(4):771–778. doi:10.1016/j.ijrobp.2015.02.057

[16] Craft D, Hong T, Shih H, Bortfeld T. Improved planning time and plan quality through multicriteria optimization for intensity-modulated radiotherapy. *Int J Radiat Oncol Biol Phys*. 2015;82(1):e83–e90. doi:10.1016/j.ijrobp.2010.12.007

[17] Barth N. An inverse problem in radiation therapy. *Int J Radiat Oncol Biol Phys*. 1990; 18(2):425–431. doi:10.1016/0360-3016(90)90111-v

[18] Goitein M. The inverse problem. *Int J Radiat Oncol Biol Phys*. 1990;18:489–491

[19] Xing L, Chen G. Iterative methods for inverse treatment planning. *Phys Med Biol*. 1996;41(10):2107–2123. doi:10.1088/0031-9155/41/10/018

[20] Oelfke U, Bortfeld T. Inverse planning for photon and proton beams. *Med Dosim*. 2001; 26(2):113–124. doi:10.1016/s0958-3947(01)00057-7

[21] Hamacher H, Küfer KH. Inverse radiation therapy planning — a multiple objective optimization approach. *Discrete Appl Math*. 2002;118(2):145–161. doi:10.1016/S0166-218X(01)00261-X

[22] Orton C, Bortfeld T, Niermierko A, Unkelback J. The role of medical physicists and the aapm in the development of treatment planning and optimization. *Med Phys*. 2008; 35(11):4911–4923. doi:10.1118/1.2990777

[23] Liu H, Dong P, Xing L. Using measurable dosimetric quantities to characterize the inter-structural tradeoff in inverse planning. *Phys Med Biol*. 2017;62(16):6804. doi: 10.1088/1361-6560/aa6fcb

[24] Ezzell GA, Galvin JM, Low D, et al. Guidance document on delivery, treatment planning, and clinical implementation of IMRT: report of the IMRT subcommittee of the AAPM radiation therapy committee. *Med Phys*. 2003;30(8):2089–2115. doi: 10.1118/1.1591194

[25] Xing L, Li J, Donaldson S, Le Q, Boyer A. Optimization of importance factors in inverse planning. *Phys Med Biol*. 1999;44(10):2525–2536. doi:10.1088/0031-9155/44/10/311

[26] Bortfeld T. IMRT: A review and preview. *Phys Med Biol*. 2006;51(13):R363–R379. doi:10.1088/0031-9155/51/13/R21

[27] Wu Q, Djajaputra D, Wu Y, Zhou J, Liu H, Mohan R. Intensity-modulated radiotherapy optimization with gEUD-guided dose-volume objectives. *Phys Med Biol*. 2003; 48(3):279–291. doi:10.1088/0031-9155/48/3/301

[28] Prior P, Chen X, Botros M, et al. MRI-based IMRT planning for MR-linac: comparison between CT- and MRI-based plans for pancreatic and prostate cancers. *Phys Med Biol*. 2016;61(10):3819. doi:10.1088/0031-9155/61/10/3819

[29] Heerkens H, Hall W, Li X, et al. Recommendations for MRI-based contouring of gross tumor volume and organs at risk for radiation therapy of pancreatic cancer. *Pract Radiat Oncol*. 2017;7(2):126–136. doi:10.1016/j.prro.2016.10.006

Chapter 3

MRI super resolution reconstruction for MRI-guided adaptive radiotherapy using cascaded deep learning: In the presence of limited training data and unknown translation model

3.1 Introduction

Magnetic resonance imaging-guided radiation therapy (MR-IGRT) exploits MRI's superior soft tissue contrast and real-time tracking and delineation of the tumor and organs at risk (OARs) during the course of treatment.[1,2] Such unique features enabled online adaptive radiation therapy (ART).[3,4] In online ART, the treatment plan is re-optimized according to anatomical or positional changes in OARs on a daily basis, significantly improving the accuracy of radiation dose delivery and a substantial reduction in the irradiated volume of normal tissue. In this way, online ART facilitates local dose escalation and the reduction of normal tissue toxicities through real-time, risk-adaptive, and high precision radiotherapy.[5–8]

Despite the superior features of MRI compared to other image guiding modalities (e.g. CT or cone beam CT), it faces the persistent problem of limited spatial resolution.[9–12] High spatial resolution in MRI comes at the expense of longer scanning time, reduced field of view (FOV), and reduced signal to noise ratio (SNR).[13,14] In MR-IGRT, the resolution of MRI fundamentally determines the setup uncertainty, target delineation, and tracking accuracy. Therefore, exploring methods to minimize the scanning time while maximizing the spatial resolution of MRI is an important issue. Reducing the scan time while maintaining sufficient spatial resolution is crucial during breath-hold MRI scans for patients with moving targets (e.g. lung or liver). The patient must maintain breath-hold at the peak inhale/exhale phase for the duration of the MRI scanning period (17-24 s), which is often not feasible. Another example is 4D-MRI, where there is a growing interest to replace real-time cine 2D planar imaging for online tumor tracking.[15,16] 4D-MRI uses fast 3D cine MRI sequences with parallel imaging and echo sharing techniques to minimize the scanning and reconstruction time, and increase temporal resolution. However, limitations in current hardware and software make it challenging to acquire high-resolution scans, and such characteristics of 4D-MRI are the critical factors limiting its implementation during MR-IGRT.

To resolve such a persistent problem in MRI, the image post-processing technique known as super-resolution (SR) may be utilized to significantly improve the spatial resolution of MRI without changing hardware or scanning components.[2,17–24] The aim of SR reconstruction is to reconstruct high-resolution (HR) images from a single or a set of low-resolution (LR) images to improve the visibility of, or recover, image details. The first application of SR in MRI was proposed by Peled et al., where multiples of spatially shifted, single-shot, diffusion-weighted brain images were fused to generate a new image with improved resolution and finer detail.[20] Since then, various advanced SR techniques established in MRI have offered the possibility to efficiently improve the image resolution and increase the diagnostic potential.[2,18,21–25]

There are generally three methods to achieve image SR in MRI: 1) interpolation-,[9,26,27] 2) reconstruction-,[28–30] and 3) machine learning-based.[31–35] Interpolation-based techniques assume that points/regions in an LR image can be expanded into corresponding points/regions in the SR reconstruction using polynomial or interpolation functions with some smoothness priors,[9,26,27] which is not valid in inhomogeneous regions.[25] Moreover, the actual LR sampled points represent a non-ideal sampling where the sampled points represent the intermediate value of the underlying HR points that exist within the LR points. Hence, SR through interpolation results in a blurred version of the corresponding HR reference images.

Reconstruction approaches, which are based upon image down sampling and a degradation model, solve an ill-posed inverse recovery problem from LR to HR images.[28–30] The reconstruction-based SR methods solve an optimization problem incorporating two terms: the fidelity term, which penalizes the difference between a degraded SR image and an observed LR image, and the regularization term, which promotes sparsity and inherent characteristics of recovering the SR signal (e.g. edge gradient). However, the performance of these techniques becomes suboptimal especially in the high frequency region when the input data becomes too sparse or the model becomes even slightly inaccurate.[25] Such drawbacks limit the effectiveness of these methods to small magnification factors that are less than 4.

Machine learning techniques, particularly deep learning (DL)-based SR approaches, have recently attracted considerable attention for achieving state-of-the-art performance in SR for natural images.[33,34] In contrast to the former approaches, the DL-based method does not require the construction of a transformation model, but instead learns the direct mapping based on information from previously scanned data sets. Among these approaches, the convolutional neural network (CNN) is popular on account of its simple network structure and high accuracy.[35–38]

However, conventional DL-based methods have three major limitations. First, since MRIs have relatively lower image quality and more complex tissue structures than natural images, it is more challenging to restore high frequency information. The optimization of the conventional method employs the minimization of a pixel-wise difference (e.g. mean squared error between SR images and ground truth HR images), that is often limited when attempting to capture high texture detail. The second major limitation is the availability of sufficient training data, which in this case are perfectly matched pairs of LR and HR images. Although HR MRIs from patients previously scanned in the clinic are abundant, the corresponding LR images are not available. Current approaches in this case include preparing the training data set from abundant HR images utilizing a simple translation model such as bicubic interpolation or k-space down sampling.[35–38] However, these approaches do not reflect real LR scans that are directly generated from MRI scanners, as properties of the signal (e.g. sampling bandwidth or noise) are often not feasible to model with a simple down-sampling method. The third limitation is the performance of SR reconstruction in the presence of noise. In general, the SR network cannot distinguish noise from useful features and hence the noise is amplified in the generated HR images, degrading the resulting image quality. Intuitively, this issue can be resolved by denoising the LR images prior to feeding them into the network. However, conventional image denoising techniques (e.g. non-local means filter (NLM)) can be time consuming, limiting their use in a real-time imaging framework.

In this study, we present a robust SR MRI reconstruction framework using the novel cascaded DL framework that resolves such limitations in an efficient and practical manner. Our DL framework was comprised of three stages: 1) we used a CNN-based denoising autoencoder (DAE) at the front end of the SR network as the image processing unit. The DAE was trained with pairs of noisy and denoised LR MRIs from the author-selected image denoising model. 2) The denoised LR MRIs were paired with HR MRIs to train a CNN-based down-sampling network (DSN) to derive an unknown translation model between the LR and HR images. The DSN was designed to have a small and simple network structure that can be sufficiently trained even with a limited amount of paired LR/HR images acquired from physical phantoms and/or volunteers. The trained DSN was then applied to abundant HR MRI scans to generate perfectly matched pairs of LR and HR images. 3) Finally, we developed a deep generative adversarial network (GAN)-based MRI SR network that provides detail preserving HR MRI for the final output.

3.2 Materials and Methods

3.2.1 Model

In this study, the translation model between LR and HR MRIs was formulated as the following equation:

$$\widehat{LR_k} = T_s\left(HR_k\right) + \eta_k, \tag{3.1}$$

where HR_k are the desired HR MRIs to be recovered from the given set of observed LR MRIs $\left\{\widehat{LR_k}\right\}_{k=1,\ldots,K}$, T_s describes the down-sampling operator that decreases the resolution by a factor of s, and η_k is the additive noise. For under-sampled k-space measurements, the system of Eq. 3.1 is underdetermined and hence the inversion process is ill-defined.

Therefore, the goal of the SR problem is to recover a HR image given its down-sampled version such that:

$$\min_{HR_k} E\left[HR_k, \widehat{LR_k}\right], \tag{3.2}$$

where $E[\cdot]$ is the energy function that represents the optimization objectives.

Reconstruction-based methods model this energy function as a convex optimization problem to find a plausible HR solution while balancing regularization and/or *a priori* terms simultaneously, as is the case in total variation reconstruction. On the other hand, DL-based methods construct the energy function on the basis of mean absolute error (MAE), which is the measure of data fidelity between generated SR images from the trained neural network and the corresponding HR images. The DL-based method uses a parametric CNN model to learn the non-linear mapping function that minimizes the Manhattan norm between SR images reconstructed from LR images $\widehat{LR_k}$ and the corresponding ground truth HR images HR_k. However, as mentioned earlier, there are two unknown variables: the down-sampling operator T_s and the noise η_k shown in Eq. 3.1, which are both ill-posed inverse problems. To design a robust SR model with the physical LR MRI scan as an input, it is crucial to estimate and subtract these unknown terms considering they directly relate to the quality of the training data.

To overcome this problem, we utilize a novel splitting technique we have named as cascaded DL. This algorithm splits the SR reconstruction process into three stages: 1) construction of an image denoising autoencoder (DAE) to subtract noise η_k from a noisy LR image input, 2) construction of a down-sampling network (DSN) to model the unknown down-sampling operator T_s from a subset of paired denoised LR and HR MRI data, and 3) construction of a SR generative (SRG) model using numerous pairs of HR and LR MRIs generated from the estimated T_s. Mathematically, the proposed cascaded DL model can be

formulated as:

$$\text{Step 1.} \qquad \min_{\theta} \left\{ \sum_{q} \left\| DAE\left(\widehat{LR_q}; \theta\right) - \mu_q \right\|_2^2 \right\}, \tag{3.3}$$

where $DAE\left(\widehat{LR_q}\right)$ is the denoising autoencoder network with input LR MRI data, μ_q are denoised LR images from $\widehat{LR_q}$ using the author-selected NLM filter,[39] $DAE\left(\widehat{LR_q}; \theta\right)$ is a CNN parameterized by θ, $q = 1, ..., Q$ is the subset of training data from total number of training sets $k = 1, ..., K$ that had physical scans of LR MRI data acquired from physical phantoms and/or volunteers,

$$\text{Step 2.} \qquad \min_{\rho} \left\{ \sum_{p} \left\| DSN\left(HR_p; \rho\right) - DAE\left(\widehat{LR_p}\right) \right\|_1 \right\}, \tag{3.4}$$

where $DSN\left(HR_p\right)$ is the down-sampling network with input HR MRI data that are equivalent to the first term of Eq. 3.1, $p = 1, ..., P$ is the subset of training data from the total number of training sets $k = 1, ..., K$ that had close pairs of LR and HR images acquired from physical phantoms and/or volunteers, and $DSN\left(HR_p; \rho\right)$ is a CNN parameterized by ρ, and finally,

$$\text{Step 3.} \qquad \min_{\vartheta} E\left[SRG\left(DSN\left(HR_k\right); \vartheta\right), HR_k\right], \tag{3.5}$$

where $SRG\left(DSN\left(HR_k\right)\right)$ is the SR generative model with input $DSN\left(HR_k\right)$, $E\left[\cdot\right]$ is the GAN energy loss function that will be illustrated in Section 3.2.5, and $SRG\left(DSN\left(HR_k\right); \vartheta\right)$ is a GAN parameterized by ϑ.

3.2.2 Overview of framework

Figure 3.1 illustrates the proposed MRI SR reconstruction framework. In general, the deep learning process is divided into two parts: training and inferencing. As shown in Eq. 3.3, the first step in training is to develop a CNN-based DAE as the front end preprocessing unit. The motivation for using the DL-based DAE is that conventional image denoising techniques can be time consuming and computationally expensive to implement, and thus it may not be feasible to fit them into a real-time imaging framework. Since the inferencing step of the trained network requires only a simple matrix multiplication process, the computation time required for applying a sophisticated denoising algorithm can be reduced significantly. In this study, the DAE was trained with pairs of noisy and denoised LR MRIs that were preprocessed using the non-local means filter (NLM), which is a popular and efficient image processing technique for denoising MRIs.[39]

After the DAE was trained from physically scanned LR MRI data in Step 1, we selected closely paired LR and HR MRIs that were acquired in a phantom and/or volunteers, and trained the down-sampler. As mentioned earlier, manual down-sampling approaches that derive LR images from clinical HR MRIs are not capable of generating the realistic LR images that are directly reconstructed from the MRI system. Therefore, we used a CNN-based DSN characterized by a relatively small size and simple structure to learn the actual down-sampling process using a limited amount of paired LR and HR images acquired from a phantom or volunteers. In this manner, the large set of training data required to sufficiently train the SRG model can be prepared by applying the trained DSN to abundant clinical HR MRI scans.

The third step in training was to train the SRG model. During the SRG model training, the network aims to learn the image prior for inversely mapping the LR image to the reference HR image. Specifically, an HR image was constructed iteratively to best explain the given data set of LR images by minimizing the differences between the given HR images and the

generated HR version of LR images fed into the network through the front end DAE. The SRG model was based on the high level GAN architecture proposed by Goodfellow et al.[40] and trained using image pairs generated by the DSN trained in Step 2. Finally, once the framework was trained according to the cascaded training model that we have defined, the inference system inferred SR MRIs based on the noisy LR input data.

3.2.3 Denoising autoencoder (DAE)

The network structure of the DAE used in this study is shown in Figure 3.2. The basic framework of the DAE consisted of an encoder that maps the noisy input image to some hidden representation and a decoder that maps this hidden representation back to the reconstructed version of the denoised input image. The parameters of the DAE were learned to minimize the construction error measured by the loss function defined in Eq. 3.3.

In this study, we implemented a CNN-based DAE using a cascade of convolutional filters paired with nonlinear activation functions. The network consisted of six convolutional layers (encoders) with 4×4 filters and six deconvolutional layers (decoders) with 4×4 filters. Each layer consisted of a single convolutional/deconvolutional filter with stride 2. The input images (64×64 pixels) were convolved with successive filters until feature reduction is performed to extract 1×512 features at the deepest layer. These features were processed with subsequent deconvolutional filters to recover the original dimensions of the input images. At the end of each convolutional and deconvolutional layer, we used leaky and standard rectified linear unit (ReLU) nonlinear activation functions of the form Leaky $\mathrm{ReLU}(x) = \max(0.2x, x)$ and $\mathrm{ReLU}(x) = \max(0, x)$ respectively to ensure that the output of a layer is a nonlinear representation of the input.[41]

The DAE was trained using 480 LR breath-hold images. We first utilized the NLM filter to obtain the noise-free LR images noted by μ in Eq. 3.3. Instead of directly training the CNN based on the noisy observations and the corresponding denoised LR images that is less

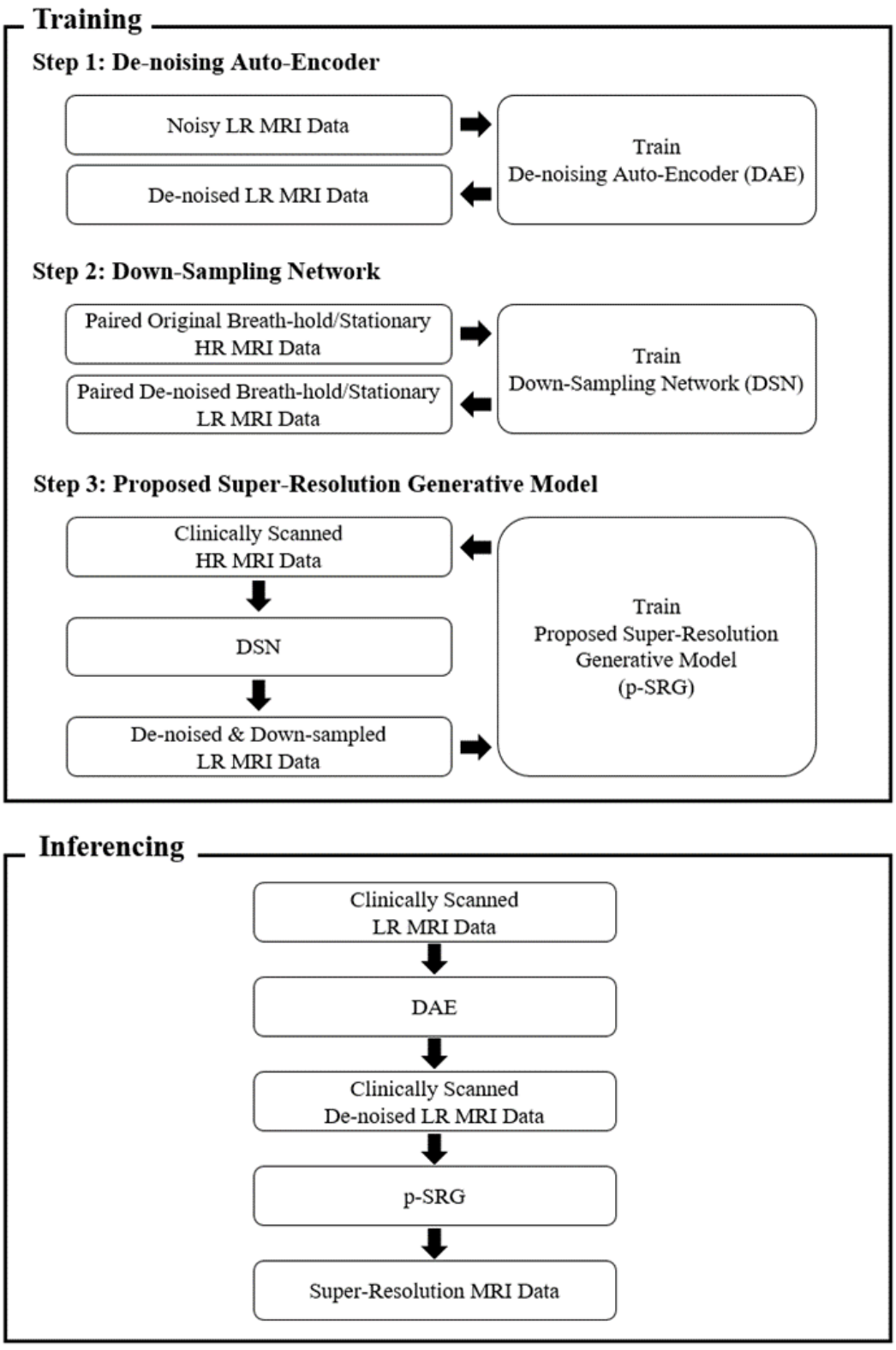

Figure 3.1. Proposed framework for SR MRI reconstruction using cascaded deep learning.

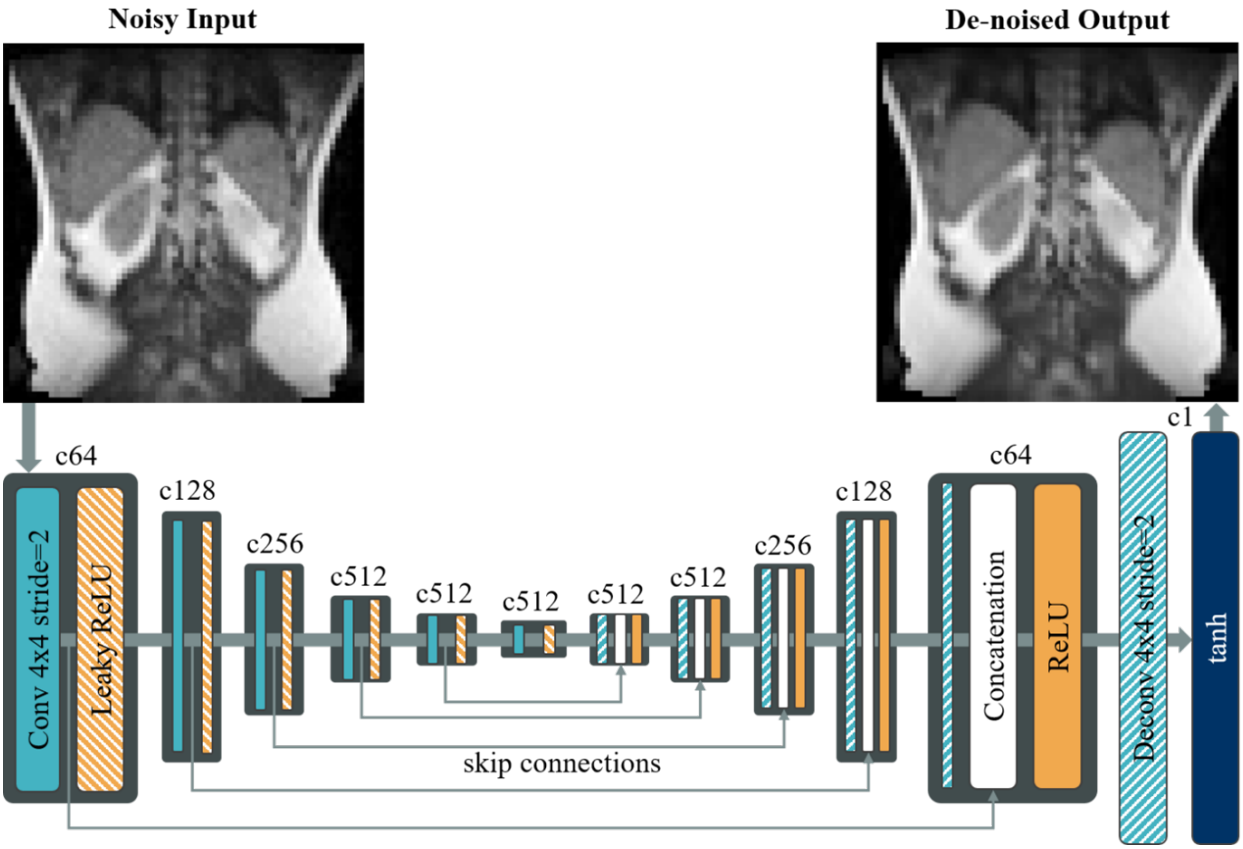

Figure 3.2. Denoising autoencoder (DAE) network architecture.

robust due to the insufficient amount of data, we applied the traditional DAE[42,43] training methodology by generating training samples through adding noise to the preprocessed images μ. The noise in an MRI is governed by a Rician distribution, and is nearly Gaussian if the SNR > 2.[44] Hence, we utilized Rician noise in the experiment. The noncentrality and scale parameters describing the Rician distribution were obtained by analyzing the noise removed by means of the NLM filter, which could be observed by subtracting NLM filtered images from the original, noisy images. Parameter optimization during training was performed using the stochastic gradient descent method (Adam Optimizer[45]) embedded in TensorFlow[46] with a leaning rate starting at 0.0002. The training was performed over 200 epochs using a GeForce GTX 1080 Ti GPU (NVIDIA, Santa Clara, CA, USA). Five-fold cross-validation was performed to test and demonstrate the robustness of the DAE.

3.2.4 Down-sampling network (DSN)

The main goal of the DSN was to inference LR MRIs from corresponding, abundant scans of HR clinical MRIs in order to maximize the number of training sets available to train the complex structured SRG network. Since the training data sets for the DSN require a perfect pairing of LR and HR MRIs that are limited in number, it is crucial to design the network to be as simple as possible while ensuring the resulting images reflect true LR scans acquired directly from the MRI scanners.

The overview of the structure of the proposed DSN is presented in Figure 3.3. The DSN was comprised of a simple, single encoder that consists of two down-sampling layers, two residual blocks, and an output layer. The down-sampling layer consisted of a single 3×3 convolutional filter of stride 2 followed by a ReLU activation. In each down-sampling layer, the dimensions of the output features were reduced by half of the original size. The residual block consisted of two 3×3 convolutional filters separated by a ReLU activation and followed by an element-wise sum operator that integrates the extracted features from the preceding

down-sampling layer.

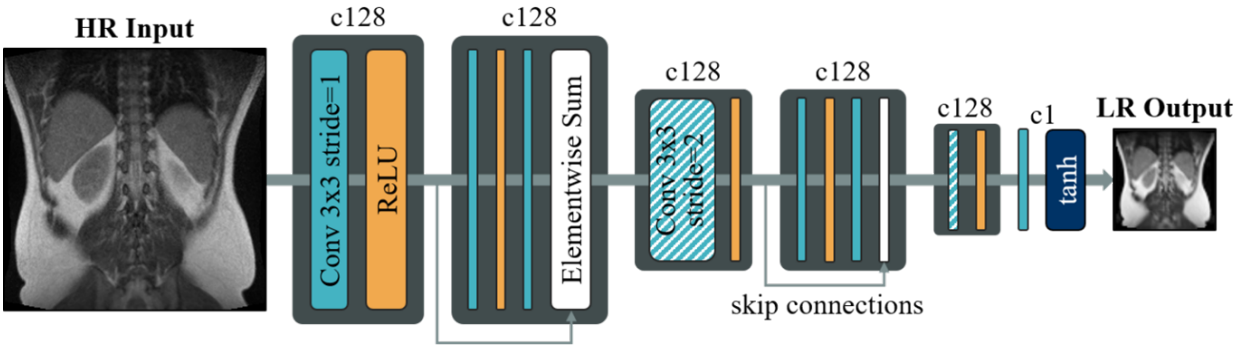

Figure 3.3. Down-sampling network (DSN) network architecture.

Training data for the DSN was collected from serial LR and HR scans of a phantom and volunteers acquired in a single breath-hold, resulting ultimately in 480 data pairs. Since the performance of the SRG model is highly dependent on the robustness of the DSN, we manually selected the data pairs to be used by rejecting or repetitively scanning the volunteers until the HR and LR scans were perfectly paired. Specifically, the patients were instructed to breath hold during the HR and LR MRI pair scanning while breathing motion was recorded with the Physiologic Monitoring Unit (PMU). The acquired respiratory signal from the PMU was then extracted from the console unit and the global minimum and maximum amplitudes within the normalized range of the PMU from 0 to 4095 were calculated. The paired scans were rejected and rescanned when the relative difference between the maximum and minimum (RDMM) signal values was larger than 5% of the mean value. The RDMM is defined here as

$$RDMM = \frac{PMU_{max} - PMU_{min}}{PMU_{mean}}, \tag{3.6}$$

where PMU_{max}, PMU_{min}, and PMU_{mean} are the maximum, minimum, and mean PMU values recorded during the breath-hold, respectively. The 5% threshold was chosen empirically based on the series of experiments in order to accommodate variation of the PMU signal

due to electrical noise and residual motion occurring despite the breath-hold. To ensure the accepted scans were anatomically compatible, HR and LR images were fused together by enlarging the LR images with bicubic interpolation and each slice was visually inspected by checking the overlay of the anatomical features including the diaphragm and liver.

The HR training data sets were cropped to a size of 192 × 192 pixels (1.5 × 1.5 mm per pixel) with the corresponding output LR size of 48 × 48 pixels (6.0 × 6.0 mm). Uncropped HR MRIs of 256 × 256 pixels (1.5 × 1.5 mm) and the corresponding LR images of 64 × 64 pixels (6.0 × 6.0 mm) were used in the testing and inferencing steps. The optimization of parameters during training was performed using the gradient descent method as described in Section 3.2.3 with the learning rate starting at 0.0002. The MAE L1 loss function defined in Eq. 3.4 was selected as the energy function to minimize at each epoch. The model was trained over 200 epochs using the GPU and five-fold cross-validation was performed to test and demonstrate the robustness of the DSN.

3.2.5 Super-resolution generative (SRG) model

Once the DSN was trained from the small amount of paired LR and HR images acquired from the phantom/volunteers, this trained network was utilized to generate LR MRIs from abundant clinical HR MRI data sets to form the full training set for building the SRG model. As mentioned earlier, the fundamental architecture of the SRG model used in this study is the GAN. Specifically, we modified the Generative Adversarial Network for Image Super Resolution (SRGAN) proposed by Ledig et al., who first introduced the GAN framework to the CNN-based image SR problem for photo-realistic, SR color images.[47]

The GAN is a framework for generative model estimation that consists of two models: a generative model G parametrized by θ_G that generates a sample in the generator's distribution p_g from a sample in a latent space p_z, and a discriminative model D parameterized by θ_D that determines whether a given sample is drawn from the data distribution p_{data} or

generated by G with a certain probability (p_g). The two models are trained simultaneously in an adversarial manner by solving the following minimax problem:

$$\min_{\theta_G} \max_{\theta_D} \{\mathbb{E}_{x \sim p_{data}} \log D_{\theta_D}(x) + \mathbb{E}_{z \sim p_z} \log\left(1 - D_{\theta_D}\left(G_{\theta_G}\left(z\right)\right)\right)\} \tag{3.7}$$

such that if an equilibrium is achieved during training, we have a global optimum for $p_g = p_{data}$, meaning that the discriminator D is not able to distinguish the differences between the generated data and the true samples.

In the context of SRGAN, the goal was to train a CNN-based generator G that will generate HR images based on corresponding LR images. Meanwhile, we utilized another CNN-based discriminator model D that serves as a binary classifier that helps to train the generator G such that the HR images generated by G resemble the images drawn from the true distribution of HR images.

The cost function of the generator G came from two sources. Mathematically, the cost function of G can be formulated as

$$\begin{aligned} l_{loss} &= l_{mae} + \lambda l_{adversarial} \\ &= \left\| G\left(I_{lr}\right) - I_{hr} \right\|_1 - \lambda \log\left(D\left(G\left(I_{lr}\right)\right)\right), \end{aligned} \tag{3.8}$$

where l refers to the loss function and $I_{lr} \sim p_{lr}$ and $I_{hr} \sim p_{hr}$ are LR and HR images from the training data set. In Eq. 3.8, the first loss term comes from the reconstruction error, or content loss, between the generated image and the ground truth image. The mean absolute error (MAE) was selected as the metric for content loss as it shows better performance

compared to the conventional L2 norm. The second loss term, denoting adversarial loss derived from the discriminator D, drives the generated image to pursue more high frequencies in order to more closely resemble the real images. Note that to provide a sufficient gradient we used $-\log\left(D\left(G\left(I_{lr}\right)\right)\right)$ instead of $\log\left(1-D\left(G\left(I_{lr}\right)\right)\right)$.[40] Based on Eqs. 3.7 and 3.8, the optimization problem for the SRG network can be formulated as:

$$\theta_G = \arg\min_{\theta_G} \mathbb{E}_{I_{lr},I_{hr}\sim p_{lr,hr}} \left\{ \left\| G_{\theta_G}\left(I_{lr}\right) - I_{hr} \right\|_1 - \lambda \log\left(D_{\theta_D}\left(G_{\theta_G}\left(I_{lr}\right)\right)\right) \right\} \tag{3.9}$$

$$\theta_D = \arg\min_{\theta_D} \mathbb{E}_{I_{lr},I_{hr}\sim p_{lr,hr}} \left\{ -\log\left(D_{\theta_D}\left(I_{hr}\right)\right) - \log\left(1 - D_{\theta_D}\left(G_{\theta_G}\left(I_{lr}\right)\right)\right) \right\}, \tag{3.10}$$

where the problem is solved by alternatingly updating the discriminator D and the generator G in each sub-problem.

The network structure of our GAN-based SRG model is illustrated in Figure 3.4. The generator consisted of a total of 12 layers including eight residual blocks, two up-sampling layers, and two output layers. Each of the residual blocks was comprised of two 3×3 convolutional filters separated by a ReLU activation with an element-wise sum operator attached at the end of the layer. The output block consisted of a 3×3 convolutional filter, ReLU activation, and a sub-pixel operator. The sub-pixel operator aggregated the feature maps from LR space and built the SR image in a single step.[48] The discriminator consisted of six 3×3 convolution layers and two 1D dense layers separated by a Leaky ReLU activation. In order to convert 2D residual features into 1D, the flattening operator was attached at the exit of the convolution layer.

In this study, we tested the validity of the proposed SR MRI in two different applications: 1) 3D LR breath-hold MRI that enables the reconstruction of axial SR MRIs from the set of axial LR MRIs acquired over a short breath-hold interval (<3 s) and 2) SR 4D-MRI that enhances the resolution of a LR 4D-MRI volume series acquired at rate of 2 volumes/s.

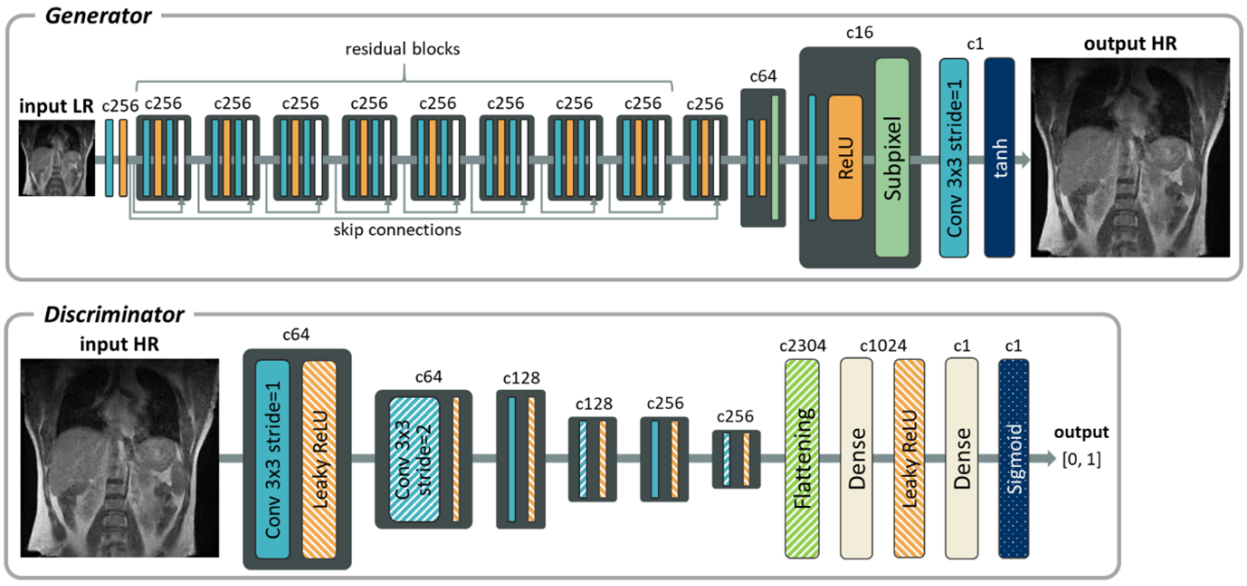

Figure 3.4. Network structures of the components of the super-resolution generative (SRG) model based on the generative adversarial network (GAN) framework.

In each of these applications, the front end DAE and DSN were trained using a common collection of 480 paired data sets from physical scans of a phantom/volunteers. The SRG network for each application was then trained with a total of 4,955 or 1,855 data pairs, respectively, a total that consists of the combination of 1) clinically available HR MRIs and the corresponding LR MRIs generated by the trained DSN, totaling to 4,475 or 1,375 images, respectively, and 2) the 480 paired data sets mentioned previously. Like in training, the LR images were cropped to a size of 48×48 pixels (6.0×6.0 mm per pixel) with a corresponding output HR size of 192×192 pixels (1.5×1.5 mm). The original input sizes of 64×64 pixels (6.0×6.0 mm) for LR images and 256×256 pixels (1.5×1.5 mm) for HR images were used for testing and inferencing.

3.2.6 Data acquisition

All data used in this study were acquired from the first commercially available MR-IGRT system (ViewRay Inc., Oakwood Village, Ohio, USA) that integrates a split-bore 0.35 T

whole body MRI system with a three head Co-60 radiation therapy delivery system. MRIs of the patients and volunteers were acquired using torso phased array receiver coils. Both breath-hold 3D and free-breathing 3D cine (or 4D) volumetric images in three orientations (transverse, sagittal, and coronal) were acquired using the 3D true fast imaging with steady-state precession (TrueFISP) sequence. The generalized auto calibrating partially parallel acquisition (GRAPPA) technique was used to accelerate image acquisition.

3.3 Results

3.3.1 Validation of denoising autoencoder (DAE)

Figure 3.5 displays the reference denoised LR MRI obtained by applying the NLM filter, the denoised output of the DAE compared with additional methods (Gaussian and median filtering), and the noisy input. It is noted here that all test data sets that we present in Section 3.3 are independent data sets that were excluded from the training data sets. As can be seen in Figure 3.5, the test output shows that the DAE can reproduce image quality that is comparable to the reference denoised image without major structural loss, whereas the results of Gaussian or median filtering are characterized especially by errors at tissue boundaries.

To evaluate the performance of the DAE more comprehensively, we used the peak signal-to-noise ratio (PSNR), structural similarity index (SSIM),[49] and normalized root mean square error (RMSE) indexes as quantitative measures. Values of the PSNR, SSIM, and normalized RMSE calculated for the DAE output over five validation folds and those of additional denoising filters are included in Table 3.1. The metrics were calculated using the NLM-denoised images as the reference for each of the 480 images pairs. Values of 32.4 dB (PSNR), 0.935 (SSIM), and 0.025 (RMSE) were observed for Gaussian filtering. For the DAE, the results consistently showed more than 96% structural agreement, along with mean

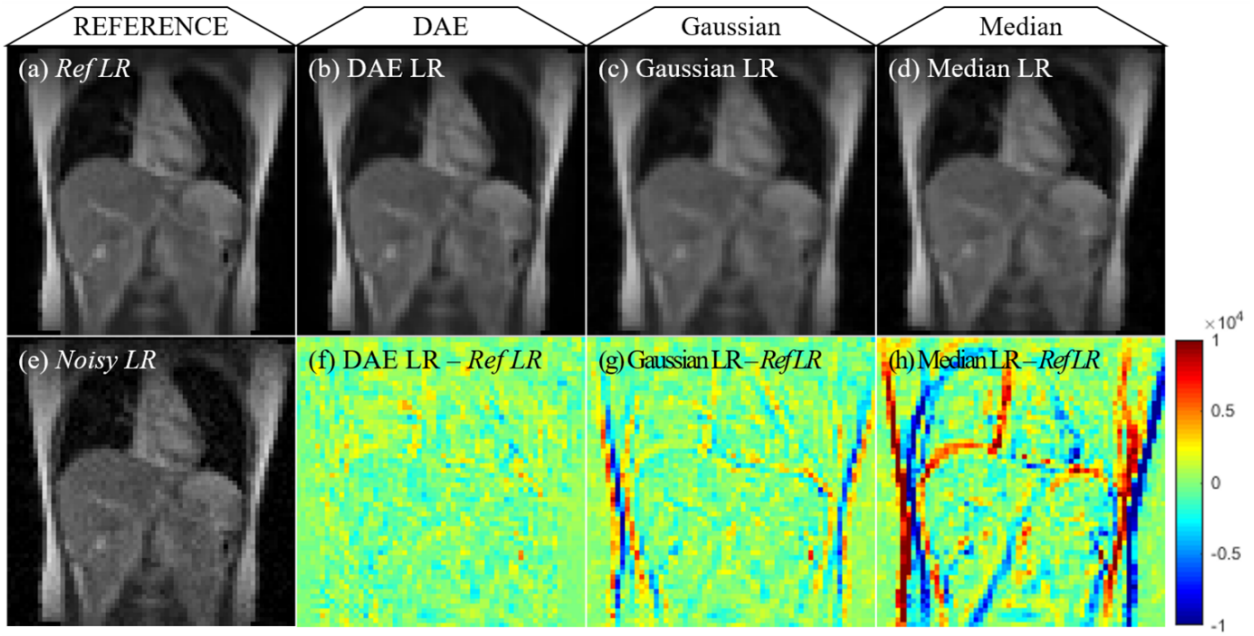

Figure 3.5. Denoised LR images derived from various denoising methods. Reference LR image denoised using the NLM filter (a) and noise-added LR image (e), DAE-filtered LR image (b) and difference map (f), Gaussian-filtered image (c) and difference map (g), median-filtered image (d) and difference map (h).

values of 36.5 dB for PSNR and 0.015 for RMSE.

3.3.2 Validation of down-sampling network (DSN)

Figure 3.6 shows the down-sampled result of the proposed DSN compared with the manual down-sampling methods (k-space down-sampling, bicubic interpolation, and nearest neighborhood sampling) used to generate LR and HR image pairs for training the SRG model. LR images from MRI pairs that were physically scanned during a single breath-hold were used as the ground truth reference. It is clear to see from the figure that the difference between the proposed DSN-generated LR and the ground truth reference LR images is much less than the difference observed for the manual down-sampling methods.

Box plots displayed in Figure 3.7 show the statistics over five-fold cross-validation for the calculated values of PSNR, SSIM, and normalized RMSE for the proposed DSN-generated

Table 3.1. Statistics for values of PSNR, SSIM, and RMSE computed over five validation folds using the proposed DAE, Gaussian filtering, and median filtering. Images denoised with the NLM filter were used as the reference.

	DAE			Gaussian			Median		
	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
Mean	36.5	0.963	0.015	32.4	0.935	0.025	25.8	0.849	0.054
Std	1.2	0.01	0.002	2.5	0.016	0.006	2.6	0.018	0.014
Min	34.3	0.918	0.009	28.6	0.848	0.012	22	0.8	0.019
Median	36.3	0.965	0.015	31.7	0.938	0.026	24.9	0.848	0.057
Max	41.1	0.979	0.019	38.6	0.957	0.037	34.3	0.885	0.079

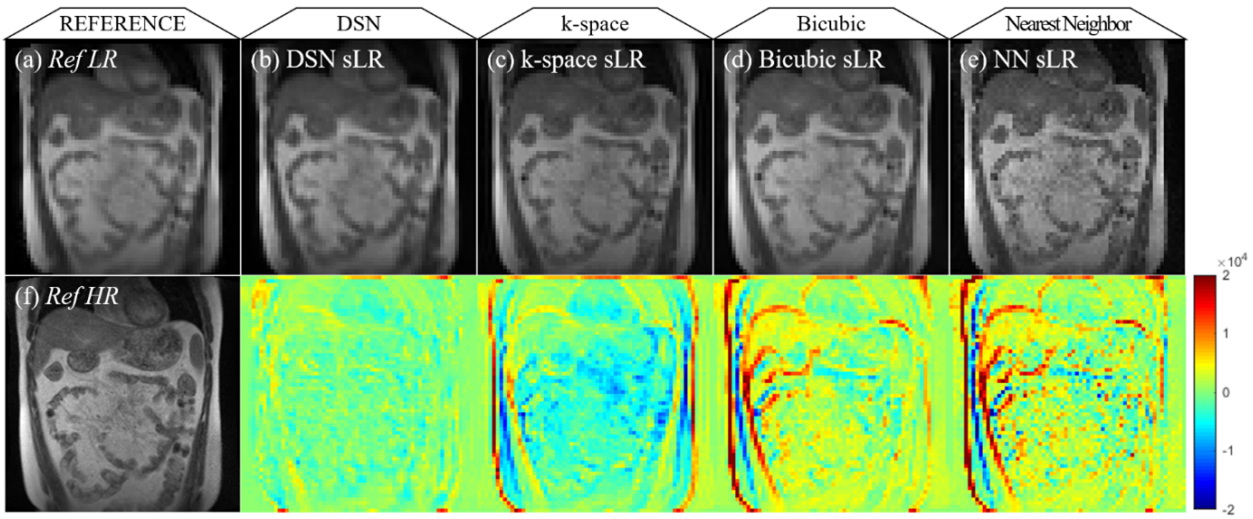

Figure 3.6. Down-sampled LR images derived from various down-sampling methods. Ground truth LR (a) and HR (f) images acquired from physical scans, DSN-generated synthetic LR (sLR) image (b) and difference map (g), k-space down-sampled image (c) and difference map (h), bicubic down-sampled image (d) and difference map (i), and nearest neighbor down-sampled image (e) and difference map (j).

LR images compared to the manually down-sampled LR images. Again, LR MRIs acquired from physical scans were used as the reference. These statistics are also presented in Table 3.2. The values of 24.4 dB (PSNR), 0.826 (SSIM), and 0.062 (RMSE) were observed for k-space down-sampling—the most accurate method amongst the conventional down-sampling methods. For the DSN, the results were consistently improved by 19.7%, 7.8%, and 74.7%, measuring 29.2 dB, 0.891, and 0.036, respectively.

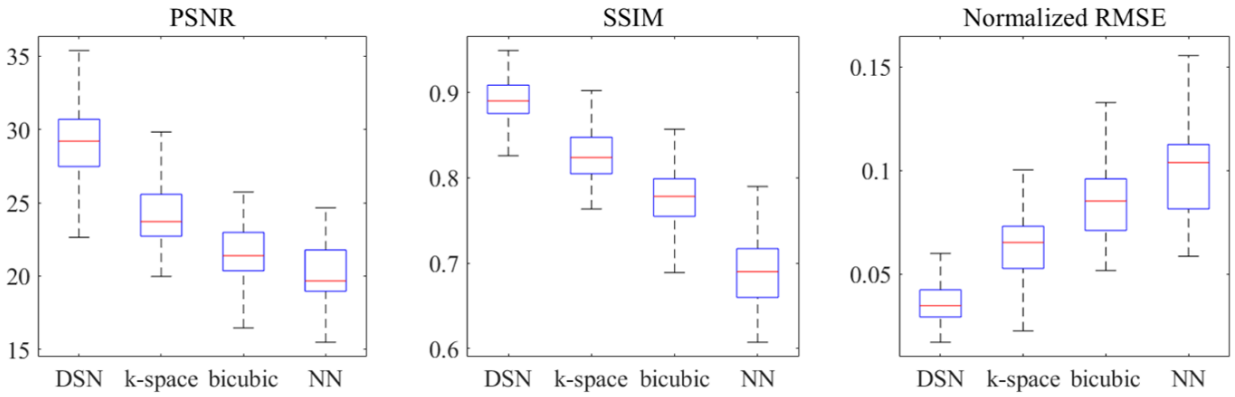

Figure 3.7. Five-fold cross-validation results of PSNR, SSIM, and normalized RMSE for the 480 outputs of four down-sampling approaches: proposed DSN, k-space down-sampling, bicubic down-sampling, and nearest neighbor (NN) down-sampling. LR images acquired from physical scans are used as the reference.

Table 3.2. Statistics of computed values of PSNR, SSIM, and RMSE for 480 outputs of the proposed DSN, k-space down-sampling, bicubic down-sampling, and nearest neighbor down-sampling.

	DSN			k-space			Bicubic			Nearest Neighbor		
	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE	PSNR	SSIM	RMSE
Mean	29.2	0.891	0.036	24.4	0.826	0.062	21.6	0.777	0.086	20.2	0.691	0.100
Std	2.2	0.027	0.009	2.5	0.029	0.015	2.1	0.036	0.021	1.9	0.038	0.021
Min	22.3	0.795	0.017	20.0	0.763	0.019	15.6	0.665	0.052	15.5	0.607	0.059
Median	29.2	0.890	0.035	23.7	0.824	0.065	21.4	0.778	0.085	19.7	0.690	0.104
Max	35.4	0.949	0.077	34.4	0.902	0.100	25.7	0.857	0.166	24.7	0.790	0.168

3.3.3 Validation of proposed super-resolution generative model (p-SRG)

Figure 3.8 shows the result of the proposed SRG (p-SRG) model trained with pairs of HR (256×256 pixels, 1.5×1.5 mm) images and LR (64×64 pixels, 6.0×6.0 mm) images generated by the DSN compared to the conventional SRG (c-SRG) model trained with pairs of HR images and LR images generated through k-space down-sampling. LR and HR MRIs acquired during a single breath-hold were used as the ground truth reference. As shown in the figure, the SR image generated by the c-SRG model is generally blurry, exhibiting large discrepancies compared to the reference ground truth HR image. On the other hand, the p-SRG does well to preserve the detail and contrast present in the ground truth image since the training dataset created through the DSN well approximates the down-sampling process of the physically scanned LR MRI data from the corresponding HR MRI data.

The box plots presented in Figure 3.9 display the values of PSNR, SSIM, and normalized RMSE calculated over five-fold cross-validation for the p-SRG model compared to the c-SRG model. HR images acquired from physical scans were used as the reference. Table 3.3 presents the detailed list of statistics. In the case of the c-SRG, values of 22.1 dB (PSNR), 0.625 (SSIM), and 0.080 (RMSE) were measured. For the p-SRG, the results were consistently improved by 21.7%, 12.6%, and 73.4%, measuring 26.9 dB, 0.704, and 0.046, respectively.

Table 3.3. Statistics of values of PSNR, SSIM, and RMSE for 480 results of the p-SRG and c-SRG models calculated over five validation folds.

	p-SRG			c-SRG		
	PNSR	SSIM	RMSE	PSNR	SSIM	RMSE
Mean	26.9	0.704	0.046	22.1	0.625	0.080
Std	2.1	0.078	0.011	2.0	0.093	0.016
Min	21.4	0.592	0.020	18.6	0.484	0.033
Median	26.8	0.684	0.046	21.7	0.605	0.083
Max	34.0	0.892	0.085	29.6	0.853	0.117

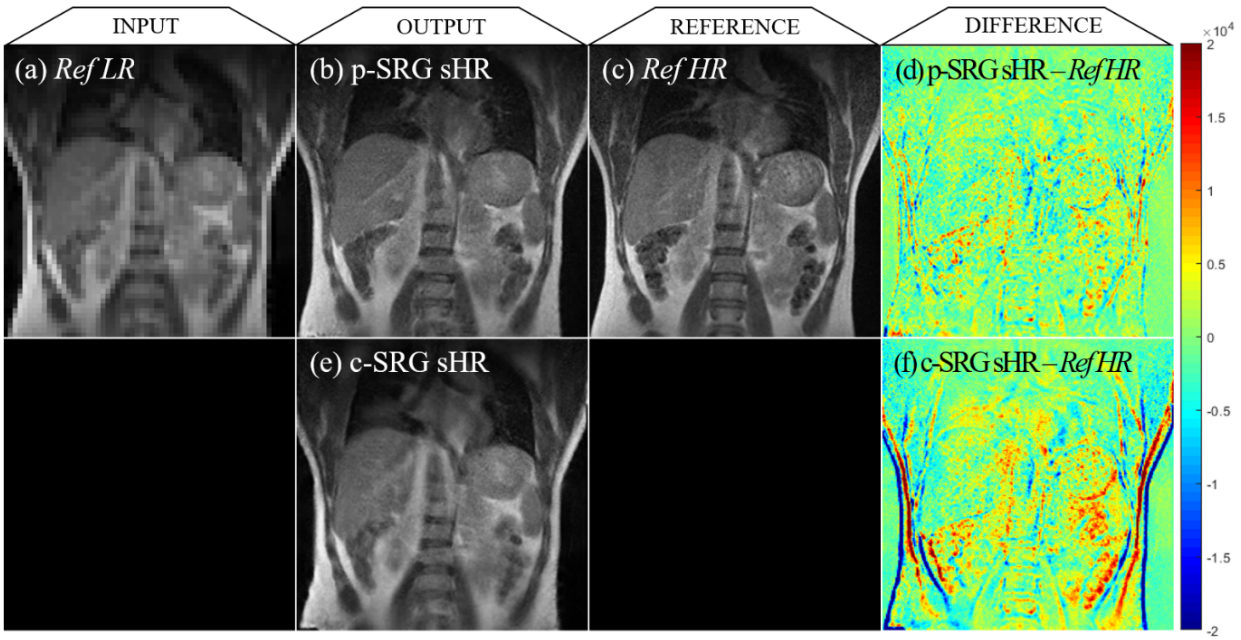

Figure 3.8. SR reconstruction results from the proposed SRG (p-SRG) and k-space down-sampling-based conventional SRG (c-SRG) models. Input LR image (a) (64×64 pixels, 6.0×6.0 mm) from a physical scan, SR output of the p-SRG (b) and c-SRG (e) models, reference HR image (c) (256×256 pixels, 1.5×1.5 mm per pixel) from a physical scan, and difference maps between the p-SRG (d) or c-SRG (f) output and the reference HR image.

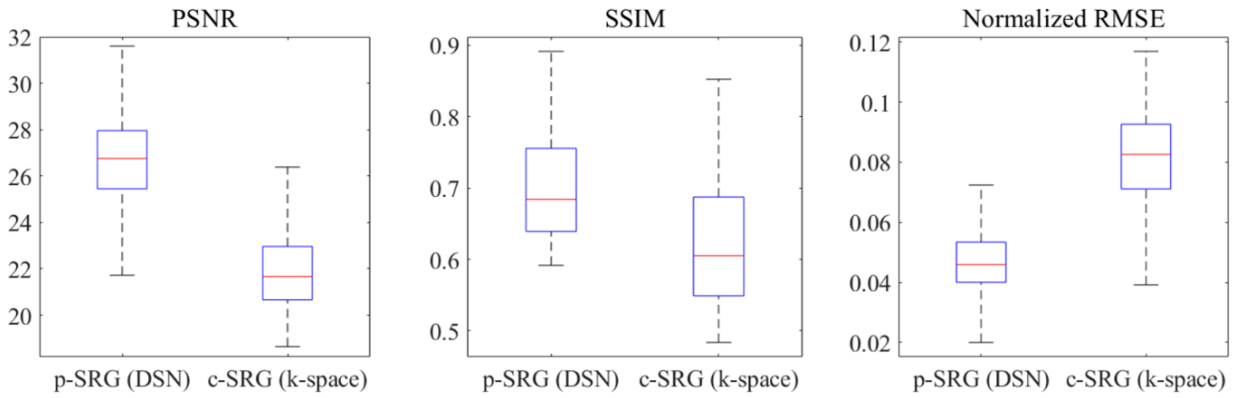

Figure 3.9. Five-fold cross-validation results of PSNR, SSIM, and normalized RMSE for the 480 outputs of the DSN-based p-SRG model and the k-space down-sampling-based c-SRG model. HR images from physical scans were used as the reference.

3.3.4 SR reconstruction of 3D LR breath-hold MRI

Results for the SR reconstruction of 3D LR breath-hold MRI scans are shown in Figure 3.10 for the p-SRG and c-SRG models. In addition, we included the paired data-only SRG (po-SRG) model that was trained with 480 data pairs to demonstrate the effectiveness of the DSN. Visually improvements in image quality are observable for the output of the p-SRG model and even that of the po-SRG model compared to that of the c-SRG model. Upon closer inspection, it is also observed that the boundaries of the vertebral body and the right kidney in the p-SRG output are clearly distinctive, while both structures are severely blurred in the c-SRG output and are not clear in the po-SRG output.

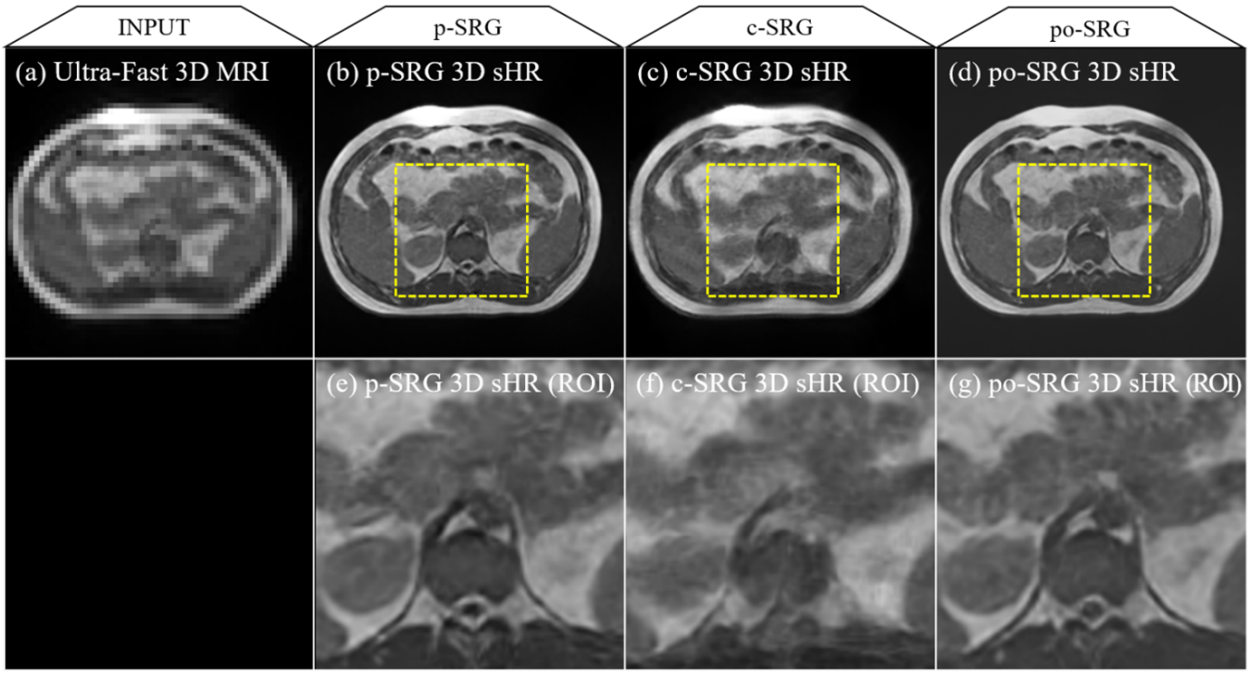

Figure 3.10. Axial SR reconstruction results for 3D LR breath-hold MRIs. (a) LR input from physical scan (2.5 s/vol., 64×64 pixels, 6.0×6.0 mm per pixel), (b) p-SRG output (256×256 pixels, 1.5×1.5 mm per pixel) and (e) magnified ROI, (c) c-SRG output and (f) magnified ROI, and (d) Paired data-only SRG (po-SRG) output and (g) magnified ROI.

To evaluate the performance of three SRG models in the absence of ground truth reference HR images, the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE)[50] metric

was computed (Figure 3.11). The BRISQUE index uses scene statistics of locally normalized luminance coefficients to quantify possible losses of "naturalness" in the image due to the presence of distortions, thereby leading to a holistic measure of quality. A smaller score indicates better perceptual quality. The c-SRG output scored 42.0 ± 4.6, while the score for the p-SRG improved by 42.6%, measuring 24.1 ± 3.8. The po-SRG output scored 30.9 ± 3.9. The statistics of the computed values of the BRISQUE index are presented in Table 3.4.

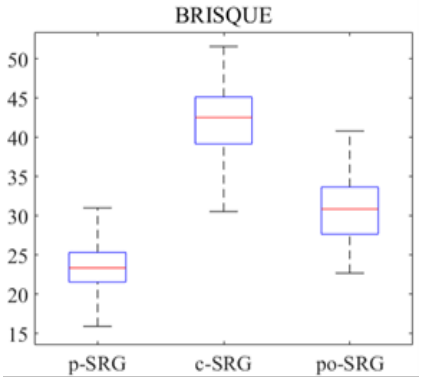

Figure 3.11. BRISQUE index calculated for 632 SR outputs of p-SRG, c-SRG, and po-SRG tested on 3D LR breath-hold MRIs.

Table 3.4. Statistics of the computed BRISQUE scores for p-SRG, c-SRG, and po-SRG outputs for 632 3D LR breath-hold MRIs.

	p-SRG 3D sHR	c-SRG 3D sHR	po-SRG 3D sHR
Mean	24.1	42.0	30.9
Std	3.8	4.6	3.9
Min	15.4	28.9	22.7
Median	23.4	42.5	30.9
Max	35.7	51.6	40.8

Figure 3.12 displays axial slices of the 3D LR breath-hold MRI acquisitions and the corresponding SR 3D MRIs generated using the full cascaded DL framework. The overall qualities of the cascaded DL results were equivalent to the physically scanned HR breath-hold MRIs.

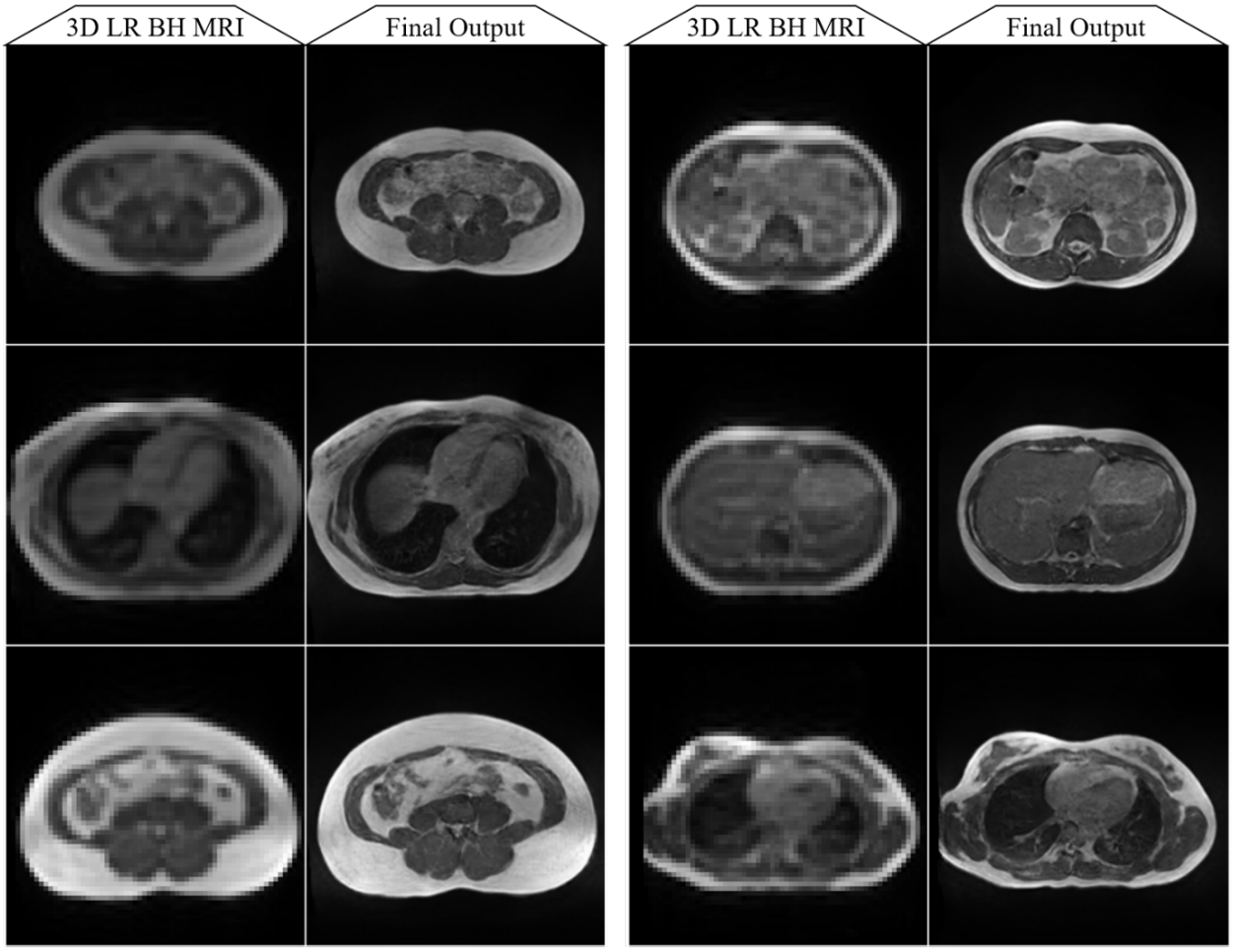

Figure 3.12. Input (left; 2.5 s/vol., 64×64 pixels, 6.0×6.0 mm per pixel) and output (right; 0.022 s/vol. processing time, 256×256 pixels, 1.5×1.5 mm per pixel) for 3D LR breath-hold (BH) scans using the proposed cascaded deep learning framework.

3.3.5 SR reconstruction of LR free-breathing 4D-MRI

Test results for the LR free-breathing 4D-MRI case are shown in Figure 3.13 for both the p-SRG and c-SRG model. Based on the results of the axial cases, the p-SRG model outperformed the c-SRG model. In a magnified view of the marked region of interest (ROI), the boundaries of diaphragm positions in p-SRG images are clearly distinctive compared to the c-SRG outputs.

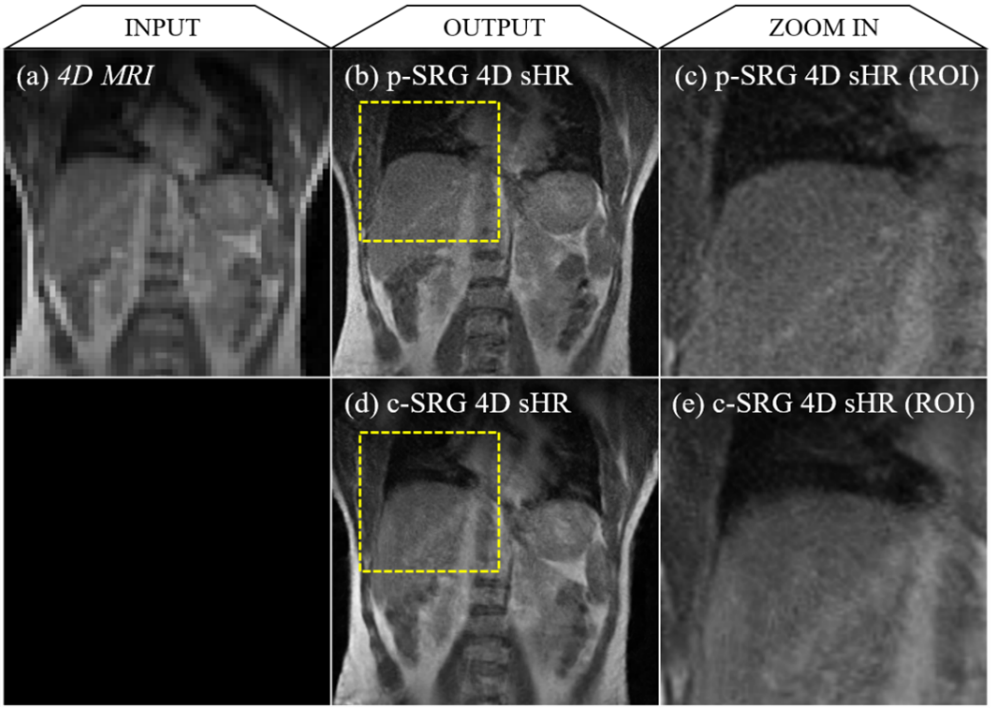

Figure 3.13. Comparison of coronal SR reconstructions using the p-SRG and c-SRG models tested on a free-breathing 4D-MRI scan. (a) Physically scanned LR free-breathing 4D-MRI input (0.5 s/vol., 64×64 pixels, 6.0×6.0 mm per pixel), (b) p-SRG and (d) c-SRG output (256×256 pixels, 1.5×1.5 mm per pixel), and magnified ROIs (c) and (e).

Figure 3.14 displays the results of the BRISQUE index computed for SR 4D-MRI data generated by the p-SRG and c-SRG models. In the case of the c-SRG, the images were scored 27.6 ± 8.1, while for the p-SRG the results were consistently improved by 19.5%, measuring 22.2 ± 5.7. The associated statistics are presented in Table 3.5.

Reconstruction results for a LR 4D-MRI scan with and without the DAE are displayed

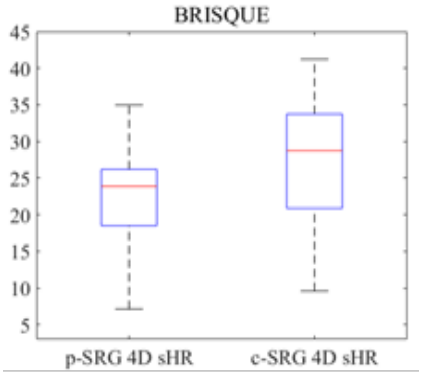

Figure 3.14. BRISQUE index calculated for 800 SR outputs of p-SRG and c-SRG test on LR free-breathing 4D-MRI scans.

Table 3.5. Statistics of the computed BRISQUE scores for 800 p-SRG and c-SRG outputs for LR free-breathing 4D-MRI scans.

	p-SRG 4D sHR	c-SRG 4D sHR
Mean	22.2	27.6
Std	5.7	8.1
Min	1.1	9.6
Median	23.9	28.8
Max	35.0	41.2

in Figure 3.15. Without the use of DAE, the noise was amplified in the generated SR images since the SRG network cannot distinguish the noise features from the input LR data. The difference was prominent in the magnified view of the vertebral and intervertebral structures (Figure 3.15c,f), where the boundary became unclear due to the degradation in image quality. By contrast, no severe degradation was present with the use of DAE. This confirms that the DAE served an important role in preserving anatomical detail during the SR reconstruction process.

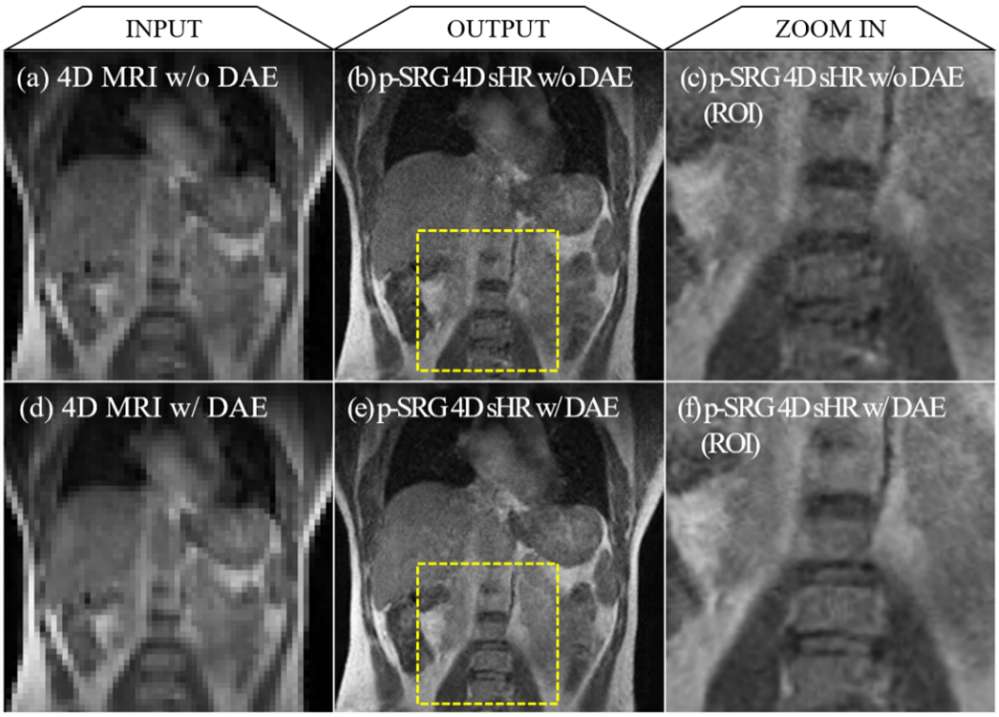

Figure 3.15. Comparison of coronal SR reconstructions of LR 4D-MRI acquisitions without (top) and with (bottom) the front end DAE. (a) Original, physically scanned LR free-breathing 4D-MRI acquisition (0.5 s/vol., 64×64 pixels, 6.0×6.0 mm per pixel), (b) p-SRG output with noisy input (a), (c) magnified ROI from noisy reconstruction (b), (d) de-noised output of DAE applied to (a), (e) p-SRG output with de-noised input (d), and (f) magnified ROI from reconstruction (e).

Figure 3.16 displays coronal slices of a LR 4D-MRI scan along with the corresponding SR reconstruction results using the full cascaded DL framework at 10%, 30%, 50%, 70%, and 90% breathing phases. Diaphragm motion at each breathing phase is observed relative

to the dotted yellow line in Figure 3.16. The overall quality of the cascaded DL results was equivalent to the physically scanned HR MRIs at each phase.

3.4 Discussion

In this study, we presented a robust MRI SR reconstruction method based on our novel cascaded DL technique and successfully showed that the resolution of physically scanned LR MRIs can be enhanced four times without compromising image quality.

The main innovation of our cascaded DL-based MRI SR can be summarized in two points. 1) We introduced a simply structured DSN trained with a limited number of LR and HR MRI data in pairs and used it to generate perfectly paired LR images from abundant clinical HR MRI data for the training of the SRG network. Previous studies of DL-based MRI SR approaches simulated the LR MRIs through known down-sampling methods such as bicubic or k-space down-sampling. However, as we demonstrated in Section 3.3, the c-SRG that depends on the accuracy of the assumptions inherent in the use of these approaches cannot be utilized with physically scanned LR MRI data since the translation models for the physical scan and the simulations are different. We observed that appropriate model estimation is the most vital step that can profoundly influence the performance of the DL-based MRI SR algorithm. 2) We introduced the DAE at the front end of our framework to suppress noise amplification during the SR reconstruction process. In MRI, the noise is randomly generated from the measurement process and is typically modeled as a Rician distribution due to the presence of Gaussian noise in both the real and imaginary parts of the raw k-space data. In this study, we selected the NLM filter as the de-noising operator to approximate using the DAE. We chose the NLM method because it provided a powerful framework for MRI de-noising compared to other state-of-the-art methods. However, the process is computationally expensive, making it impractical to use the operator directly in

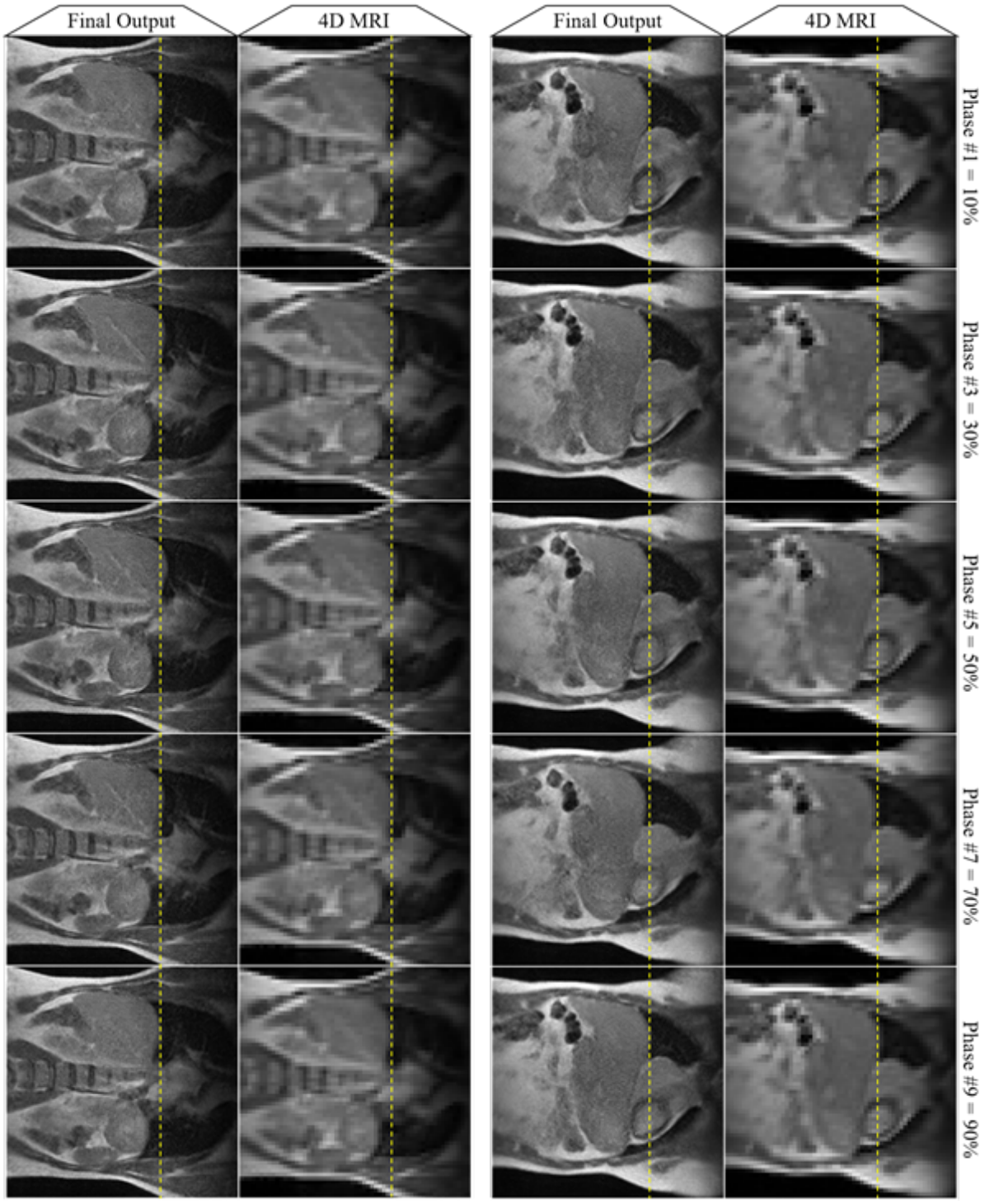

Figure 3.16. Input (right; 2.5 s/vol., 64×64 pixels, 6.0×6.0 mm per pixel) and output (left; 0.022 s/vol. processing time, 256×256 pixels, 1.5×1.5 mm per pixel) of coronal 4D-MRI acquisition shown for two views at five breathing phases using the proposed cascaded deep learning framework.

a real-time imaging framework.[51,52] During the experiments, it was found that the speed of the NLM processing was 10.217 seconds per 100 frames on the CPU (Intel® Core™ i7-8700K) whereas the DAE required 0.044 seconds per 100 frames. Note that GPU implementation and parallel processing for NLM is currently an active research area that is beyond the scope of our study. Nonetheless, we observed that the DAE can act as an excellent replacement or estimate of computationally sophisticated image processing algorithms when the bandwidth of data processing is time limited.

The main motivation of the method proposed here is to reduce the scanning time required for HR MR images. True HR MRI images (256×256 pixels, 1.5×1.5 mm per pixel) used in this study were acquired at an average rate of 17 seconds per volume of 100 slices, while LR MRI (64×64 pixels, 6.0×6.0 mm per pixel) scans required only 2.5 seconds for the same volume. The clinical implementation of our SR framework, illustrated in Figure 3.1, would involve passing clinically scanned, LR MR images through the pre-trained DAE and p-SRG to generate reconstructed HR MR images. Using a single GPU and the TensorFlow backend, 0.016 seconds per 100 slices are required to pass through the DAE, and 0.029 seconds per 100 slices are required for the p-SRG, meaning processing in the proposed framework can be achieved in near real-time. Ultimately, time savings of 14.455 seconds in the 3D case can be achieved by implementing the proposed framework, enabling 4D HR MRI.

The clinical relevance of these savings may be seen in considering breath-hold requirements during LR and HR scans. The 2.5 seconds required for a LR acquisition are a reasonably achievable breath-hold period for a vast majority of patients in a radiation oncology clinic. When that period is extended to 17 seconds, however, a more significant portion of the patient population may experience trouble in maintaining a breath-hold, giving rise to image artifacts exemplified by those shown in Figure 3.17. In this way, the acquisition of LR images in a clinically feasible time frame followed by processing using our proposed framework allows for HR MR images without the requirement of a potentially strenuous

breath-hold.

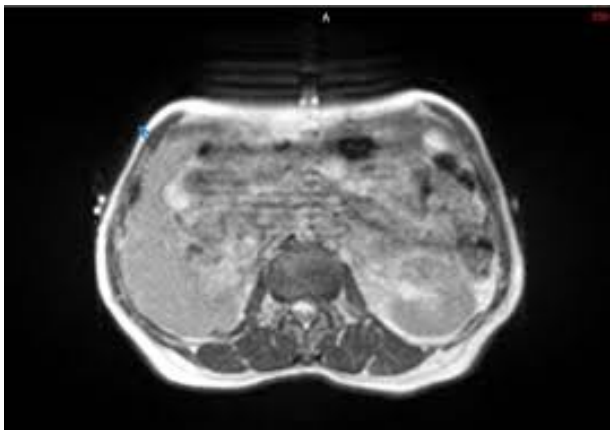

Figure 3.17. Representative axial slice of an HR scan demonstrating breathing artifacts arising from a failure to maintain breath-hold in the 17 second scan period.

In this study, we limited our experiments and investigations to a 2D network. Constructing a 3D network requires an extensively larger network structure and a vastly increased number of parameters. Training such an expanded network would require a much larger amount of data that is of limited availability in a clinical setting. In this study, the proposed network aimed to improve in-plane resolution, leaving the slice thickness the same as the original. Based on our experiments, a 2D network is more efficient and less time consuming to process in such a scenario since 3D MRI data can be considered as a collection of independent 2D image slices stacked orthogonally to the in-plane direction. Despite the inherent limitations of the 2D network, the method does demonstrate promise in clinically relevant scenarios. In addition to the impact on breath-hold requirements discussed previously, some diagnostic tasks may benefit from the proposed method. Figure 3.18 illustrates the SR reconstruction of fine structures in the lung compared to reference HR images of the same ROI. The output SR images demonstrate the fair ability of the framework to reconstruct small lung nodules, even when the structures are washed out due to undersampling in LR acquisitions. Reconstructions in this scenario are challenged by two factors in the present implementation: 1) the majority of the training data sets are abdominal scans that contain

little data in the region of the lungs, rendering the training set insufficient to achieve truly accurate reconstructions in this region, and 2) the inherent limitations of the 2D network discussed here. By design, a 2D framework does not capture the same rich structural information a 3D network does that aids in the reconstruction task. The imperfect reconstruction of the lung nodules shown in Figure 3.18 highlights the importance of transitioning to a 3D framework to test the validity of our method in all planar directions. This will require additional data acquisition from volunteers to construct the 3D DSN as well as collecting more HR clinical data to train the more sophisticated generative network.

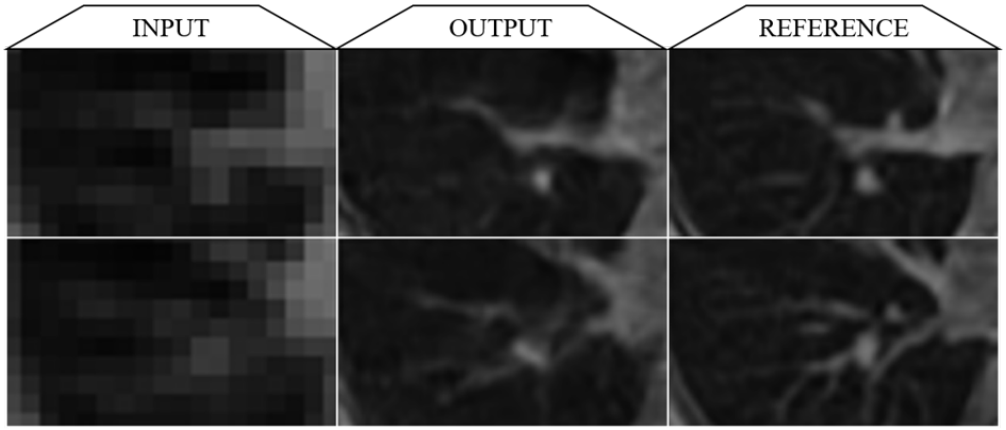

Figure 3.18. Magnified inputs (left; 2.5 s/vol., 6.0×6.0 mm per pixel), outputs (middle; 0.029 s/vol. processing time, 1.5×1.5 mm per pixel) and reference images (right; 17 s/vol., 1.5×1.5 mm per pixel) that are focused on small lung nodules for 3D LR breath-hold scans using the proposed cascaded deep learning framework.

Recently, there was a considerable interest in advancing GAN-based networks into various image processing problems.[53–55] Although the purpose of these works is for different applications, we anticipate that a similar strategy can be applied to the MRI SR reconstruction problem as well. The style transfer framework exemplified by CycleGan[53] could act as a replacement for the DSN or SRG to enable HR to LR transformations or vice versa without necessitating the collection of paired data sets. The issue of data availability remains, however; collecting such a large amount of LR data would require the inclusion of LR

acquisitions into routine clinical MRI scan protocols, which may not be clinically feasible. Inversely, as many recent innovations in DL require abundant perfectly matched pairs of training data that are often not available, it will be an exciting topic to seek an idea that is similar to our use of the DSN to translate one "mode" of data to another for the purpose of approximating a translation model like we did here.

The performance of a clinical MRI scan is limited by numerous factors including long scanning time, poor SNR, and uncontrolled intermittent motion. Validation studies and experiments showed that our newly developed SR technique is extremely beneficial under such circumstances. The reconstruction model of our cascaded deep learning is general and could be effectively applied for all types of MRI (i.e., various field strengths) and pulse sequences. Therefore, our next step is to test this framework on various clinical sites, not only limited to MR-IGRT but also to diagnostic applications where resolution is limited by the aforementioned factors. In MR-IGRT applications, our next study will be to test the proposed method on clinical patient data cases with moving anatomical sites such as the lung, liver, or pancreas. It is anticipated that the improved reconstruction capability of SR MRI enables enhanced tumor visualization and organ delineation while facilitating further research in the areas of automated tumor tracking, modeling of organ-specific motion, and dose accumulation. All in all, we anticipate that our proposed cascaded deep learning algorithm is a promising new technique to reconstruct high quality, high-resolution MRI images without sacrificing SNR or increasing scanning time.

Finally, we remark that our cascaded DL approach is the first DL-based MRI SR reconstruction technique that formulates SR MRI as the blind image super-resolution problem. Our studies have shown that high quality SR reconstruction can be achieved in the absence of a LR to HR translation model and in the presence of noise. This brings us one step closer to making MRI SR technology and real-time MRI applications more practical.

3.5 Conclusions

In this paper, we developed an end-to-end framework to generate detail preserving HR MRI images from physically scanned LR MRI images using a novel cascaded DL technique. The results have shown that our proposed technique enables 3D LR breath-hold MRI as well as 4D-MRI within a clinically feasible time frame without sacrificing spatial resolution or image quality. Such techniques will facilitate the target delineation and tracking process during MR-IGRT. This makes our technique fully useful in precision MR-IGRT.

References

[1] Fallone B, Murray B, Rathee S, et al. First MR images obtained during megavoltage photon irradiation from a prototype integrated linac-MR system. *Med Phys*. 2009; 36(6):2084–2088. doi:10.1118/1.3125662

[2] Cao Y, Tseng C, Balter J, Teng F, Parmar H, Sahgal A. MR-guided radiation therapy: transformative technology and its role in the central nervous system. *Neuro Oncol*. 2017; 19:ii16–ii29. doi:10.1093/neuonc/nox006

[3] Wu Q, Li T, Wu Q, Yin F. Adaptive radiation therapy: technical components and clinical applications. *Cancer J*. 2011;17(3):182–189. doi:10.1097/PPO.0b013e31821da9d8

[4] Olberg S, Green O, Cai B, et al. Optimization of treatment planning workflow and tumor coverage during daily adaptive magnetic resonance image guided radiation therapy (MR-IGRT) of pancreatic cancer. *Radiat Oncol*. 2018;13(1):51. doi:10.1186/s13014-018-1000-7

[5] Kupelian P, Sonke J. Magnetic resonance-guided adaptive radiotherapy: a solution to the future. *Semin Radiat Oncol*. 2014;24(3):227–232. doi: 10.1016/j.semradonc.2014.02.013

[6] Liang F, Qian P, Su K, et al. Abdominal, multi-organ, auto-contouring method for online adaptive magnetic resonance guided radiotherapy: An intelligent, multi-level fusion approach. *Artif Intell Med*. 2018;90:34–41. doi:10.1016/j.artmed.2018.07.001

[7] Tyran M, Cao M, Raldow A, et al. Stereotactic magnetic resonance-guided online adaptive radiotherapy for oligometastatic breast cancer: A case report. *Cureus*. 2018; 10(3):e2368. doi:10.7759/cureus.2368

[8] Vestergaard A, Hafeez S, Muren L, et al. The potential of MRI-guided online adaptive re-optimisation in radiotherapy of urinary bladder cancer. *Radiother Oncol*. 2016; 118(1):154–159. doi:10.1016/j.radonc.2015.11.003

[9] Manjón J, Coupé P, Buades A, Fonov V, Louis Collins D, Robles M. Non-local MRI upsampling. *Med Image Anal*. 2010;14(6):784–792. doi:10.1016/j.media.2010.05.010

[10] Carmi E, Liu S, Alon N, Fiat A, Fiat D. Resolution enhancement in MRI. *Magn Reson Med*. 2006;24(2):133–154. doi:10.1016/j.mri.2005.09.011

[11] Tieng Q, Cowin G, Reutens D, Galloway G, Vegh V. MRI resolution enhancement: how useful are shifted images obtained by changing the demodulation frequency? *Magn Reson Med*. 2011;65(3):664–672. doi:10.1002/mrm.22653

[12] Yap P, An H, Chen Y, Shen D. A generative model for resolution enhancement of diffusion MRI data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 527–534

[13] Li G, Wei J, Kadbi M, et al. Novel super-resolution approach to time-resolved volumetric 4-dimensional magnetic resonance imaging with high spatiotemporal resolution for multi-breathing cycle motion assessment. *Int J Radiat Oncol Biol Phys*. 2017;98(2):454–462. doi:10.1016/j.ijrobp.2017.02.016

[14] Plenge E, Poot D, Bernsen M, et al. Super-resolution methods in MRI: can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time? *Magn Reson Med*. 2012;68(6):1983–1993. doi:10.1002/mrm.24187

[15] Cai J, Chang Z, Wang Z, Paul SW, Yin F. Four-dimensional magnetic resonance imaging (4D-MRI) using image-based respiratory surrogate: a feasibility study. *Med Phys*. 2011; 38(12):6384–6394. doi:10.1118/1.3658737

[16] Deng Z, Pang J, Yang W, et al. Four-dimensional MRI using three-dimensional radial sampling with respiratory self-gating to characterize temporal phase-resolved respiratory motion in the abdomen. *Magn Reson Med*. 2016;75(4):1574–1585. doi: 10.1002/mrm.25753

[17] Greenspan H, Oz G, Kiryati N, Peled S. MRI inter-slice reconstruction using super-resolution. *Magn Reson Imaging*. 2002;20(5):437–446. doi:10.1016/s0730-725x(02)00511-8

[18] Gholipour A, Estroff J, Warfield S. Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain MRI. *IEEE Trans Med Imaging*. 2010; 29(10):1739–1758. doi:10.1109/TMI.2010.2051680

[19] Bhatia M. Evaluation of ischemic heart disease and viability by cardiac MRI. *Indian Heart J*. 2014;66(1):143–144. doi:10.1016/j.ihj.2013.10.006

[20] Peled S, Yeshurun Y. Superresolution in MRI: application to human white matter fiber tract visualization by diffusion tensor imaging. *Magn Reson Med*. 2001;45(1):29–35. doi:10.1002/1522-2594(200101)45:1¡29::aid-mrm1005¿3.0.co;2-z

[21] Calamante F, Oh S, Tournier J, et al. Super-resolution track-density imaging of thalamic substructures: comparison with high-resolution anatomical magnetic resonance imaging at 7.0T. *Hum Brain Mapp*. 2013;34(10):2538–2548. doi:10.1002/hbm.22083

[22] Jog A, Carass A, Prince J. Self super-resolution for magnetic resonance images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 553–560

[23] Kawamura N, Yokota T, Hontani H. Super-resolution of magnetic resonance images via convex optimization with local and global prior regularization and spectrum fitting. *Int J Biomed Imaging*. 2018;doi:10.1155/2018/9262847

[24] Peng J, Xu Q, Feng Y, Lv Q, Chen W. Adaptive regularized super-resolution reconstruction for magnetic resonance images. *J Southern Medical University*. 2011;31(10):1705–1708

[25] Luo J, Mou Z, Qin B, et al. Fast single image super-resolution using estimated low-frequency k-space data in MRI. *Magn Reson Imaging*. 2017;40:1–11. doi:10.1016/j.mri.2017.03.008

[26] Jafari-Khouzani K. MRI upsampling using feature-based nonlocal means approach. *IEEE Trans Med Imaging*. 2014;33(10):1969–1985. doi:10.1109/TMI.2014.2329271

[27] Mahmoudzadeh A, Kashou N. Interpolation-based super-resolution reconstruction: effects of slice thickness. *J Med Imaging (Bellingham)*. 2014;1(3):034007. doi:10.1117/1.JMI.1.3.034007

[28] Rueda A, Malpica N, Romero E. Single-image super-resolution of brain MR images using overcomplete dictionaries. *Med Image Anal*. 2013;17(1):113–132. doi:10.1016/j.media.2012.09.003

[29] Tourbier S, Bresson X, Hagmann P, Thiran J, Meuli R, Cuadra M. An efficient total variation algorithm for super-resolution in fetal brain MRI with adaptive regularization. *Neuroimage*. 2015;118:584–597. doi:10.1016/j.neuroimage.2015.06.018

[30] Zhang D, He J, Zhao Y, Du M. MR image super-resolution reconstruction using sparse representation, nonlocal similarity and sparse derivative prior. *Comput Biol Med*. 2015;58:130–145. doi:10.1016/j.compbiomed.2014.12.023

[31] Dong C, Loy C, He K, Tang X. Image super-resolution using deep convolutional networks. *IEEE Trans Pattern Anal Mach Intell*. 2016;38(2):295–307. doi:10.1109/TPAMI.2015.2439281

[32] Shi J, Li Z, Ying S, et al. MR image super-resolution via wide residual networks with fixed skip connection. *IEEE J Biomed Health Inform*. 2019;23(3):1129–1140. doi: 10.1109/JBHI.2018.2843819

[33] Shi J, Liu Q, Wang C, Zhang Q, Ying S, Xu H. Super-resolution reconstruction of MR image with a novel residual learning network algorithm. *Phys Med Biol*. 2018; 63(8):085011. doi:10.1088/1361-6560/aab9e9

[34] Jog A, Carass A, Prince J. Improving magnetic resonance resolution with supervised learning. In *IEEE International Symposium on Biomedical Imaging*. pp. 987–990. doi: 10.1109/ISBI.2014.6868038

[35] Zeng K, Zheng H, Cai C, Yang Y, Zhang K, Chen Z. Simultaneous single- and multi-contrast super-resolution for brain MRI images based on a convolutional neural network. *Comput Biol Med*. 2018;99:133–141. doi:10.1016/j.compbiomed.2018.06.010

[36] Chaudhari A, Fang Z, Kogan F, et al. Super-resolution musculoskeletal MRI using deep learning. *Magn Reson Med*. 2018;80(5):2139–2154. doi:10.1002/mrm.27178

[37] Dong C, Loy C, He K, Tang X. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*. pp. 184–199. doi:10.1007/978-3-319-10593-2_13

[38] Kim J, Lee J, Lee K. Deeply-recursive convolutional network for image super-resolution. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. pp. 1637–1645

[39] Buades A, Coll B, Morel J. A non-local algorithm for image denoising. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. pp. 60–65

[40] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. *ArXiv*. 2014;

[41] Nair V, Hinton G. Rectified linear units improve restricted Boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning*. pp. 807–814

[42] Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol P, Bottou L. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J Mach Learn Res*. 2010;11(12):3371–3408

[43] Vincent P, Larochelle H, Bengio Y, Manzagol P. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning*. pp. 1096–1103

[44] Gudbjartsson H, Patz S. The Rician distribution of noisy MRI data. *Magn Reson Med*. 1995;34(6):910–914. doi:10.1002/mrm.1910340618

[45] Kingma D, Ba J. Adam: A method for stochastic optimization. *ArXiv*. 2017;:1–15

[46] Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-scale machine learning on heterogeneous systems. 2015. Software available from tensorflow.org

[47] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 4681–4690

[48] Shi W, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1874–1883

[49] Wang Z, Bovik A, Sheikh H, Simoncelli E. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13(4):600–612. doi: 10.1109/tip.2003.819861

[50] Mittal A, Moorthy A, Bovik A. No-reference image quality assessment in the spatial domain. *IEEE Trans Image Process*. 2012;21(12):4695–4708. doi: 10.1109/TIP.2012.2214050

[51] Wiest-Daesslé N, Prima S, Coupé P, Morrissey S, Barillot C. Rician noise removal by non-local means filtering for low signal-to-noise ratio MRI: Applications to DT-MRI. In *International Conference on Medical Image Computing and Computer-assisted Intervention*. pp. 171–179

[52] Manjón J, Carbonell-Caballero J, Lull J, García-Martí G, Martí-Bonmatí L, Robles M. MRI denoising using non-local means. *Med Image Anal*. 2008;12(4):514–523. doi: 10.1016/j.media.2008.02.004

[53] Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. *ArXiv*. 2018;:1–18

[54] Zhang H, Goodfellow I, Metaxas D, Odena A. Self-attention generative adversarial networks. In *International conference on machine learning*. pp. 7354–7363

[55] Karras T, Aila T, Laine S, Lehtinen J. Progressive growing of GANs for improved quality, stability, and variation. *ArXiv*. 2017;

Chapter 4

Synthetic CT reconstruction using a deep spatial pyramid convolutional framework for MR-only breast radiotherapy

4.1 Introduction

The use of magnetic resonance imaging (MRI) in radiation therapy (RT) has grown in recent years due to a desire to take advantage of the attractive properties of the modality, especially with the advent of MRI-based delivery systems.[1–7] Compared to x-ray computed tomography (CT), MRI offers superior soft tissue contrast without exposing the patient to ionizing radiation.[8–12] Taking full advantage of these attractive properties with MRI-based therapy delivery systems allows for numerous applications including motion tracking and management as well as online adaptive therapy.[1,5,6,13] The increasing use of MRI for tissue and target delineation makes an MRI-only RT workflow, one in which MRI is the sole imaging modality employed for treatment planning and dose calculations, a favorable option for MR image-guided RT (MR-IGRT) applications. Presently, however, such a workflow has not been widely adopted in a clinical setting.

In the conventional MR-IGRT treatment planning workflow, the patient undergoes both MRI and CT simulation before planning commences.[5,8,14] MRI is the primary imaging modality used to guide treatment, but the electron density information required for dose calculations necessitates that CT simulation be a component of the conventional workflow.[12,14,15] No correlation between the nuclear magnetic properties on which MRI relies and the radiological properties of tissue used in dose calculations has been demonstrated.[16,17] As such, CT images must be acquired and registered to MR images to provide the necessary electron density information.

Advancing towards MRI-only RT avoids a number of issues that arise in a multi-modality workflow. Eliminating CT simulation and setup scans reduces the radiation dose delivered unnecessarily to the patient throughout the course of their treatment and lessens the cost of treatment.[8,14,15,17,18] Additionally, inter-modality registration errors that can represent a clinically significant barrier to the accurate delivery of treatment can be avoided in an

MRI-only workflow.[9,19–23]

Achieving an MRI-only RT workflow requires the creation of a so-called substitute or synthetic CT (sCT) image set that provides the electron density information necessary for dose calculations based on predicted HU values. A number of previous approaches at creating these sCTs fall generally into two broad categories, briefly covered here but reviewed in detail elsewhere: atlas-based approaches and voxel-based approaches.[17,24–26] In an atlas-based approach,[18,27–33] pairs of co-registered MRI and CT scans from a patient database are used to create MRI and CT atlases that constitute a typical anatomy as represented by each imaging modality. Determining the deformation between a patient's MRI scan and the MRI atlas allows for the creation of a sCT scan by applying this deformation to the CT atlas. Atlas-based approaches can involve an average atlas[27] for MRI and CT scans composed of the entire database or a multi-atlas[32,33] in which CT numbers are assigned based on a weighted average of values from multiple deformed CT atlases. Such approaches can quickly become computationally intense as the number of deformations and atlas complexity increases.[17,26] Additionally, atlas-based approaches struggle in cases where the incoming patient's anatomy differs from that represented by the atlas due to missing tissues, air cavities, or surgical implants.[26,28,31] Alternatively, voxel-based approaches involve the assignment of CT numbers through a number of methods, including segmentation with standard or specialized MRI sequences,[34–42] statistical methods,[43–46] and learning-based approaches.[47] The simplest and most widely used voxel-based approaches use MRI voxel intensities from standard sequences to segment general tissue classes (air, fat, soft tissue, etc.) that are then bulk-assigned corresponding CT numbers. While voxel-based approaches perform better than atlas-based approaches in the case of atypical anatomy, the ambiguity between regions of bone and air due to a lack of signal in both regions can present a challenge.[17,36]

More recently, deep learning (DL)-based approaches using convolutional neural networks (CNNs) have been proposed to solve the problem of sCT generation for sites in the head and

pelvis.[33,48–51] Han[48] employed a pre-trained U-Net architecture[52] to generate sCTs for brain tumor cases. Similarly, Fu et al[50] and Chen et al[53] generated sCTs of the male pelvis using 2D and 3D U-Nets. Viewing the generation of sCTs from MR images as an image-to-image translation problem, Maspero et al[49] implemented a generative adversarial network[54] (GAN) named pix2pix[55] that has demonstrated success in this space to produce sCTs for general pelvis cases.

In the present study, we evaluate the robustness of this ubiquitous pix2pix framework and our proposed framework that utilizes a convolutional process termed deep spatial pyramid pooling. In this novel framework, we exploit atrous convolution to effectively capture and encode large-scale structural information that aids in the sCT generation task while also maintaining a compact network architecture that is effectively trained with a limited dataset. The performance of these GAN architectures was evaluated against the size of the training data set to determine the time required to train the model as well as the quality of the resulting image output of a trained model. Additionally, sCT images generated using the novel deep spatial pyramid convolutional framework were used to perform MRI-based dose calculations, which were compared to plans calculated based on true CT images using the gamma index.[56]

4.2 Materials and Methods

4.2.1 Model and Loss Formulation

The task of sCT generation may be viewed simply using the following formulation of a forward problem:

$$g = \boldsymbol{T}(m), \tag{4.1}$$

where m is the input MR image data, $\boldsymbol{T}$ is the transformation operator that defines the mapping from MR to CT images, and g is the observable output of generated CT data. Considering the one-to-many or many-to-one nature of mapping that arises due to similar MRI signal intensities in regions of bone and air, for example, determining the form of the operator $\boldsymbol{T}$ is non-trivial. The task is ill-posed and the goal in applying DL to the problem is to estimate a suitable operator $\boldsymbol{T_e}$ such that a generic cost function C is minimized between the synthetic output $s = \boldsymbol{T_e}(m)$ and the ground truth CT g:

$$\min_{s} C[s, g] \tag{4.2}$$

The GAN frameworks explored here consist of two components, each of which is discussed in more detail in the following sections: 1) a generative model G that generates a sample in the data distribution p_{data} in which true CT image data resides from a sample in a latent space and 2) a discriminator D that distinguishes whether a given sample is drawn from the data distribution or generated by G with a certain probability. The generator G and discriminator D undergo training in alternating minimization steps of their respective loss functions. In these frameworks, the generic cost minimization presented in Eq. 4.2 is the driving force of learning during the training process of the generator G. The discussion of the specific form of $C[\cdot]$ relies on the definition of sigmoid cross entropy loss:[57]

$$\begin{aligned} L &= \vec{z} * -\log\left(S(\vec{x})\right) + (1 - \vec{z}) * -\log\left(1 - S(\vec{x})\right) \\ &\text{with } S(\vec{x}) = \frac{1}{1 + \exp\left(-\vec{x}\right)} \text{ simplifies to} \\ &= \vec{x} - \vec{x} * \vec{z} + \log\left(1 + \exp\left(-\vec{x}\right)\right), \end{aligned} \tag{4.3}$$

where the elements of $\vec{x}$ are the true or predicted image logits computed by the discriminator D and $\vec{z}$ are the labels corresponding to true (1) or predicted (0).

The generative models discussed in the later sections share the same loss function g_{loss} defined by

$$g_{loss} = L_{adv} + 100 * l_{mae}, \tag{4.4}$$

where the adversarial loss L_{adv} is the sigmoid cross entropy loss (Eq. 4.3) with predicted image logits $\vec{x}$ and labels $\vec{z} = \vec{1}$, a ones vector, and the mean absolute error (MAE) loss l_{mae} is simply the mean of the absolute difference of the predicted images I_{pred} and true images I_{true}:

$$l_{mae} = \text{mean}\left(|I_{pred} - I_{true}|\right). \tag{4.5}$$

With this formulation, the complete loss function g_{loss} penalizes predictions that are not computed as true data distributions by D through the adversarial loss term L_{adv} and attempts to ensure pixel-wise agreement between predicted and true images with the MAE loss term l_{mae}. It should be noted that each term is reduced in dimensions by computing the mean before summing in Eq. 4.4 to yield a single value. The expressions have been simplified here for presentation.

The aim of the discriminator D in the GAN framework differs from that of G, resulting in a loss function as follows:

$$d_{loss} = L_{pred} + L_{true}, \tag{4.6}$$

where predicted loss L_{pred} and true loss L_{true} are each the sigmoid cross entropy loss (Eq. 4.3) with predicted image or true image logits $\vec{x}$ computed by D and labels $\vec{z} = \vec{0}$ or $\vec{1}$, respectively. The formulation of L_{pred} and L_{true} in d_{loss} differ from that of L_{adv} in g_{loss} (Eq. 4.4) in that the labels corresponding to predicted image logits in L_{adv} are unity, as G strives to produce sCT outputs that are evaluated as true CT images by D.

Training G and D involves alternating minimization steps of each loss function, during which the parameters that define operations within each model are optimized. Optimization for G was performed using the Adam gradient-based stochastic optimization algorithm[58] with an initial learning rate of 0.0002, $\beta_1 = 0.7$, $\beta_2 = 0.999$, and $\hat{\epsilon} = 10^{-8}$. For D, TensorFlow's[57] gradient descent optimizer was used with an initial learning rate of 0.00002. In each case, the learning rates decayed according to a staircase exponential function every 100000 steps with a base decay rate of 0.95.

4.2.2 Network Architectures

As mentioned previously, the GAN framework that is the general subject of this work consists of two adversarial CNNs: G and D. The focus here is placed on the architecture of the generative model G. Two network architectures for G are explored: the conventional U-Net architecture implemented in the pix2pix framework,[49,55] referred to as "pix" here and discussed in Section 4.2.2, and the proposed deep spatial pyramid convolutional framework we present as an alternative, referred to as "aspp" and discussed in Section 4.2.2. D, the second half of the GAN framework, is of the same form in both implementations, the details of which are discussed in Section 4.2.2.

Conventional U-Net: pix

The conventional pix architecture, illustrated in Figure 4.1, consists of a stacked encoder-decoder U-Net structure (using notation introduced by Isola et al):[55]

encoder:

C64–C128–C256–C512–C512–C512–C512–C512

decoder:

CD512–CD1024–CD1024–C1024–C1024–C512–C256–C128,

where Ck denotes a convolution-BatchNorm-ReLU layer with k convolutional filters with the

exception of the first layer in the encoder in which batch normalization is not performed. ReLUs on the encoder side of the network are leaky with a slope of 0.2. CDk denotes a convolution-BatchNorm-Dropout-ReLU layer with a dropout rate of 50%.[59] The final layer in the decoder applies a convolution to map to the desired number of output channels (one here), followed by a hyperbolic tangent function. Each convolution is a 4x4 filter applied with a stride of 2. At each layer on the encoder side of the network, convolutions downsample by a factor of 2. On the decoder side, convolutions upsample by a factor of 2. Skip connections that share activations from a layer in the encoder with a corresponding layer in the decoder allow for the transfer of shared, underlying structural features like prominent edges from the input MRI to the generated sCT.[55]

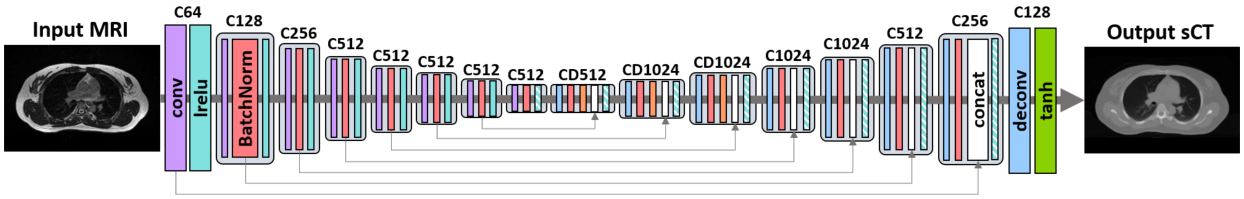

Figure 4.1. The conventional pix architecture. The input MR image is encoded sequentially as a feature map of reduced spatial dimension and increased depth as it travels through the encoder layers on the left side of the network. The process is reversed as the decoder layers recover spatial information and reconstruct the output sCT image. Skip connections between corresponding encoder/decoder layers, represented as grey lines at bottom here, allow shared structural features to move across the network efficiently.

In total, the conventional pix architecture consists of 54,408,832 trainable parameters. These parameters, which include the weights and biases of the convolutional filters applied at each layer of the encoder and decoder, are updated at every training step according to the Adam stochastic optimization algorithm mentioned in Section 4.2.1.

Atrous Spatial Pyramid Pooling: aspp

As an alternative to the conventional pix framework, we explore the application of the Atrous Spatial Pyramid Pooling (ASPP) method to the U-Net encoder-decoder architecture

as discussed by Chen et al.[51,60] The ASPP method employs atrous or dilated convolution, illustrated in Figure 4.2, to encode multi-scale contextual information. Pixels included in the atrous convolution kernel are spaced away from the central pixel based on the rate of convolution, where larger rates correspond to a larger field of view for the convolution kernel without changing the size of the kernel.

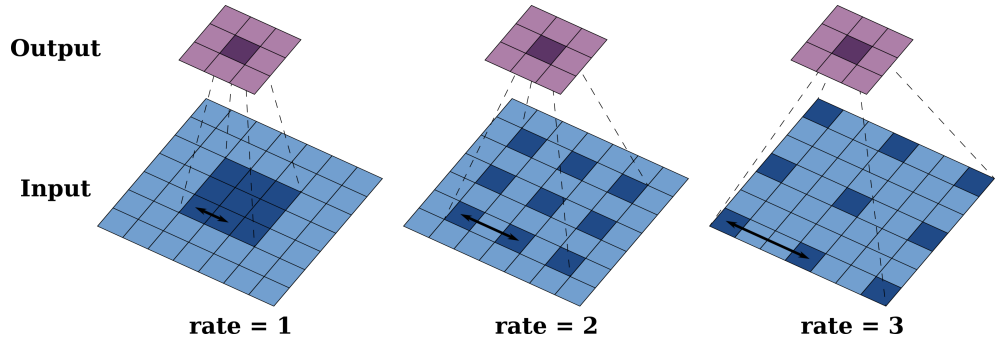

Figure 4.2. Atrous convolution applied to an input feature map with a 3x3 kernel and increasing rate. Increasing the rate widens the field of view without changing the size of the kernel, capturing context at a larger scale than standard convolution ($rate = 1$).

By performing atrous convolution at multiple rates in parallel, multi-scale features can be exploited to characterize a single pixel. Shown in Figure 4.3, an example ASPP module extracts contextual features from multiple fields of view, allowing the encoder to capture rich multi-scale information that aids in the image-to-image translation task.

Building from the conventional pix framework, we implement the ASPP module in a similar U-Net architecture to create G. The diagram in Figure 4.4 illustrates the network's structure, using the same notation presented in the previous section for direct comparison and AC to denote atrous convolution:

encoder:

C64–C128–C256–C512

ASPP module:

(C512, AC512 rate 3, AC512 rate 6, AC512 rate 9, average pool–C512)–C512

decoder:

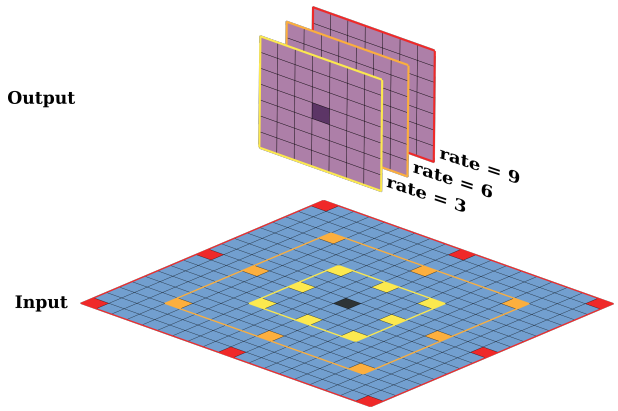

Figure 4.3. Atrous Spatial Pyramid Pooling.[64] Atrous convolution filters of increasing rate, shown in different colors, are applied in parallel to the input feature map. The resulting output feature maps that capture contextual features at fields of view of increasing size are concatenated together. For simplicity, output feature maps are not necessarily to scale and padding for convolution is not shown.

C512–C512–C256–C128.

Again, all convolutions are 4x4 spatial filters with stride 2 except for convolutions within the ASPP module. The four encoder layers that precede the ASPP module have an output stride of 16 compared to 256 in the case of the 8 encoder layers in the pix framework. The ASPP module implemented here consists of 5 convolution operations performed in parallel, all with stride 1: 1) 1x1 convolution; 2) 4x4 atrous convolution with rate 3; 3) 4x4 atrous convolution with rate 6; 4) 4x4 atrous convolution with rate 9; and 5) 1x1 convolution following an average pooling operation. The feature maps resulting from each of these operations is then concatenated before another set of 1x1 convolutional filters is applied. The successive decoder layers upsample the resulting feature maps to recover spatial information and construct the sCT output. As in the conventional pix framework, skip connections between corresponding encoder/decoder layers allow for the transfer of underlying structural features shared by the input and output images.

In contrast to the conventional pix framework, the elimination of 8 filter-dense layers from the interior of the network in favor of the ASPP module results in a significant reduc-

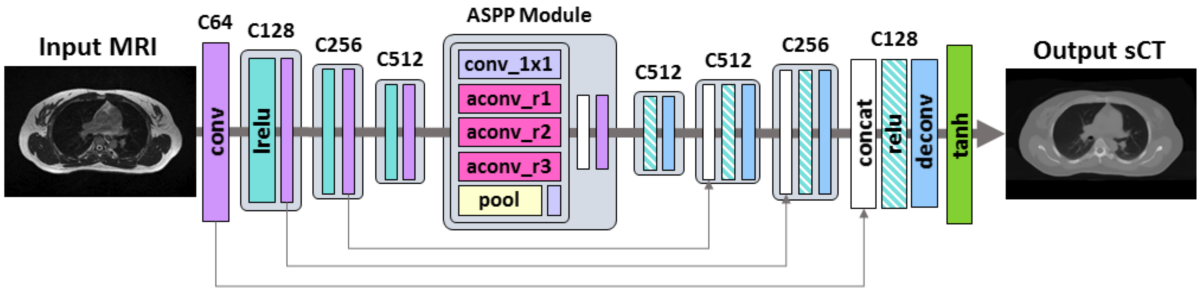

Figure 4.4. The proposed aspp architecture. Like in the conventional pix framework, the input MR image is encoded as richer feature maps as it travels through the encoder side of the network. The ASPP module then applies five convolution operations in parallel before concatenation and another convolution. The decoder layers then recover spatial information as they upsample the resulting feature maps and reconstruct the output sCT image.

tion in the total number of trainable parameters from 54,408,832 for the conventional pix architecture to 15,080,733 for the proposed aspp architecture. These parameters are updated during training by the Adam optimizer discussed in Section 4.2.1.

Discriminator

The architecture of the discriminator, illustrated in Figure 4.5, did not change with the generator architecture in each GAN implementation. Using the same notation introduced previously, the structure of D is as follows:

C64–C128–C256–C512,

with the exception of the first layer, where no batch normalization is performed. Again, leaky ReLUs are utilized with slope 0.2. Following the last layer, a final convolution is applied to map to one channel before applying a sigmoid function. The result is values between zero and one that denote the probability that the input image is a true CT image.

The total number of trainable parameters in this discriminator architecture is 2,763,520. In training steps alternating with those of G, the parameters of D are updated using the gradient descent optimization algorithm discussed in Section 4.2.1.

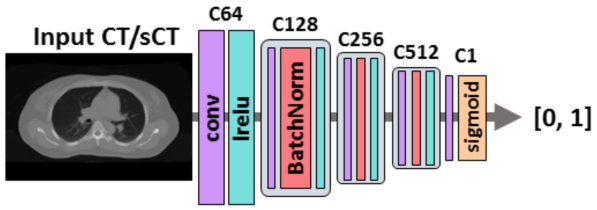

Figure 4.5. The discriminator architecture shared in each GAN framework. An input CT or sCT is encoded as a set of feature maps before being mapped to one channel by the last convolution operation. The sigmoid function is applied to yield values between zero and one that denote the probability that the input image is a true CT image.

4.2.3 Training Data

To determine the robustness of each G network architecture against the size of the training data set, both implementations were trained from scratch in five separate instances using image data from 10, 20, 30, 40, and 48 breast patients previously treated at our institution and evaluated after each completed training. 50 paired CT/MR images from a single CT and MRI scan for each patient were selected for training, yielding a training data set of 500, 1000, 1500, 2000, and 2400 paired images in each case respectively. 0.35T MRI setup scans (nominally 334 x 334 x 240 or 288 matrix, 1.5 x 1.5 x 1.5 mm^3) acquired using the first generation MRIdian system (ViewRay Inc., Oakwood Village, OH) with a T1w GR sequence were used in the present study. CT simulation scans (nominally 512 x 512 x 118–154 matrix, 0.977 x 0.977 x 3 mm^3) acquired using a dedicated simulation machine (Brilliance CT, Philips Medical Systems, Andover, MA) were downsampled to match the resolution of the corresponding MRI scan and deformably registered using the ViewRay treatment planning system before being exported for training. The 2D images used in each case were padded to a matrix size of 512 x 512 prior to training. Histogram matching was performed on both sets of images prior to thresholding images of each modality at specified values and rescaling the intensities to fill the entire unsigned 16-bit integer range to create a more uniform data set.

For every training set size and network architecture pairing, the framework was trained

for 2500 epochs using TensorFlow[57] v1.7.0 running on a 12 GB Titan Xp GPU (NVIDIA, Santa Clara, CA). The time required for training is presented in Section 4.3.

4.2.4 Evaluation

We evaluate the performance of both network architectures using a number of metrics. First, 10-fold cross validation was conducted for the aspp architecture to establish the performance of the network. We utilize three quantitative metrics to evaluate image quality: the root mean square error (RMSE) to capture voxel-wise agreement (Eq. 4.7), the structural similarity index (SSIM)[61] that evaluates structural agreement (Eq. 4.8), and the peak signal-to-noise ratio (PSNR) to measure the quality of sCT reconstructions (Eq. 4.12). The RMSE may be calculated as

$$RMSE = \sqrt{\frac{\sum_N \left(I_{pred} - I_{true}\right)^2}{N}}, \tag{4.7}$$

using the square of the pixel-wise difference between the predicted and true image over a region of interest containing N pixels. The SSIM considers comparisons in luminance, contrast, and structure to compute a similarity score between two images. Briefly, the metric may be computed between two images $\vec{x}$ and $\vec{y}$ as[61]

$$SSIM(\vec{x}, \vec{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{4.8}$$

where μ_x is the mean intensity of image $\vec{x}$:

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{4.9}$$

σ_x is an estimate of the signal contrast:

$$\sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu_x)^2 \right)^{1/2}, \tag{4.10}$$

σ_{xy} is correlation coefficient between $\vec{x}$ and $\vec{y}$:

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y), \tag{4.11}$$

and C_1 and C_2 are constants. Finally, the PSNR is defined as

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right), \tag{4.12}$$

where MAX_I is the maximum possible pixel value that depends on the data type of the image of interest and MSE is the mean square error calculated as the square of the value obtained using the same steps in Eq. 4.7.

A subsequent comparison between the sCT images generated by both architectures and true CT images for 12 test patients was performed using the RMSE, SSIM, and PSNR metrics. Additionally, the time required to complete 2500 training epochs was recorded for both architectures and each training data set size. Dose distributions calculated based on the electron density information derived from a full stack of sCT images generated with the proposed deep spatial convolutional framework were compared to those based on true CT images for 4 test patients using the 3D gamma index with 2%/2 mm criterion.[56] Dose calculations were performed in each case using the ViewRay treatment planning system integrated Monte Carlo algorithm without a magnetic field for 2.4 million histories with a 0.3 cm dose resolution.

4.3 Results

4.3.1 Cross Validation of ASPP

The performance of the deep spatial pyramid convolutional framework was evaluated using 10-fold cross validation. At each iteration, 54 data sets were used to train the architecture from scratch. The trained model was then applied to 5 data sets – approximately 500 images per iteration – for evaluation. Over all 10 folds, average values of 27.6 $\pm$ 7.2 HU, 0.9985 $\pm$ 0.0011, and 67.8 $\pm$ 2.2 were observed for the RMSE, SSIM, and PSNR metrics, respectively.

4.3.2 Training Performance Evaluation

The time required to complete 2500 training epochs – consisting of passes over the entire training set – is plotted for each architecture and training data set size in Figure 4.6, along with the generator loss (Eq. 4.4) calculated over the validation set at each epoch for the pix48 and aspp48 models. Shorter training times were observed for the aspp architecture compared to those of the pix architecture at every training set size, ranging from 21-87 and 37-207 hours, respectively, depending on the size of the data set.

1042 images from 12 test patients were generated using each architecture and training data set size. Values of the RMSE calculated within the body contour for each architecture and training data set size are plotted in Figure 4.7. The lowest mean value of 17.7 $\pm$ 4.3 HU is observed for the aspp architecture trained with data from 48 patients. Figure 4.8 displays the calculated values of the SSIM metric over this test set for each architecture and training data set size. The aspp48 model again demonstrates the best observed value of 0.9995 $\pm$ 0.0003, although relatively stable performance is observed in this metric. Finally, calculated results for the PSNR metric are presented in Figure 4.9. The highest observed value of 71.7 $\pm$ 2.3 is achieved by the aspp48 model. Additional statistics for each metric are presented in Tables 4.1, 4.2, and 4.3.

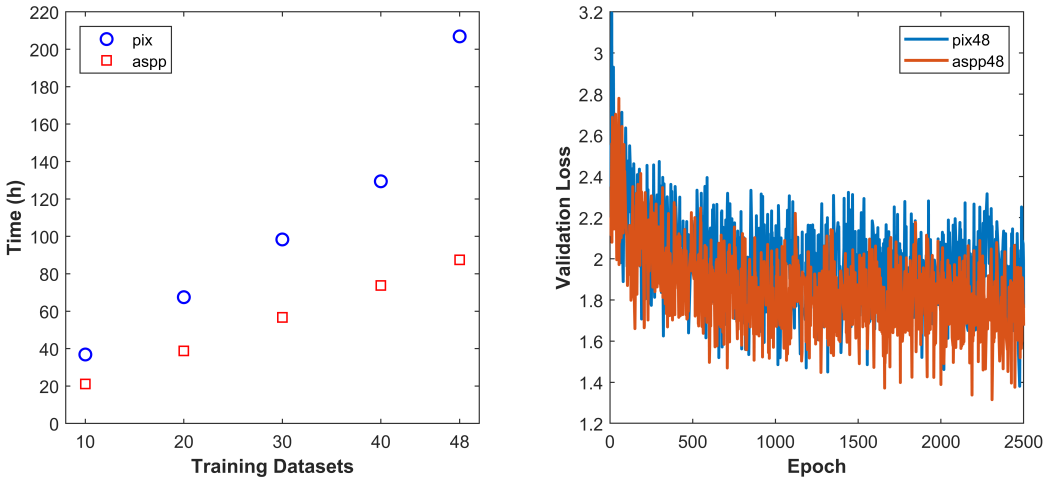

Figure 4.6. At left, the time to complete 2500 training epochs for the pix (blue) and aspp (red) architectures is plotted against the size of the training data set. At right, loss (Eq. 4.4) calculated over the validation set is plotted against training epoch for the pix48 (blue) and aspp48 (red) models.

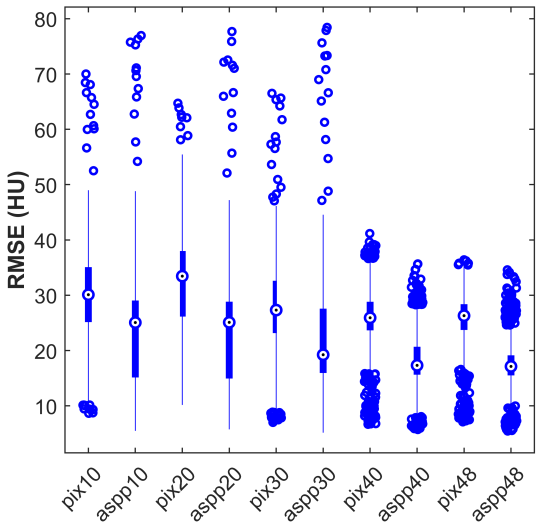

Figure 4.7. RMSE values calculated over 1042 test images generated by each architecture and training data set size. Statistical measures are included in Table 4.1.

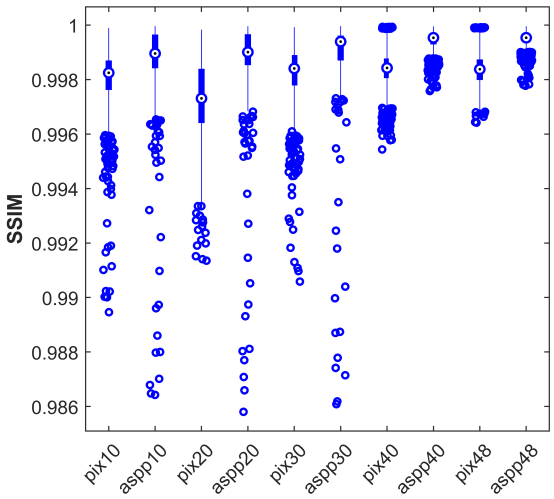

Figure 4.8. Values of the SSIM metric calculated over 1042 test images generated by each architecture and training data set size. Statistical measures are included in Table 4.2.

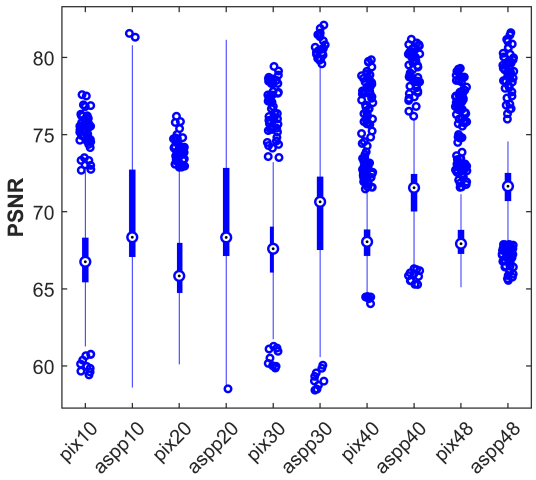

Figure 4.9. PSNR values calculated over 1042 test images generated by each architecture and training data set size. Statistical measures are included in Table 4.3.

Table 4.1. RMSE statistics accompanying Figure 4.7. Reported measures are in units of HU.

	pix10	aspp10	pix20	aspp20	pix30	aspp30	pix40	aspp40	pix48	aspp48
Mean	29.8 ± 8.6	23.9 ± 9.7	32.4 ± 8.5	23.8 ± 9.5	27.8 ± 8.1	22.0 ± 9.0	25.8 ± 5.8	18.6 ± 5.2	25.4 ± 5.2	17.7 ± 4.3
Median	30.1	25.1	33.5	25.1	27.3	19.2	25.9	17.3	26.3	17.1
Min	8.6	5.5	10.2	5.7	7.0	5.1	6.7	5.7	7.1	5.4
Max	70.0	77.0	64.7	77.7	66.5	78.4	41.2	35.7	36.4	34.6

Table 4.2. SSIM statistics accompanying Figure 4.8.

	pix10	aspp10	pix20	aspp20	pix30	aspp30	pix40	aspp40	pix48	aspp48
Mean	0.9980 ± 0.0013	0.9988 ± 0.0013	0.9972 ± 0.0015	0.9988 ± 0.0013	0.9982 ± 0.0012	0.9990 ± 0.0013	0.9984 ± 0.0008	0.9994 ± 0.0004	0.9984 ± 0.0007	0.9995 ± 0.0003
Median	0.9982	0.9990	0.9973	0.9990	0.9984	0.9994	0.9984	0.9995	0.9984	0.9995
Min	0.9895	0.9864	0.9914	0.9858	0.9906	0.9861	0.9954	0.9976	0.9964	0.9978
Max	0.9999	1.0000	0.9998	1.0000	0.9999	1.0000	0.9999	1.0000	0.9999	1.0000

Table 4.3. PSNR statistics accompanying Figure 4.9.

	pix10	aspp10	pix20	aspp20	pix30	aspp30	pix40	aspp40	pix48	aspp48
Mean	67.3 ± 2.8	69.5 ± 3.6	66.5 ± 2.6	69.5 ± 3.5	67.9 ± 2.9	70.1 ± 3.4	68.4 ± 2.5	71.3 ± 2.6	68.5 ± 2.4	71.7 ± 2.3
Median	66.8	68.3	65.8	68.3	67.6	70.6	68.1	71.6	67.9	71.7
Min	59.4	58.6	60.1	58.5	59.9	58.4	64.0	65.3	65.1	65.6
Max	77.6	81.6	76.2	81.1	79.4	82.1	79.9	81.2	79.3	81.6

4.3.3 Image Comparison

Image comparisons for a representative axial slice and a central coronal view of the full image stacks are presented in Figure 4.10. RMSE values of 17.7 HU and 27.3 HU for the sCT reconstructions of the selected slice for the aspp48 and pix48 results, respectively, are comparable to the mean values presented in Table 4.1. The input MR image is shown alongside the sCT reconstruction generated by the aspp48 and pix48 models as well as the corresponding ground truth CT image. Difference maps illustrate the difference in HU values in the true CT image and the sCT image for each architecture.

As an additional note, evaluation of the deep spatial pyramid convolutional framework has been extended to images of other MRI sequences and field strengths. Figure 4.11 presents the sCT reconstruction for a selected slice from an ASPP-based model trained with 1.5T mDixon MR images (Ingenia MR, Philips Medical Systems, Andover, MA) and the corresponding CT images of the breast. Training was performed in the same manner described previously with 1750 paired images from 35 patients previously treated at our institution.

4.3.4 Dose Calculation

Dose calculations were performed for 4 test patients using optimization parameters selected in clinical simulation plans and electron density information derived from aspp-generated sCT images. Dose distributions from the clinical and proposed plans are presented in Figure 4.12

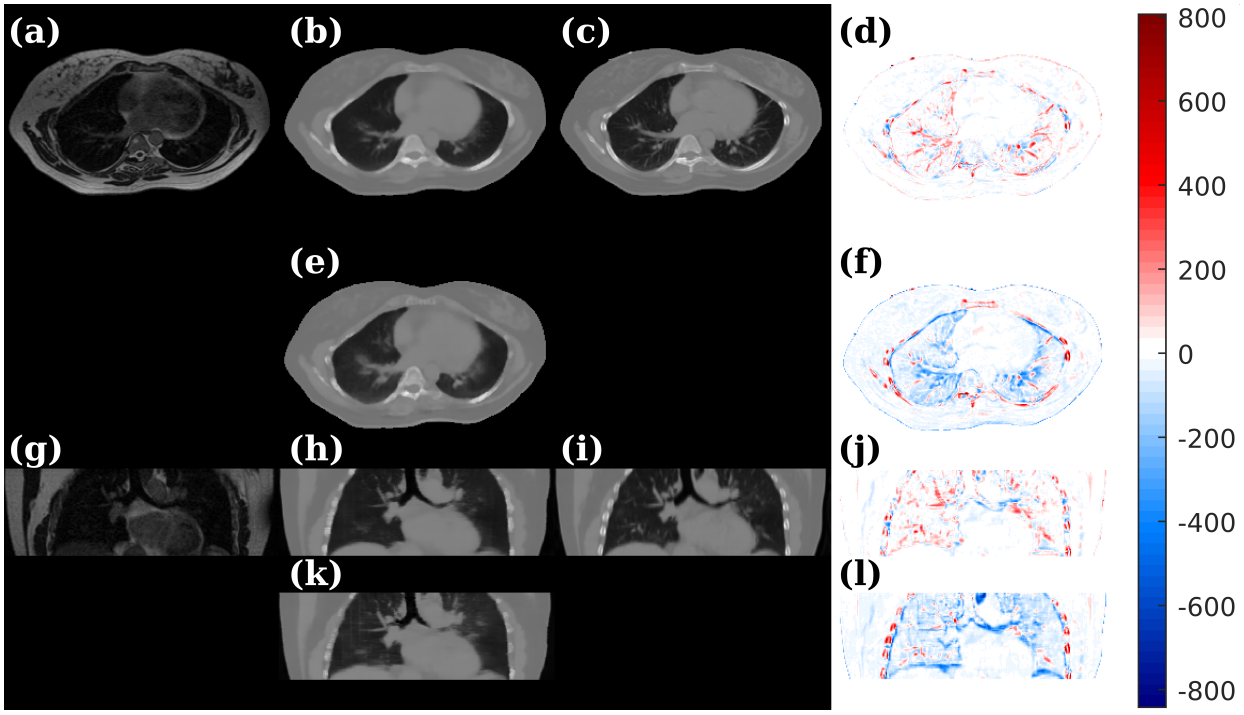

Figure 4.10. Input MR image (a), sCT image generated by the aspp48 model (RMSE = 17.7 HU) (b), true CT image (c), difference map (d) between (c) and (b) for a representative axial slice. Similarly, (e) and (f) are results for the same slice based on the pix48 (RMSE = 27.3 HU) model. Central coronal view of the full MRI stack (g), aspp48-generated sCT stack (h), true CT stack (i), difference map (j) between (i) and (h). Similarly, (k) and (l) are results for the same stack based on the pix48 model. Values in the difference maps are in units of HU.

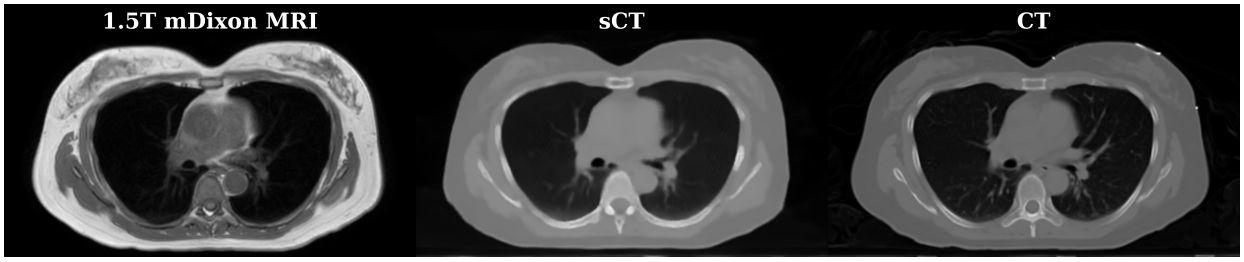

Figure 4.11. Input MR image (left) acquired in a 1.5T MRI scanner with an mDixon sequence, ASPP-based sCT image (center), and corresponding CT image (right) for a breast case.

for comparison along with the corresponding dose-volume histograms (DVHs). Calculated dose metrics including the percentage of the PTV covered by 95% of the prescribed dose (D_{95}), PTV max and mean dose, left and right lung max and mean dose, percent volume of the heart receiving 2 Gy or more (V_2), and heart max and mean dose for the clinical and proposed plans are presented in Table 4.4. D_{95} for the PTV calculated in the proposed plan varies by less than 1% from that calculated in the clinical plan in each test case. Similar agreement is observed in the other calculated metrics. Disagreements of greater than 1% arise from dose differences of less than 1 Gy, where discrepancies are largely inconsequential. The 3D gamma index comparing the clinical and proposed dose distributions yields passing rates equal to or greater than 98% for all patients using a 2%/2 mm criterion.

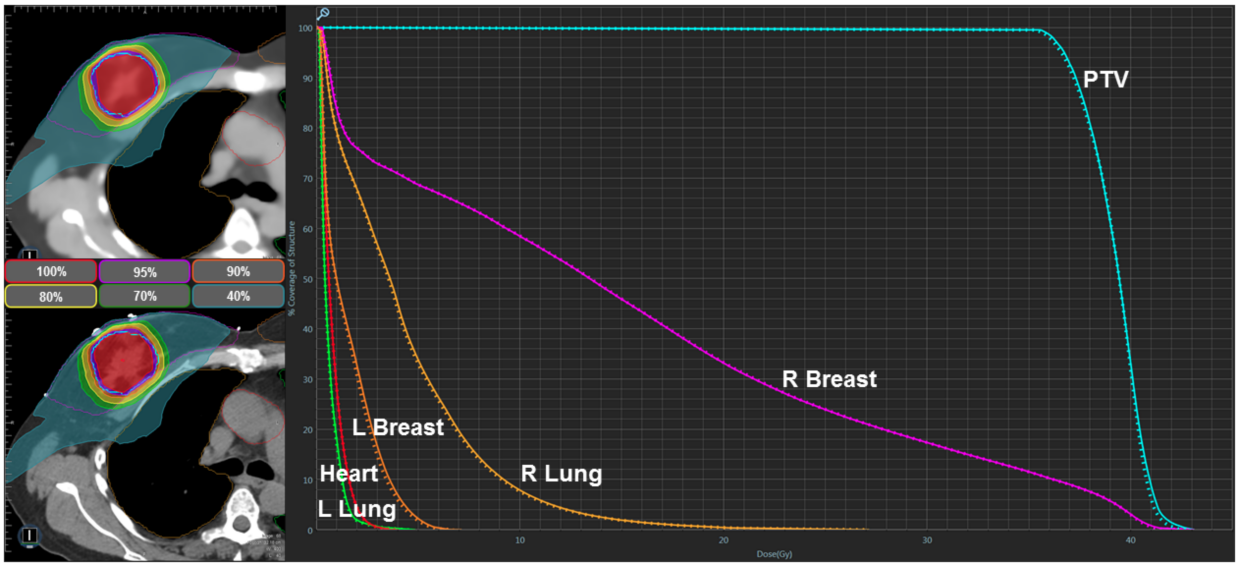

Figure 4.12. Dose distributions for the sCT-based (top left) and clinical (bottom left) plans alongside the plan DVHs for the proposed (solid lines) and clinical (dashed lines) plans.

4.4 Discussion

The value of our novel deep spatial pyramid framework that uses ASPP in a U-Net architecture is readily observed in the training time and image evaluation results. Looking

Table 4.4. Selected dose metrics compared between clinical (c.) simulation plans and proposed (p.) plans calculated based on aspp-generated sCTs for four test patients. Max and mean doses are reported in units of Gy.

Pt.		PTV			Left Lung		Right Lung		Heart		
		D_{95} (%)	Max	Mean	Max	Mean	Max	Mean	V_2 (%)	Max	Mean
1	c.	56.18	22	19.28	1.09	0.23	6.88	0.9	0	1.35	0.21
	p.	56.12	22	19.26	1.14	0.24	6.87	0.89	0	1.37	0.2
2	c.	97.88	42.12	39.27	4.15	0.83	21.8	3.85	5	4.32	0.69
	p.	98.66	42.59	39.51	4.22	0.85	21.97	3.9	5.09	4.32	0.68
3	c.	95.14	42.89	39.2	4.72	0.6	26.97	4.29	4.36	3.87	0.8
	p.	96.05	43.12	39.26	4.86	0.61	27.15	4.35	4.46	4.03	0.8
4	c.	97.27	43.33	39.46	3.1	0.228	29.36	4.16	4.25	4.52	0.9
	p.	97.45	43.2	39.48	3.31	0.29	29.45	4.21	4.48	4.54	0.91

specifically at results for networks trained with data from 48 patients, a drastic reduction in the time required to train the model is observed. The aspp48 network completed 2500 training epochs in 87.5 hours, nearly two-and-a-half times faster than the 207 hours required for the pix48 architecture to complete the same number of epochs. This reduction in training time came with nearly a 40% improvement in the MAE computed across all test images when the ASPP module is used instead of the conventional pix architecture. Ultimately, an MAE value of 16.1 $\pm$ 3.5 HU is observed over 1042 test images. These observed results compare favorably with those reported in the literature. Kraus et al[13] used an atlas-related method to achieve MAE values ranging from 29.9-66.6 HU in regions of bone, bladder, and soft tissue within the pelvic region. Other atlas-based techniques report values ranging from 40-200 HU for sites including the pelvis and cranium.[17,26] Han[48] used a simple U-Net architecture similar to the pix architecture discussed here consisting of 8 total layers to achieve an MAE of 84.8 $\pm$ 17.3 HU in intracranial cases. More recently, Maspero et al[49] used the pix2pix framework[55] in the general pelvis, reporting an MAE value of 61 $\pm$ 9 HU. Fu et al[50] applied a 3D CNN architecture in the male pelvis to achieve an MAE of 37.6 $\pm$ 5.1 HU. The improved performance of the deep spatial pyramid framework proposed here in image generation and

training time relative to pix-like frameworks may be explained with two reasons.

First, adopting the ASPP module offers a significant reduction in the total number of trainable parameters compared to the pix framework. In the clinical environment, where access to data for training may be difficult – especially when perfectly paired data is required – this reduction is crucial to achieving acceptable performance. The entire GAN framework that includes the pix architecture for G is parameterized by 57,172,352 total trainable parameters, 54,408,832 of which are related to G itself. The encoder layers of the pix architecture carry 19,536,640 parameters, while the decoder layers require 34,872,192 parameters. Looking now at the entirety of the proposed deep spatial pyramid framework, a reduction of nearly 70% in the total number of parameters is observed: from 57,172,352 to 17,844,353. Encoder layers in the aspp architecture for G require only 2,754,496 parameters. The ASPP module implemented here is parameterized by 8,915,968 values. Finally, the decoder layers require 3,410,369 parameters. In total, the aspp generator architecture, parameterized by 15,080,733 parameters, requires less than half of the number of parameters than the decoder in the pix architecture alone. The elimination of filter-dense layers from the interior layers of the encoder and decoder in favor of the ASPP module offers a significant reduction in the number of parameters that in turn drastically reduces the time and amount of data required to effectively train the model.

Second, employing atrous convolution widens the receptive fields of the convolutional kernels at a rate that is not possible in the conventional architecture. While the field of view of a 3x3 standard convolution kernel increase linearly in size as the feature maps are progressively downsampled in an encoder, applying atrous convolution drastically widens the receptive field even at low rates. Applying kernels of increasing rate in parallel to feature maps at the interior-most portion of the network allows for the effective capture of large-scale structural information in a way that is not possible in a conventional framework of the same depth. This effective encoding paired with the significant reduction in the number of

trainable parameters yields a robust architecture that is well suited to applications where training data is limited, which is often true for problems faced in the clinic.

It is important to note that with the reduction in the total number of trainable parameters may also come the concern of inaccurate fitting by the trained model. This potential issue has been monitored in the present study in two ways. First, 10-fold cross validation of the proposed framework was performed to establish the stable performance of the method with varied training and testing data sets. Second, prediction error in the validation set was observed in the course of training. In this way, model performance during training may be monitored and an appropriate stopping criterion may be established in order to avoid inaccurate model fitting.

A number of barriers still stand in the way of the full-scale adoption of methods for sCT generation like the one proposed here. Most notably, learning-based methods that require paired image data for training face the challenge of multi-modality image registration that is further complicated by considerations of anatomical changes with time. Anatomical variations that occur even in short time frames result in acquired images that are notably different between successive CT and MRI simulation scans.[62] This challenge is readily apparent in abdominal cases in which intestinal gas in the bowel may change size and position relative to the target frequently. As a result, the corresponding CT and MRI simulation scans are made largely incompatible, necessitating intensive manual intervention in the form of contouring for bulk density overrides for the purpose of dose calculations or image pairing in this case. Techniques for propagating intestinal gas between modalities may be applied to mitigate one facet of this problem.[41,49] Additionally, style-transfer methods exemplified by CycleGAN that use unpaired CT and MR images to learn the mapping between CT and MR "styles" may be applied to the problem of sCT generation in an effort to avoid the requirement of multi-modality image registration and the challenges that come with curating paired image data.[63]

Despite these challenges, the deep spatial pyramid framework proposed here demonstrates promise in the task of sCT generation for an MR-only RT workflow. Trials using higher quality 1.5T MR images are ongoing at our institution. Future studies should pursue the external validation of the proposed architecture in varied settings beyond the scope of the present study, examining the robustness of the approach against disease site, tumor shape and size, and an extensive patient population. Success in these endeavors would open the door to conducting a rigorous clinical validation of the MR-only radiation therapy workflow enabled by the method proposed here.

4.5 Conclusions

In this study, we have evaluated the robustness of the conventional pix2pix GAN framework that is ubiquitous in the image-to-image translation task as well as the novel deep spatial pyramid framework we propose here. The proposed framework demonstrates improved performance in metrics of training time and image quality, even in cases when training data is limited. The success of the framework in sCT generation is a promising step towards an MR-only RT workflow that eliminates the need for CT simulation and setup scans while enabling online adaptive therapy applications that are becoming ever more prevalent in MR-IGRT.

References

[1] Lagendijk JJW, Raaymakers BW, Van den Berg CAT, Moerland MA, Philippens ME, van Vulpen M. MR guidance in radiotherapy. *Phys Med Biol.* 2014;59(21):R349–R369. doi:10.1088/0031-9155/59/21/R349

[2] Mutic S, Dempsey JF. The ViewRay system: Magnetic resonance–guided and controlled radiotherapy. *Semin Radiat Oncol.* 2014;24(3):196–199. doi: 10.1016/j.semradonc.2014.02.008

[3] Fallone BG. The rotating biplanar linac-magnetic resonance imaging system. *Semin Radiat Oncol.* 2014;24:200–202. doi:10.1016/j.semradonc.2014.02.011

[4] Keall PJ, Barton M, Crozier S. The Australian magnetic resonance imaging-linac program. *Semin Radiat Oncol*. 2014;24:203–206. doi:10.1016/j.semradonc.2014.02.015

[5] Fischer-Valuck BW, Henke L, Green O, et al. Two-and-a-half-year clinical experience with the world's first magnetic resonance image guided radiation therapy system. *Adv Radiat Oncol*. 2017;2(3):485–493. doi:10.1016/j.adro.2017.05.006

[6] Pollard JM, Wen Z, Sadagopan R, Wang J, Ibbott GS. The future of image-guided radiotherapy will be MR guided. *Br J Radiol*. 2017;90(1073):20160667. doi: 10.1259/bjr.20160667

[7] Raaymakers BW, Jürgenliemk-Schulz IM, Bol GH, et al. First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. *Phys Med Biol*. 2017;62(23):L41–L50. doi:10.1088/1361-6560/aa9517

[8] Beavis AW, Gibbs P, Dealey RA, Whitton VJ. Radiotherapy treatment planning of brain tumours using MRI alone. *Br J Radiol*. 1998;71(845):544–548. doi: 10.1259/bjr.71.845.9691900

[9] Chen L, Price RA, Wang L, et al. MRI-based treatment planning for radiotherapy: Dosimetric verification for prostate IMRT. *Int J Radiat Oncol*. 2004;60(2):636–647. doi:10.1016/j.ijrobp.2004.05.068

[10] Rasch C, Steenbakkers R, van Herk M. Target definition in prostate, head, and neck. *Semin Radiat Oncol*. 2005;15(3):136–145. doi:10.1016/j.semradonc.2005.01.005

[11] Fiorentino A, Caivano R, Pedicini P, Fusco V. Clinical target volume definition for glioblastoma radiotherapy planning: magnetic resonance imaging and computed tomography. *Clin Transl Oncol*. 2013;15(9):754–758. doi:10.1007/s12094-012-0992-y

[12] Schmidt MA, Payne GS. Radiotherapy planning using MRI. *Phys Med Biol*. 2015; 60(22):R323–R361. doi:10.1088/0031-9155/60/22/R323

[13] Kraus KM, Jäkel O, Niebuhr NI, Pfaffenberger A. Generation of synthetic CT data using patient specific daily MR image data and image registration. *Phys Med Biol*. 2017;62(4):1358–1377. doi:10.1088/1361-6560/aa5200

[14] Nyholm T, Jonsson J. Counterpoint: Opportunities and challenges of a magnetic resonance imaging–only radiotherapy work flow. *Semin Radiat Oncol*. 2014;24(3):175–180. doi:10.1016/j.semradonc.2014.02.005

[15] Kapanen M, Collan J, Beule A, Seppälä T, Saarilahti K, Tenhunen M. Commissioning of MRI-only based treatment planning procedure for external beam radiotherapy of prostate. *Magn Reson Med*. 2013;70(1):127–135. doi:10.1002/mrm.24459

[16] Rank CM, Tremmel C, Hünemohr N, Nagel AM, Jäkel O, Greilich S. MRI-based treatment plan simulation and adaptation for ion radiotherapy using a classification-based approach. *Radiat Oncol*. 2013;8(1):51. doi:10.1186/1748-717X-8-51

[17] Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. *Radiat Oncol*. 2017;12(1):28. doi:10.1186/s13014-016-0747-y

[18] Wang C, Chao M, Lee L, Xing L. Mri-based treatment planning with electron density information mapped from CT images: A preliminary study. *Technol Cancer Res Treat*. 2008;7(5):341–347. doi:10.1177/153303460800700501

[19] Roberson PL, McLaughlin PW, Narayana V, Troyer S, Hixson GV, Kessler ML. Use and uncertainties of mutual information for computed tomography/magnetic resonance (CT/MR) registration post permanent implant of the prostate. *Med Phys*. 2005; 32(2):473–482. doi:10.1118/1.1851920

[20] Dean CJ, Sykes JR, Cooper RA, et al. An evaluation of four CT–MRI co-registration techniques for radiotherapy treatment planning of prone rectal cancer patients. *Br J Radiol*. 2012;85(1009):61–68. doi:10.1259/bjr/11855927

[21] Nyholm T, Nyberg M, Karlsson MG, Karlsson M. Systematisation of spatial uncertainties for comparison between a MR and a CT-based radiotherapy workflow for prostate treatments. *Radiat Oncol*. 2009;4(1):54. doi:10.1186/1748-717X-4-54

[22] Daisne JF, Sibomana M, Bol A, Cosnard G, Lonneux M, Grégoire V. Evaluation of a multimodality image (CT, MRI and PET) coregistration procedure on phantom and head and neck cancer patients: accuracy, reproducibility and consistency. *Radiother Oncol*. 2003;69(3):237–245. doi:10.1016/j.radonc.2003.10.009

[23] Ulin K, Urie MM, Cherlow JM. Results of a multi-institutional benchmark test for cranial CT/MR image registration. *Int J Radiat Oncol*. 2010;77(5):1584–1589. doi: 10.1016/j.ijrobp.2009.10.017

[24] Bezrukov I, Mantlik F, Schmidt H, Schölkopf B, Pichler BJ. MR-based PET attenuation correction for PET/MR imaging. *Semin Nucl Med*. 2013;43:45–59. doi: 10.1053/j.semnuclmed.2012.08.002

[25] Mehranian A, Arabi H, Zaidi H. Vision 20/20: Magnetic resonance imaging-guided attenuation correction in PET/MRI: Challenges, solutions, and opportunities. *Med Phys*. 2016;43:1130–1155. doi:10.1118/1.4941014

[26] Johnstone E, Wyatt JJ, Henry AM, et al. Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging–only radiation therapy. *Int J Radiat Oncol*. 2018;100(1):199–217. doi:10.1016/j.ijrobp.2017.08.043

[27] Dowling JA, Lambert J, Parker J, et al. An atlas-based electron density mapping method for magnetic resonance imaging (mri)-alone treatment planning and adaptive mri-based prostate radiation therapy. *Int J Radiat Oncol Biol Phys*. 2012;83(1):e5–e11. doi:10.1016/j.ijrobp.2011.11.056

[28] Uh J, Merchant TE, Li Y, Li X, Hua C. MRI-based treatment planning with pseudo CT generated through atlas registration. *Med Phys*. 2014;41(5):051711. doi: 10.1118/1.4873315

[29] Sjölund J, Forsberg D, Andersson M, Knutsson H. Generating patient specific pseudo-CT of the head from MR using atlas-based regression. *Phys Med Biol*. 2015;60(2):825–839. doi:10.1088/0031-9155/60/2/825

[30] Arabi H, Koutsouvelis N, Rouzaud M, Miralbell R, Zaidi H. Atlas-guided generation of pseudo-CT images for MRI-only and hybrid PET–MRI-guided radiotherapy treatment planning. *Phys Med Biol*. 2016;61(17):6531–6552. doi:10.1088/0031-9155/61/17/6531

[31] Demol B, Boydev C, Korhonen J, Reynaert N. Dosimetric characterization of MRI-only treatment planning for brain tumors in atlas-based pseudo-CT images generated from standard T1-weighted MR images. *Med Phys*. 2016;43(12):6557–6568. doi: 10.1118/1.4967480

[32] Ren S, Hara W, Wang L, et al. Robust estimation of electron density from anatomic magnetic resonance imaging of the brain using a unifying multi-atlas approach. *Int J Radiat Oncol*. 2017;97(4):849–857. doi:10.1016/j.ijrobp.2016.11.053

[33] Arabi H, Dowling JA, Burgos N, et al. Comparative study of algorithms for synthetic CT generation from MRI: Consequences for MRI-guided radiation planning in the pelvic region. *Med Phys*. 2018;45:5218–5233. doi:10.1002/mp.13187

[34] Jonsson JH, Karlsson MG, Karlsson M, Nyholm T. Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions. *Radiat Oncol*. 2010;5(1):62. doi:10.1186/1748-717X-5-62

[35] Korhonen J, Kapanen M, Keyriläinen J, Seppälä T, Tenhunen M. A dual model HU conversion from MRI intensity values within and outside of bone segment for MRI-based radiotherapy treatment planning of prostate cancer. *Med Phys*. 2013;41(1):011704. doi: 10.1118/1.4842575

[36] Paulus DH, Quick HH, Geppert C, et al. Whole-body PET/MR imaging: Quantitative evaluation of a novel model-based mr attenuation correction method including bone. *J Nucl Med*. 2015;56(7):1061–1066. doi:10.2967/jnumed.115.156000

[37] Kim J, Garbarino K, Schultz L, et al. Dosimetric evaluation of synthetic CT relative to bulk density assignment-based magnetic resonance-only approaches for prostate radiotherapy. *Radiat Oncol*. 2015;10(1):239. doi:10.1186/s13014-015-0549-7

[38] Prior P, Chen X, Botros M, et al. MRI-based IMRT planning for MR-linac: comparison between CT- and MRI-based plans for pancreatic and prostate cancers. *Phys Med Biol*. 2016;61(10):3819–3842. doi:10.1088/0031-9155/61/10/3819

[39] Wang H, Chandarana H, Block KT, Vahle T, Fenchel M, Das IJ. Dosimetric evaluation of synthetic CT for magnetic resonance-only based radiotherapy planning of lung cancer. *Radiat Oncol*. 2017;12(1):108. doi:10.1186/s13014-017-0845-5

[40] Kim SW, Shin HJ, Hwang JH, et al. Image similarity evaluation of the bulk-density-assigned synthetic CT derived from MRI of intracranial regions for radiation treatment. *PLoS One*. 2017;12(9):e0185082. doi:10.1371/journal.pone.0185082

[41] Maspero M, van den Berg CAT, Landry G, et al. Feasibility of MR-only proton dose calculations for prostate cancer radiotherapy using a commercial pseudo-CT generation method. *Phys Med Biol*. 2017;62(24):9159–9176. doi:10.1088/1361-6560/aa9677

[42] Wang H, Du K, Qu J, Chandarana H, Das IJ. Dosimetric evaluation of magnetic resonance-generated synthetic CT for radiation treatment of rectal cancer. *PLoS One*. 2018;13(1):e0190883. doi:10.1371/journal.pone.0190883

[43] Johansson A, Garpebring A, Karlsson M, Asklund T, Nyholm T. Improved quality of computed tomography substitute derived from magnetic resonance (MR) data by incorporation of spatial information – potential application for MR-only radiotherapy and attenuation correction in positron emission tomography. *Acta Oncol (Madr)*. 2013; 52(7):1369–1373. doi:10.3109/0284186X.2013.819119

[44] Rank CM, Hünemohr N, Nagel AM, Röthke MC, Jäkel O, Greilich S. MRI-based simulation of treatment plans for ion radiotherapy in the brain region. *Radiother Oncol*. 2013;109(3):414–418. doi:10.1016/j.radonc.2013.10.034

[45] Edmund JM, Kjer HM, Van Leemput K, Hansen RH, Andersen JA, Andreasen D. A voxel-based investigation for MRI-only radiotherapy of the brain using ultra short echo times. *Phys Med Biol*. 2014;59(23):7501–7519. doi:10.1088/0031-9155/59/23/7501

[46] Gudur MSR, Hara W, Le QT, Wang L, Xing L, Li R. A unifying probabilistic Bayesian approach to derive electron density from MRI for radiation therapy treatment planning. *Phys Med Biol*. 2014;59(21):6595–6606. doi:10.1088/0031-9155/59/21/6595

[47] Yang X, Lei Y, Shu HK, et al. Pseudo CT estimation from MRI using patch-based random forest. p. 101332Q. doi:10.1117/12.2253936

[48] Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys*. 2017;44(4):1408–1419. doi:10.1002/mp.12155

[49] Maspero M, Savenije MHF, Dinkla AM, et al. Fast synthetic CT generation with deep learning for general pelvis MR-only radiotherapy. *ArXiv*. 2018;:1–14

[50] Fu J, Yang Y, Singhrao K, Ruan D, Low DA, Lewis JH. Male pelvic synthetic CT generation from T1-weighted MRI using 2D and 3D convolutional neural networks. *ArXiv*. 2018;:1–13

[51] Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. *ArXiv*. 2018;:1–12

[52] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *ArXiv*. 2015;:1–8

[53] Chen S, Qin A, Zhou D, Yan D. Technical Note: U-net-generated synthetic CT images for magnetic resonance imaging-only prostate intensity-modulated radiation therapy treatment planning. *Med Phys*. 2018;45:5659–5665. doi:10.1002/mp.13247

[54] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. *ArXiv*. 2014;:2672–2680

[55] Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. *ArXiv*. 2017;:1–17

[56] Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. *Med Phys*. 1998;25(5):656–661. doi:10.1118/1.598248

[57] Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-scale machine learning on heterogeneous systems. 2015. Software available from tensorflow.org

[58] Kingma D, Ba J. Adam: A method for stochastic optimization. *ArXiv*. 2017;:1–15

[59] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15:1929–1958

[60] Chen LC, Papandreou G, Schroff F, Adam H. Rethinking atrous convolution for semantic image segmentation. *ArXiv*. 2017;:1–14

[61] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13:1–14

[62] Olberg S, Green O, Cai B, et al. Optimization of treatment planning workflow and tumor coverage during daily adaptive magnetic resonance image guided radiation therapy (MR-IGRT) of pancreatic cancer. *Radiat Oncol*. 2018;13:1–8

[63] Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. *ArXiv*. 2018;:1–18

[64] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *ArXiv*. 2017;:1–14

Chapter 5

Abdominal synthetic CT reconstruction with intensity projection prior for MRI-only adaptive radiotherapy

5.1 Introduction

The proliferation of magnetic resonance imaging (MRI)-based radiotherapy delivery systems in recent years has pushed applications of MRI-guided radiation therapy (MRgRT) to the forefront of RT.[1–7] The superior soft tissue contrast of MRI compared to that of x-ray computed tomography (CT) improves target delineation in many sites, enabling near real-time motion tracking and management during treatment.[5,8] The advantages of MRI also lend themselves well to managing interfractional changes in a patient's anatomy, as MRI setup scans acquired in an adaptive radiotherapy (ART) workflow capture the anatomy of the day without exposing the patient to additional ionizing radiation.[4] The need for electron density information in dose calculations, however, necessitates that these setup scans be registered to a CT simulation scan that may have been acquired weeks prior to a given treatment fraction.[9] Issues may arise in some cases when the anatomy represented in each scan is incompatible due to changes in the geometry and position of organs of interest, which is a challenge that is especially relevant to gastrointestinal (GI) structures and intestinal gas pockets considered during ART in the abdomen. In these cases, the ideal, unconstrained MRgRT workflow would involve intensive manual contouring so that density overrides could be performed in order to approximate in the CT simulation scan the anatomy of the day captured in the MRI setup scan.

It is this potential for challenge and uncertainty that comes with multi-modal image registration and intensive contouring that makes an MRI-only workflow—one in which MRI is the sole imaging modality used for planning and guidance—an attractive alternative to the conventional MRgRT workflow. The primary challenge in an MRI-only workflow is generating synthetic CT (sCT) data that yields the electron density information used in dose calculations. Many existing approaches to this task have been summarized as belonging to three general classes: atlas-based, voxel-based, and learning-based methods.[10,11] Recent

investigations have focused primarily on approaches belonging to the last category, namely deep learning (DL)-based approaches in which convolutional neural networks are used to approximate a mapping between MRI and CT images.[12]

Studies in this space have primarily been limited to regions of relatively static anatomy, including the head & neck[13] and general pelvis.[14–16] More recently, our group has extended these investigations into the thorax, where registration is challenged by pulmonary and cardiac motion, by generating sCT data for use in MRI-only breast RT.[17] When considering applications in the more dynamic region of the abdomen, a primary challenge to employing these deep learning strategies becomes readily apparent: the majority of frameworks require paired data for training. The presence and hard-to-characterize motion of intestinal gas gives rise to observable differences in bowel filling and position on numerous time scales: seconds in the course of a single scan, minutes during treatment delivery, and hours between MRI and CT scans used for treatment planning.[18–22] This challenge is even more relevant clinically when considering setup scans in MRI-guided adaptive treatments. In these situations, MRI setup scans acquired the day of treatment are registered to the CT simulation scan that may have been acquired weeks previously. When large discrepancies in bowel filling and position exist, the MRI setup scan and simulation CT scan may be rendered largely incompatible. Corrective measures that could be undertaken in an online adaptive workflow represent a potentially significant delay to treatment delivery while the patient remains on-table.[23] As an alternative to the paired data approaches, one may consider the application of an unpaired style-transfer approach exemplified by CycleGAN in settings in which abundant, unpaired data exists.[12,24] However, the authors of the original CycleGAN paper acknowledge a gap between the paired and unpaired results that is hard or even impossible to close in some settings, especially those in which there exists some inherent ambiguity.[24] This conclusion is mirrored in other studies of the sCT reconstruction task; Peng et al.[25] conclude that the conditional paired approach was "preferable if high-quality MR-CT pairs were available" in

the nasopharynx, which is another area challenged by the presence of air, and Fu et al.[26] demonstrate no benefit to adopting the CycleGAN in the upper abdomen for liver cancer patients.

In light of these challenges, many existing studies on sCT reconstruction in the abdomen have abandoned the DL-based approaches in favor of various classification or thresholding-based approaches wherein manual steps are taken in the process of generating sCT data to account for the presence of air. Bredfeldt et al[27] and Hsu et al[28] utilized a fuzzy c-means clustering algorithm to classify tissue classes based on multiple MRI volumes captured using different sequences, taking care to threshold image regions where air is expected to be found before applying the classification algorithm. Alternatively, Ahunbay et al[29] opted to use deformable image registration between the daily MRI and simulation CT scans to transfer electron density information while using thresholding operations in manually contoured regions to identify the presence of air. Guerreiro et al[30] explored a hybrid atlas and intensity-based conversion algorithm[31–33] in which contoured regions of air were transferred directly from the simulation CT after Hounsfield units (HU) were assigned. Most recently, both Cusumano et al.[34] and Qian et al.[35] have reported experiences with DL-based sCT reconstruction in the abdomen. Neither study, however, deals explicitly with the issue of air discrepancies between corresponding MRI and CT scans. Cusumano et al.[34] instead employed exclusion criteria to select only MR and CT images with "high correspondence in terms of air pockets location between the two images" for training and excluded a further six patients from the test set for the reason of poor correspondence of air between the two modalities. Similarly, Qian et al.[35] do not discuss results for patients with poorly matched representations of air.

In the present study, we return to paired data-driven DL with a novel hybrid approach to the abdominal sCT reconstruction task enabled by the creation of a training data set that is clinically unavailable. As was previously discussed, the primary barrier to the adoption

of many DL-based algorithms in this setting is the requirement for paired training data. Mismatches in the presence of intestinal gas between corresponding MRI and CT scans render the collection of a training data set of sufficient size an impossible task. Considering this and the challenges faced with an unpaired approach, we first utilize automated thresholding and morphological reconstruction operations to identify and propagate regions of air between corresponding MR and CT images to produce a well-matched training data set. We present here the preliminary evaluation of our paired data DL-based approach to sCT reconstruction in the abdomen enabled by the novel utilization of the intensity projection prior with a focus on showcasing the effects of intestinal gas differences on dose calculations. Dosimetric comparisons are made between two classes of test patients separated by a qualitative distinction made at the time of data collection prior to any evaluation: Class 1, consisting of well-matched patients demonstrating little involvement of intestinal gas, and Class 2, consisting of patients characterized by notable differences in the presence of intestinal gas in corresponding MRI and CT scans. Dosimetric accuracy is established using the patients of Class 1 while comparisons of target coverage between the sCT-based plans and the simulation CT-based clinical plans for patients of Class 2 highlight the complications posed by intestinal gas during MRI-only ART in the abdomen.

5.2 Materials and Methods

5.2.1 Patient population

Data sets used in the present study were retrospectively collected from a population of pancreatic cancer patients previously treated at our institution using MRgRT. In each case, patients underwent CT and MRI simulation scans prior to treatment planning. Scans were acquired in treatment position using an Alpha Cradle (Smithers Medical Products Inc., North Canton, OH) with no additional immobilization devices. The nominal prescription in

the selected population was a total dose of 50 Gy delivered to 95% of the PTV in 5 fractions. A small number of cases deviated from this prescription with total prescribed doses ranging from 30–40 Gy delivered in 5 fractions. For the nominal prescription, satisfying dose-volume constraints of <0.5 cm^3 at 36 Gy for notable structures including the duodenum, stomach, small bowel, and large bowel was prioritized over target coverage. In any case in which one of these constraints was violated, the calculated dose was normalized to satisfy the violated constraint. The same normalization applied in the clinical case was also applied in the evaluation of our proposed method as discussed later. In all patient cases involved here, dose calculations in the clinical plan were performed using electron density data derived directly from the simulation CT scan with no additional density overrides included to account for the presence of air.

5.2.2 Training and testing data

A total of 89 patient data sets (one pair of corresponding MRI and CT scans per patient) were used in the present study, randomly assigned in the following splits: 53 train / 3 valid / 33 test. Ten-fold validation of the framework in the sCT reconstruction task has been previously carried out by our group, and an additional 3-holdout validation is performed here.[17] The 33 test patients were qualitatively subdivided at the time of data collection prior to testing into the two classes mentioned previously: 13 well-matched patients in Class 1 and 20 patients characterized by notable discrepancies in bowel filling in Class 2. CT simulation scans acquired using a dedicated simulation machine (Brilliance CT, Philips Medical Systems, Andover, MA) were registered with 0.35 T MRI scans (nominally $276 \times 276 \times 80$ matrix, $1.63 \times 1.63 \times 3$ mm^3) acquired using the MRIdian system (ViewRay Inc., Oakwood Village, OH) with a bSSFP sequence before being exported for pre-processing. Processing yielded a training data set of 2017 paired images, which were padded to dimensions of 520×520 before training via 320×320 random crops. The framework was trained for 1500 epochs

with a batch size of 1 using TensorFlow[36] v1.7.0 in Python running on a 12 GB Titan Xp GPU (NVIDIA, Santa Clara, CA).

5.2.3 Pre-processing

A primary challenge in training a generative model to solve an image-to-image translation task such as this is constructing a set of training data consisting of well-matched pairs of images. This process becomes even more complicated in the abdomen when the variable presence of intestinal gas must be considered. Corresponding MRI and CT scans used in treatment planning may demonstrate notable mismatches in bowel filling and position that present a barrier to accurate dose calculations. While these mismatches could be handled during the treatment planning process in the clinical setting through intensive manual contouring to enable electron density overrides, we opt to avoid this entirely through the creation of an intensity projection prior. Here we adopt a novel approach to data augmentation for paired data DL applications in which incompatible representations of intestinal gas in corresponding MRI and CT scans are made compatible through the propagation of air from MRI to CT images. The handling of intestinal gas proceeded in the following steps:

(1) Corresponding MRI and CT scans were rigidly registered in the ViewRay treatment planning system (ViewRay Inc., Oakwood Village, OH) to achieve a gross alignment, primarily of bony anatomy.

(2) Using automated thresholding and morphological reconstruction operations, regions of air in each scan were identified. For CT images, Otsu's method[37] was used to compute a single threshold value with which the image could be quantized to produce a body mask that excludes most notably the couch. Within this mask, regions of air were thresholded with a histogram shape-based method by selecting intensities falling within an offset (defined as 7 bin widths here, nominally 140 HU) around the lower mode. Finally, binary erosion and dilation operations were performed to eliminate small, noisy regions and produce a smoother

segmentation, respectively. For MR images, a similar approach was taken to produce a body mask. In identifying regions of air within this mask, extra precautions were required to avoid selecting low-signal regions not containing air (e.g. vertebral body). To this end, the erosion of a quantized image based on a five-level thresholding achieved with Otsu's method was used as the basis for the morphological reconstruction[38] of segmented regions within the abdominal cavity as defined by the space surrounded by the body wall and excluding the region around the vertebral body.

(3) These regions in the CT images were infilled with realistic texture via harmonic inpainting to produce a CT image with no air-containing regions.[39,40]

(4) Regions of air identified in the MR images in step (2) are then propagated to the corresponding CT images. As such, regions of air originally represented in the CT image but not the corresponding MR image maintain the infilled texture from step (3). Finally, the scans are deformably registered by way of a statistical dependence measure algorithm.[41]

5.2.4 Model and loss formulation

The task of sCT reconstruction viewed as a forward mapping from MRI to CT has been previously discussed.[17] Briefly, the goal in establishing a generative model is to estimate a suitable operator that maps from MRI to CT, which is challenged by the many-to-one correspondence of pixel intensities between the two modalities. We approach the image-to-image translation task using a generative adversarial network (GAN) framework consisting of two competing networks: (1) a generative model G that produces sCT samples residing in the same space as true CT data and (2) a discriminator D that attempts to distinguish between samples generated by G and true samples. During training, G and D undergo alternating minimization steps of their respective loss function, each of which depends on the generalized definition of sigmoid cross entropy loss:[36]

$$L = \vec{x} - \vec{x} * \vec{z} + \log\left(1 + \exp\left(-\vec{x}\right)\right), \tag{5.1}$$

where $\vec{x}$ is the true or predicted image logits computed by D and $\vec{z}$ is a label corresponding to true ($\vec{1}$) or predicted ($\vec{0}$). Using this definition of sigmoid cross entropy loss, the generator loss function g_{loss} is defined by

$$g_{loss} = L_{adv} + l_{mae}, \tag{5.2}$$

where the adversarial loss L_{adv} is the sigmoid cross entropy loss (Eq. 5.1) with predicted image logits $\vec{x}$ assigned a true label ($\vec{z} = \vec{1}$) and the mean absolute error (MAE) loss l_{mae} is the mean of the absolute difference between true images I_{true} and predicted images I_{pred}:

$$l_{mae} = \text{mean}\left(\left|I_{pred} - I_{true}\right|\right). \tag{5.3}$$

Minimizing such a loss formulation yields synthetic images that are computed as true images by the discriminator through the adversarial term L_{adv} while also maintaining pixel-wise agreement between the generated and true images through the MAE term l_{mae}.

While true labels are assigned to predicted images in g_{loss}, the discriminator aims to correctly identify true and predicted images. As such, the discriminator loss function d_{loss} depends only on the sigmoid cross entropy loss:

$$d_{loss} = L_{pred} + L_{true}, \tag{5.4}$$

where L_{true} and L_{pred} are the sigmoid cross entropy loss (Eq. 5.1) with true or predicted image logits $\vec{x}$ and corresponding labels $\vec{z} = \vec{1}$ or $\vec{z} = \vec{0}$, respectively. While G strives to generate outputs computed as true images by D through the adversarial loss term L_{adv} with predicted image logits assigned a true label, the expected true and false labels are instead

used for the true and predicted CT images in the formulation of d_{loss}.

Trainable paramaters describing the various operations in each of the competing networks are optimized through alternating minimization steps of the respective loss functions. For d_{loss}, TensorFlow's[36] gradient descent optimizer is used with an initial learning rate of 0.00002. In the following minimization step, g_{loss} is minimized utilizing the Adam gradient-based stochastic optimization algorithm[42] with an initial learning rate of 0.0002, $\beta_1 = 0.7$, $\beta_2 = 0.999$, and $\hat{\epsilon} = 10^{-8}$. In both cases, learning rates decay every 10000 steps subject to a staircase exponentional function with a decay rate of 0.95.

5.2.5 Network architecture

Generator

The fully convolutional DenseNet[43] employed here, illustrated in Figure 5.1, consists of individual dense blocks arranged to form a stacked encoder-decoder U-net[44] structure. The constituent dense blocks, which resemble residual blocks[45] in that intermediate feature maps are iteratively concatenated, are built using the following components: Batch Normalization (BN), ReLU activation, 3×3 same convolution, and dropout with probability $p = 0.2$. The growth rate k of the layer ($k = 16$ in this case) dictates the number of feature maps computed by each layer. Feature maps computed by each of these intermediate layers are iteratively concatenated to form the output of the dense block itself, granting a degree of convergence-aiding supervision due to the short paths to all feature maps in an architecture that is ultimately quite parameter efficient.[43] Transition down (TD) operations in the encoder path, which serve to reduce the dimensionality of feature maps, consist of BN, followed by ReLU activation, 1×1 convolution, dropout with probability $p = 0.2$, and 2×2 max pooling with stride 2. To recover the input dimensions on the decoder path, transition up (TU) layers perform 3×3 transpose convolution with stride 2. Skip connections between corresponding

layers of the encoder and decoder sides of the network transfer structural information that aids in the reconstruction of fine detail as the full input resolution is recovered along the decoder path.

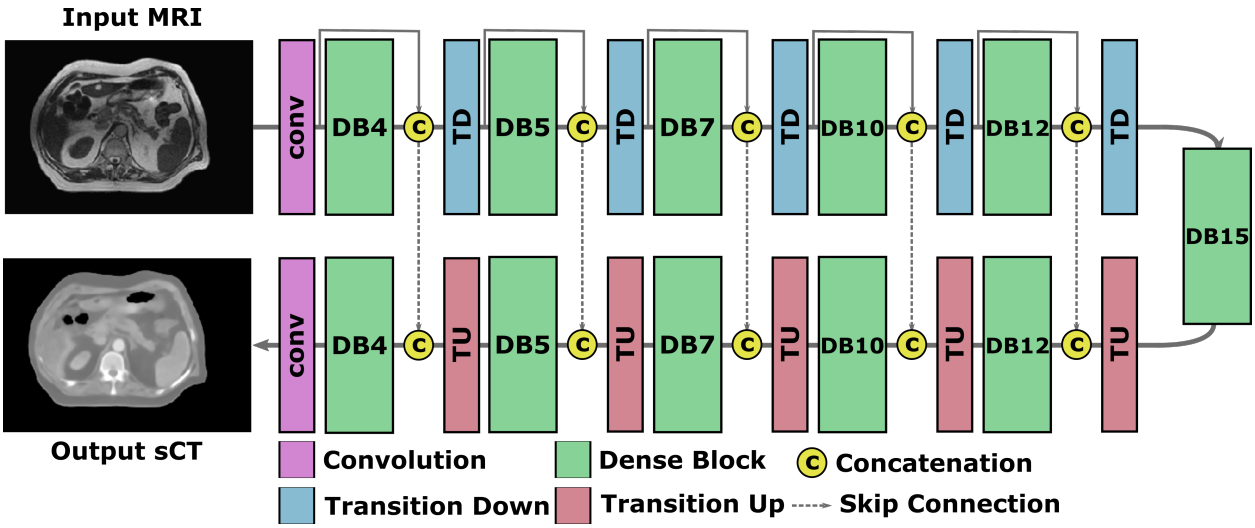

Figure 5.1. The DenseNet architecture. Individual dense blocks are arranged to form a stacked encoder-decoder U-net structure. DBn denotes a dense block consisting of n intermediate layers. The input MR image is encoded as a set of feature maps that grows progressively deeper as it travels through the encoder layers. Transpose convolution operations in the decoder recover the input spatial resolution, reconstructing details in the output sCT image with the aid of skip connections that transfer structural information from the encoder.

Discriminator

The architecture of the discriminator is unchanged from the previous application to sCT generation in the breast.[17] D is a straightforward encoder consisting of five convolutional layers that ultimately applies the sigmoid function to yield the probability of the evaluated image being a true CT image.

5.2.6 Evaluation

The proposed approach to sCT reconstruction is evaluated in two primary ways, each with a focus on the two classes of patients previously discussed. Pixel-wise image comparisons are made between true CT images and reconstructed sCT images for patients belonging to each class using the MAE and mean absolute percentage error (MAPE) measured in regions within the body contour not containing air:

$$\text{MAE} = \frac{\sum_{i=1}^{n} |CT_i - sCT_i|}{n}, \text{ and} \tag{5.5}$$

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{CT_i - sCT_i}{CT_i} \right|, \tag{5.6}$$

where n is the number of pixels not containing air in both the CT reference image and the generated sCT image. These image comparisons are made for sCT outputs of both the model proposed here and a "blind" model trained with the same patient data only without the pre-processing treatment of regions of air, which represents the conventional DL-based approach to the present problem. Considering the fact that the reference CT image for patients of Class 2 may be largely incompatible with the corresponding MRI image, we also evaluate the degree of overlap of regions of air in input MR images and reconstructed sCT outputs of each model using the Dice similarity coefficient (DSC):

$$\text{DSC} = \frac{2\,|X \cap Y|}{|X| + |Y|}, \tag{5.7}$$

where X and Y are the sets of pixels in air masks of an MR image and corresponding sCT image. Additionally, the structural similarity index (SSIM), which assesses similarity through three distinct luminance, contrast, and structure terms,[17] is calculated between the input MR image and reconstructed sCT output of each model in these air-containing regions.

A subsequent dosimetric evaluation compares dose distributions calculated at the first treatment fraction in simulation CT-based clinical plans to those recalculated using sCT-derived electron density information and the same optimization parameters used in the clinical plans for each of the 33 test patients. Full dose-volume histograms (DVHs) for the target and surrounding tissues of well-matched patients in Class 1 are used to first establish the baseline accuracy of the proposed approach to sCT reconstruction. The same comparison is made for patients in Class 2 demonstrating notable discrepancies in the presence of intestinal gas between corresponding MRI and CT scans to explore the effect of these discrepancies. For both patient classes, we examine differences in prescribed dose coverage of the PTV between the clinical CT-based plans and the proposed sCT-based plans. The 3D gamma index with a 3%/3 mm criterion is computed to evaluate agreement between the CT-based and sCT-based dose distributions for patients in each class.[46] Additionally, mean DVH differences for each structure of interest are computed and evaluated for statistical significance in each patient class. For patients of Class 2, the full width at half maximum (FWHM) of the profile of the difference in calculated target coverage is used to evaluate the uncertainty in high-dose coverage of the target due to the involvement of intestinal gas. Dose calculations in each case were performed using the ViewRay treatment planning system integrated Monte Carlo algorithm in the presence of a magnetic field with a dose grid resolution of 0.3 cm and calculation uncertainty of 1%.

5.3 Results

5.3.1 Image comparison

Completing 1500 epochs during training required 126 hours in total. At deployment, inference requires approximately 0.26 s/slice.

Each of the following comparisons shows an input MR image along with the sCT re-

construction produced by the blind and proposed models and the corresponding true CT image. Difference maps illustrate differences in pixel intensities between the true CT image and sCT image in units of HU. Axial slices for representative patients belonging to Class 1 are displayed in Figure 5.2. Difference maps (Figure 5.2e-f) between the true CT image and the sCT reconstruction for each patient show a general agreement in the bulk of the soft tissue represented in each image in the case of the proposed model and a failure to accurately reproduce HU values in the case of the blind model. This is reflected in the MAE values computed for patients of Class 1; the blind and proposed approaches achieve values of 143 ± 29 HU (MAPE $= 14 \pm 3\%$) and 90 ± 29 HU (MAPE $= 9 \pm 3\%$), respectively. Also included in the last row of this comparison is an example of the relatively rare case of a notable presence of intestinal gas that is well-matched in corresponding MRI and CT scans.

In contrast, Figure 5.3 shows the same comparison made for patients of Class 2 in which notable discrepancies in the presence of intestinal gas between corresponding MRI and CT scans are observed. These discrepancies give rise to pixel-wise disagreements of the order of $\pm$ 800 HU in the involved gas-containing regions. At the same time, the failure on the part of the blind model to produce accurate HU values in regions of soft tissue is observed. The increased likelihood of discrepancies in soft tissue positions between corresponding scans in patients of Class 2 causes an increase in the MAE computed in regions not containing air in either image: up to 164 ± 41 HU (MAPE $= 18 \pm 5\%$) for the blind model and 112 ± 41 HU (MAPE $= 14 \pm 4\%$) for the proposed model.

The overlap of regions of air in MR images and the corresponding sCT reconstructions was evaluated in a total of 158 images from the four patients included in Figure 5.3 using the DSC. The average DSC improved from 0.56 ± 0.22 for the blind model to 0.80 ± 0.21 for the proposed model. In these air-containing regions, the SSIM computed between the input MR image and reconstructed sCT image improved from 0.14 ± 0.06 to 0.34 ± 0.07 with the adoption of the proposed model over the blind model.

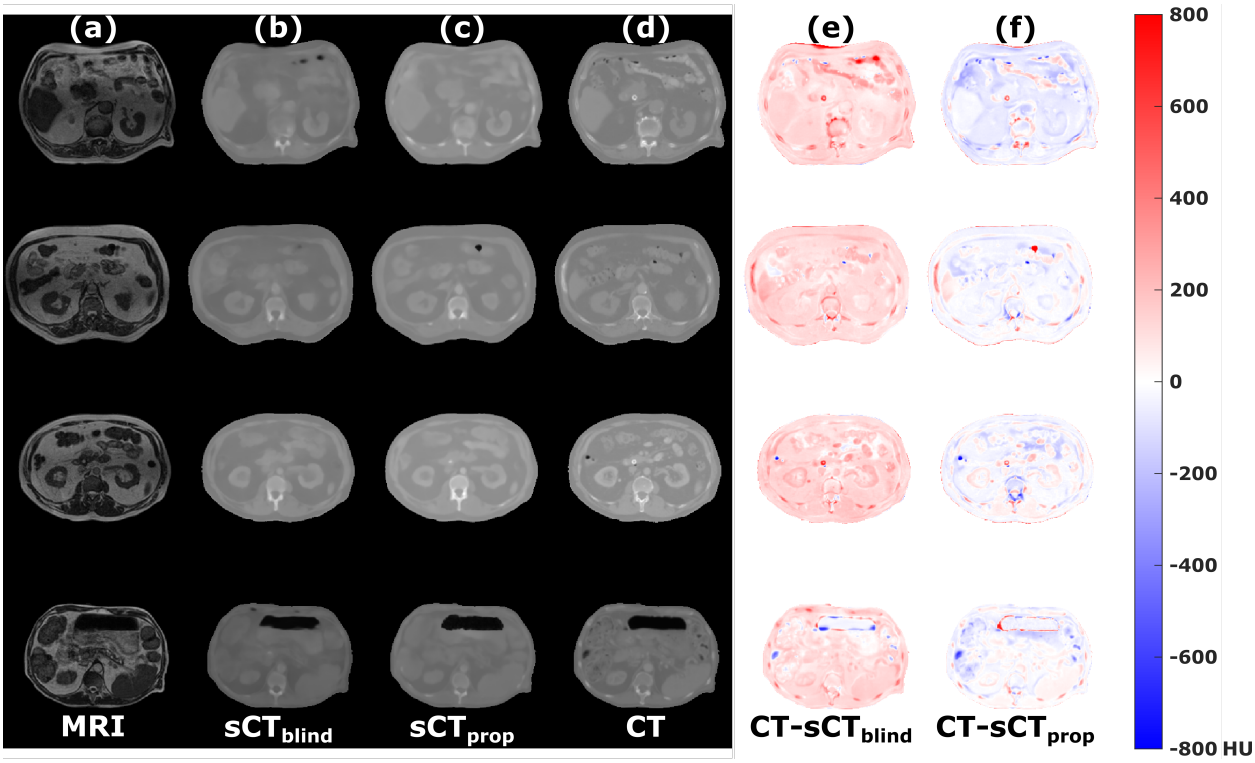

Figure 5.2. Image comparisons for representative slices of well-matched patients of Class 1. Input MR images (a), output sCT images for the blind model (b) and proposed model (c), true CT images (d), and difference maps (e-f) between the true CT images and generated sCT images for the blind and proposed model, respectively. Values in the difference maps are in units of HU. The final row illustrates the rare case of a relatively well-matched slice with a notable presence of intestinal gas.

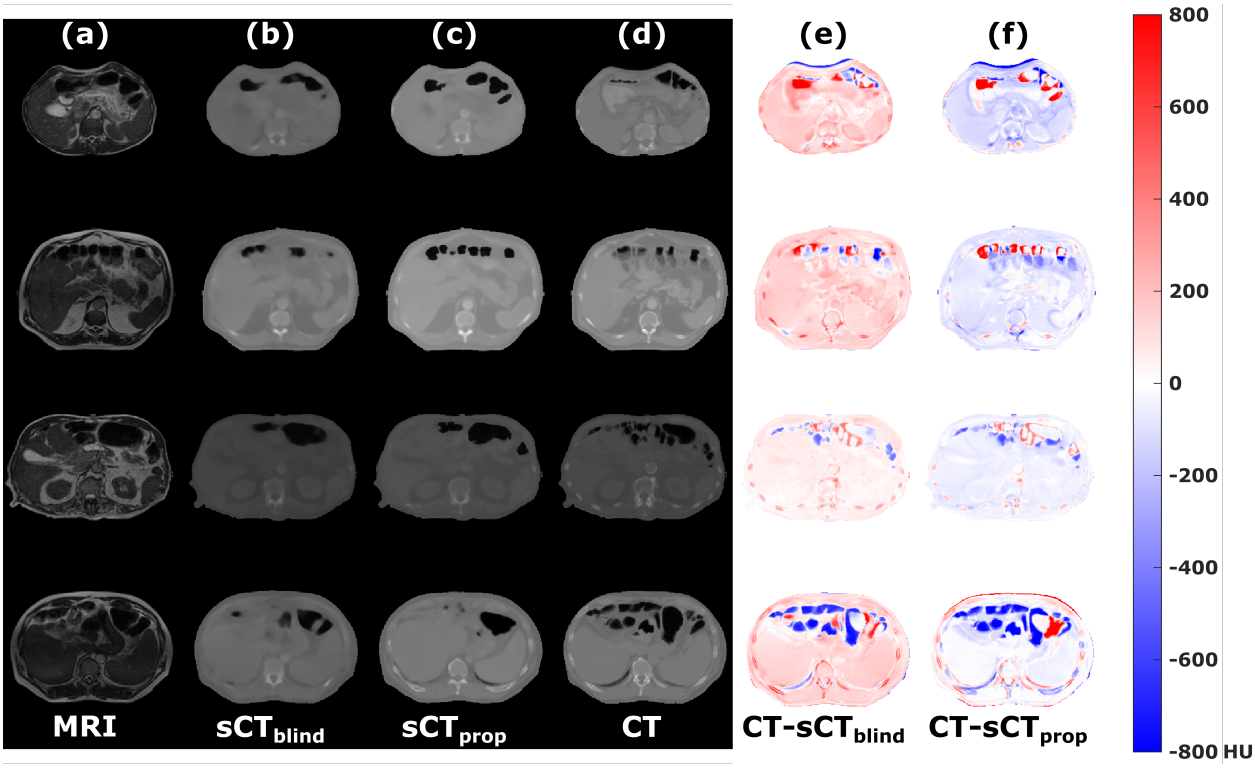

Figure 5.3. Image comparisons for representative slices of patients in Class 2 characterized by notable differences in the presence of intestinal gas between corresponding MRI and CT scans. Input MR images (a), output sCT images for the blind model (b) and proposed model (c), true CT images (d), and difference maps (e-f) between the true CT images and generated sCT images for the blind and proposed model, respectively. Values in the difference maps are in units of HU.

5.3.2 Dosimetric evaluation

Similarly, the dosimetric evaluation of the proposed approach to sCT reconstruction focused on the two distinct classes of patients. In both cases, optimization parameters selected in the CT-based clinical plans were used to recalculate dose distributions based on electron density information derived from the generated sCT images. Used as a baseline point of reference to establish the dosimetric accuracy of the proposed reconstructions, the thirteen well-matched patients of Class 1 demonstrate differences in the prescribed dose coverage of the PTV (V_{100}) of $1.3 \pm 2.1\%$ between CT-based clinical plans and the sCT-based plans with a gamma pass rate of $98.3 \pm 1.3\%$ using 3%/3 mm criterion. The representative DVHs for patients belonging to Class 1 shown in Figure 5.4 demonstrate close agreement in calculated target coverage and doses to surrounding tissues between the CT-based clinical plans and the sCT-based recalculations.

For the twenty poorly-matched patients of Class 2, notable discrepancies in the representation of intestinal gas between corresponding MRI and CT scans result in sizeable and variable differences in PTV V_{100} coverage: $13.3 \pm 11.0\%$ on average. Due to these differences, the gamma pass rate is reduced to $93.9 \pm 9.8\%$ using the same 3%/3 mm criterion. These differences in target coverage, along with small discrepancies in the dose to closely involved tissues like the duodenum, are observable in the representative DVHs included in Figure 5.5. Also plotted in Figure 5.5 is the difference in target coverage at each point, which yields an approximately Gaussian profile. The FWHM of this profile for all patients in Class 2 covers an average range of 51.4(SD=1.3)–58.2(1.6) Gy. Mean DVH differences between the CT-based clinical plans and the sCT-based plans for all structures of interest are plotted in Figure 5.6 for patients of each class. For patients in Class 1, differences between the CT-based and sCT-based plans are computed to be statistically insignificant (distributed with a median of zero) using a two-sided Wilcoxon signed rank test[47] for each of the duodenum (p=.34), large bowel (p=.62), liver (p=.52), small bowel (p=.38), spinal cord (p=.91), stom-

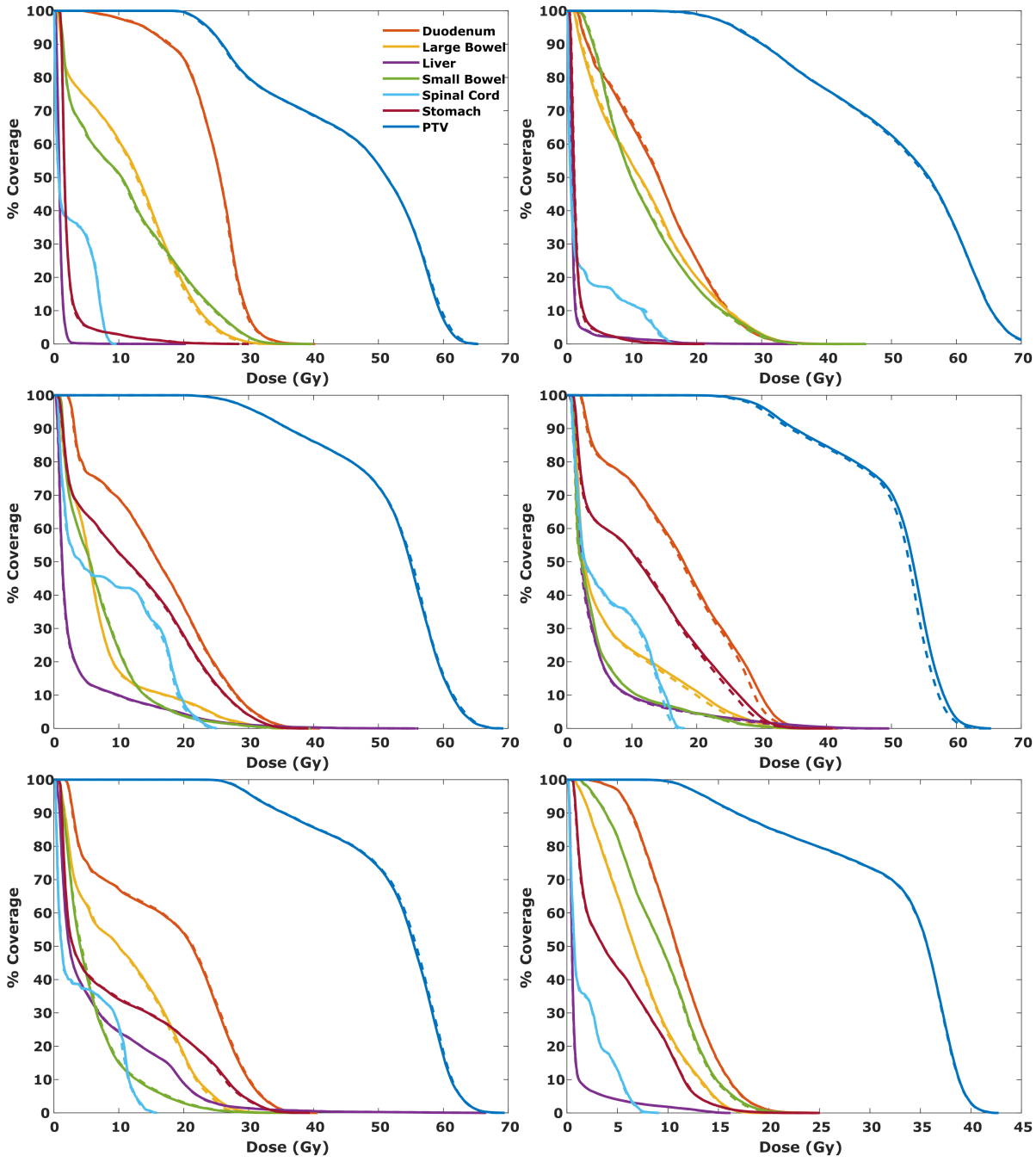

Figure 5.4. Representative DVHs for well-matched test patients of Class 1 comparing the CT-based clinical plans (dashed) and sCT-based plans (solid) recalculated using the same plan parameters. The prescribed dose was 50 Gy in all but the last case. The legend at top-left is applicable to all sub-figures.

ach (p=.91), and PTV (p=.20). For patients in Class 2, differences were instead shown to be statistically significant (not distributed with a median of zero) using a two-sided Wilcoxon signed rank test for each of the duodenum (p=.002), large bowel (p=.007), liver (p<.001), small bowel (p<.001), spinal cord (p=.01), stomach (p<.001), and PTV (p<.001).

5.4 Discussion

The potential value of generating synthetic CT data for MRI-only ART in the abdomen is multifaceted. Although therapeutic gains may be achieved, adopting an online adaptive workflow introduces additional time burdens to the process of treatment delivery including re-contouring, re-planning, and quality assurance—all of which must occur while the patient remains on-table. Re-contouring, which must be undertaken to accommodate changes in normal tissue volumes and—ideally—the variable presence of intestinal gas that is our current focus, represents a significant portion of the total on-table time per fraction: up to 24 minutes in the worst case.[23] By utilizing the proposed approach to sCT reconstruction explored here in which the focus was placed on producing a clinically unavailable data set of well-matched representations of intestinal gas, one is able to rapidly produce (0.26 s/slice) sCT data in the clinical setting that accurately reflects both the presence of intestinal gas shown in a patient's daily MRI scan and HU values present in a true CT scan. In this way, one of the primary concerns prompting re-contouring in the adaptive setting is potentially eliminated. When paired with an auto-contouring strategy designed for MRI-guided ART,[48] the time burden associated with re-contouring may become negligible.

Another primary motivation in adopting an adaptive workflow is to achieve dose escalation under shifting anatomic conditions.[49] Plan adaptation is often performed to increase OAR sparing while also increasing target coverage.[23] The observed underdosing of the target for patients of Class 2 characterized by mismatched representations of intestinal gas is

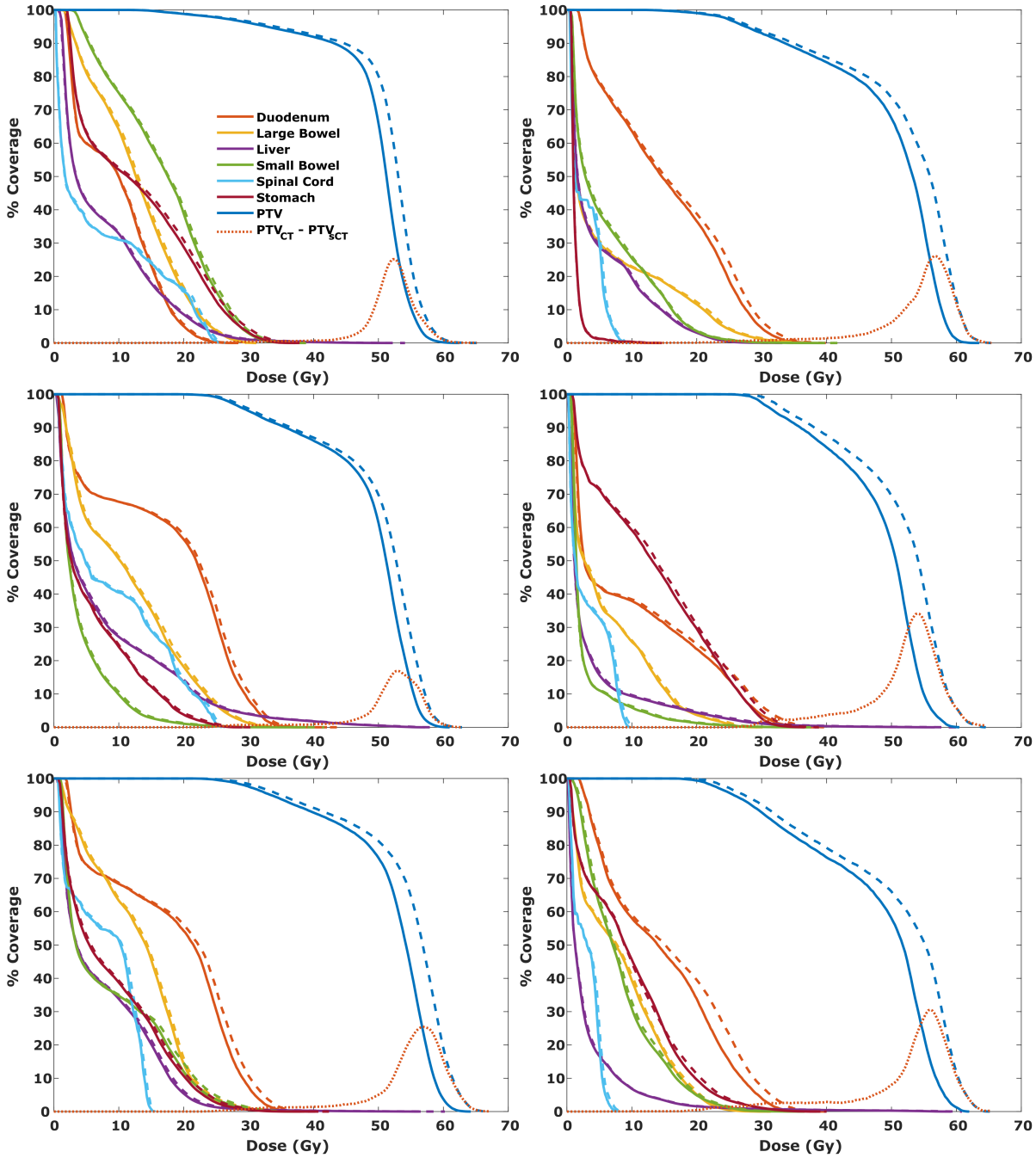

Figure 5.5. Representative DVHs for patients of Class 2 characterized by notable differences in the presence of intestinal gas between corresponding MRI and CT scans comparing the CT-based clinical plans (dashed) and sCT-based plans (solid) recalculated using the same plan parameters. Differences in calculated target coverage are plotted at each point (dotted) to yield an approximately Gaussian curve. The FWHM of these difference profiles for all patients of Class 2 covers an average range of 51.4–58.2 Gy. The legend at top-left is applicable to all sub-figures.

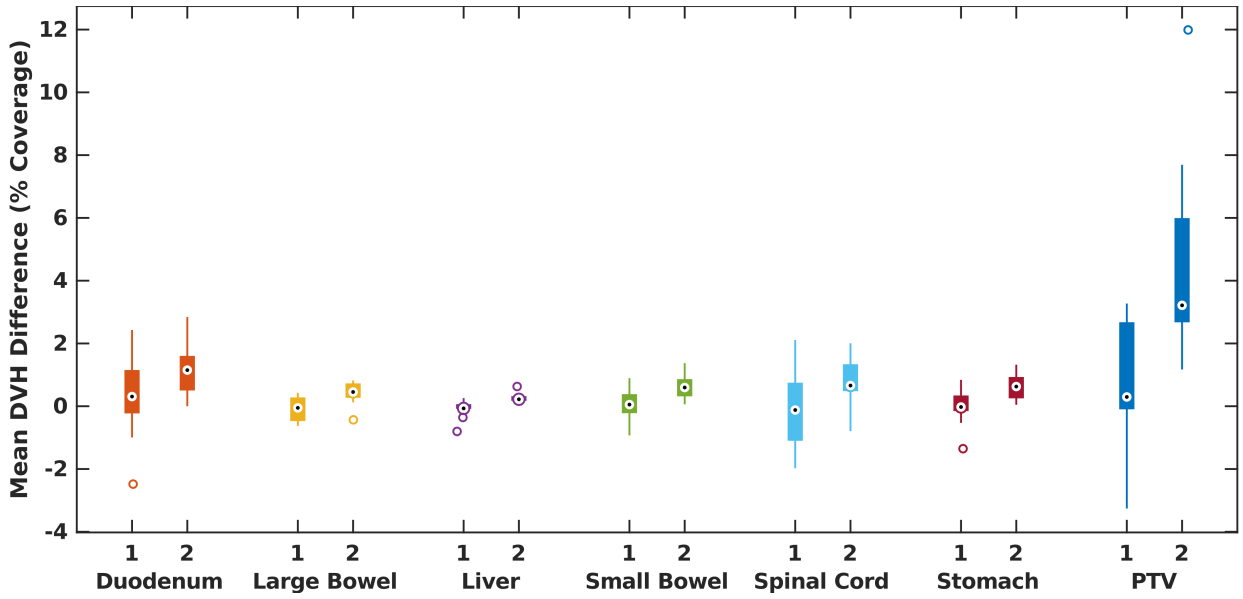

Figure 5.6. Summary of mean DVH differences between CT-based clinical plans and sCT-based recalculated plans for each structure of interest in well-matched patients of Class 1 ($n = 13$) and patients of Class 2 ($n = 20$) characterized by notable differences in the presence of intestinal gas between corresponding MRI and CT scans.

especially relevant in these scenarios when dose escalation is a specific aim of pursuing plan adaptation. The higher calculated target coverage in clinical plans in which mismatches between the planning CT and setup MRI are not accounted for compared to the sCT-based plans and uncertainty in high-dose regions (Figure 5.5) represent a barrier to any escalation that is pursued in these scenarios. In the present study, we have explored this effect at the first treatment fraction for 33 test patients, demonstrating that even at the first fraction, a non-negligible portion of the patient population may experience uncertainties in simulation CT-based dose calculations due to the involvement of intestinal gas. This concern becomes even more important when considering the accumulation of dose over the course of a treatment in which each fraction is adapted and these differences accrue. However, it is important to note that the DVHs computed for a given treatment fraction and presented in the ViewRay treatment planning system are full rather than fractional DVHs. Considering this, the DVHs examined in this study convey the overall effect in the case that the

magnitude of the discrepancies observed at the first fraction are carried forward through each subsequent fraction. Considering the hard-to-characterize nature of gas motion, we do not examine here whether there exists some sort of interplay effect throughout the course of treatment.

We acknowledge a number of additional limitations to the present study. First and foremost, the image-to-image translation approach employed here struggles fundamentally in a situation in which the anatomy represented in corresponding training image pairs differs. Although we took care in the present study to propagate regions of air, geometric differences in soft tissues surrounding the target are not always handled sufficiently by multi-modal deformable image registration. These general effects are observed in image comparisons, where outputs of the blind model appear to be uniformly lower in intensity compared to the true CT images. We attribute this to the increased ambiguity in pixel correspondences in the training image pairs. In effect, more pixel intensities are pushed towards the lower intensity of air. In the proposed case with the pre-processed training data set, the limiting ambiguity becomes the difference in soft tissue representation in the corresponding MR and CT images. The model is reasonably robust to the variations present in the training data set, but struggles to faithfully reconstruct the most dynamic tissues and structures. Style-transfer methods that do not rely on matched pairs of training data as exemplified by CycleGAN[24] may be of particular use in this application to overcome the limitations of multi-modal deformable image registration. The implementation of an unpaired approach may still benefit from semi-supervised data produced in the manner described here to overcome the limitations of the unpaired approach in cases with inherent ambiguity, as is the case for MRI signal intensity. A second limitation of the trained model is reconstruction inaccuracies at the superior and inferior extremes of a patient's image stack that stem from the make-up of the training data set. While every patient data set was roughly centered on the target and surrounding tissues, slices containing views of the lungs and diaphragm or inferior

portions of the abdomen were not as equally represented. These reconstruction inaccuracies may be ameliorated by adopting a more robust 3D network architecture over the relatively lightweight 2D architecture at the expense of heavily increased memory usage, which may not be a feasible trade-off in all settings. An additional concern regarding MR image quality in the abdominal sCT reconstruction task is the issue of susceptibility artifacts in the GI tract. The effects of these artifacts are lessened due to the lower field strength of the MRgRT platform utilized here,[50] but certain circumstances involving the ingestion of fortified foods prior to treatment—although not encountered in the present study—have been reported and are thus an important consideration in this application.[51]

An inherent challenge in this space is the issue of image evaluation when the underlying premise is that the "ground truth" simulation CT data is incompatible with the setup MRI data used due to the involvement of intestinal gas and the motion of GI structures. In computing both the MAE and MAPE, direct pixel-wise comparisons are made between the sCT images and the corresponding CT images. In a similar way, the multi-modality image similarity comparison made between MR images and sCT reconstructions by way of the SSIM is challenged by modality-specific differences in contrast and luminance. As such, the image comparisons made here are imperfect comparisons and we instead rely more heavily on the improvement in the representation of air in our sCT reconstructions as measured by the DSC considering that is the primary focus of this work. Nonetheless, a comparison to other methods is warranted. Ahunbay et al.[29] achieved an MAE of approximately 25 HU in the abdomen using a method entirely reliant of multiple deformations of true CT images. Closer to the realm of true image synthesis, multiple atlas-based techniques have reported values ranging from 40–200 HU for sites including the pelvis, cranium, and general torso.[10,11,30] Finally, results of DL-based approaches at various sites quantified using various metrics have been collected and reported by Spadea et al.[12] for general comparison with the results achieved here. For abdominal cases demonstrating little involvement of intestinal

gas, Cusumano et al.[34] report an MAE of 78.71 ± 18.46 within the body contour. In a similar cohort, Qian et al.[35] report MAE values ranging from approximately 42–79 HU for two DL-based methods. In the MAE values we report, we do not distinguish between regions of soft tissue and bone, but we do compare competitively to existing methods.

Finally, the separation of test patients into two separate classes performed here relied on the qualitative assessment of the relative involvement of intestinal gas and the degree to which representations of intestinal gas matched between corresponding MRI and CT scans. In some cases, the characterization of the patient was clear—there were easily observable discrepancies or gas was entirely uninvolved—but the categorization was more challenging in other cases. As such, the furthering of this and related work would benefit from some quantitative approach to the characterization of the involvement and similarity of representations of intestinal gas that would in turn enable the exploration of trends in distinct patient groups.

5.5 Conclusions

The approach to sCT reconstruction in the abdomen evaluated here highlights the challenges posed by the presence of intestinal gas throughout the MRI-guided ART workflow. Eliminating the burden of handling intestinal gas from the clinical setting through the creation of a clinically unavailable training data set for training a paired data generative model offers the potential to streamline a time-intensive portion of the adaptive treatment workflow. These time savings are gained while also enabling accurate dose calculations in adaptive treatments despite the variable presence of intestinal gas at each stage of treatment planning and delivery during MRI-only ART in the abdomen.

References

[1] Mutic S, Dempsey JF. The ViewRay system: Magnetic resonance–guided and controlled radiotherapy. *Semin Radiat Oncol*. 2014;24(3):196–199. doi: 10.1016/j.semradonc.2014.02.008

[2] Fallone BG. The rotating biplanar linac-magnetic resonance imaging system. *Semin Radiat Oncol*. 2014;24:200–202. doi:10.1016/j.semradonc.2014.02.011

[3] Keall PJ, Barton M, Crozier S. The Australian magnetic resonance imaging-linac program. *Semin Radiat Oncol*. 2014;24:203–206. doi:10.1016/j.semradonc.2014.02.015

[4] Fischer-Valuck BW, Henke L, Green O, et al. Two-and-a-half-year clinical experience with the world's first magnetic resonance image guided radiation therapy system. *Adv Radiat Oncol*. 2017;2(3):485–493. doi:10.1016/j.adro.2017.05.006

[5] Raaymakers BW, Jürgenliemk-Schulz IM, Bol GH, et al. First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment. *Phys Med Biol*. 2017;62(23):L41–L50. doi:10.1088/1361-6560/aa9517

[6] Wen N, Kim J, Doemer A, et al. Magnetic resonance guided linear accelerator for stereotactic radiosurgery treatment. In *Proceedings of the American Society for Radiation Oncology*. p. E479. doi:j.ijrobp.2018.07.1369

[7] Pollard JM, Wen Z, Sadagopan R, Wang J, Ibbott GS. The future of image-guided radiotherapy will be MR guided. *Br J Radiol*. 2017;90(1073):20160667. doi: 10.1259/bjr.20160667

[8] Noel CE, Parikh PJ, Spencer CR, et al. Comparison of onboard low-field magnetic resonance imaging versus onboard computed tomography for anatomy visualization in radiotherapy. *Acta Oncol (Madr)*. 2015;54(9). doi:10.3109/0284186X.2015.1062541

[9] Schmidt MA, Payne GS. Radiotherapy planning using MRI. *Phys Med Biol*. 2015; 60(22):R323–R361. doi:10.1088/0031-9155/60/22/R323

[10] Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. *Radiat Oncol*. 2017;12(1):28. doi:10.1186/s13014-016-0747-y

[11] Johnstone E, Wyatt JJ, Henry AM, et al. Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging–only radiation therapy. *Int J Radiat Oncol*. 2018;100(1):199–217. doi:10.1016/j.ijrobp.2017.08.043

[12] Spadea MF, Maspero M, Zaffino P, Seco J. Deep learning-based synthetic-CT generation in radiotherapy and PET: a review. *ArXiv*. 2021;:1–52

[13] Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys*. 2017;44(4):1408–1419. doi:10.1002/mp.12155

[14] Arabi H, Dowling JA, Burgos N, et al. Comparative study of algorithms for synthetic CT generation from MRI: Consequences for MRI-guided radiation planning in the pelvic region. *Med Phys*. 2018;45:5218–5233. doi:10.1002/mp.13187

[15] Chen S, Qin A, Zhou D, Yan D. Technical Note: U-net-generated synthetic CT images for magnetic resonance imaging-only prostate intensity-modulated radiation therapy treatment planning. *Med Phys*. 2018;45:5659–5665. doi:10.1002/mp.13247

[16] Maspero M, Savenije MHF, Dinkla AM, et al. Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy. *Phys Med Biol*. 2018;63:1–11. doi:10.1088/1361-6560/aada6d

[17] Olberg S, Zhang H, Kennedy WR, et al. Synthetic CT reconstruction using a deep spatial pyramid convolutional framework for MR-only breast radiotherapy. *Med Phys*. 2019;doi:10.1002/mp.13716

[18] Mostafaei F, Tai A, Omari E, et al. Variations of MRI-assessed peristaltic motions during radiation therapy. *PLoS One*. 2018;13:e0205917. doi:10.1371/journal.pone.0205917

[19] Nakamoto Y, Chin BB, Cohade C, Osman M, Tatsumi M, Wahl RL. PET/CT: artifacts caused by bowel motion. *Nucl Med Commun*. 2004;25:221–225

[20] Feng M, Balter J, Normolle D, et al. Characterization of pancreatic tumor motion using cine MRI: surrogates for tumor position should be used with caution. *Int J Radiat Oncol Biol Phys*. 2009;74:884–891. doi:10.1016/j.ijrobp.2009.02.003

[21] Kumagai M, Hara R, Mori S, et al. Impact of intrafractional bowel gas movement on carbon ion beam dose distribution in pancreatic radiotherapy. *Int J Radiat Oncol Biol Phys*. 2009;73:1276–1281. doi:10.1016/j.ijrobp.2008.10.055

[22] Corradini S, Alongi F, Andratschke N, et al. MR-guidance in clinical reality: current treatment challenges and future perspectives. *Radiat Oncol*. 2019;14:92. doi: 10.1186/s13014-019-1308-y

[23] Henke L, Kashani R, Robinson C, et al. Phase I trial of stereotactic MR-guided online adaptive radiation therapy (SMART) for the treatment of oligometastatic or unresectable primary malignancies of the abdomen. *Radiother Oncol*. 2018;126(3):519–526. doi:10.1016/j.radonc.2017.11.032

[24] Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. *ArXiv*. 2018;:1–18

[25] Peng Y, Chen S, Qin A, et al. Magnetic resonance-based synthetic computed tomography images generated using generative adversarial networks for nasopharyngeal carcinoma radiotherapy treatment planning. *Radiother Oncol*. 2020;150:217–224. doi: 10.1016/j.radonc.2020.06.049

[26] Fu J, Singhrao K, Cao M, et al. Generation of abdominal synthetic CTs from 0.35T MR images using generative adversarial networks for MR-only liver radiotherapy. *Biomed Phys Eng Express*. 2020;6:015033. doi:10.1088/2057-1976/ab6e1f

[27] Bredfeldt JS, Liu L, Feng M, Cao Y, Balter JM. Synthetic CT for MRI-based liver stereotactic body radiotherapy treatment planning. *Phys Med Biol*. 2017;62:2922–2934. doi:10.1088/1361-6560/aa5059

[28] Hsu SH, Peng Q, Tomé WA. On the generation of synthetic CT for a MRI-only radiation therapy workflow for the abdomen. *J Phys Conf Ser*. 2019;1154:1–4. doi:10.1088/1742-6596/1154/1/012011

[29] Ahunbay EE, Thapa R, Chen X, Paulson E, Li XA. A technique to rapidly generate synthetic computed tomography for magnetic resonance imaging-guided online adaptive replanning: An exploratory study. *Int J Radiat Oncol*. 2019;103(5):1261–1270. doi: 10.1016/j.ijrobp.2018.12.008

[30] Guerreiro F, Koivula L, Seravalli E, et al. Feasibility of MRI-only photon and proton dose calculations for pediatric patients with abdominal tumors. *Phys Med Biol*. 2019; 64:1–13. doi:10.1088/1361-6560/ab0095

[31] Korhonen J, Kapanen M, Keyriläinen J, Seppälä T, Tenhunen M. A dual model HU conversion from MRI intensity values within and outside of bone segment for MRI-based radiotherapy treatment planning of prostate cancer. *Med Phys*. 2014;41(1):1–12. doi:10.1118/1.4842575

[32] Koivula L, Wee L, Korhonen J. Feasibility of MRI-only treatment planning for proton therapy in brain and prostate cancers: Dose calculation accuracy in substitute CT images. *Med Phys*. 2016;43(8):4643–4642. doi:10.1118/1.4958677

[33] Koivula L, Kapanen M, Seppälä T, et al. Intensity-based dual model method for generation of synthetic CT images from standard T2-weighted MR images – generalized technique for four different MR scanners. *Radiother Oncol*. 2017;125(3):411–419. doi: 10.1016/j.radonc.2017.10.011

[34] Cusumano D, Lenkowicz J, Votta C, et al. A deep learning approach to generate synthetic CT in low field MR-guided adaptive radiotherapy for abdominal and pelvic cases. *Radiother Oncol*. 2020;153:205–212. doi:10.1016/j.radonc.2020.10.018

[35] Qian P, Xu K, Wang T, et al. Estimating CT from MR abdominal images using novel generative adversarial networks. *J Grid Computing*. 2020;18:211–226. doi: 10.1007/s10723-020-09513-3

[36] Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-scale machine learning on heterogeneous systems. 2015. Software available from tensorflow.org

[37] Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern*. 1979;9:62–66

[38] Vincent L. Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE Trans Image Process*. 1993;2(2):176–201. doi: 10.1109/83.217222

[39] Schönlieb CB. *Partial Differential Equation Methods for Image Inpainting*. Cambridge, UK: Cambridge University Press. 2015. doi:10.1017/CBO9780511734304

[40] Parisotto S, Schönlieb C. MATLAB/Python codes for the Image Inpainting Problem. GitHub repository, MATLAB Central File Exchange. 2016

[41] Shi J. *Statistical dependence measure based multi-modal image registration and registration assisted non-parametric image segmentation*. Ph.D. thesis. University of Florida. 2013

[42] Kingma D, Ba J. Adam: A method for stochastic optimization. *ArXiv*. 2017;:1–15

[43] Jégou S, Drozdzal M, Vazquez D, Romero A, Bengio Y. The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation. *ArXiv*. 2017;:1–9

[44] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *ArXiv*. 2015;:1–8

[45] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *ArXiv*. 2015;:1–12

[46] Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. *Med Phys*. 1998;25(5):656–661. doi:10.1118/1.598248

[47] Wilcoxon F. Individual comparisons by ranking methods. *Biometrics Bulletin*. 1945; 1:80–83

[48] Fu Y, Mazur TR, Wu X, et al. A novel MRI segmentation method using CNN-based correction network for MRI-guided adaptive radiotherapy. *Med Phys*. 2018;doi: 10.1002/mp.13221

[49] Henke L, Contreras J, Green O, et al. Magnetic resonance image-guided radiotherapy (MRIgRT): a 4.5-Year clinical experience. *Clin Oncol (R Coll Radiol)*. 2018;30(11):720–727. doi:10.1016/j.clon.2018.08.010

[50] Ginn J, Agazaryan N, Cao M, et al. Characterization of spatial distortion in a 0.35 T MRI-guided radiotherapy system. *Phys Med Biol*. 2017;62:4525–4540. doi:10.1088/1361-6560/aa6e1a

[51] Green O, Henke L, Parikh P, Roach M, Michalski J, Gach H. Practical implications of ferromagnetic artifacts in low-field MRI-guided radiotherapy. *Cureus*. 2018;10:e2359. doi:10.7759/cureus.2359

Chapter 6

Ensemble learning and personalized training for the improvement of unsupervised deep learning-based synthetic CT reconstruction

6.1 Introduction

The growing adoption of integrated MR-linac systems has placed an increased focus on the capabilities and practicalities of the MRI-guided radiation (RT) therapy workflow.[1–7] Compared to the x-ray CT examinations used in the conventional image-guided RT setting, MRI demonstrates an improved soft tissue contrast that is useful for the delineation of targets and organs at risk (OARs) throughout the body.[8–10] Additionally, the functional imaging capabilities of MRI have opened the door to biological adaptation in a way that is not possible in the conventional CT-based setting.[9,10] These expanded and attractive capabilities carry practical limitations, however. A key limitation in this MRI-guided setting is that these nascent workflows still rely on a simulation CT scan to provide electron density information for dose calculations. The inclusion of this CT simulation scan brings concerns related not only extra imaging dose but also challenges with multi-modality image registration that can be excessively limiting, particularly for sites in which large anatomical variations are observed on a regular basis.[10–15] In light of these challenges related to a reliance on CT simulation in this setting, the possibility of adopting an MRI-only RT workflow in which MRI is the sole image modality used for treatment guidance and planning has attracted a significant amount of attention.[7,16,17]

A primary technical hurdle to enabling such a workflow is the creation of synthetic CT (sCT) data that provides electron density information for dose calculations, thus eliminating the need for a simulation CT scan. Approaches to this problem of sCT reconstruction have undergone various stages of development as the MRI-guided workflow has garnered more interest over time but may be broadly categorized into three groups: bulk density assignment, atlas-based / hybrid approaches, and learning-based approaches.[18–20] The broad strokes bulk density assignment approaches involve assigning uniform HU values based on drawn tissue contours and are thus better suited in practice for relatively homogeneous sites with fewer

tissue types such as the breast.[21] The atlas-based or hybrid approaches are relatively more complex, as tissue contours or HU values directly are achieved through the registration of an incoming patient's MRI data with a large patient atlas.[19,22] Recent experiences with the sCT reconstruction task have seen deep learning-based approaches falling into the third category be the overwhelmingly popular choice in this setting since they were introduced in 2017 in the context of the brain.[23] The expressive power of these deep network architectures makes them well suited for solving the MRI-to-CT image-to-image translation task that sCT reconstruction may be posed as, leading to the explosion of varying techniques and approaches to the problem of sCT reconstruction in recent years.[20]

One notable axis for the categorization of these deep learning-based approaches is the required structure of the training data set. Considering the straightforward framing of the sCT reconstruction task as an image-to-image translation problem in which a mapping from MRI to CT is learned, a training data set consisting of paired MR and CT images showing the same anatomy as represented by each modality is an intuitive starting point that is indeed used by many frameworks.[20] Naturally, these paired-data approaches rely on image pairs that demonstrate good agreement in the represented anatomy, which is not always so easily satisfied. Training data in this sCT reconstruction task is often collected from corresponding MRI and CT simulation scans performed prior to treatment. The time delay that often exists between the respective acquisition of each of these simulation scans gives rise to normal anatomical variations on various timescales due to respiration or OAR filling, for example.[24–31] The discrepancies in corresponding MRI and CT scans resulting from these anatomical variations have an obvious clinical impact in the MRI-guided setting, but also represent a technical hurdle to the successful adoption of these paired-data approaches in some disease sites.[32]

In order to avoid the data availability challenges presented by these discrepancies, one may instead adopt highly expressive unpaired-data approaches that do not rely on a struc-

tured training data set consisting of paired MR and CT images like the paired-data approaches. In this setting, the image-to-image translation problem is further posed as a problem of style transfer rather than a simple mapping from MRI to CT based on observed pixel correspondences in related images. The sCT output of a network that solves this style transfer problem is an answer to the question "What would this MR image look like if it were instead 'painted' in the style of a CT image?" Instead of relying on pixel-wise differences between a sCT image and the ground truth CT image during training, an unpaired-data approach exemplified by CycleGAN[33] enforces a cycle consistency loss in an image that has been transformed in both the forward (MRI-to-CT) and backward (CT-to-MRI) directions, eliminating the need for a ground truth CT image in the loss formulation. Avoiding this requirement for a structured training data set consisting of well-matched image pairs presents an obvious advantage of an unpaired-data approach in the low-data radiation oncology setting. These expressive unpaired-data approaches are not without their drawbacks, however. More so than their paired-data counterparts, unpaired-data approaches fundamentally hallucinate features in the reconstructed output in a way that goes unchecked by any pixel-wise comparison to a ground truth image.[34] Along this line, the authors of the seminal CycleGAN paper acknowledge a gap between the unpaired- and paired-data approaches that may be difficult to close in practice, a finding that has also been echoed in sCT reconstruction studies that follow.[33,35–37]

Considering this gap, our interest in this work is in improving the quality of CycleGAN outputs in the sCT reconstruction task through two distinct methods. The first approach we employ involves the well established concept of ensemble learning in which multiple models are used to ultimately achieve better performance than can be achieved by a lone model.[38] In this case, we adopt a stacking or cascading approach specifically, applying a paired-data approach to the sCT outputs of the unpaired-data CycleGAN to learn a mapping from sCT to CT as an additional fine-tuning step. The alternative to this more established ensemble ap-

proach that we explore here involves a patient-specific training strategy designed to produce personalized models that demonstrate improved task performance. The approach, termed Intentional Deep Overfit Learning (IDOL),[39] was developed with paired-data approaches in mind, so the novel question we seek to answer here is whether or not an unpaired-data approach benefits from some degree of implicit supervision as imparted by such a personalized training strategy. We explore this question in the context of the sCT reconstruction task in the male pelvis, comparing the paired-data fully convolutional DenseNet (FCDN)[40] with and without the IDOL strategy to the unpaired-data CycleGAN framework in its unimproved, ensemble, and IDOL forms.

6.2 Materials and Methods

6.2.1 Training and testing data

Image data in this retrospective study was collected from a total of 25 prostate cancer patients previously treated at our institution using conventional C-arm linacs. In each case, CT and MRI simulation scans were acquired prior to treatment delivery for planning and guidance. 1.5 T MRI scans (864 $\times$ 864 $\times$ 444 matrix, 0.64 $\times$ 0.64 $\times$ 1.35 mm^3) acquired using a dedicated MR simulation machine (Ingenia Ambition X, Philips Healthcare, Andover, MA) with a T2w SE sequence were registered to the corresponding CT simulation scans (nominally 512 $\times$ 512 $\times$ 100-150 matrix, 1.17 $\times$ 1.17 $\times$ 3 mm^3) acquired using a dedicated CT simulator (Big Bore, Philips Healthcare, Andover, MA). Deformable registration and resampling was performed using a stand-alone software suite (Velocity, Varian Medical Systems, Palo Alto, CA) prior to exporting the images for training and testing.

Training and evaluation of the paired-data (described in Section 6.2.3), unpaired-data (described in Section 6.2.4), and cascade ensemble (described in Section 6.2.5) architectures was performed using a five-fold cross validation scheme. Models were trained in each case

for 120 epochs with a batch size of 1 using TensorFlow[41] v2.0 in Python running on a 12 GB Titan Xp GPU (NVIDIA, Santa Clara, CA). The personalized IDOL training approach (described fully in Section 6.2.6) involved a secondary stage of training of 60 additional epochs for each of the base paired-data and unpaired-data models for the five patients in fold 3, chosen at random for the validation of this approach.

6.2.2 Model and loss formulation

The fundamental approach to sCT reconstruction as an image-to-image translation task has been previously described.[21] As in that case, the paired- and unpaired-data approaches we adopt here (Figure 6.1) fall into the category of the generative adversarial network (GAN), which consists of two networks: a generative model that produces sCT outputs from MRI inputs and a discriminative model that learns to differentiate between true CT images and generated sCT images. During training, the loss functions of these competing networks are optimized through alternating training steps such that as one network performs better at its respective task it reinforces the training of the other network. In other words, as the discriminator does a better job of differentiating between true CT images and generated sCT images, the generator in turn does a better job of producing convincing sCT images, ultimately resulting in a generative model that provides an adequate solution to the MRI-to-CT translation task.

The discussion of the loss functions used by the various approaches we explore here relies in part on the definition of sigmoid cross entropy loss:[41]

$$L = \vec{x} - \vec{x} * \vec{z} + \log\left(1 + \exp\left(-\vec{x}\right)\right), \tag{6.1}$$

where $\vec{x}$ is the vector of logits computed by the discriminator and $\vec{z}$ is a labels vector containing true ($\vec{1}$) or false ($\vec{0}$) labels. Using this general definition and starting with the

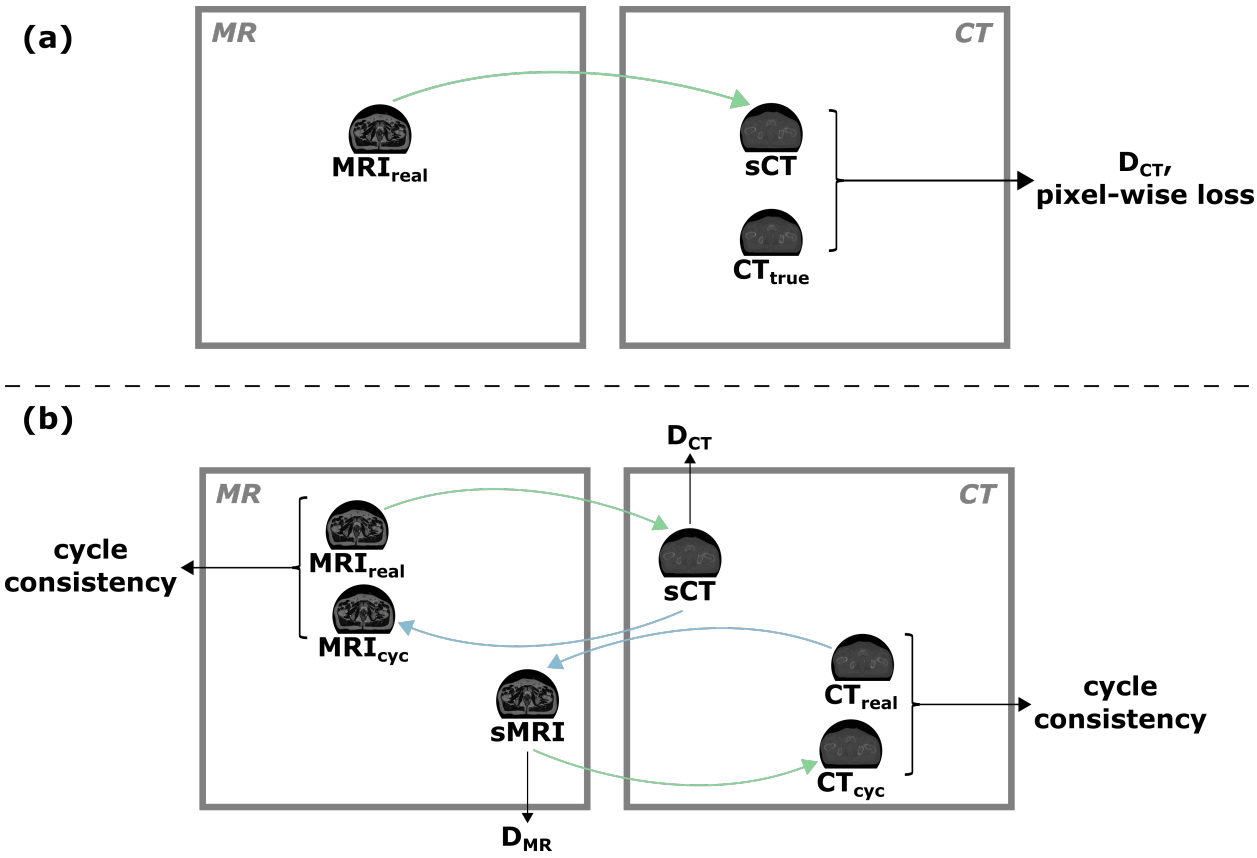

Figure 6.1. Conceptual representations of the paired-data and unpaired-data approaches used in this study. In the paired-data approach (a), a forward translation from MRI to CT is applied by a generative model (green). Generator loss consists of adversarial loss computed by the discriminative model D_{CT} and pixel-wise loss between the sCT output and the ground truth CT. In the unpaired-data approach (b), two sets of generative and discriminative models perform a similar process in both the forward (MRI-to-CT, green) and backward (CT-to-MRI, blue) directions. The key distinction is the inclusion of a cycle consistency loss between real and cycled images rather than a comparison to a ground truth image. (b) adapted from Zhu et al.[33]

relatively more straightforward paired-data approach (Figure 6.1a, described fully in Section 6.2.3), we define the loss function of the generator g_{loss} as

$$g_{loss} = L_{adv} + \alpha \cdot l_{MAE}, \tag{6.2}$$

with adversarial loss L_{adv} and pixel-wise mean absolute error (MAE) loss l_{MAE} weighted by factor $\alpha = 10$ in this case. The adversarial loss term, intended to penalize predictions that are not observed to be true CT images by the discriminator, is the cross entropy loss (Eq. 6.1) computed with predicted image logits $\vec{x}$ paired with the true label of $\vec{z} = 1$. Pixel-wise agreement between ground truth CT images CT_{true} and sCT predictions CT_{pred} is enforced through the MAE:

$$l_{mae} = \text{mean}\left(|CT_{true} - CT_{pred}|\right). \tag{6.3}$$

The discriminator in turn seeks to correctly distinguish between true CT images and sCT images produced by the generator. As such, the discriminator loss function d_{loss} relies only on the cross entropy loss defined in Eq. 6.1:

$$d_{loss} = L_{pred} + L_{true}, \tag{6.4}$$

with separate terms for the predicted sCT images L_{pred} and true CT images L_{true}. In contrast to the adversarial loss formulation for the generator in which predicted images are paired with a contradictory true label, predicted or true image logits are assigned the "correct" label of false or true, respectively, in the computation of d_{loss}.

In the absence of ground truth CT images in the unpaired-data setting, the unpaired-data framework utilized here and described fully in Section 6.2.4 extends the GAN concept to include both a forward translation (MRI-to-CT) and a backward translation (CT-to-MRI) using two sets of adversarial generators and discriminators (Figure 6.1b). The loss

formulation for the generators in this setting maintains the cross entropy-based adversarial loss term L_{adv} with predicted image logits paired with a true label as previously described, but implements additional cycle consistency loss l_{cyc} and identity loss l_{id} terms in place of a pixel-wise comparison between predictions and ground truth images:

$$g_{loss} = L_{adv} + l_{cyc} + l_{id}. \tag{6.5}$$

The concept of cycle consistency loss can perhaps be best understood through the lens of language: one desires that an English text translated to Spanish and then back to English be identical to the original English text. In the same way, the cycle consistency loss term l_{cyc} aims to enforce that completely cycled images that have gone through the forward and backward translations (i.e. $CT_{real} - sMR - CT_{cyc}$ or $MR_{real} - sCT - MR_{cyc}$) are identical to the original images:

$$l_{cyc} = \lambda \cdot \left[\text{mean}\left(|CT_{real} - CT_{cyc}|\right) + \text{mean}\left(|MR_{real} - MR_{cyc}|\right)\right], \tag{6.6}$$

with weighting factor $\lambda = 10$ in this case. Finally, the identity loss term l_{id} aims simply to enforce similarity between a real image I_{real} already belonging to the target class of a given generator and its translation by that generator I_{same}:

$$l_{id} = 0.5 \cdot \lambda \cdot \text{mean}\left(|I_{real} - I_{same}|\right), \tag{6.7}$$

again with weighting factor $\lambda = 10$. In other words, the forward mapping from MRI-to-CT should not modify a CT image input that already belongs to the destination class.

The identifying true versus fake objective of the CT and MR discriminators in the unpaired-data framework is identical to that of the CT discriminator in the paired-data approach; as such, the discriminator loss formulation for these discriminators is unchanged from the general equation presented in Eq. 6.4.

6.2.3 Paired-data architecture: FCDN

The paired-data approach employed here is the FCDN architecture previously used by our group in application to the sCT reconstruction task in the abdomen.[32] The primary building block of the FCDN is the dense block, an example of which is shown in Figure 6.2. Dense blocks are composed themselves of intermediate convolutional layers consisting of ReLU activation, same convolution with 3×3 kernels computing a total of 16 feature maps per layer in this case, and dropout with probability $p = 0.2$. The characteristic feature of the dense block is the degree of connectivity: the input of the block is concatenated before each intermediate layer to the output of the preceding layer and the output of the block itself is formed by the concatenation of the output of each intermediate layer. Ultimately, these connections provide a short path to each layer in the network, improving its convergence behavior in the course of training.[40]

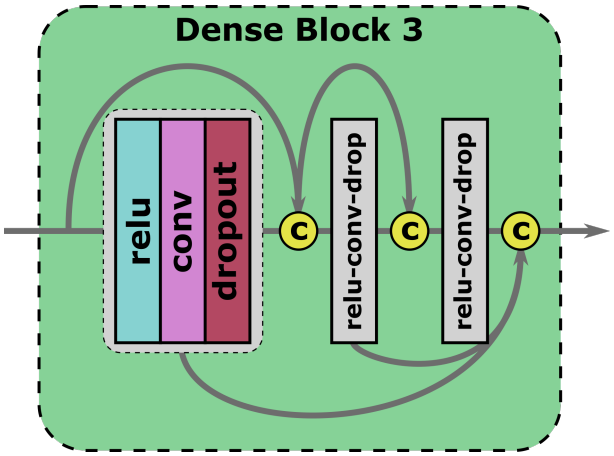

Figure 6.2. Example dense block consisting of three internal ReLU-convolution-dropout layers. The input of the block is concatenated before each intermediate layer and the output of the block itself is the concatenation of the output of each layer.

Figure 6.3 illustrates the U-net[42] structure constructed with these dense blocks, which consists of stacked encoder and decoder paths connected by skip connections that transfer structural features that aid in the reconstruction task from the encoder to the decoder. As

the input MR image travels through the encoder path it is represented as an increasingly rich set of feature maps as computed by the convolutional layers making up the encoding dense blocks. Transition down (TD) layers in the encoder path progressively reduce the dimension of these feature maps through ReLU activation, 1×1 same convolution, dropout, and 2×2 max pooling operations. On the decoder path, the dimensionality of the input is progressively recovered through transition up (TU) layers consisting of 3×3 transpose convolution operations with stride 2.

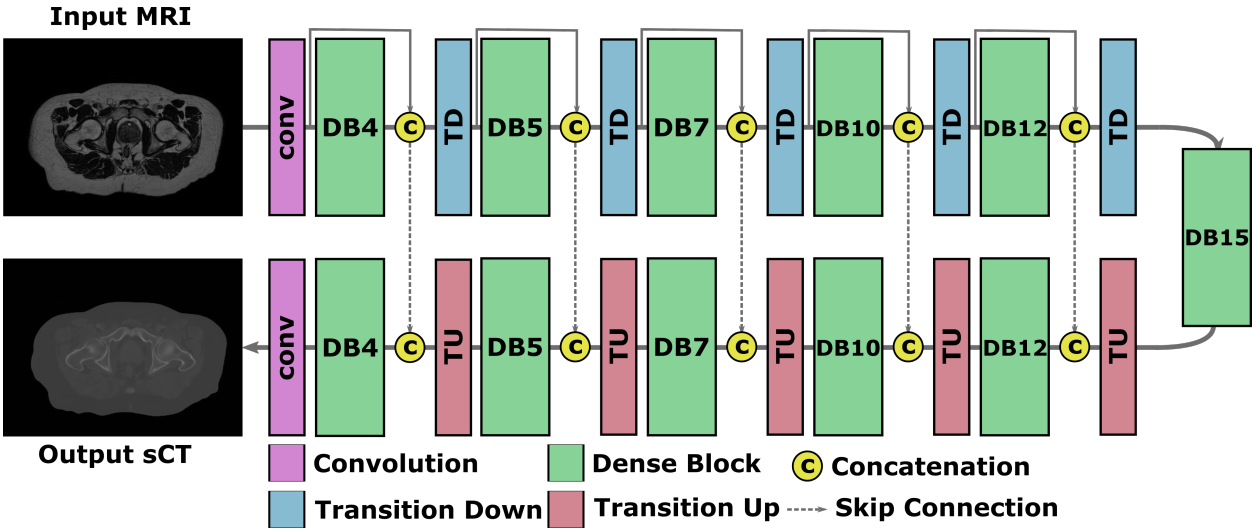

Figure 6.3. FCDN architecture. The encoder and decoder paths making up the U-net structure are connected through skip connections that transfer reconstruction-aiding structural features from the encoder to the decoder. DBn denotes a dense block with n convolutional layers in which intermediate outputs are concatenated to form the output of the dense block itself. Adapted from Olberg et al.[32]

The adversarial discriminator paired with the FCDN generator is by comparison a relatively simple architecture consisting only of 5 convolution-BatchNorm-ReLU layers that apply 64, 128, 256, 512, and 1 filters, respectively.[21] Layers in the discriminator utilize 4×4 kernels and leaky ReLU activation with slope 0.2.

Optimization of the trainable parameters describing the computations performed in the generator and discriminator of the paired-data FCDN architecture is performed in alternating

steps throughout the course of training. For g_{loss}, the Adam optimizer[43] is used with an initial learning rate of 2×10^{-4} subject to staircase exponential decay of rate 0.95 at intervals of 10000 steps, $\beta_1 = 0.7$, $\beta_2 = 0.999$, and $\hat{\epsilon} = 10^{-8}$. The optimization of d_{loss} utilizes TensorFlow's gradient descent optimizer with an initial learning rate of 2×10^{-5} subject to the same decay described previously.

6.2.4 Unpaired-data architecture: CycleGAN

Due to the increased memory requirements of the CycleGAN architecture that consists of two sets of generative and discriminative models, the generator architecture[44] adopted for both the forward and backward models in this unpaired-data approach is relatively simple compared to the FCDN architecture described in Section 6.2.3. The U-net structure with skip connections between the encoder and decoder paths is maintained, but the constituent dense blocks on the encoder path are replaced instead with 6 layers consisting of 4×4 same convolution with stride 2 followed by instance normalization[45] and leaky ReLU activation with slope 0.3. Layers in this encoder path apply 64, 128, 256, 512, 512, and 512 filters, respectively. Following a bottleneck layer computing another 512 feature maps, the decoder path again replaces the dense blocks with 6 layers that perform 4×4 same transpose convolution with stride 2, instance normalization, dropout with probability $p = 0.5$ in the first two layers, and ReLU activation. Similarly, these layers in the decoder path apply 512, 512, 512, 256, 128, and 64 filters, respectively. A final convolutional layer with hyperbolic tangent activation maps to the desired number of channels to produce the ultimate output.

The discriminator architecture utilized in the unpaired-data approach is comparable to that of the paired-data approach described in Section 6.2.3 with the key distinction being the adoption of instance normalization[45] rather than batch normalization.[46] The batch size of 1 chosen during training for both the paired- and unpaired-data approaches renders this distinction moot in practice, but it is an important consideration if a larger batch size is

chosen.

Similar to the process described in the paired-data case, optimization of the trainable parameters contained in the generative and discriminative models making up the unpaired-data CycleGAN architecture is performed in a sequential manner throughout the course of training. For both g_{loss} and d_{loss} for the forward and backward models, optimization is performed using the Adam optimizer with learning rate 2×10^{-4}, $\beta_1 = 0.5$, $\beta_2 = 0.999$, and $\hat{\epsilon} = 10^{-7}$.

6.2.5 Ensemble

Borrowing from the concept of ensemble learning, the first approach to improving upon the unpaired-data results we adopt may be categorized broadly as a stacking approach in which multiple models are used to produce the final output.[38] Although stacking approaches have most often been applied in image classification tasks using the average of predictions made by multiple models in parallel,[47–49] a sequential cascade approach more akin to the one we adopt here has been applied to an image reconstruction task.[50] In a similar way, the ensemble approach used here utilizes unpaired-data and paired-data models applied sequentially to produce the final output. The sCT output of the trained CycleGAN model is treated as an intermediate result that is in turned used along with the ground truth CT image as the training input for the FCDN architecture applied to learn a secondary sCT-to-CT mapping.

Optimization of the trainable parameters of the secondary paired-data models utilized in this cascade ensemble approach is performed in the same manner as described for the MRI-to-CT models in Section 6.2.3.

6.2.6 IDOL

As an alternative to the more established ensemble approach, the second approach to improving upon the unpaired-data results we explore here is the IDOL training strategy previously introduced by our group, which aims to produce patient-specific models by leveraging the behavior of a model intentionally overfitted to a personalized training data set. The formulation and backing of the IDOL framework has been described in detail elsewhere.[39] Briefly, the approach is motivated by the focus on individual patients characteristic of the idea of precision medicine, an idea that runs counter to the conventional wisdom in deep learning settings in which the goal is to produce a model that generalizes well to all future patients. In the IDOL approach, this conventional generalized model is further fine-tuned through a secondary training process that utilizes patient- and task-specific prior information as illustrated in Figure 6.4. In application to the sCT reconstruction task, this secondary training data set consists of deformed image pairs achieved through the repeated application of randomized deformations to a patient's simulation MRI and CT scans.

Evaluation of the IDOL approach in the paired- and unpaired-data settings was performed for the five patients in fold 3, chosen at random. For both the paired-data and unpaired-data approaches, the same training data set consisting of 30 deformed image sets derived from the application of random deformation vector fields to the patient-of-interest's MRI and CT simulation scans was used in the second stage of training for each patient. These deformed image sets were produced as follows:

1. Seed pixels were selected at random from a binary mask of muscle-containing regions in the source MR image.

2. The sign and magnitude of the deformation vector at each of these seed pixels was randomized, with x and y components of the magnitude being controlled by user-defined strength and offset parameters.

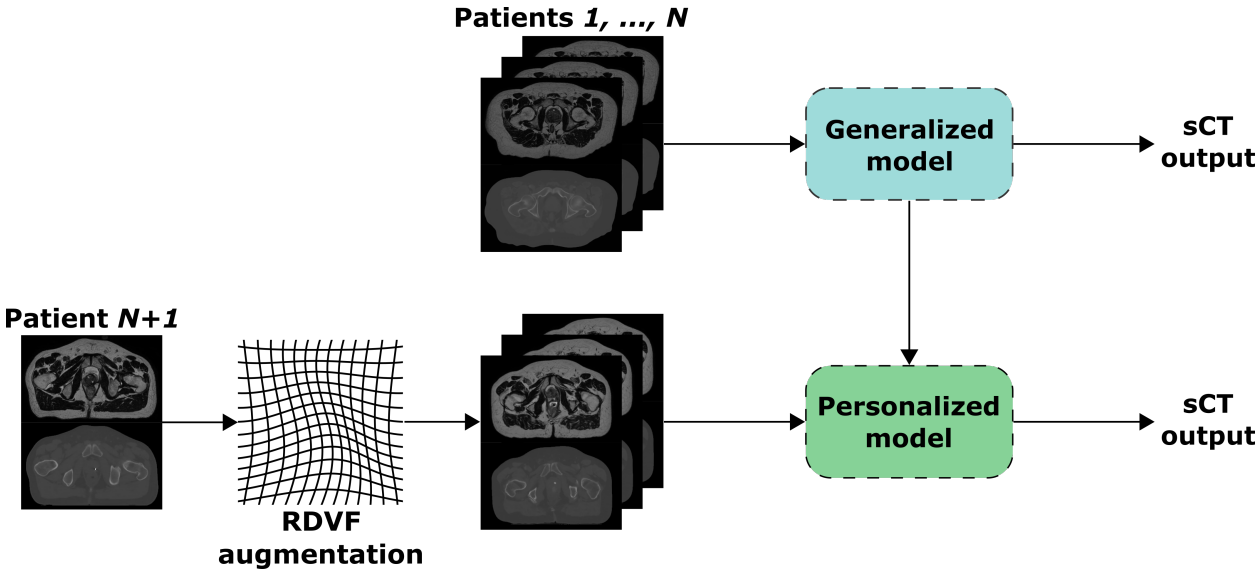

Figure 6.4. Two stage training process of the IDOL approach. In the first stage, a generalized model is trained in the conventional manner with data from the first N patients. This generalized model is fine-tuned to produce a personalized model tailored to patient $N+1$ in the second stage of training. The secondary training data set is constructed through the repeated application of random deformation vector fields (RDVF) to the simulation image data of the patient of interest. Adapted from Chun et al.[39]

3. A 41×41 Gaussian filter with $\sigma = 10$ was applied to produce a continuous deformation vector field.

4. Pixel values in the resulting deformed MR and CT images were assigned as the weighted average of a four pixel neighborhood surrounding the source coordinates.

Personalized models for each of these five patients were produced through 60 epochs of additional training with this patient-specific training data set, subject to the same optimization processes described in Sections 6.2.3 and 6.2.4 for the paired- and unpaired-data approaches, respectively.

6.2.7 Evaluation

Evaluation of the various methods explored here is focused on comparisons of image quality. Visual comparisons between the sCT outputs of each model are accompanied by difference maps between the true CT image and sCT outputs. Subsequent quantitative comparisons make use of the MAE and peak signal-to-noise ratio (PSNR):

$$\text{MAE} = \frac{\sum_{i=1}^{n} |CT_i - sCT_i|}{n}, \tag{6.8}$$

where n is the total number of pixels within the region of interest, and

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{\text{MSE}} \right), \tag{6.9}$$

where MAX_I is the maximum possible pixel value governed by the data type of the input image and the mean square error (MSE) between the ground truth CT and predicted sCT is calculated as

$$\text{MSE} = \frac{\sum_{i=1}^{n} (CT_i - sCT_i)^2}{n}, \tag{6.10}$$

with n being the total number of pixels within the body contour.

Pixel-wise agreement as measured by the MAE is evaluated both within the body contour and within a binary mask containing only regions of bone (HU $>$ 200). The evaluation of image quality using the PSNR is performed within the body contour. Finally, the evaluation of statistical significance between methods is performed with the left- or right-tailed Wilcoxon signed rank test[51] where appropriate.

6.3 Results

Completing 120 epochs of training for the paired-data and unpaired-data base models required approximately 13 hours and 30 hours, respectively, on average across the five folds. An additional 60 epochs of training using the IDOL approach required another 8 hours and 19 hours for the paired- and unpaired-data models, respectively, on average across the five validation patients. Finally, the 120 epochs of secondary training undertaken for the ensemble approach required an average of 13 hours across the five folds, like in the case of training the base paired-data models. Considering the time required to train the initial unpaired-data model in the cascade, the ensemble approach required an average of 43 hours on average to train. At the time of deployment, inference is performed in a sub-second time frame per slice in each case.

The visual comparison in Figure 6.5 presents the input MR image, corresponding ground truth CT image, and sCT outputs of the various approaches explored here at a representative slice. Difference maps between the ground truth CT and sCT outputs illustrate differences in pixel intensities in units of HU. These difference maps demonstrate general trends observable across the validation folds. Stark improvements are seen with the adoption of the IDOL training strategy in the paired-data setting (Figure 6.5d) and to a lesser extent in the unpaired-data setting, with the CycleGAN IDOL output (Figure 6.5g) demonstrating improvements at tissue boundaries and in larger regions of bone. Similarly, Figure 6.5e illustrates the smoothing effect characteristic of the cascade ensemble approach that leads to improvements in the bulk soft tissue and within regions of bone but also acts to blur certain boundaries.

Observations from this visual comparison reflect what is shown in the distributions of the quantitative metrics computed over the data set. MAE values calculated over all five folds for the paired-data FCDN models, unpaired-data CycleGAN models, and cascade ensemble

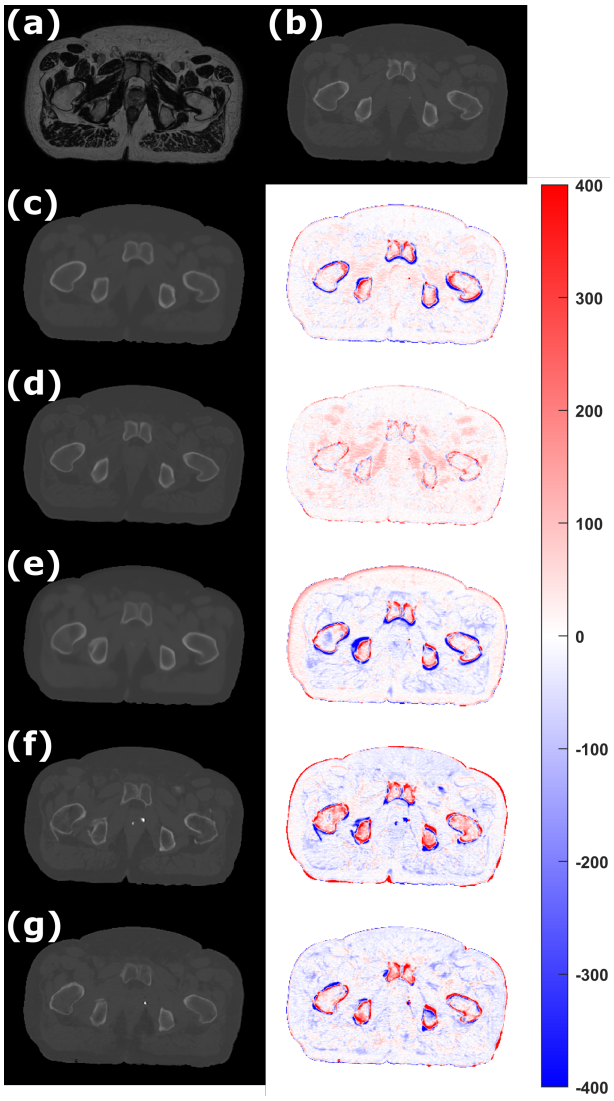

Figure 6.5. Image comparisons for a representative slice showing the input MR image (a) and true CT image (b) along with sCT outputs and $CT - sCT$ difference maps for the paired-data FCDN (c), FCDN IDOL (d), cascade ensemble (e), unpaired-data CycleGAN (f), and CycleGAN IDOL (g) models. Values in the difference maps are in units of HU.

models are presented in Figure 6.6 alongside the subset of patients in fold 3 for whom the IDOL approach was performed to produce personalized paired- and unpaired-data models. Similarly, Figure 6.7 and 6.8 present the same comparison for MAE values calculated within regions of bone and the PSNR, respectively. Mean values for these three metrics are presented for each model in Table 6.1 for comparison.

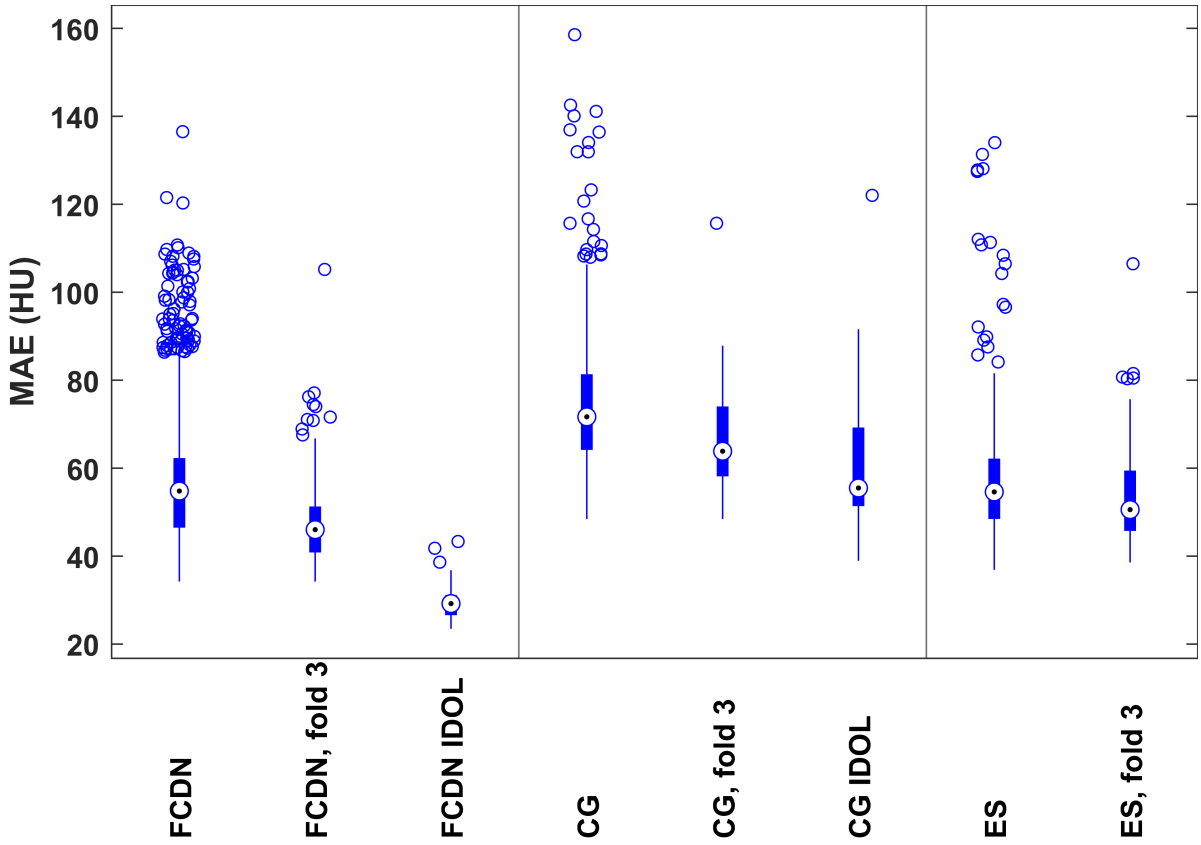

Figure 6.6. MAE values calculated within the body contour. Values over 1369 images for the base paired-data FCDN, unpaired-data CycleGAN (CG), and cascade ensemble (ES) models are shown at left in each region, followed by values over 290 images for patients of fold 3 for whom personalized IDOL models were also produced. Mean values are included in Table 6.1 for comparison.

Statistical significance between approaches was established as follows, all with $p < .001$ unless otherwise noted:

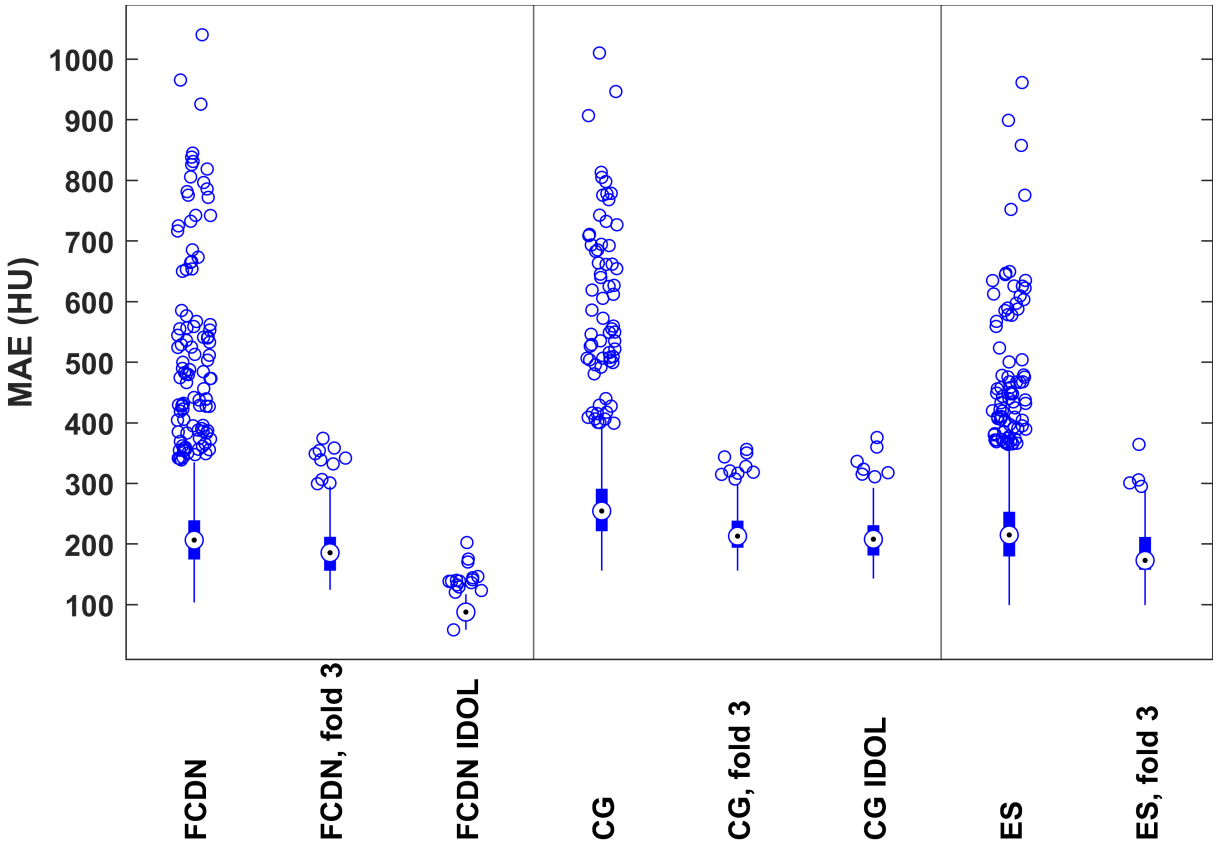

Figure 6.7. MAE values calculated only within regions of bone. Values over 1369 images for the base paired-data FCDN, unpaired-data CycleGAN (CG), and cascade ensemble (ES) models are shown at left in each region, followed by values over 290 images for patients of fold 3 for whom personalized IDOL models were also produced. Mean values are included in Table 6.1 for comparison.

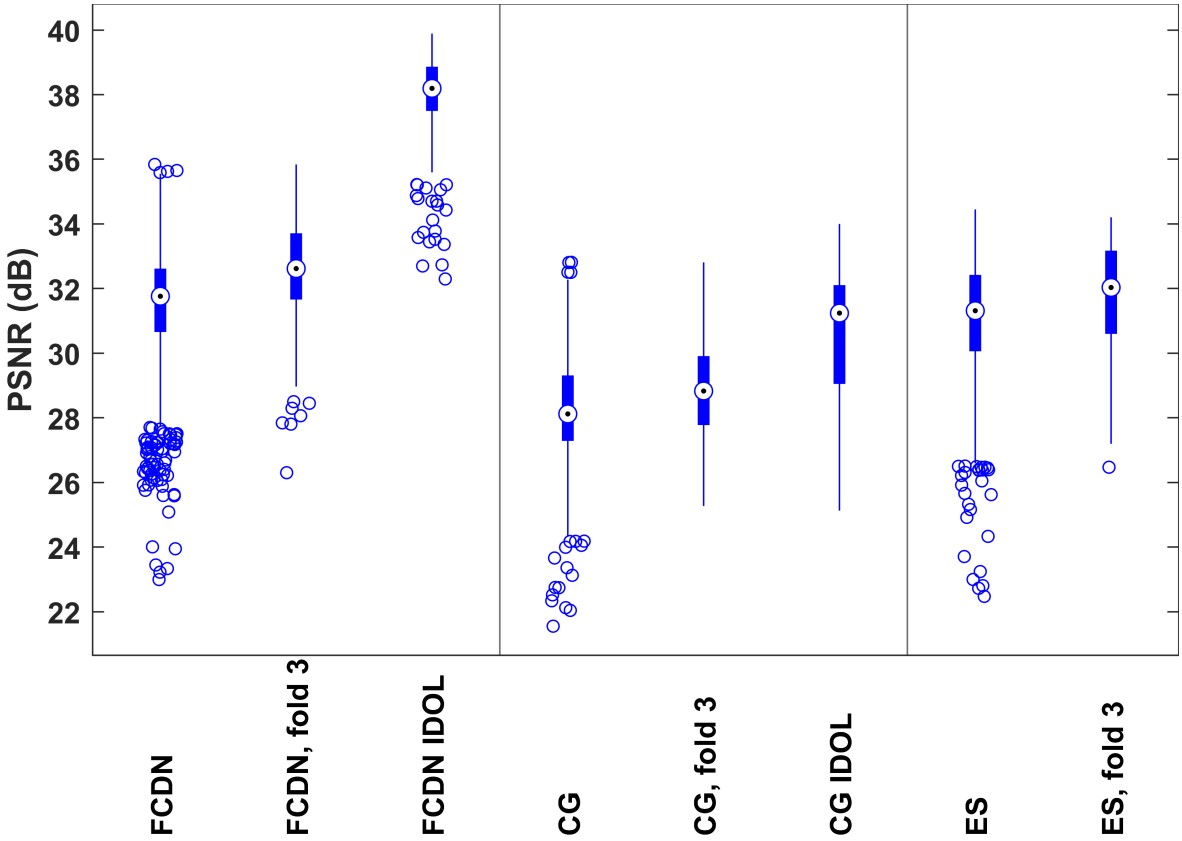

Figure 6.8. PSNR values calculated within the body contour. Values over 1369 images for the base paired-data FCDN, unpaired-data CycleGAN (CG), and cascade ensemble (ES) models are shown at left in each region, followed by values over 290 images for patients of fold 3 for whom personalized IDOL models were also produced. Mean values are included in Table 6.1 for comparison

1. Adopting the IDOL approach for patients in fold 3 offers significant improvements in body MAE, bone MAE, and PSNR for both the paired- and unpaired-data approaches.

2. The cascade ensemble offers additional and significant benefits over the unpaired-data IDOL models in all three metrics for patients in fold 3.

3. Further, the ensemble cascade models demonstrate significant improvements over the base CycleGAN models in all three metrics.

4. The paired-data approach demonstrates significantly better performance in all metrics than the unpaired-data approach in the base case and better performance than the ensemble approach in bone MAE and PSNR, but differences in overall body MAE are statistically insignificant ($p = .06$).

Table 6.1. Mean ± standard deviation listed for the MAE calculated within the body contour and in regions of bone as well as the PSNR within the body contour. Evaluations in the base case for the paired-data FCDN, unpaired-data CycleGAN (CG), and cascade ensemble (ES) models are made over 1369 images. Evaluations of the personalized IDOL approach for patients in fold 3 are made over 290 images.

	FCDN ($n = 1369$)	FCDN, fold 3 ($n = 290$)	FCDN IDOL ($n = 290$)	CG ($n = 1369$)	CG, fold 3 ($n = 290$)	CG IDOL ($n = 290$)	ES ($n = 1369$)	ES, fold 3 ($n = 290$)
MAE, body (HU)	57 ± 14	47 ± 9	29 ± 3	74 ± 14	66 ± 10	60 ± 12	56 ± 11	53 ± 10
MAE, bone (HU)	229 ± 105	189 ± 45	90 ± 18	271 ± 93	220 ± 36	208 ± 39	232 ± 87	185 ± 41
PSNR, body (dB)	31 ± 2	33 ± 2	38 ± 1	28 ± 2	29 ± 1	31 ± 2	31 ± 2	32 ± 2

6.4 Discussion

As the MRI-guided RT environment and its horizons have expanded, the technical challenge of sCT reconstruction has prompted a great deal of development ranging from early explorations of atlas- and classification-based approaches[18,19] to the recent focus on deep learning approaches to the problem.[20] A defining characteristic of many previous studies in this space is the choice of network architecture and the accompanying requirement (or lack thereof) for paired training data. Surveyed studies present a notable slant towards the paired-data setting, with approximately 84% of approaches falling into this category.[20] These paired-data approaches represent an intuitive entry point to the problem of deep learning-based sCT reconstruction as the correspondence between images in the training data set enables the use of direct comparisons between true and predicted images in the choice of loss function, rendering the training process relatively transparent.[52,53] However, the challenge of collecting a training data set of sufficient size and quality under the restriction of this paired-data requirement can be limiting in the low-data radiation oncology setting, which is a motivating factor in the adoption of the unpaired-data approaches exemplified by CycleGAN. Despite this attractive advantage of the unpaired-data approach, the gap between paired-data and unpaired-data results can be difficult to close, which is an observation that has been reported in the original CycleGAN paper[33] and other studies in this space.[35–37,54] We have again demonstrated that gap here, but we also have explored two distinct approaches that are successful in bridging the gap between the paired- and unpaired-data approaches to some degree.

First among these approaches is the cascade ensemble approach in which the unpaired-data CycleGAN and paired-data FCDN architectures are applied in sequence to produce the ultimate sCT output. Adopting the ensemble approach offers significant improvements in image quality and reconstruction fidelity in regions of bone over the unpaired-data approach

alone. It is important to note, however, that the implementation explored here is ultimately a paired-data approach, which may not always be a feasible design choice considering the challenges of data availability that we have already discussed. The design of the ensemble itself is a topic deserving of further study with a focus on determining optimal choices for the size of the ensemble, the complexity of the constituent networks, and the relationships between networks in the ensemble. We have adopted a simple sequential structure here as a point of comparison, but there are certainly a wide array of design choices that could have been made instead, taking inspiration from other common ensemble learning strategies like bagging or boosting.[38] A final note on the ensemble approach regards the time required for inference at deployment: although it was still a sub-second process to generate images with the ensemble implemented here, as the structure of the ensemble becomes more complex with more constituent models, the latency associated with producing the ultimate output may become a concern in time-sensitive settings like the online adaptive therapy workflow.

In our exploration of the second approach to improving the unpaired-data results we sought to answer a novel question: does an unpaired-data approach benefit from a personalized training strategy initially designed with paired-data approaches in mind? The IDOL approach aims to leverage the behavior of a model that has been intentionally overfitted to a patient-specific data set, taking advantage of the strong supervision granted with paired training samples. The significant improvements observed in the unpaired-data results with the adoption of the IDOL training strategy suggest that unpaired-data approaches may in fact benefit from some degree of implicit supervision that exists in a tailored, patient-specific training data set. This result echoes the tangentially related finding of previous studies that even imperfect registration or otherwise pairing data in the training data set is beneficial in the unpaired-data setting.[55–58] In this study, we maintain a similar degree of supervision with the paired training data set in the unpaired-data setting, but also demonstrate that the additional implicit supervision granted with the patient-specific IDOL training strategy

offers the potential for significant improvements in this setting.

Although the IDOL approach as implemented in the present study did utilize paired data, the approach to its benefit does not strictly require implementing a paired-data architecture as in the case of the cascade ensemble. In furthering this point, an area of future interest is the implementation of the IDOL training strategy in the unpaired-data setting without perfectly corresponding training image pairs. Designed to operate in the conventional early MRI-guided workflows that still rely on CT simulation, the IDOL approach does however require MRI and CT data for a given patient, which limits its applicability in a pure MRI-only setting. Considering this, approaches that may be trained with existing retrospective databases of MR and CT images may be more beneficial in the MRI-only workflows of the future. An additional limitation of the IDOL approach as implemented in the present study is specific and inherent to the treatment setting of patients involved in this study. In contrast to the MRI-guided setting utilizing a combination MR-linac in which MR setup images are acquired at each fraction, the treatments in this study were performed with pre-treatment MRI and CT simulation for guidance but using conventional C-arm linacs with on-board CBCT imaging capabilities. In the MR-linac case, MR setup images may be used as test cases for personalized models trained with data sets derived from the simulation image sets.[39] Owing to the fact that setup images in this case are CBCT images, the only corresponding MRI and CT scans available in this study are the simulation images. As such, IDOL training was performed with deformations of the simulation MR data that was used in testing for each patient. There is a lesser degree of separation in this case than in the ideal setting in which IDOL was introduced, but we maintain the principle of testing a network on unseen data nevertheless.

A final limitation of each of the models implemented here is the single plane 2D architecture utilized in each case. The loss of through-plane structural information may hinder a model in the reconstruction task, prompting the adoption of multi-plane 2D approaches at

the cost of increased computational expense or model complexity.[59,60] Even further, fully 3D architectures may be utilized to completely capture the complexity of a patient's anatomy, but these architectures face their own challenges of heavily increased memory usage that introduce design compromises (e.g. small patch-based reconstruction), leading to trade-offs that may not be feasible or desirable in all situations.[20]

6.5 Conclusions

The methods explored here demonstrate the gap between paired- and unpaired-data approaches to the sCT reconstruction task that can be limiting in the MRI-only RT setting. The cascade ensemble that combines multiple generative models and the personalized IDOL training strategy each offer a means of closing this gap to a certain extent, and each method carries its own benefits and requirements for implementation. Ultimately, these methods are a step towards enabling the full utilization of powerfully expressive unpaired-data approaches to sCT reconstruction, avoiding the requirement for paired training data that can be fundamentally limiting in the low-data radiation oncology setting.

References

[1] Mutic S, Dempsey JF. The ViewRay system: Magnetic resonance–guided and controlled radiotherapy. *Semin Radiat Oncol.* 2014;24(3):196–199. doi: 10.1016/j.semradonc.2014.02.008

[2] Fallone BG. The rotating biplanar linac-magnetic resonance imaging system. *Semin Radiat Oncol.* 2014;24:200–202. doi:10.1016/j.semradonc.2014.02.011

[3] Keall PJ, Barton M, Crozier S. The Australian magnetic resonance imaging-linac program. *Semin Radiat Oncol.* 2014;24:203–206. doi:10.1016/j.semradonc.2014.02.015

[4] Raaymakers BW, Jürgenliemk-Schulz IM, Bol GH, et al. First patients treated with a 1.5 T MRI-Linac: clinical proof of concept of a high-precision, high-field MRI

guided radiotherapy treatment. *Phys Med Biol*. 2017;62(23):L41–L50. doi:10.1088/1361-6560/aa9517

[5] Fischer-Valuck BW, Henke L, Green O, et al. Two-and-a-half-year clinical experience with the world's first magnetic resonance image guided radiation therapy system. *Adv Radiat Oncol*. 2017;2(3):485–493. doi:10.1016/j.adro.2017.05.006

[6] Wen N, Kim J, Doemer A, et al. Magnetic resonance guided linear accelerator for stereotactic radiosurgery treatment. In *Proceedings of the American Society for Radiation Oncology*. p. E479. doi:j.ijrobp.2018.07.1369

[7] Pollard JM, Wen Z, Sadagopan R, Wang J, Ibbott GS. The future of image-guided radiotherapy will be MR guided. *Br J Radiol*. 2017;90(1073):20160667. doi: 10.1259/bjr.20160667

[8] Noel CE, Parikh PJ, Spencer CR, et al. Comparison of onboard low-field magnetic resonance imaging versus onboard computed tomography for anatomy visualization in radiotherapy. *Acta Oncol (Madr)*. 2015;54(9). doi:10.3109/0284186X.2015.1062541

[9] Chandarana H, Wang H, Tijssen R, Das IJ. Emerging role of MRI in radiation therapy. *J Magn Reson Imaging*. 2018;48(6):1468–1478. doi:10.1002/jmri.26271

[10] Otazo R, Lambin P, Pignol JP, et al. MRI-guided radiation therapy: An emerging paradigm in adaptive radiation oncology. *Radiology*. 2020;298(2):248–260. doi: 10.1148/radiol.2020202747

[11] Corradini S, Alongi F, Andratschke N, et al. MR-guidance in clinical reality: current treatment challenges and future perspectives. *Radiat Oncol*. 2019;14:92. doi: 10.1186/s13014-019-1308-y

[12] Murphy MJ, Balter J, Balter S, et al. The management of imaging dose during image-guided radiotherapy: Report of the AAPM Task Group 75. *Med Phys*. 2007;34(10):4041–4063. doi:10.1118/1.2775667

[13] Bruijnen T, Stemkens B, Terhaard C, Lagendijk J, Raaijmakers C, Tijssen R. MRI-based radiation therapy: Intrafraction motion quantification of head and neck tumors using cine magnetic resonance imaging. *Int J Radiat Oncol Biol Phys*. 2018;100(5):1358. doi:10.1016/j.ijrobp.2017.12.134

[14] Shirato H, Seppenwoolde Y, Kitamura K, Onimura R, Shimizu S. Intrafractional tumor motion: lung and liver. *Semin Radiat Oncol*. 2004;14(1):10–18. doi: 10.1053/j.semradonc.2003.10.008

[15] Jadon R, Pembroke C, Hanna C, et al. A systematic review of organ motion and image-guided strategies in external beam radiotherapy for cervical cancer. *Clin Oncol (R Coll Radiol)*. 2014;26(4):185–196. doi:10.1016/j.clon.2013.11.031

[16] Jonsson J, Nyholm T, Söderkvist K. The rationale for MR-only treatment planning for external radiotherapy. *Clin Transl Radiat Oncol*. 2019;18:60–65. doi: 10.1016/j.ctro.2019.03.005

[17] Tyagi N, Zelefsky MJ, Wibmer A, et al. Clinical experience and workflow challenges with magnetic resonance-only radiation therapy simulation and planning for prostate cancer. *Phys Imaging Radiat Oncol*. 2020;16:43–49. doi:10.1016/j.phro.2020.09.009

[18] Edmund JM, Nyholm T. A review of substitute CT generation for MRI-only radiation therapy. *Radiat Oncol*. 2017;12(1):28. doi:10.1186/s13014-016-0747-y

[19] Johnstone E, Wyatt JJ, Henry AM, et al. Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging–only radiation therapy. *Int J Radiat Oncol Biol Phys*. 2018;100(1):199–217. doi: 10.1016/j.ijrobp.2017.08.043

[20] Spadea MF, Maspero M, Zaffino P, Seco J. Deep learning based synthetic-CT generation in radiotherapy and PET: A review. *Med Phys*. 2021;48(11):6537–6566. doi: 10.1002/mp.15150

[21] Olberg S, Zhang H, Kennedy WR, et al. Synthetic CT reconstruction using a deep spatial pyramid convolutional framework for MR-only breast radiotherapy. *Med Phys*. 2019;46(9):4135–4147. doi:10.1002/mp.13716

[22] Dowling JA, Sun J, Pichler P, et al. Automatic substitute computed tomography generation and contouring for magnetic resonance imaging (MRI)-alone external beam radiation therapy from standard MRI sequences. *Int J Radiat Oncol Biol Phys*. 2015; 93(5):1144–1153. doi:10.1016/j.ijrobp.2015.08.045

[23] Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys*. 2017;44(4):1408–1419. doi:10.1002/mp.12155

[24] Ali AN, Rossi PJ, Godette KD, et al. Impact of magnetic resonance imaging on computed tomography-based treatment planning and acute toxicity for prostate cancer patients treated with intensity modulated radiation therapy. *Pract Radiat Oncol*. 2013; 3(1):e1–e9. doi:10.1016/j.prro.2012.04.005

[25] Seymour ZA, Fogh SE, Westcott SK, et al. Interval from imaging to treatment delivery in the radiation surgery age: How long is too long? *Int J Radiat Oncol Biol Phys*. 2015; 93(1):126–132. doi:10.1016/j.ijrobp.2015.05.001

[26] Tyagi N, Fontenla S, Zelefsky M, et al. Clinical workflow for MR-only simulation and planning in prostate. *Radiat Oncol*. 2017;12(119):1–12. doi:10.1186/s13014-017-0854-4

[27] Kerkmeijer L, Maspero M, Meigher G, van der Voort van Zyp J, de Boer H, van den Berg C. Magnetic resonance imaging only workflow for radiotherapy simulation and planning in prostate cancer. *Clin Oncol*. 2018;30(11):692–701. doi:10.1016/j.clon.2018.08.009

[28] Das IJ, McGee KP, Tyagi N, Wang H. Role and future of MRI in radiation oncology. *Br J Radiol*. 2019;92(1094):20180505. doi:10.1259/bjr.20180505

[29] van der Heide UA, Frantzen-Steneker M, Astreinidou E, Nowee ME, van Houdt PJ. MRI basics for radiation oncologists. *Clin Transl Radiat Oncol*. 2019;18:74–79. doi: 10.1016/j.ctro.2019.04.008

[30] Glide-Hurst CK, Paulson ES, McGee K, et al. Task group 284 report: magnetic resonance imaging simulation in radiotherapy: considerations for clinical implementation, optimization, and quality assurance. *Med Phys*. 2021;48(7):e636–e670. doi: 10.1002/mp.14695

[31] Zhang Y, Kashani R, Cao Y, Lawrence TS, Johansson A, Balter JM. A hierarchical model of abdominal configuration changes extracted from golden angle radial magnetic resonance imaging. *Phys Med Biol*. 2021;66:045018. doi:10.1088/1361-6560/abd66e

[32] Olberg S, Chun J, Choi BS, et al. Abdominal synthetic CT reconstruction with intensity projection prior for MRI-only adaptive radiotherapy. *Phys Med Biol*. 2021;66:204001. doi:10.1088/1361-6560/ac279e

[33] Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. *ArXiv*. 2018;:1–18

[34] Cohen JP, Luck M, Honari S. Distribution matching losses can hallucinate features in medical image translation. *ArXiv*. 2018;:1–11

[35] Peng Y, Chen S, Qin A, et al. Magnetic resonance-based synthetic computed tomography images generated using generative adversarial networks for nasopharyngeal carcinoma radiotherapy treatment planning. *Radiother Oncol*. 2020;150:217–224. doi: 10.1016/j.radonc.2020.06.049

[36] Fu J, Singhrao K, Cao M, et al. Generation of abdominal synthetic CTs from 0.35T MR images using generative adversarial networks for MR-only liver radiotherapy. *Biomed Phys Eng Express*. 2020;6:015033. doi:10.1088/2057-1976/ab6e1f

[37] Li Y, Li W, Xiong J, Xia J, Xie Y. Comparison of supervised and unsupervised deep learning methods for medical image synthesis between computed tomography and magnetic resonance images. *Biomed Res Int*. 2020;2020(5193707):1–9. doi: 10.1155/2020/5193707

[38] Ganaie M, Hu M, Malik A, Tanveer M, Suganthan P. Ensemble deep learning: A review. *ArXiv*. 2022;:1–49

[39] Chun J, Park JC, Olberg S, et al. Intentional deep overfit learning (IDOL): A novel deep learning strategy for adaptive radiation therapy. *Med Phys*. 2022;49(1):488–496. doi:10.1002/mp.15352

[40] Jégou S, Drozdzal M, Vazquez D, Romero A, Bengio Y. The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation. *ArXiv*. 2017;:1–9

[41] Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-scale machine learning on heterogeneous systems. 2015. Software available from tensorflow.org

[42] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. *ArXiv*. 2015;:1–8

[43] Kingma D, Ba J. Adam: A method for stochastic optimization. *ArXiv*. 2017;:1–15

[44] Isola P, Zhu JY, Zhou T, Afros AA. Image-to-image translation with conditional adversarial networks. *ArXiv*. 2016;:1–17

[45] Ulyanov D, Vedaldi A, Lempitsky V. Instance normalization: The missing ingredient for fast stylization. *ArXiv*. 2016;:1–6

[46] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *ArXiv*. 2015;:1–11

[47] Streeter M. Approximation algorithms for cascading prediction models. In *Proceedings of the 35th International Conference on Machine Learning*. pp. 4752–4760

[48] Kondratyuk D, Tan M, Brown M, Gong B. When ensembling smaller models is more efficient than single large models. *ArXiv*. 2020;:1–5

[49] Lobacheva E, Chirkova N, Kodyran M, Vetrov D. On power laws in deep ensembles. *ArXiv*. 2021;:1–22

[50] Schlemper J, Caballero J, Hajnal JV, Price A, Rueckert D. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *ArXiv*. 2017;:1–14

[51] Wilcoxon F. Individual comparisons by ranking methods. *Biometrics Bulletin*. 1945; 1:80–83

[52] Zhao H, Gallo O, Frosio I, Kautz J. Loss functions for neural networks for image processing. *ArXiv*. 2015;:1–11

[53] Largent A, Barateau A, Nunes JC, et al. Comparison of deep learning-based and patch-based methods for pseudo-CT generation in MRI-based prostate dose planning. *Int J Radiat Oncol Biol Phys*. 2019;105(5):1137–1150. doi:10.1016/j.ijrobp.2019.08.049

[54] Klages P, Benslimane I, Riyahi S, et al. Patch-based generative adversarial neural network models for head and neck MR-only planning. *Med Phys*. 2020;47(2):626–642. doi:10.1002/mp.13927

[55] Harms J, Lei Y, Wang T, et al. Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. *Med Phys*. 2019;46(9):3998–4009. doi: 10.1002/mp.13656

[56] Jin CB, Kim H, Liu M, et al. Deep CT to MR synthesis using paired and unpaired data. *Sensors (Basel)*. 2019;19(10):2361. doi:10.3390/s19102361

[57] Lei Y, Harms J, Wang T, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. *Med Phys*. 2019;46(8):3565–3581. doi: 10.1002/mp.13617

[58] Yang H, Sun J, Carass A, et al. Unsupervised MR-to-CT synthesis using structure-constrained CycleGAN. *IEEE Trans Med Imaging*. 2019;39(12):4249–4261. doi: 10.1109/TMI.2020.3015379

[59] Spadea MF, Pileggi G, Zaffino P, et al. Deep convolution neural network (DCNN) multiplane approach to synthetic CT generation from MR images—application in brain proton therapy. *Int J Radiat Oncol Biol Phys*. 2019;105(3):495–503. doi: 10.1016/j.ijrobp.2019.06.2535

[60] Maspero M, Bentvelzen LG, Savenije MH, et al. Deep learning-based synthetic CT generation for paediatric brain MR-only photon and proton radiotherapy. *Radiother Oncol*. 2020;153:197–204. doi:10.1016/j.radonc.2020.09.029

Chapter 7

Conclusions

The works detailed in this dissertation address shortcomings of the MRI-guided ART workflow at three critical stages: 1) the expensive re-planning process that occurs when the anatomy observed at treatment differs drastically from that represented in the simulation treatment plan; 2) the acquisition of MR images for treatment guidance, a process limited by the trade-off between spatial resolution and the time required for a given scan; and 3) the challenge of synthetic CT (sCT) reconstruction to enable an MR-only RT workflow in which the requirement for a secondary CT simulation scan is eliminated. An integrated software suite that seamlessly connects the functionalities of the various approaches discussed here could represent a means to achieving a streamlined and effective MR-only RT workflow in a way that has not previously been possible. Investigating these integrated solutions is an important direction of future interest considering the time-critical nature of the adaptive setting—reducing the time required at each stage of the adaptive therapy process reduces the time a patient must remain on the couch in treatment position.

The OAR grouping method discussed in Chapter 2 was applied with the aim of improving target coverage in adapted fractions during MRI-guided daily ART of pancreatic cancer. The motivation of the method is to render simulation treatment plans used throughout the course of treatment more robust to anatomical variations observed daily, especially variations in the size and position of the primary critical structures relative to the target. In the conventional approach, these four primary structures—the stomach, duodenum, small bowel, and large bowel—are considered separately in the optimizer, meaning that the weighting applied to a structure at simulation may not be applicable at a later treatment fraction. By considering these four structures as a single combined structure, the OAR grouping method greatly simplifies the treatment planning process in this setting and ultimately led to improved prescription dose coverage of the target in 84% of adapted fractions.

Chapter 3 presents a framework for MRI super-resolution (SR) reconstruction that enables the use of fast (~ 3 s), low-resolution (LR, 64×64 pixels, 6.0×6.0 mm per pixel) scans

for treatment guidance through fourfold upsampling. Acquiring true high-resolution (HR) images of the same specifications as the SR outputs required an average scan period of 17 s, meaning an ultimate time savings of over 14 seconds could be achieved. Considering the detrimental effects of motion artifacts that can arise in even the relatively short 17 s acquisition time, this time savings translates directly into important clinical impacts of improving image quality and patient comfort. The primary innovation of the proposed framework is the focus placed on producing a training data set of paired LR and HR images that reflects clinical reality through the use of the down-sampling network (DSN). As opposed to a more conventional approach such as k-space down-sampling that represents an idealized transformation, the DSN was trained with true LR and HR image pairs from a small number of volunteers to learn a true mapping from the HR domain to the LR domain. Applying this mapping to abundant HR data allowed for the creation of a data set of sufficient size and quality for the training of the ultimate SR generative model that was successfully applied to the SR reconstruction task for breath-hold 3D scans and free-breathing 4D-MRI acquisitions.

Approaches to the various technical and site-specific challenges encountered in the sCT reconstruction task discussed in Chapters 4, 5, and 6 represent steps towards achieving an MR-only RT workflow in which MRI is the sole imaging modality used for treatment planning and guidance, eliminating the requirement for CT simulation in this setting. Compared to a ubiquitous and parameter-dense network architecture commonly used in the sCT task, the parameter-efficient deep spatial pyramid pooling architecture utilized in Chapter 4 offers considerable improvements in the time required for training while also improving resulting image quality as measured by the mean absolute error by nearly 40%.

In Chapter 5, the proposed preprocessing approach aimed to enable the use of a supervised, paired-data framework for sCT reconstruction in the abdomen, where the primary barrier is the variable presence of intestinal gas in corresponding MRI and CT scans. These discrepancies present an obvious clinical impact on the accuracy of dose calculations in this

setting, giving rise to differences in calculated target coverage of approximately 13% on average and a degree of uncertainty in the high-dose region when unaccounted for that may limit the effectiveness of any dose escalation efforts pursued through plan adaptation. An important point of future interest in the abdomen in particular is gaining an understanding of any interplay effects that may arise from the complex interactions between anatomical motion, imaging, and dose delivery on an inter- or intrafractional basis and how these effects relate to the sCT reconstruction task. The results discussed here grant an incomplete view of a concept central to the practical challenge of RT but represent an important first step towards illuminating the bigger picture.

Finally, the exploration of supervised, paired-data and unsupervised, unpaired-data approaches to sCT reconstruction in Chapter 6 illustrates the gap between the paired- and unpaired-data results that can be difficult to close. The methods proposed for the improvement of the unpaired-data sCT results offered significant improvements in reconstruction fidelity and image quality, and each method carries unique benefits and applications. The cascade ensemble approach in which the unpaired- and paired-data architectures are applied in sequence to produce the final sCT output is ultimately a paired-data approach as implemented here, but exploring a variety of ensemble techniques is a promising area for future studies. Of particular interest was the application of the personalized training strategy originally designed for use in the paired-data setting to the unpaired-data architecture. The improvements in image quality and accuracy offered by the personalized models demonstrate that the unpaired-data approach does benefit from some degree of implicit supervision granted by a tailored, patient-specific training data set, cementing the promise of the training approach in both the paired- and unpaired-data settings. Ultimately, the contributions described in this dissertation represent steps towards enabling an MR-only RT workflow that aims to utilize the attractive capabilities that integrated MRI-guided RT platforms offer to the fullest extent.