The Texas Medical Center Library

DigitalCommons@TMC

Dissertations and Theses (Open Access)

MD Anderson UTHealth Houston Graduate School

6-2023

# Ai-Enabled Online Plan Adaptation For Mr-Guided Stereotactic Ablative Radiotherapy (Sabr) Of Head And Neck Cancer

Yao Zhao

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations

Part of the Health and Medical Physics Commons

## Recommended Citation

Zhao, Yao, "Ai-Enabled Online Plan Adaptation For Mr-Guided Stereotactic Ablative Radiotherapy (Sabr) Of Head And Neck Cancer" (2023). *Dissertations and Theses (Open Access)*. 1283. https://digitalcommons.library.tmc.edu/utgsbs_dissertations/1283

This Dissertation (PhD) is brought to you for free and open access by the MD Anderson UTHealth Houston Graduate School at DigitalCommons@TMC. It has been accepted for inclusion in Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digcommons@library.tmc.edu.

The TMC LIBRARY
Health Sciences Resource Center

# AI-ENABLED ONLINE PLAN ADAPTATION FOR MR-GUIDED STEREOTACTIC ABLATIVE RADIOTHERAPY (SABR) OF HEAD AND NECK CANCER

by

*Yao Zhao, M.S.*

APPROVED:

Jinzhong Yang, Ph.D.
Advisory Professor

Laurence Court, Ph.D.

Xin Wang, Ph.D.

He Wang, Ph.D.

Jack Phan, M.D., Ph.D.

Tinsu Pan, Ph.D.

APPROVED:

Dean, The University of Texas
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

# AI-ENABLED ONLINE PLAN ADAPTATION FOR MR-GUIDED STEREOTACTIC ABLATIVE RADIOTHERAPY (SABR) OF HEAD AND NECK CANCER

A
DISSERTATION

Presented to the Faculty of

The University of Texas
MD Anderson Cancer Center UTHealth
Graduate School of Biomedical Sciences

in Partial Fulfillment
of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

by

Yao Zhao, M.S.
Houston, Texas
May 2023

# Dedication

To my parents, Chuanming Zhao and Xiaojuan Wu, words cannot express how much your love and support have meant to me throughout this journey. Your support and encouragement have made it possible for me to reach this point in my life, and for that, I am forever grateful. I dedicate this thesis to both of you with all my love and appreciation for everything you have done for me. Thank you for being my guiding lights and for always believing in me.

To my wife Ping Jiang, you have been my constant source of inspiration, support, and encouragement throughout my educational journey. Your unwavering love, patience, and understanding have been the foundation of my accomplishments. I am grateful for every moment we have shared together and for all the sacrifices you have made to ensure that I had the time and space to pursue my dreams. Thank you for being my partner in life and for sharing in my dreams.

# Acknowledgements

First, I would like to express my deep appreciation to UTHealth and MD Anderson for the outstanding academic support and research resources that have enabled me to achieve my research goals. The past several years has been an incredibly enriching experience, and I feel fortunate to have been a part of the MDA community.

I would also like to express my profound gratitude to my advisor, Dr. Jinzhong Yang, for your invaluable mentorship and guidance throughout my PhD studies. I am grateful for the countless hours you spent discussing my research and challenging me to think more critically. Your guidance and support have been essential to my academic and professional growth, and I appreciate the feedback you offered to help me refine my research skills and achieve my goals. I feel honored to have had the opportunity to work with you, and your influence will continue to inspire and guide me throughout my future endeavors.

To my advisory committee members: Dr. Laurence Court, Dr. Xin Wang, Dr. He Wang, Dr. Jack Phan, and Dr. Tinsu Pan, I want to express my sincere appreciation for your support and guidance throughout my academic journey. Your constructive feedback, insightful perspectives, and extensive expertise have been instrumental in shaping my research and professional growth. I would also like to take this opportunity to individually thank each of you for your contributions to my dissertation process. Dr. Court, thank you for your encouragement and support throughout my doctoral studies, and for always being willing to share your knowledge and expertise. Dr. Xin Wang, your ongoing support and advice on career development have been invaluable to me, and I am grateful for the time you have spent sharing your experience with me. Dr. He Wang, your expertise and insights into my research, as well as your willingness to provide feedback and guidance whenever I needed it, have been crucial in shaping my research direction. Dr. Phan, I would like to express my gratitude for your dedication to helping me better understand the clinical aspects of my research. Your willingness to share your clinical knowledge and expertise has been crucial in helping me to understand the practical applications of my work. And Dr. Pan, thank you for your valuable feedback and for always being available to discuss any questions or concerns I had during my research. Your contributions have been essential in improving the quality of my work.

To every member in Court Lab, your contributions, support, and assistance have been invaluable to me throughout my time here. I am grateful for the opportunity to work alongside such a talented and dedicated team of researchers. Your commitment to excellence and teamwork has been inspiring, and I have learned so much from each of you. I look forward to our continued friendship and professional connections.

To my program classmates, I have many memories of the days and nights we spent together at the first-year office. Your help and support were instrumental in my successful completion of all the coursework. I couldn't have done it without you, and I am grateful for the collaboration that we shared.

# Abstract

## AI-ENABLED ONLINE PLAN ADAPTATION FOR MR-GUIDED STEREOTACTIC ABLATIVE RADIOTHERAPY (SABR) OF HEAD AND NECK CANCER

Yao Zhao, M.S.
Advisory Professors: Jinzhong Yang, Ph.D.

Head and neck cancer (HNC) is a prevalent cancer type worldwide. Stereotactic ablative radiotherapy (SABR) has emerged as an effective treatment for HNC, delivering highly conformal doses to the tumor target while sparing surrounding normal tissues with a sharp dose gradient. However, the accuracy of the treatment delivery is limited by setup errors, anatomical changes, and intra-/inter-fraction organ motion. The emergence of MR-guided adaptive radiotherapy (ART) has the potential to further improve the SABR of HNC, by providing superior visualization of soft tissue and enabling real-time plan adaptation based on the daily anatomical changes of patients. This novel technology has the potential to achieve improved local control and normal tissue sparing in HNC patients. However, its clinical implementation is challenged by the available technologies of accurate delineation of gross tumor volume (GTV) for treatment planning and the real-time accurate contouring for MR-based adaptive planning. The purpose of this study is to address the challenges associated with MR-guided online ART for HNC SABR by automating the key steps in the clinical workflow using deep learning methods.

First, we developed an advanced auto-segmentation framework to automate the GTV delineation for SABR treatment planning. The framework was specifically designed to simulate the GTV delineation process performed by radiation oncologists, and multimodality images (CT, PET, and MRI) were included to improve the contouring accuracy. We found that more than 95% of the automatically generated GTVs were clinically acceptable.

Next, we automated the deformable image registration between planning and daily images, enabling accurate and rapid contour propagation for the MR-guided online ART. A novel hierarchical registration framework was proposed and validated for CT-to-MR and MR-to-MR deformable image registration. Our evaluation demonstrated superior performance than traditional registration tool implemented in current clinical practice.

The last component of this project was to automate the generation of high-quality synthetic CT from daily MR images for MR-based adaptive planning and dose calculation. To achieve this, we developed a novel deep learning model based on Cycle Generative Adversarial Network. We validated both image quality and dosimetric accuracy of the generated synthetic CT images by comparing them to their corresponding real CT images.

In conclusion, we have developed and validated the advanced deep learning methods for GTV autosegmentation, deformable image registration, and synthetic CT generation. These tools enabled automation of key steps in the MR-guided online ART workflow for HNC SABR, improving the accuracy and efficiency of treatment planning and delivery.

## Contents

# List of Illustrations

# List of Tables

# Chapter 1: Introduction

## 1.1 Background

Head and neck cancers (HNC) represent a significant global health challenge, ranking seventh among cancer types in terms of incidence. The disease causes approximately 890,000 new cases and 450,000 deaths worldwide annually (Figure 1Figure 7)[1]. In the United States, it accounts for 3% of all cancer cases and about 1.5% of all cancer-related deaths [2]. The HN region is anatomically complex and encompasses a diverse range of structures, including the oral cavity, pharynx (including the nasopharynx, oropharynx, and hypopharynx), larynx, paranasal sinuses, nasal cavity, and salivary glands, etc. This anatomical diversity leads to a marked heterogeneity of tumors arising in the HN region, with each tumor type exhibiting unique clinical and pathological characteristics [3].

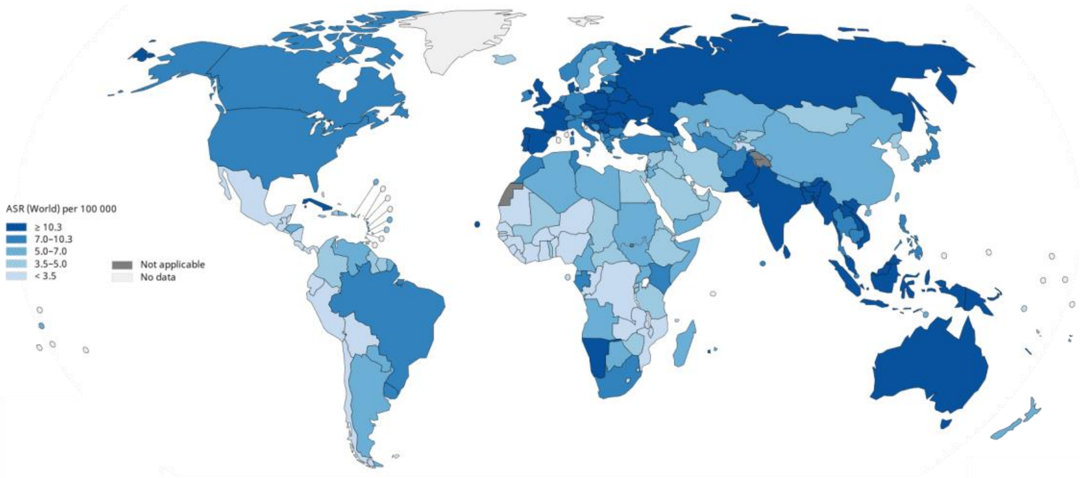


Figure 1: The estimated age-standardized rates (ASRs) of head and neck cancers incidence worldwide. The map was generated using the GLOBOCAN website ([https://gco.iarc.fr/](https://gco.iarc.fr/)) mapping tool by selecting the ‘hypopharynx’, ‘larynx’, ‘lip, oral cavity’, ‘nasopharynx’ and ‘oropharynx’ cancer sites.

HNC predominantly affect the elderly and are closely associated with the use of tobacco and alcohol consumption [4]. The global incidence of these cancers is slowly declining, in part due to the decreased use of tobacco[1]. In recent years, infection with cancer-causing types of human papillomavirus (HPV), especially HPV Type 16, has emerged as another significant risk factor for oropharyngeal cancer[3], [5]. Apart from the abovementioned factors, the use of paan (betel quid)[6], occupational exposure to wood dust, nickel dust, or formaldehyde[7], radiation exposure[8], [9], Epstein-Barr virus infection[10], Asian ancestry[7], and underlying genetic disorders[11], have also been reported as risk factors for certain types of HNCs.

The treatment of HNC typically involves a combination of surgery, chemotherapy, and radiation therapy (RT). Surgery may be a viable treatment option for patients with early stage HNC, but it is often accompanied by significant morbidity. It may result in postoperative functional deficits and physical deformities, which can have a negative impact on elderly patients[12]. For patients with advanced stage cancers or those who are not suitable for surgical treatment, RT with or without chemotherapy is the standard treatment option. However, the addition of chemotherapy increases the risk of toxicity and can be particularly challenging for patients with multiple medical comorbidities. Such patients are at higher risk of experiencing treatment intolerance, which may result in hospitalization and interruptions to the treatment course[13]. Therefore, high precision RT has become one of the most effective treatments for HNC and yields better functional outcomes when compared to other approaches[14], [15]. It has been reported that approximately 75% of all HNC patients receive RT, either for curative or palliative intent [16]. However, even with state-of-the-art techniques like intensity modulated radiation therapy (IMRT) or volumetric modulated arc therapy (VMAT), the benefits of RT must be weighed against the potential side effects from the treatment owing to the delivered dose to adjacent critical structures, including the brain stem, cranial nerves, and major blood vessels.

To enhance local tumor control while minimizing toxicity, stereotactic ablative radiotherapy (SABR) has been increasingly used to treat various HN tumors[17], [18]. It typically delivers highly conformal dose to the tumor target with improved local control in limited treatment fractions (typically 1-5 fractions in <2 weeks)[19]. Due to the extreme high dose in each fraction, accurate patient setup is vital to reduce possible high dose delivered to organs at risk that are very close to the target tumor[20]. The recently developed MR guided linear accelerator (MR-Linac) has the potential to substantially reduce treatment setup error by enabling MR-guided radiotherapy (MRgRT) for HNC[21]. The integrated MR-Linac provides an excellent platform for high-quality online adaptive radiotherapy (ART) to take advantage of superior MR-based soft tissue visualization before and during SABR treatment. Compared to conventional CT-based image-guidance, the advanced MRgRT allows more accurate patient setup, smaller target margins, and improved normal tissue sparing. Therefore, MR-guided online adaptive RT has the potential for dose escalation and better normal tissue sparing. However, its clinical usability is limited by the available technology of fast and accurate contouring for online adaptive planning and accurate dose accumulation of adaptive plans for treatment evaluation [22]. Furthermore, the challenge in accurate definition of the gross tumor volume (GTV) in treatment planning further impedes the clinical usability of MR-guided online adaptive RT for HNC SABR  [23]–[27].

SABR of HNC begins with the delineation of target volumes by radiation oncologists. However, defining the true GTV in an accurate and consistent manor remains a major challenge in HNC [28]–[30]. Studies have showed that integrated multi-modal imaging can result in more accurate and consistent delineation of HNC GTV [31]–[35]. In our routine SABR clinic, physicians delineated GTV manually based on a combination of simulation CT, Positron Emission Tomography (PET), and Magnetic Resonance Imaging (MRI). The manual delineation is time-consuming and subjective and suffers from low reproducibility and high inter-observer variability. Recent publications and our studies demonstrated a great potential of using deep learning in HNC GTV segmentation [36]–[40]. However, to the best of our knowledge, there does not exist a robust auto-segmentation tool for HNC GTV delineation, particularly for SABR having a small GTV, with a comprehensive clinical evaluation using a combination of CT, PET and MR images.

The key enabling technology for MR-guided online ART is deformable image registration (DIR) between the simulation CT scan, where the radiation treatment is initially planned, and the daily MRI scans [41]. The DIR establishes the relationship between the CT and MRI scans so that the original treatment plan can be adapted and re-optimized based on daily anatomy to achieve optimal treatment delivery. However, DIR between CT and MRI has long been a difficult task for extracranial sites due to significant contrast differences and anatomical variations over time [42]. Traditional DIR methods based on maximizing image similarity between two images is computationally intensive and inefficient to provide accurate and fast contour propagation for online ART. Additionally, another challenge for MR-guided online ART is that treatment planning relies on CT scans to provide electron density for dose calculation. However, the electron density is not available on the MR-based adaptive planning [43]. Most current clinical implementations use simple bulk electron density override to generate synthetic CT from daily MR, which might compromise the dosimetric accuracy due to the approximation of electron density on the synthetic CT [44]. In comparison, deep leaning methods developed in recent years have been demonstrated to have superior performance in image registration and synthetic CT generation compared to the traditional methods [45]. Despite their strengths, deep learning methods have yet to be explored to enable automated and accurate MR-guided online ART for HNC.

### 1.2 Stereotactic Ablative Radiotherapy (SABR)

SABR, also known as stereotactic body radiotherapy (SBRT), has been explored as a potential treatment option for HNC in various settings. SABR is a highly precise radiation therapy technique that uses multiple fields and advanced image guidance system to deliver highly conformal dose to the tumor target while sparing dose to surrounding normal tissues with a sharp dose fall-off. Unlike conventional radiation therapy, which involves spreading small daily doses over several weeks of treatment, SABR delivers high dose per fraction (typically > 5-8 Gy per fraction) with steep dose gradients in shorter treatment durations (1-5 fractions) [46]. Previous studies have shown that it has been used for the treatment of previously untreated HNC[47], or as a boost following conventional external beam RT[48]. Nevertheless, SABR has been mainly reported to be used as retreatment after locoregional recurrence in HNC[49], [50]. Additionally, it has been utilized for oligometastatic tumors in HN region[51], [52]. Although further accumulation of long-term clinical outcome data is necessary, SABR has demonstrated particular efficacy in patients with recurrent or unresectable HNC malignancies, leading to improved local control outcomes while minimizing exposure to normal tissues[49], [50], [52]–[54].

One primary objective of SABR is to reduce the unwanted dose to surrounding normal tissue by minimizing the planning target volume (PTV) margins and facilitating rapid dose fall-off beyond the target volume. Due to the high-gradient doses and precise targeting involved in SABR, a highly reproducible and accurate immobilization system is needed to effectively reduce patient movement and anatomical uncertainties in the treatment position. To ensure the delivery of the prescribed radiation dose to the clinical target volume (CTV), the PTV is designed to account for uncertainties in treatment delivery, and plays a crucial role in the treatment process [55]. In conventional IMRT, the recommended PTV margins range from 5-8 mm with daily guidance under image-guided radiation therapy (IGRT). However, due to the extremely high doses delivered in each treatment fraction, PTV margins for SABR of HNC are typically less than 3mm and require online IGRT performed before and during treatment delivery [56]. State-of-the-art SABR systems utilize a combination of immobilization devices, IGRT and advanced treatment planning and delivery software, which can decrease the PTV margins to 0-2 mm [56]. Figure 2 shows an example of an HNC patient with a skull base recurrence treated with SABR at MD Anderson Cancer Center [56]. The treatment process started with a pretreatment MRI for diagnostic purpose. Subsequently, the planning CT and the high-resolution planning MRI were obtained in the treatment position for accurate GTV declination and treatment planning. Immobilization devices, including a customized posterior cushion, mask, and bite block, were utilized to minimize the patient motion during treatment. A SABR plan with a prescription dose of 45 Gy was delivered in five fractions every other day. Additionally, a follow-up MRI was taken six months after treatment to evaluate the outcome.

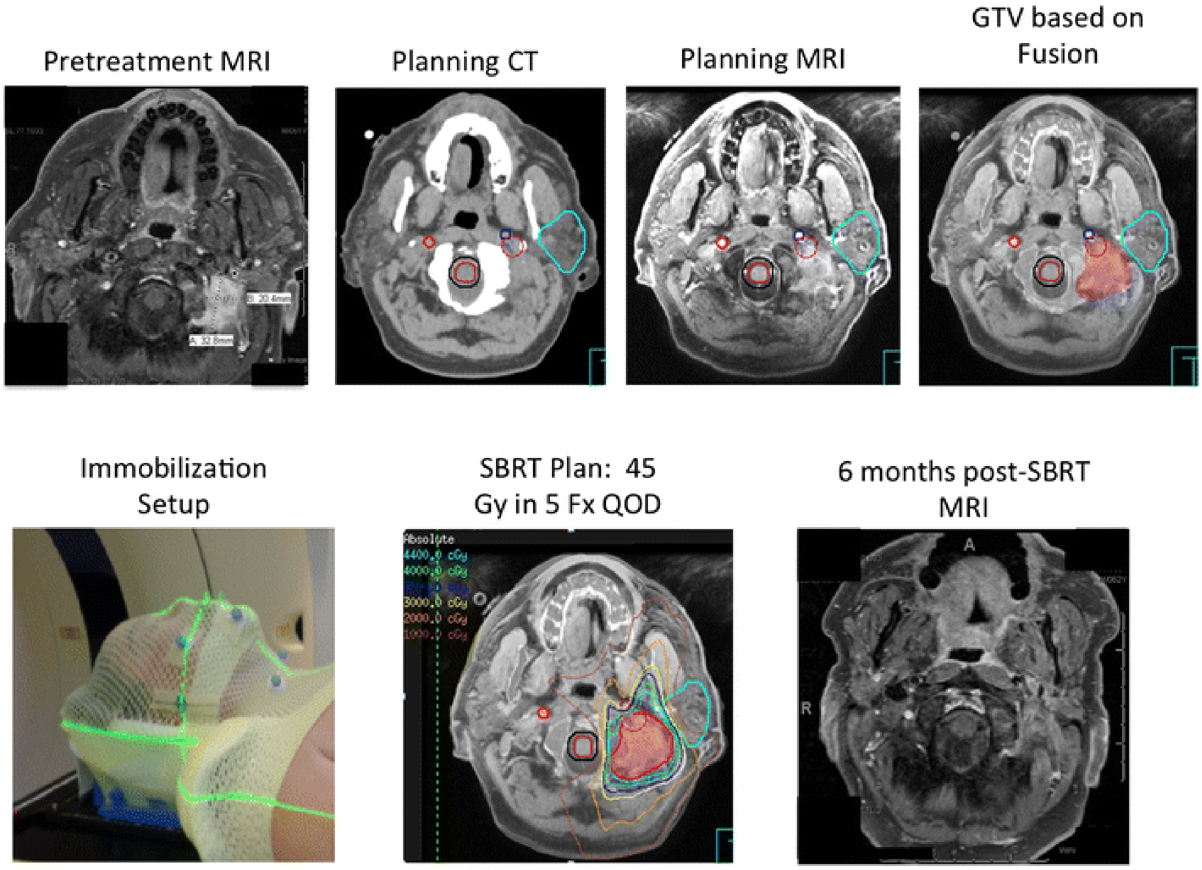


Figure 2: Example of an HNC patient treated with SABR at MD Anderson Cancer Center. A total dose of 45 Gy was delivered in five fractions. Image courtesy of Ho et al.[56].

The implementation of SABR offers several advantages that make it an attractive option for treating HNCs. One advantage of SABR is its ability to reduce the number of treatment fractions and shorten the overall treatment course. In comparison to conventional RT, which typically requires 6-7 weeks of treatment, SABR can be completed in less than two weeks. This shorter treatment course not only enhances patient convenience but also can result in financial benefits for both patients and healthcare systems. Additionally, the reduced treatment course of SABR may allow for a shorter interval for systemic therapy in patients at high risk for regional or metastatic disease, potentially improving overall treatment outcomes [56]. Although the radiobiology of SABR is not yet fully understood, it has been shown to provide a biological advantage by delivering large doses in a few fractions [57]. The hypofractionation with high dose-per-fraction in SABR has the potential to leverage additional mechanisms of cell kill, which could be especially beneficial in the recurrent HNCs where treatment-resistant clonogens are enriched. Furthermore, the shorter treatment duration of SABR may also address the issue of accelerated repopulation and lead to better outcomes in HNCs [58]. Despite the potential benefits of SABR, there are concerns that the higher dose-per-fraction may impede sublethal damage repair processes in health tissue, resulting in an increased risk of late toxicity [59]. The emergence of hybrid MR-Linac provides a new platform to further minimize this risk [60]. MR-Linac can acquire high-quality images of the tumor and surrounding soft tissue before and during treatment. It also provides an online plan adaption for accurate and precise targeting of the tumor while minimizing radiation exposure to healthy tissue, thereby reducing the risk of late toxicity. Overall, SABR offers a promising option for the treatment of HNCs, and the use of advanced image guidance systems can further enhance its effectiveness and safety.

### 1.3 Auto-segmentation in HNC Radiotherapy

RT treatment planning for HNCs begins with the delineation of tumor volumes and surrounding organs at risk (ORAs) based on the simulation planning CT. The accuracy of target and OARs delineation is crucial for effective and safe RT for HNCs. However, this process is resource-intensive, time-consuming, and heavily relies on expert knowledge, which may lead to inter- and intra-observer variability [61]. Automated and accurate segmentation of target and OARs is important to streamline this process and reduce human errors, resulting in more consistent and precise RT planning for HNCs.

Accurate and automatic segmentation of HNC CT images poses significant challenges due to the complexity and variability of anatomical structures involved. Additionally, some structures have small volumes, and the poor contrast of soft tissues in CT images further complicates the segmentation process [62]. In recent years, deep learning (DL) has been widely investigated and demonstrated as effective segmentation approaches for HN normal tissue delineation [61], [63]–[65]. Studies have shown that DL models can achieve segmentation accuracy comparable to manual segmentation by radiation oncologists [61], [66]–[68]. These approaches have the potential to reduce the time and resource required in the segmentation process and improve the consistency and reproducibility of segmentation results.

Recent developments in the automation of workflow, normal tissue auto-contouring, and automatic dose optimization have made fully automating radiation treatment planning more achievable [69]–[72]. The last obstacle to achieving a full automation of the entire planning process is the auto-contouring of GTV, which remains a major challenge for HNC. However, the availability of multiple imaging modalities, such as PET and MRI, has provided clinicians with additional information to improve GTV delineation. As shown in Figure 3, these modalities offer complementary information to CT, such as functional and metabolic activity of tissues, as well as superior soft tissue contrast, respectively [73]. By combining the information from these different modalities, clinicians can achieve a completer and more accurate picture of the tumor volume. Studies have demonstrated that the use of PET and MRI in combination with CT can substantially improve the accuracy and consistency of GTV delineation [31], [32], [35]. Furthermore, several studies have investigated DL-based autosegmentation methods that utilized multimodality images for HN GTV delineation [38], [39], [74], [75]. These DL-based methods have shown promising results in improving the accuracy and reproducibility of GTV, especially when compared to the auto-segmentation based solely on CT images. Therefore, the integration of multimodal imaging and DL-based segmentation algorithms have the potential to enable more effective and efficient GTV delineation in HNC treatment planning.

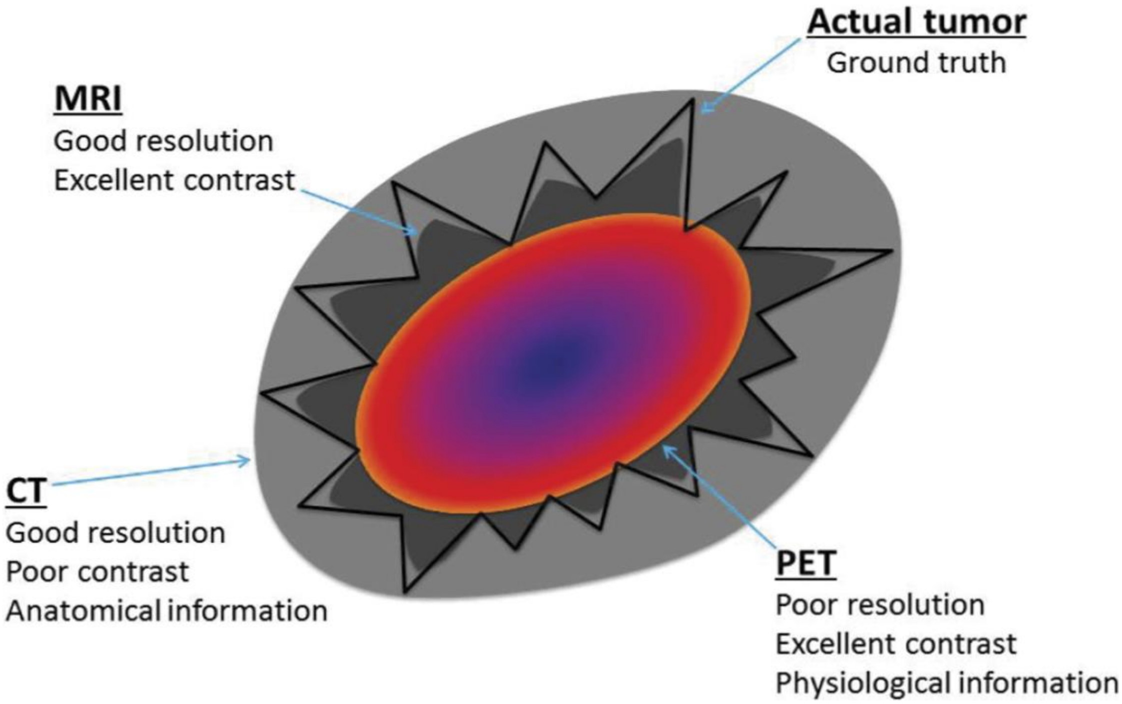


Figure 3: A diagram showing that different imaging modalities provide different types of information of information about the tumor volume and surrounding structures. Image courtesy of Yang et al. [73]

Although DL-based autosegmentation is promising, one of the main challenges in the development of these DL models is the need for a large amount of well-curated training data. The availability of such data is often limited in many clinical applications, and even when available, the large variation in image quality due to differences in image acquisition protocols, machines, and institutions can lead to significant challenges [61]. Accumulation of consistent and high-quality data remains a significant bottleneck for improving the segmentation performance of deep learning models. Although data augmentation techniques such as scaling, rotation, or random deformation have been widely used to increase the amount of available training data, the effectiveness of these methods can be limited when the available data is severely limited. The reason is that these techniques cannot accurately represent actual patient population variations, which can lead to poor model performance in real clinical scenarios [76]. Consequently, effective data augmentation approaches are needed to address these challenges and enable the development of more robust DL models for auto-segmentation.

### 1.4 MR-guided Adaptive Radiation Therapy

ART has been validated as a promising technique for improving the quality of treatment delivery, by accounting for anatomical variations such as changes in tumor volume, normal tissue deformation, weight loss, and variations in patient position from day to day [77]–[79]. ART involves modifications to the original treatment plan based on repeated daily imaging of the patient during the course of treatment. By adapting the original treatment plan, ART aims to increase the effectiveness of radiotherapy and reduce toxicity to surrounding healthy tissues. The current standard approach for implementing ART relies on in-room imaging techniques, including electronic portal imaging, kilovoltage radiographs, cone-beam computed tomography (CBCT), and CT-on-rails [80], [81]. However, these imaging modalities have limited soft tissue contrast which can result in reduced accuracy in localizing HNC tumors during IGRT. To address this limitation, the integration of MRI into ART workflows has been proposed as a solution to enhance the visualization of HN targets and OARs throughout the course of radiotherapy [82], [83]. Compared to CT, MRI offers superior soft tissue contrast for better visualization of anatomical changes. With the development of hybrid magnetic resonance imaging-linear accelerator (MR-linac) machines, daily on-board MR-guided adaptive radiotherapy for HNC has become clinically feasible [84], [85]. A study conducted on patients with nasopharyngeal cancer found that those treated with MRgRT had a higher 2-year disease-free survival rate of 93.6% compared to those treated with IMRT, which had a rate of 87.5% [86]

### 1.4.1 MR-Linac

To date, four MR-linac systems have been developed by different companies and institutions [87]–[93]. Table 1 provides a summary of the distinct features of each MR-linac machine, including differences in magnetic field strength, beam type and energy, and the orientation of the radiation beam and magnetic field (perpendicular or inline). The first two MR-linacs, the MRIdian and Unity, have received both CE mark and FDA 510k(k) clearance and have been implemented for RT clinic in many institutions [94].

At MD Anderson, the Elekta Unity MR-linac was installed in 2014 and used to treat our first patient in 2019 [95], [96]. This state-of-the-art device combines a modified 1.5 T Philips wide-bore MRI system with a 7MV flattening filter-free Elekta Linac. Compared to other MR-linac systems, Elekta Unity features a high-field magnet (1.5 T) that can provide superior image quality, allowing for precise targeting of tumors with high doses of radiation [94].

Table 1. A summary of different MR-linac systems. [94], [97]

| System | Filed strength | Beam configuration | Bore diameter | First patient treatment |
|---|---|---|---|---|
| ViewRay MRIdian [87] | 0.35 T | $^{60}$Co or 6 MV, Perpendicular | 70 cm | February 2014 ($^{60}$Co), July 2017 (Linac) |
| Elekta Unity [88], [89] | 1.5 T | 7 MV, Perpendicular | 70 cm | May 2017 |
| MagnetTx Aurora [90], [91] | 0.5 T | 6 MV, In-line | 110×60 cm | February 2023 |
| Australian MRI-Linac [92], [93] | 1.0 T | 4 and 5 MV, In-line with perpendicular option | 62 cm | Has not yet occurred |

### 1.4.2 On-line Adaptive Workflow

The clinical introduction of MR-Linac has facilitated the implementation of daily MR-guided ART for HNC treatment. Similar to conventional RT, MRgRT begins with the acquisition of both simulation CT and simulation MR images. These images are then fused together for delineation of the primary and nodal GTVs, followed by the generation of clinical target volume (CTV) and planning target volume (PTV) using margin expansion by following physician’s planning order. In addition, OARs are contoured for dose sparing during treatment plan optimization. A reference  plan is created on the planning CT as the base for daily plan adaptation [80], [95].

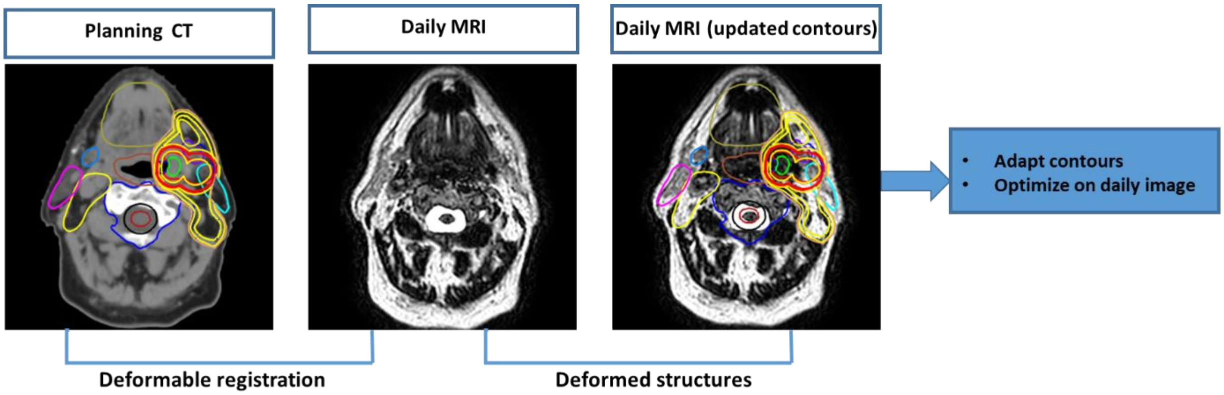


Figure 4: The workflow of MR-guided adaptive radiotherapy. The contours are propagated from the planning CT to the daily MRI by deformable image registration. The adaptive plan is then optimized based on the daily MRI with the updated contours.

On the day of treatment, patient is positioned on the MR-linac treatment table and imaged with the daily MR imaging protocol to verify the patient position. Following the acquisition of the daily MRI, the initial treatment plan is re-optimized based on the anatomy captured on the daily MR image. Specifically, the daily MR image is first rigidly registered to the planning CT to obtain the treatment isocenter shift, and then contours on the planning CT were deformed to match the anatomy on the daily MRI using DIR, as demonstrated in Figure 4 [60]. This allows for accurate adaptation of the treatment plan to account for daily changes in patient anatomy. The adaptive treatment plan is reoptimized to achieve clinical goals, thus ensuring optimal treatment delivery with minimal toxicity to surrounding healthy tissues. One of the key steps to enable the MR-guided ART is the accurate DIR between the planning CT and daily MRI. However, the current DIR tool used in clinic is not reliable for MR-guided ART. According to a recent study, the median Dice similarity coefficient (DSC) value was found to be below the recommended threshold of 0.8 by AAPM Task Group 132 [98], when utilizing the Monaco treatment planning system to perform CT-to-MR DIR for MRgRT [95]. Extensive manual adjustments are necessary to ensure accurate deformation of the contours to account for daily changes in patient anatomy. However, this process is labor-intensive and subject to user bias and interpretation, which can introduce additional dosimetric uncertainties [99]. Thus, further advancements in DIR technology are necessary to enhance the reliability of MR-guided ART in clinical practice.

### 1.4.3 MR-based Planning

As discussed in Section 1.4.2, CT image acquisitions are necessary for dose calculations in MRgRT treatment planning due to the lack of electron density information on MR images. However, since the adaptive plan is based on daily MR images, a synthetic CT generated from the daily MR image is needed for plan optimization and dose calculation. In Monaco treatment planning system, this process involves propagating the contours from the planning CT to the daily MR images, followed by bulk density assignment, where the average electron density value for each structure is measured on planning CT and assigned to corresponding voxels in the structure on the daily MR image [100].

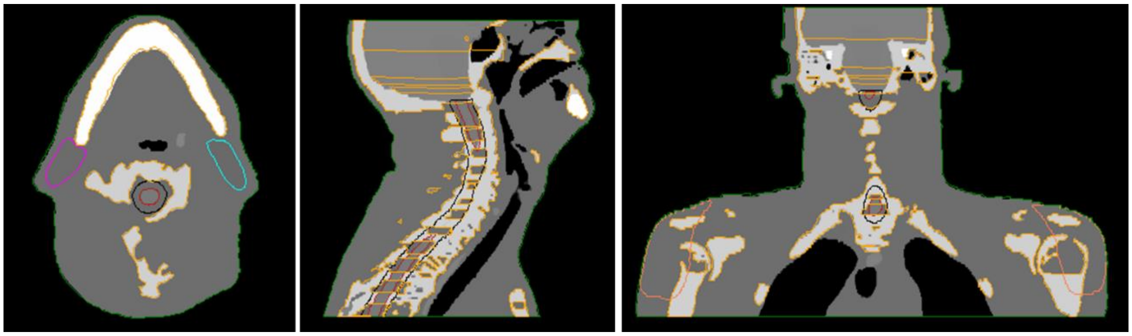


Figure 5: An illustration of the synthetic CT generated from daily MRI by bulk density assignment method in Monaco treatment system. The average electron density is assigned to each structure based on the deformed contours.

However, this method has several limitations that can affect the accuracy of dose calculations [101], [102]. One of the limitations is that it assumes uniform electron density within each structure, which may not reflect the actual density variations within the tissue.

Additionally, this method relies on the accuracy of the deformation of the contours from the planning CT to the daily MR images, which can be influenced by numerous factors such as motion artifacts, patient setup errors, and inter-observer variability. Figure 5 demonstrates an example of the synthetic CT generated from daily MRI using the bulk density assignment method in Monaco. It shows the uniform density within each structure.

In summary, generating a synthetic CT from daily MRI images is essential for MRgRT treatment planning. Despite the limitations of the bulk density assignment method, it remains widely used in clinical practice due to its simplicity and easy implementation. Other methods for generating synthetic CT images from daily MRI images have been investigated to streamline the MRgRT workflow and reduce uncertainties caused by image registration and uniform density overrides [103]–[107].

### 1.5 Summary

SABR is a highly precise and effective treatment option for HNC, which has been demonstrated with ability to improve local tumor control. However, due to the high doses and steep dose gradients used in SABR, accurate targeting of the tumor and sparing the healthy tissue are critical for minimizing side effects. MR-guided adaptive RT has emerged as a promising tool to improve the accuracy and precision of SABR treatment delivery [108]–[111]. By using the daily MRI for adaptive planning, MRgRT can deliver highly personalized and precise treatment to patients. The use of MR-guided RT for SABR has several advantages over conventional CBCT-based IGRT, including the ability to visualize soft tissues and tumors more clearly, improved contouring accuracy, and reduced inter- and intra-fractional uncertainties. This approach allows for monitoring of treatment response and adaptation of the treatment plan throughout the course of treatment, enabling clinicians to make informed decisions about changes to the plan based on the patient's individual response to treatment. Additionally, the use of adaptive planning helps to reduce the risk of under-dosing or over-dosing, which can have detrimental effects on the patient's outcome.

Overall, MR-guided ART has the potential to significantly improve the accuracy and precision of SABR treatment delivery, while minimizing the risk of radiation-related toxicity. The objective of this project is to fulfill several workflow challenges and gaps in knowledge to improve MR-guided ART for SABR of HNC. We expect to using deep leaning methods to achieve automated GTV delineation, improved online contours propagation and accurate synthetic CT generation, which can be used collectively or separately to assist the MR-guided RT in different steps. Ultimately, this project is expected to improve both treatment outcomes resulted from potential dose escalation and reduced toxicity due to improved normal tissue sparing by inter- and intrafraction adaptation for HNC patients. Our contribution will be important to bridge the gap between current clinical practice and the future of workflow automation in MR-guided online ART for SABR.

# Chapter 2: Central Hypothesis, Specific Aims, and Dissertation Organization

The *long-term goal* of this project is to improve the local tumor control and quality of life after HNC RT by using advanced artificial intelligence (AI) technologies to enable MR-guided online ART for SABR. The *goal of this particular project* is to develop DL-based auto-segmentation, deformable registration, and synthetic CT generation algorithms, to enable automated accurate GTV delineation for treatment planning, accurate contour propagation for online adaptive planning, and accurate dose calculation for MR-based adaptive planning.

## 2.1 Central Hypothesis

The central hypothesis is that the AI tools can streamline MR-guided online ART workflow for HNC by outperforming existing workflow in the following aspects: GTV auto-delineation for treatment planning with clinically-acceptable ratio over 95%, contour propagation for online adaptive planning with improved accuracy over current clinical tools in 95% of all cases, and synthetic CT generation for MR-based planning with the difference of dosimetric parameters less than 1% within clinical target volume (CTV). The clinical-acceptable contour means that the contour does not need any editing or need only minor editing for clinical use.

## 2.2 Specific Aims

**Aim 1.** Automate GTV delineation from fusion of multi-modal images for treatment planning

Hypothesis: DL models can learn the HNC GTV delineation performed by physicians using a combination of multi-modal images (CT/PET/MRI) and fully automate the entire delineation process with the outcome having a clinical acceptance ratio >95% evaluated by HN radiation oncologists.

Study 1-1: Develop a data augmentation method for training auto-segmentation models with severely limited contoured data

Study 1-2: Develop an auto-segmentation framework for HN GTV delineation based on a combination of CT, PET, and MR images.

**Aim 2.** Automate deformable registration of simulation CT and daily MRI for adaptive planning.

Hypothesis: DL models can automate the DIR between CT and MRI and achieve significantly improved accuracy than current clinical tools in 95% of cases.

**Aim 3.** Automate synthetic CT generation from MR images for MR-based adaptive planning.

Hypothesis: DL models can generate synthetic CT from MRI with high image quality to ensure the difference of dosimetric parameters: mean dose ($D_{mean}$), $D_{95\%}$, and $D_{5\%}$, of CTV less than 1%.

### 2.3 Dissertation Organization

The main body of this dissertation is Chapters 3-6. *Chapter 3: Data Augmentation for Auto-segmentation* and *Chapter 4: Gross Tumor Volume Auto-segmentation Based on Multimodality Images* address Specific Aims 1-1 and 1-2, respectively. Chapter 5 addresses Specific Aim 2 and is entitled "*Automated Deformable Image Registration for MR-guided Adaptive Radiotherapy*". *Chapter 6: Automated Synthetic CT Generation from MRI for MR-based Treatment Planning* addresses Specific Aim 3. After the main body of work, a discussion and conclusion follow in Chapter 7: Discussion, and References cited in the dissertation are included at the end.

# Chapter 3: Data Augmentation for Auto-segmentation

This chapter is based upon the following article:

Zhao, Yao, Dong Joo Rhee, Carlos Cardenas, Laurence E. Court, and Jinzhong Yang. "Training deep-learning segmentation models from severely limited data." *Medical physics* 48, no. 4 (2021): 1697-1706. doi:10.1002/mp.14728

## 3.1 Abstract

**Purpose:** To enable generation of high-quality deep learning segmentation models from severely limited contoured cases (e.g. ~10 cases).

**Methods and Materials:** Thirty head-and-neck CT scans with well-defined contours were deformably registered to 200 CT scans of the same anatomic site without contours. Acquired deformation vector fields were used to train a principal component analysis (PCA) model for each of the 30 contoured CT scans by capturing the mean deformation and most prominent variations. Each PCA model can produce an infinite number of synthetic CT scans and corresponding contours by applying random deformations. We used 300, 600, 1000, and 2000 synthetic CT scans and contours generated from one PCA model to train V-Net, a 3D convolutional neural network architecture, to segment parotid and submandibular glands. We repeated the training using same numbers of training cases generated from 7, 10, 20, and 30 PCA models, with the data distributed evenly between each PCA model. Performance of the segmentation models was evaluated with Dice similarity coefficients between auto-generated contours and physician-drawn contours on 162 test CT scans for parotid glands and another 21 test CT scans for submandibular glands.

**Results:** Dice values varied with the number of synthetic CT scans and the number of PCA models used to train the network. By using 2000 synthetic CT scans generated from 10 PCA models, we achieved Dice values of 82.8%±6.8% for right parotid, 82.0%±6.9% for left parotid, and 74.2%±6.8% for submandibular glands. These results are comparable with those obtained from state-of-the-art auto-contouring approaches, including a deep-learning network trained from more than 1000 contoured patients and a multi-atlas algorithm from 12 well-contoured atlases. Improvement was marginal when >10 PCA models or >2000 synthetic CT scans were used.

**Conclusions:** We demonstrated an effective data augmentation approach to train high-quality deep learning segmentation models from a limited number of well-contoured patient cases.

### 3.2 Introduction

Precise image segmentation is crucial for accurate disease diagnosis and efficient radiation treatment planning.[112] Deep learning has emerged recently as an effective tool for medical image segmentation with reduced human effort and variability.[113]–[116] Modern deep learning segmentation models, such as convolutional neural networks (CNNs), have millions of trainable parameters and thus typically require an extensive set of data for training. The amount of well-curated data available for training is critically important for constructing a robust model for auto-segmentation.[117] However, obtaining manual contours for training deep learning models is a time-consuming task and requires considerable expertise.[118], [119] In many clinical applications, the availability of well-curated training data is very limited. The sparsity of data is worsened by variation in image quality from differences in image acquisition procedures used across protocols, machines, and institutions. The accumulation of consistent and high-quality data remains a bottleneck to improving the segmentation performance of deep learning models.

Varied data augmentation approaches have been introduced to overcome this challenge by artificially increasing the amount of training data.[120] Traditional data augmentation functions such as flipping or random rotation,[121], [122] random nonlinear deformation,[123] or image cropping are used to synthesize new samples from original training data.[124], [125] These functions have proven to be effective in improving the performance of trained deep learning model. However, the augmented data are unable to represent actual patient population variations,[126] and the data variation relies strongly on the manual choice of the applied transformations.[127] To eliminate the bias resulting from the manual choice, Hauberg et al.[128] developed an approach to learn digit-specific spatial transformations from learning data to create new samples. Ratner et al.[129] proposed to combine domain-specific transformations by using a generative sequence model. Cubuk et al.[130] constructed a search algorithm that could find the optimal augmentation policies to improve the classification accuracy. However, these transformations are simple and insufficient to capture many of the subtle variations in medical images. In most recent years, many deep learning methods have been developed for data augmentation. Zhao et al.[131] used a deep learning model to learn appearance and spatial transformations to create new labeled magnetic resonance imaging (MRI) scans. Further, the image-to-image generative adversarial networks (GANs) technique has been widely used for synthesizing images for many medical imaging applications.[132]–[135] For example, Frid-Adar et al.[136] applied the GAN framework as data augmentation for liver lesion classification. However, the training of GANs typically requires a substantial amount of data,[120], [136], [137] making it impractical when the availability of initial well-curated data is severely limited.

In this study, we proposed a simple but effective data augmentation approach by synthesizing labeled images with principal component analysis (PCA).[138].This approach was developed based on creating synthetic images with known deformation.[139] Instead of applying arbitrary deformations, we used a PCA-driven machine learning approach to capture actual anatomy variations in unlabeled patient populations.[140], [141] The learned anatomy changes were modeled to generate random deformations that can sample realistic variations in anatomy. Then, new synthetic images with the appearance of real patient anatomy were generated via applying the PCA-informed random deformations. These synthetic CT scans were used to train a CNN for auto-segmentation. We assessed the effectiveness of our data augmentation method for auto-segmenting parotid glands and submandibular glands on CT scans.

## 3.3 Methods and Materials

### 3.3.1 Patient Data

CT scans acquired during radiotherapy simulation and the corresponding clinical contours of 30 head and neck patients who received external photon beam radiation treatment at The

University of Texas MD Anderson Cancer Center were used as templates to synthesize new training data. These contours were reviewed and approved by radiation oncologists sub-specializing in head and neck cancer. Another 200 CT scans without contours were used to sample anatomic differences in other patients with head-and-neck cancer. All 230 CT scans had the same voxel size of $1\ mm \times 1\ mm \times 2.5\ mm$. In this proof-of-principle study, we investigated the segmentation of parotid glands (two structures, left and right parotids) and submandibular glands (as a single structure). These structures were chosen because they are important organs at risk that must be spared during radiation treatment, but manually contouring them is not straightforward.

### 3.3.2 Learning anatomy variations

To perform data augmentation, we evaluated the application of a known deformation, representing realistic anatomy variations, to a well-contoured CT scan. This deformation was learned by capturing the most prominent variations between the CT scans with contours and those without contours through PCA.[138] This learning process was developed from the active shape model.[142], [143] We first chose the 30 well-contoured CT scans as reference images, and then registered each of them to the other 200 CT scans without contours, using a dual-force demons deformable image registration algorithm.[144] Before registration, we performed histogram equalization to match the contrast of the two images, using a number of 256 bins of the histogram. The histogram equalization was performed locally by separating the images into small blocks with a size of 20 x 20 voxels. A 6-level multiresolution scheme was used to accelerate the registration and improve the robustness. In demons registration, we used an upper bound of step size as 1.25, and the Gaussian variance of 1.5 for regularization of deformation vector fields. Refer to Wang *et al*.[144] for details about this deformable registration algorithm.

The resulting deformation vector field (DVF) maps characterize the distribution of anatomic differences between the reference image and other images. Mathematically, the DVF maps can be represented by:

$$d(\vec{\omega}, i) = \left\{ d_x(\vec{\omega}, i), d_y(\vec{\omega}, i), d_z(\vec{\omega}, i) \right\} \tag{1}$$

where $d_x(.,i)$, $d_y(.,i)$, and $d_z(.,i)$ are the displacement field matrices from image *i, i = 1, 2, …, N*, to the reference image along the left-right, anterior-posterior, and superior-inferior directions, respectively, and $\vec{\omega}$ stands for the voxel position.

To create the active shape model, we vectorized the DVF for image *i* to be $\vec{d}(i)$. For each reference image, we let *G* = {$\vec{d}(1)$, $\vec{d}(2), \dots, \vec{d}(N)$} be the matrix consisting of all the DVFs resulting from registration. We defined the sample covariance matrix of *G* as $E$:

$$E = \frac{1}{N-1} \sum_{i=1}^{N} ( \vec{d}(i) - \bar{d})( \vec{d}(i) - \bar{d} )^T \tag{2}$$

where $\bar{d}$ denotes the mean deformation over N images, and *T* represents the transpose.

We calculated the eigenvalues and eigenvectors of the covariance matrix $E$ to represent anatomy variations in the training dataset. Let $\lambda$ and $\vec{\varphi}$ be the set of eigenvalues and eigenvectors of the covariance matrix $E$, respectively. They can be calculated by:

$$E\, \vec{\varphi}^{(j)} = \lambda^{(j)}\, \vec{\varphi}^{(j)}, j = 1, 2, \dots, N \tag{3}$$

According to the theory of PCA,[138] a combination of a small number of eigenvectors corresponding to the large eigenvalues can capture the most prominent variations of the deformation. These eigenvectors consist a subset of $\vec{\varphi}$ and are known as principal modes. Assuming $\lambda^{(1)} \geq \lambda^{(2)} \geq \cdots \geq \lambda^{(N)}$, the first $M^{th}$ eigenvectors are able to represent at least $\alpha$ percent of the total variations that deviate from the mean deformation. The first $M^{th}$ eigenvectors should satisfy:

$$\sum_{j=1}^{M} \lambda^{(j)} \geq \frac{\alpha}{100} \sum_{j=1}^{N} \lambda^{(j)} \tag{4}$$

where $\alpha$ stands for a value between 0 and 100. The principal modes, $\vec{\varphi}^{(1)}, \vec{\varphi}^{(2)}, \dots, \vec{\varphi}^{(M)}$, and the mean deformation, $\bar{d}$, constituted the active shape model. For this study, we set the parameter $\alpha$ to 90, which means the active shape model represented at least 90% of the total variations in that patient population.

### 3.3.3 Synthesizing new samples

The acquired active shape model, named the PCA model, enables us to sample the most variations and generate a new random deformation, denoted by *d*, by using the equation:

$$d = \bar{d} + \sum_{j=1}^{M} \beta_j \, \vec{\varphi}^{(j)} \qquad (5)$$

where $\beta_j$ is a random coefficient that quantifies the new deformation contributed by the $j^{th}$ eigenvector. To ensure that the new deformation is realistic, we set a maximum value, $D_{max}$, for the randomly generated parameter ($\beta_j$) based on the following equation:

$$\sum_{j=1}^{M} \frac{\beta_j^2}{\lambda^{(j)}} \leq D_{max}^2 \qquad (6)$$

In our implementation, $\beta_j$was generated using a uniform distribution in the range of (0, $\frac{D_{max}^2}{M\lambda^{(j)}}$). The $D_{max}$ was carefully chosen by manually reviewing the generated deformations with certain $D_{max}$ value. In this study, $D_{max}$ was set to 6 to ensure that the new deformations are reasonable and include a considerably large proportion of possible deformations.

Let the pair $(x, l_x)$ represent the reference CT scan and its corresponding contours, respectively, and let $\tau_k$ be the newly generated transformation representing the process of applying the generated random deformations (*d*). The synthetic CT scans were created by applying the transformations to the original well-contoured CT scans:

$$\hat{x}_k = \tau_k(x) \qquad (7)$$

$$\hat{l}_k = \tau_k(l_x) \tag{8}$$

where $(\hat{x}_k, \hat{l}_k)$ represent the synthetic CT scan and corresponding contours, generated by transformation $\tau_k$. Theoretically, an infinite number of synthetic CT scans can be created as the deformations can be generated randomly.

In brief, we first captured anatomic variations in the patient population by training an active shape model (PCA model). The synthetic CT scans and the corresponding contours were then created by applying random deformations to the reference CT scans. When creating the synthetic CT scans, DVFs in both directions are needed: the DVF from contoured CT scan to the random CT scan to generate contours on synthetic CT; and the DVF from the random CT scan to the contoured CT scan for backward image resampling to create random CT scan. The former DVF was the random deformation $(d)$; and the latter DVF was created by inversing the former DVF using an iterative approach, similar to the one described by Yan et al.[145] This process is summarized in Figure 6.

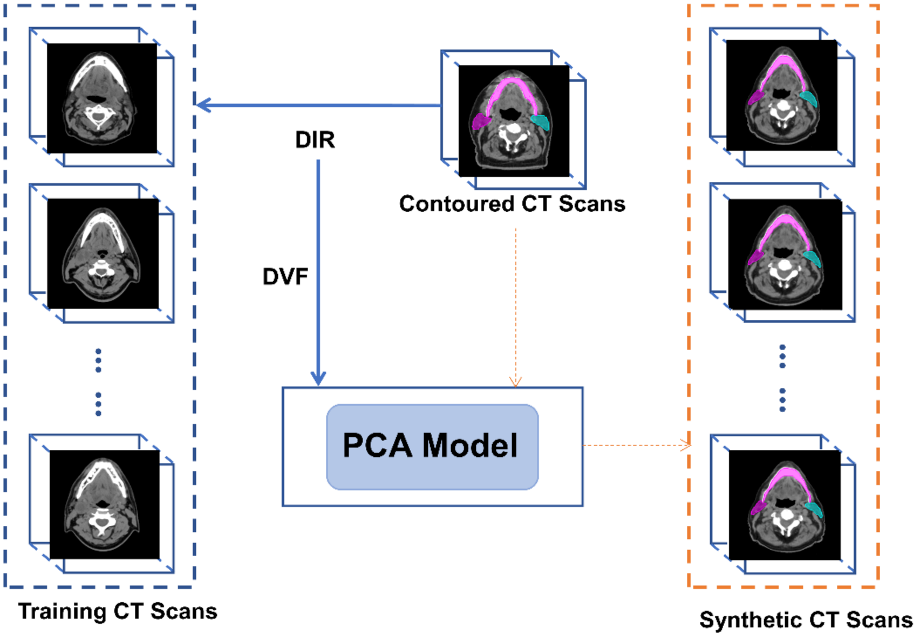


Figure 6. General workflow of the principal component analysis (PCA) approach to generate synthetic CT scans. Deformation vector fields (DVF) from deformable image registration (DIR) between a well-contoured image and other CT scans are used to create the PCA model. Then, the PCA model can simulate random deformations applied to the contoured CT scan, to create infinite number of synthetic CT scans with contours

### 3.3.4 Training data

The aforementioned process was used to create 30 PCA models by using the 30 well-contoured CT scans as the PCA model template. Each PCA model can generate a random synthetic patient within a space of patient anatomy variations represented by the 200 CT scans without contours, and each individual model can produce an infinite number of synthetic CT scans and corresponding contours.

We used 300, 600, 1000, and 2000 synthetic CT scans and contours generated from a single PCA model to train a CNN-based model described in the next section for auto-segmentation of parotid and submandibular glands. Of the synthetic CT scans, 80% were used for training and 20% were used for cross validation. We repeated this process by using the same numbers of synthetic CT scans generated from 7, 10, 20, and 30 PCA models, with the scans distributed evenly between each PCA model. Scans generated from rough 80% of the total number of PCA models (e.g. 6 out of 7 PCA models) were used for training, and the scans generated from the rest of the PCA models were used for cross validation.

In addition, the CNN-based model was trained on subsets of 200, 300, 600, 1000, 1500, 2000, 5000, and 10,000 synthetic CT scans and contours generated from 10 PCA models to investigate the impact of training data quantity on the segmentation performance.

### 3.3.5 Segmentation model

In this study, we used V-Net,[146] a three-dimensional CNN-based model, for autosegmentation. We modified the original V-Net architecture by adding batch renormalization layers to the end of each convolutional and up-sampling layer. During the network training, the parotid glands and submandibular glands were trained separately, and the patch size of the data for both training sets was set to $512 \times 512 \times 50$. A previous study showed that the dimension of the parotid glands is typically $46.3 \pm 7.7$ mm along the longitudinal axis,[147] and thus the patch size is sufficient to include the parotid glands in the longitudinal direction. The submandibular glands are smaller along the longitudinal direction ($14.3 \pm 5.7\ mm$).

In addition, the CT numbers were capped between –1000 HU to 3000 HU during training. We used the Dice similarity coefficient[148] as the loss function in our implementation, and the Adam optimizer[149] was chosen as the optimization algorithm with a learning rate of $5 \times 10^{-4}$.

### 3.3.5 Evaluation

We evaluated the performance of the trained segmentation models for parotid glands by calculating the Dice coefficients between the auto-generated contours and the physician-drawn contours on 162 independent test CT scans from patients with head-and-neck cancer who received proton radiation treatment at MD Anderson Cancer Center. Among the 162 test scans, 156 contained the left parotid contours and 147 contained the right parotid contours. This test dataset is a subset of the test dataset used in Rhee et al.[150] Other 21 CT scans with manual contours of submandibular glands were used as the test dataset for segmentation of submandibular glands.

For the comparison, we also used a paired non-parametric Wilcoxon signed-rank test at a significance level of 0.05 (defined by a $p < 0.05$) to evaluate whether a statistically significant difference exists between different models.

## 3.4 Results

### 3.4.1 Synthesized CT scans

Our proposed data augmentation method enables the generation of synthetic CT scans to simulate realistic patients. Figure 7 shows some examples of the synthetic CT scans and the corresponding contours, depicting variations in the synthetic CT scans that emulate the real patient anatomy.

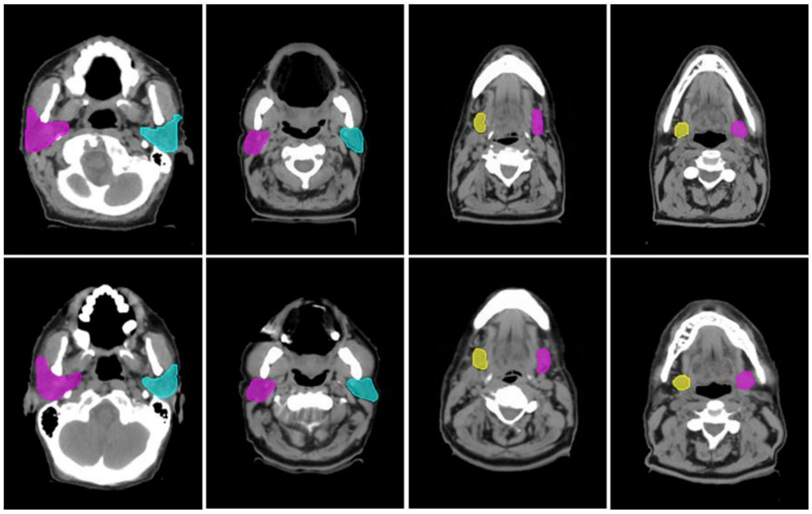


Figure 7. Comparisons between original and synthetic CT scans with contours. The top row are the original CT scans, and the bottom row are the corresponding synthetic CT scans.

### 3.4.2 Segmentation accuracy

We investigated the extent of change in segmentation accuracy with the number of PCA models used to create the synthetic CT scans for training, and the number of synthetic CT scans used to train the segmentation model. In Figure 8, the segmentation performance was plotted varying with the number of synthetic CT scans used for training the results, and the number of PCA models used to create the synthetic CT scans. For left and right parotids, the Dice values were calculated based on the average results from 162 testing patients; for the submandibular glands, the results were calculated from the average of 21 testing patients. In general, use of more PCA models and more synthetic CT scans for the model training led to more accurate segmentation. In our particular evaluation, we achieved the best segmentation accuracy as Dice coefficients of 83.1%±6.1% (left parotid), 82.9%±6.7% (right parotid), and 76.3%±7.7% (submandibular glands) when 30 PCA models and 2000 synthetic CT scans generated from these PCA models were used to train the V-net model.

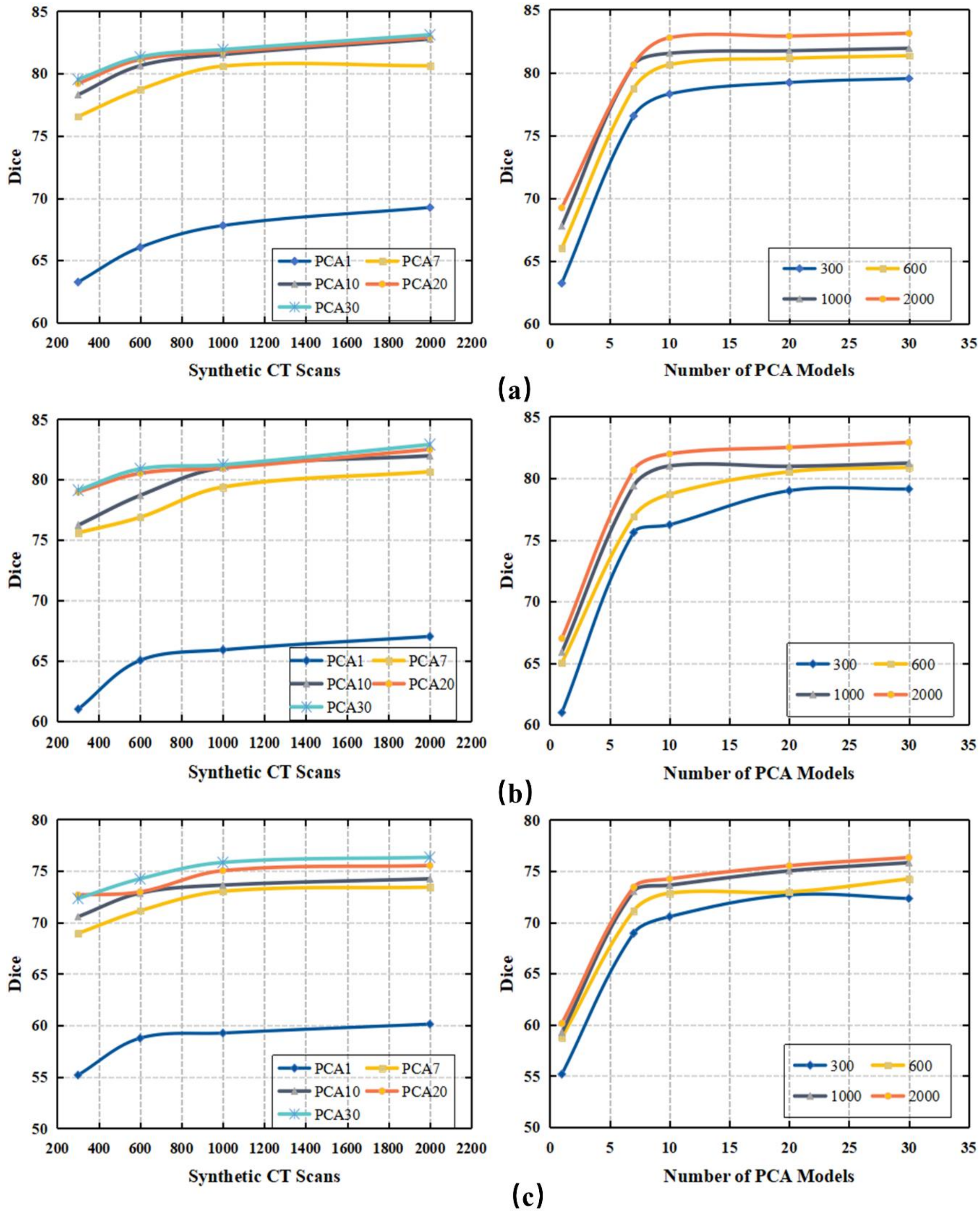


Figure 8. Dice values vary with the number of synthetic CT scans (shown in the left column) and the number of PCA models used to generate the synthetic CT scans (shown in the right column) for training the V-net. Different numbers of PCA models were used to generate the

synthetic CT scans (e.g., PCA1 represents one PCA model). Numbers of synthetic CT scans were evenly distributed between the PCA models. (a) Left parotid; (b) right parotid; (c) submandibular gland.

### 3.4.3 Impact of the number of PCA models

The number of PCA models represent the number of well-contoured patients used for the data training. The goal is to use as few well-contoured patients as possible to train a high-performance deep learning segmentation model. In our study, using 1 PCA model led to Dice coefficient values of only 68.3% ± 11.1% for the left parotid, 66.5% ± 10.5% for the right parotid, and 60.1% ±8.5% for the submandibular glands. These findings indicate that using 1 patient as the PCA template cannot capture enough variations to train a good segmentation model. As we increased the number of PCA models from 1 to 10, the segmentation performance increased significantly for all 3 structures. The segmentation performance did not show obvious improvement ($p > 0.05$) when the PCA model number was increased from 10 to 30 when a sufficient number of synthetic CT scans (e.g., 2000) were generated for training the segmentation model. This observation showed that our data augmentation approach could train a high-quality segmentation model from as few as 10 well-contoured patients.

Figure 9 shows two examples of the segmentation results from the V-Net model trained from synthetic generated CT scans with 10 PCA models, compared with the physician-drawn contours. The auto-segmented contours are quite consistent with the manually delineated contours.

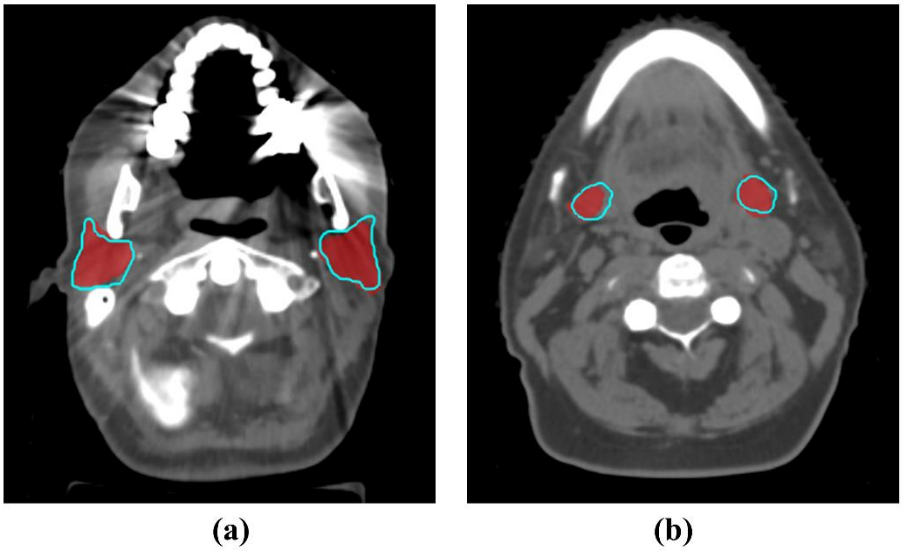


Figure 9. Physician-drawn contours (red colorwash) were compared with the auto-segmentation results (blue contours) by our networks trained on 2000 synthetic CT scans generated by 10 PCA models. (a) Parotid glands (Dice=86.9%). (b) Submandibular glands (Dice=82.0%).

### 3.4.4 Impact of the number of synthetically generated CT scans

The number of synthetic CT scans used for training affects the segmentation performance. We wished to determine a range of numbers of synthetic CT scans, with the goal of balancing the time needed for generating synthetic CT scans from PCA models and the accuracy of the segmentation. Figure 10 shows that segmentation accuracy increased considerably when the number of synthetic CT scans was increased from 300 to 1000. However, when more than 1000 synthetic CT scans were generated for training, the increase in segmentation performance slowed, with the curves becoming flat. In terms of segmentation accuracy, no statistically significant difference was found between using 1000 and 2000 synthetic CT scans for training ($p > 0.05$).

Although the difference between the models trained on 1000 and 2000 training cases was not significant ($p > 0.05$), the average Dice coefficient still increased. To validate this observation, we further evaluated the segmentation accuracy by training the V-Net model with 10,000 synthetic CT scans generated from 10 PCA models (Figure 10). The results showed that the improvement in segmentation performance was still negligible when the number of synthetic CT scans was increased from 2000 to 10,000. Our study showed that using 2000 synthetic CT scans may be sufficient to obtain good segmentation of the structures under consideration.

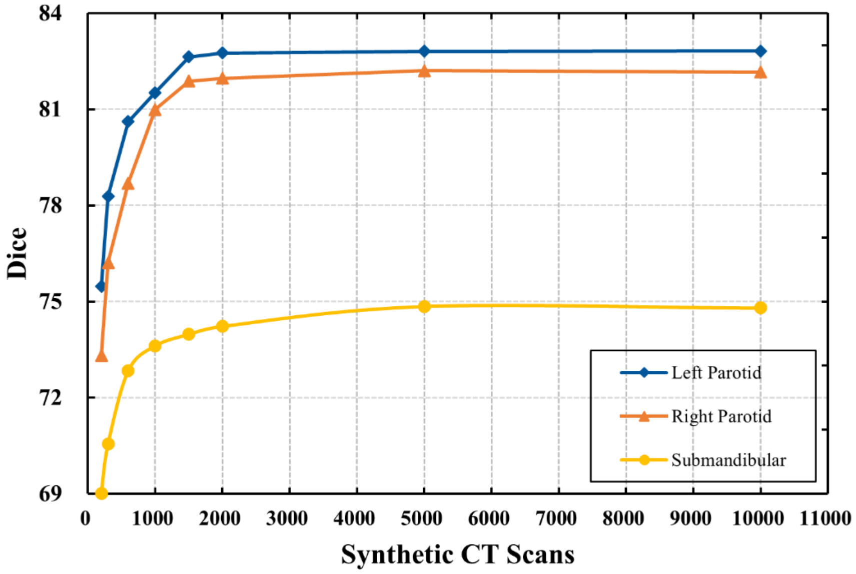


Figure 10. Dice values varied with the number of synthetic CT scans used to train the V-net. Synthetic CT scans were generated from 10 PCA models, with the number of synthetic CT scans evenly distributed between the PCA models.

### 3.4.5 Comparison with the state of the art

To further evaluate our approach, we compared our segmentation performance with one of the state-of-the-art deep learning segmentation methods[150] and a multi-atlas segmentation method, multi-atlas contouring service (MACS),[148], [151] used routinely in our clinic. The CNN-based segmentation model developed by Rhee et al[150] was trained from more than 1000 well-contoured patients to segment the parotid glands. In this model, a two-dimensional (2D) CNN-based architecture, FCN-8s[152], was modified with additional batch renormalization layers added to the end of every layer. The MACS used 12 well-contoured atlases, which are a subset of our PCA model templates, to perform multi-atlas segmentation. The same test dataset was used to evaluate the segmentation performance of parotid glands for the 3 approaches. The comparison is summarized in Table 2. This comparison showed that by training from 10 well-contoured patients, we could achieve segmentation accuracy comparable to that of the state-of-the-art CNN-based segmentation model[150] ($p > 0.05$). Compared with the MACS, our segmentation method showed comparable segmentation results when using more than 20 well-contoured patients, and the differences were not significant ($p > 0.05$).

However, the CNN-based segmentation by Rhee et al[150] was developed based on a modified FCN-8s[152] architecture, which is different from the V-Net architecture used in our approach. For a comparison, we used the same training dataset of Rhee et al[150] (more than 1000 well-contoured patients) to train our V-Net model for parotid segmentation. Using the same test dataset, the average Dice coefficient was 81.4% ± 5.9% for the right parotid, and 81.3% ± 5.5% for the left parotid. This confirmed that our proposed data augmentation approach with more than 20 well-contoured patients could achieve better performance than a CNN model trained from more than 1000 well-contoured patients ($p < 0.05$).

The CNN-based segmentation[150] was not trained to segment the submandibular glands. For comparison, we used the same test dataset (21 patients) to compare the results from multi-atlas segmentation with atlas selection (MACS-AS) [153]. The segmentation of MACS-AS was obtained by using the leave-one-out test. Among the 21 patient scans, when 1 scan was set as testing, MACS-AS chose 12 or less most similar patients as atlases to perform a multi-atlas segmentation. This comparison is also shown in Table 1. When we used 10 PCA models, we achieved performance similar to that of the MACS-AS, with a Dice value of 74.2%±6.8%. When we used 30 PCA models for training, our approach outperformed the MACS-AS.

Table 2. Performance comparison between our models and state-of-the-art autocontouring tools. The best Dice results of PCA 10, 20, and 30 were chosen for comparison. MACS, multi-atlas contouring service; MACS-AS multi-atlas contouring service with atlas selection; V-Net means our V-Net model trained with the same training dataset of Rhee et al[150]. PCA, principal component analysis. PCA 10 means that 10 PCA models were used to create synthetic CT scans; SD: standard deviation.

| **Segmentation Method** | **Dataset** | **Right Parotid** $Dice \pm SD$ | **Left Parotid** $Dice \pm SD$ | **Submandibular** $Dice \pm SD$ |
|---|---|---|---|---|
| *CNN (Rhee et al.*[150]*)* | ~1000 (real) | 82.3±6.5 | 82.4±4.8 | NA |
| *V-Net* | ~1000 (real) | 81.4±5.9 | 81.3±5.5 | NA |
| *MACS (Yang et al*[148]*)/MACS-AS (Yang et al*[153]*)* | 12 (real)/20 (real) | 84.5±9.2 | 84.1±7.9 | 74.0±7.6 |
| *PCA 10* | 2000 (synthetic) | 82.8±6.8 | 82.0±6.9 | 74.2 ± 6.8 |
| *PCA 20* | 2000 (synthetic) | 82.9±7.1 | 82.5±7.2 | 75.5 ± 6.3 |
| *PCA 30* | 2000 (synthetic) | 83.1±6.1 | 82.9±6.7 | 76.3 ± 7.7 |

### 3.5 Discussion

In this work, we proposed an effective data augmentation method using the PCA method, which can create synthetic CT scans to train deep learning segmentation models. This approach was able to learn the anatomy variations in a patient population through the simple PCA method. We tested our approach on the segmentation of parotid and submandibular glands with great success. The segmentation accuracy was comparable with that of state-of-the-art deep learning segmentation approaches. Essentially, our approach enabled deep learning model training from a severely limited dataset (only 10 well-contoured patients), making the deep learning segmentation widely applicable in varied clinical situations with limited contoured datasets, one example being newly emerging technology for which patient data are limited, such as MR-Linac patient treatment [154]. On the other hand, using our approach can also greatly reduce the effort in curating the training data for deep learning models, making deep learning more practical for clinical use. One need only curate 20 or fewer well-contoured head-and-neck patients instead of 1000 patients to train a high-performance deep learning model for most normal tissue auto-contouring. In this study, we used V-Net to demonstrate the effectiveness of our data augmentation approach. However, this approach can be used with any other supervised deep learning segmentation networks.

All of the PCA model-training and synthetic CT scan-generating were implemented in C++ and were performed on a PC (Intel i5-8500 CPU, 32 GB RAM). Each PCA model took about 26 hours to create, and each synthetic CT scan took about 6 minutes to generate. For this study, we used a high-performance computing cluster with four GPU nodes at MD Anderson, for which each node contains four V100 GPUs. We used only two of the GPUs to train our deep-learning segmentation models because of resource sharing. Training took about 4 hours when using 300 synthetic CT scans, 9 hours when using 600 scans, 19 hours when using 1000 scans, and 36 hours when using 2000 scans. The segmentation took about 30 seconds for each patient, making the model quite practical for clinical implementation.

There is a clear trend of improvement in segmentation performance when using more PCA models or more training cases, as shown in Figure 10. However, we noticed that the Dice value decreased from 80.99% ± 7.2% to 80.96% ± 6.8% when increasing the number of PCA models from 10 to 20 with 1000 synthetic CT scans for right Parotid segmentation.

This decrease is not significant though. A possible explanation might the limited anatomic variations of the PCA models. With 1000 random synthetic CT scans, it might not be enough to capture necessary variations with more PCA models. Using more synthetic CT scans (e.g., 2000 scans) to train the CNN model can overcome this issue, which has been demonstrated in our results. In addition, using more CT scans to create PCA models could potentially address this issue also.

The performance comparison between our approach and other state-of-the-art auto-segmentation tools in Table 1 shows that our approach is comparable to or even outperform some state-of-the-art auto-segmentation approaches. Although our approach did not outperform MACS in terms of segmentation accuracy in this validation, our approach was about 20 times faster in performing a segmentation task than MACS, which normally took about 10 minutes to segment one patient. Also, Rhee et al[150] has shown that the CNN-based segmentation was more robust than MACS in some extreme cases, such as irregular patient positioning or abnormally large tumors.

We used the parotid and submandibular glands to demonstrate our approach. These two organs were selected because contouring them, either manually by clinicians or automatically by segmentation algorithms, is often challenging. Their contrast with surrounding tissues is low on CT scans, and their shape and appearance vary significantly from patient to patient. Also, dental artifacts can often affect parotid contouring. Given our success in using this approach to auto-segment the parotid and submandibular glands, we are confident that this approach can be extended to other structures in head and neck, and even other anatomic sites. Moreover, because we previously showed that the PCA model can be used to create synthetic MR images for validating deformable registration algorithms [155], [156], we are confident that we can extend this approach for MRI-based auto-segmentation as well.

The dual-force demons registration algorithm[144] was used to capture the anatomical variations between difference patients to create PCA models. The accuracy of the registration algorithm has been validated in clinical settings both quantitatively and qualitatively for head and neck cancer radiotherapy [139], [157]–[160]. However, it is worth mentioning that the registration accuracy is not an important concern in our study. In the PCA models, only prominent anatomical variations were kept to represent inter-patient variations. Detailed patient-to-patient registration could be left out from the model if the variation it represents is not typical in the patient population. Rather than registration accuracy, our approach only needs to capture meaningful and realistic anatomical variations so that the synthetic CT created from the PCA models could be close to real patient anatomy. This could be visually verified from the synthetic CT images illustrated Figure 7.

Training the PCA model and generating synthetic CT scans using the PCA model were straightforward. Our validation demonstrated the effectiveness of the PCA model in capturing anatomic variations among the patient population, and the application of this for deep learning model training. Thus, the PCA model is an attractive tool for the preparation of training data. The PCA model can use a small number of principal components to simulate a large variation space by learning the variations from a group of patients. However, there are limits to the learning capability of any PCA model, so the number of PCA models could affect the segmentation accuracy, because one PCA model may not capture enough variations. The CT images used to train the PCA model could affect the segmentation performance as well. All CT images used to train the PCA model, and the test data, were from our institution. If the trained V-Net model is used to segment CT scans obtained outside our institution, the performance might be impacted.

Recent years, many deep learning-based data augmentation methods have been proposed to overcome the data sparsity in medical imaging.[131], [136], [161], [162] Particularly, GANs received increased attention for data augmentation due to its performance. However, some researchers have reported that the synthetic image quality from GANs might not be good enough. As in a visual Turing test reported by Frid-Adar et al.[136], two experts could identify 62.5% and 58.6% of the GAN-synthesized liver lesion images. Although GANs have been improved with current cutting-edge architectures, they still could not produce a synthetic image with good fidelity.[162] In our approach, image quality of the synthetic CT scans is not an issue since they were deformed from actual CT images. Moreover, training a high-quality GAN requires a substantial amount of training data.[161] Therefore, GANs might not be a practical data augmentation approach if the initial dataset is extremely limited.

Notably, our PCA model captures variations in anatomic shape only. However, the image appearance or intensity difference could also affect the segmentation results. In our future work, we will develop an appearance model[163] based on the PCA to capture inter-scan intensity difference. By combining the shape and appearance models, we expect to include more variations in the training data, and potentially to improve the model segmentation performance. Considering the MRI scans, in which the intensity values are not calibrated like those for CT scans, the appearance model may be more useful for MR-based auto-segmentation.

### 3.6 Conclusion

We presented a simple yet effective approach to create synthetic CT scans with contours for training deep learning segmentation models and demonstrated its efficiency by developing auto-segmentation models for the parotids and submandibular glands. Our approach enables the training of high-quality deep learning segmentation models from a severely limited number of well-contoured data (as less as the number of atlases used in most multi-atlas segmentation approaches). Our approach could equal or even surpass the performance of some state-of-the-art deep learning-based auto-segmentation approaches.

Our work not only makes deep learning-based auto-segmentation applicable to more clinical situations with limited available data but also greatly reduces the effort needed for data curation for deep learning training.

# Chapter 4: Gross Tumor Volume Auto-segmentation Based on Multimodality Images

## 4.1 Abstract

**Background**: Head and neck (HN) gross tumor volume (GTV) auto-segmentation is extremely challenging due to the morphological complexity and low contrast of targets to surrounding normal tissues on planning CT images. Multi-modality images (CT, PET, and MR) are used in the routine clinic to assist radiation oncologists for accurate GTV delineation due to their different complementary characteristics.

**Purpose**: To develop a deep learning segmentation framework for automated GTV delineation of HN cancers from the combination of planning CT, PET, and MR images.

**Methods**: Two datasets were collected for this study: (1) 524 oropharyngeal cancer patients from 7 institutions with their PET/CT image pairs; (2) 128 oropharyngeal cancer patients from local institution with their planning CT, PET/CT image pairs, and T1-weighted contrasted MR images.

We developed a two-stage auto-segmentation framework that simulated the usual radiation oncologists' GTV delineation process. In stage 1 a model was trained to localize and generate an initial GTV contour using PET/CT image pairs of dataset 1. A local rigid registration was then applied to improve the local alignment among planning CT, PET/CT, and MR images, focusing on the coarse GTV region. In stage 2, a fine segmentation model was trained to generate the final GTV contour based on the combination of the initial GTV and all planning CT, PET, and MR images in dataset 2, leveraging the superior soft tissue contrast of MR images. The self-adapting nnUNet was used as segmentation models in both stages.

Twenty-five out of the 128 patients in Dataset II were randomly selected as test patients, and the proposed model was trained on the remaining patients to generate GTV auto-segmentations. This process was repeated for three times, resulting in the evaluation of 75 patients. Model performance was evaluated using Dice similarity coefficient (DSC), mean surface distance (MSD), and 95% Hausdroff Distance (HD95) between auto-segmented GTVs and manual GTVs used in clinic.

**Results**: Comparing with the coarse GTV segmentation in Stage 1 with PET/CT, the final GTV segmentation in Stage 2 with PET/CT/MR improved mean DSC from 0.60±0.14 to 0.71±0.11, MSD from 2.9±1.6 mm to2.3±1.4 mm, and HD95 from 10.4±6.5 mm to 8.8±6.7 mm, respectively. Including MR images for segmentation in Stage 2 improved the segmentation performance.

**Conclusions**: We developed a novel auto-segmentation tool for HN GTV delineation and demonstrated the efficacy of incorporating multimodal images from CT/PET/MR for accurate GTV delineation. This is the first work to simulate the clinical workflow of GTV delineation by physicians.

### 4.2 Introduction

Head and neck cancer (HNC) is one of the main causes of cancer-related death worldwide [164]. Accurate delineation of the gross tumor volume (GTV) is critical for radiation therapy planning and subsequent treatment outcomes. Despite its importance, accurately and consistently defining the true GTV remains a significant challenge in radiotherapy. This challenge is particularly pronounced in HNC patients because of the complicated morphology of HNC tumors, low target to background contrast, and potential artifacts on CT images. It makes the delineation of GTV a time-consuming task and heavily relying on the experience of the radiation oncologists [165]. The inter-observer contouring variability in defining the target volume can be considerable based on different experience.[31], [32], [166], [167] Incorrect identification of the target volume during treatment planning can lead to marginal misses that ultimately compromise local disease control, the main goal of radiotherapy.[168], [169] This is especially critical with the high-precision techniques, such as intensity modulated radiotherapy (IMRT) and volumetric modulated arc therapy (VMAT), which require a precise delineation of the GTV for optimal treatment. Several studies have demonstrated that inadequate target volume determination by radiation oncologists accounts for a significant proportion of locoregional recurrences after IMRT/VMAT for HNC.[170] Recent years, stereotactic body radiotherapy (SBRT) became more and more popular for treating HNC recurrence. SBRT delivers very precise and intense radiation doses to small targets, typically in one to five fractions. The precision radiation delivery calls for even more precise delineation of the GTV.

In recent years, multi-modality imaging technology has emerged as a promising approach to address the limitations of CT imaging alone. The combination of CT, positron emission tomography (PET), and magnetic resonance imaging (MRI) has shown considerable benefits to help radiation oncologists define GTV more accurately and consistently.[31]–[35] In radiation treatment planning, CT has been the standard imaging modality for contouring the tumor targets due to its excellent anatomical information. On the other hand, MRI offers improved soft tissue contrast and enables better discrimination of tumor tissue by utilizing different sequences.[171] Additionally, PET provides functional information about the tissue metabolism, which can be useful in localizing regions of increased metabolic activity. The complementary and synergistic information obtained from these modalities can help to overcome the limitations of each imaging modality, thereby enhancing the accuracy of GTV delineation and reducing the inter-observer variability.[73] The increasing availability of multi-modality medical images make it possible to develop deep learning auto-segmentation algorithms to aid in GTV delineation for HNC treatment planning.

Several studies have investigated deep learning autosegmentation methods that utilized multimodality images for HN GTV delineation. Moe et al.[40] reported a mean Dice similarity coefficient (DSC) of 0.71 based on CT and PET images in their study involving 197 patients. They observed that PET had a more significant impact than CT and combining CT-PET modalities resulted in improved GTV delineation compared to a single modality. Andrearczyk, et al.[172] further confirmed the benefits of combining CT and PET modalities in HN GTV delineation through their study. The annual Medical Image Computing and Computer Assisted Intervention Society (MICCAI) Head and Neck Tumor Segmentation Challenge (HECKTOR)[75] has provided a platform to evaluate different auto-segmentation models for oropharyngeal GTV delineation based on PET/CT pairs. The best method proposed by Myronenko et al.[173] achieved a mean DSC of 0.8 for primary GTV auto-segmentation. Only one study from Ren et al.[39] has investigated the incorporation of all three modalities for segmentation and compared the different CT, PET, and MRI combinations for GTV auto-segmentation. They found that the combination of CT-PET achieved a mean DSC of 0.73, with the inclusion of MRI (CT-PET-MR) resulting in a slight improvement in mean DSC to 0.74. Bollen et al.[38] reported relatively lower mean DSC of

0.69 and 0.59 for CT-PET and CT-MR combinations, respectively. However, in contrast to Ren et al.'s conclusions[39], their findings suggested that the availability of MR images could improve the delineation performance for GTV providing sufficient registration accuracy of CT-MR images. However, these studies have some limitations, including the relatively small sample sizes, limited exploration of all imaging modalities, and the lack of thorough evaluation in a clinical setting.

In this study, a two-stage deep learning autosegmentation approach was proposed to improve the accuracy and consistency of HN GTV delineation from multimodality images. By leveraging the different usability of CT, PET, and MR images, our method allowed a more comprehensive and accurate GTV delineation compared to single modality-based approaches. Our segmentation approach simulates the GTV delineation process conducted by radiation oncologist in clinical practice. In the first stage, PET image is registered to corresponding planning CT image and a model was trained to localize the tumor and perform an initial GTV segmentation from CT and PET images. In the second stage, the MR images were incorporated to train another model to refine the GTV segmentation. In addition, our auto-segmentation framework was developed with the ability of handling missing modalities by intelligently utilizing the available data to generate GTV segmentation. Our model is able to perform segmentation with missing PET or/and MR images. This further enhances the clinical practicability of our proposed method. Overall, the proposed two-stage coarse-to-fine auto-segmentation approach offers a promising solution to improve the accuracy and consistency of HN GTV delineation.

### 4.3 Materials and Methods

The overall framework of our auto-segmentation method is illustrated in Figure 11. In stage 1, a deep learning model was developed to localize GTV based on planning CT and PET/CT images after a global registration. The coarse GTV generated in this stage was then used to guide the local registration among planning CT, PET/CT, and MR images with a focus on tumor in stage 2. After the local registration, a fine segmentation model was trained using the combination of the coarse GTV and all CT/PET/MR images to generate the final GTV contour.

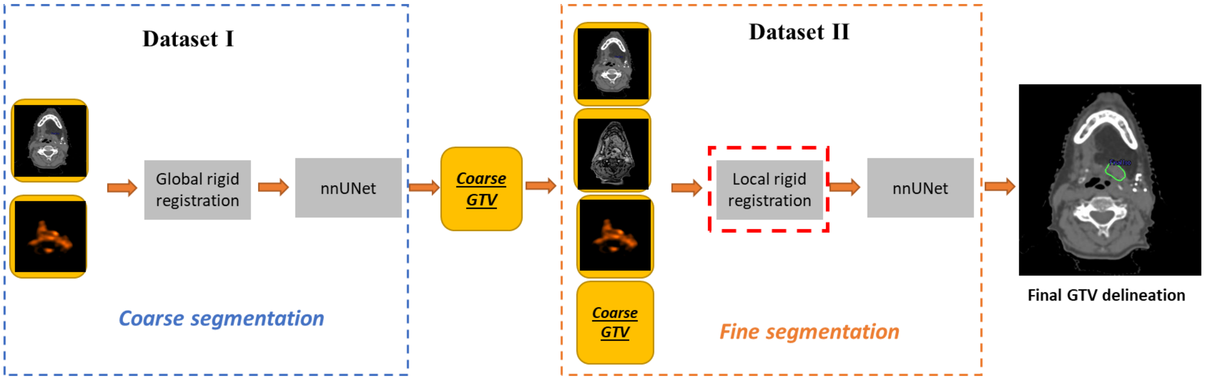


Figure 11. Framework of two-staged coarse-to-fine segmentation network for the automated gross tumor volume (GTV) delineation using planning CT, PET, and MR images.

### 4.3.1 Data acquisition and preprocessing

Two datasets were included in this study to develop the auto-segmentation models for stage 1 and 2, respectively. The segmentation model was trained for primary GTVs only. Nodal GTV segmentation is not investigated in this study.

Dataset I was comprised of PET/CT image pairs of 524 oropharyngeal cancer patients, obtained from seven distinct clinical centers and provided by the HECKTOR 2022 challenge[75]. All imaging data in the Dataset I was paired with ground truth manual segmentations of the primary tumors derived from clinical experts. The PET image intensities had been converted to Standardized Uptake Values (SUV). In Dataset I, the CT images had an average size of 512×512×200 voxels at an average resolution of 0.98×0.98×3 mm, while the PET images had an average 3D size of 200×200×200 voxels at an average resolution of 4×4×4 mm. Both CT and PET images were resampled to the same size with the resolution of 1×1×2.5 mm. This dataset was specifically utilized for the development of the auto-segmentation model in stage 1, requiring only the CT and PET images.

Dataset II was comprised of 128 HN patients who received treatment at The University of Texas MD Anderson Cancer Center from 2017 to 2022. The selection criteria for this dataset included: (1) HN patients who were histologically confirmed with tumor in the oral region, including oropharynx, tongue, base of tongue, and oral cavity; (2) planning CT, PET/CT pairs, and contrast-enhanced T1-weighted MR images were available for the patient; (3) the time interval between the different modality scans for each patient could not span more than 4 weeks and there were not extensive anatomical variations among those scans; (4) manual GTV contours were available on their planning CT images; and (5) the patients were excluded if they had received surgery or chemotherapy before radiotherapy. The MR images included in this dataset are contrast-enhanced T1-weighted out-of-phase Dixon scans, and the selection is based on our clinical experience. To mitigate the influence of intensity variations presented in MR images, a Z-score normalization was employed for MR images within the patient body. All planning CT, PET/CT and MR images were resampled to the resolution of 1×1×2.5 $mm^3$.

### 4.3.2 Two-staged auto-segmentation framework development

The two-staged auto-segmentation framework was proposed based on the GTV delineation workflow in routine clinic. When all imaging modalities available, radiation oncologists start with PET images ($GTV_m$-PET) to delineate the entire tumor extent based on FDG-uptake from PET/CT (after fusion of PET/CT with planning CT). Subsequently, GTV delineation based on MR images ($GTV_m$-MRI) are delineated to account for tumor location and extent based on anatomy that is not apparent on PET/CT (such as perineural spread), microscopic disease not evident on PET/CT and/or refinement of tumor volume due to SUV flare from PET/CT. Before the contour delineation, both treatment position PET/CT and MRI are rigidly registered to planning CT with focus on the tumor region during registration. The contours were drawn on the planning CT ($GTV_m$-Composite) with a display of the fused PET/CT and MRI images. The framework in this study was developed to replicate the clinical procedure and to provide an automated and more consistent GTV delineation.

### 4.3.3 Network architectures

The nnU-Net (“no new U-Net”) architecture[174], known for its great performance in medical imaging segmentation, was chosen as the autocontouring network for this study. This architecture is a modification of the widely popular U-Net convolutional neural network framework, and its unique self-configuring data processing and training framework allows for a robust and efficient learning process. The nnU-Net framework was employed to customize 3D U-Nets for auto-segmentation models in both stage 1 and stage 2. Specifically, the nnU-Net framework dynamically optimized the 3D U-Net network depth to ensure optimal utilization of the large patch size. The default model settings were retained for this study without any modifications. The models were trained using a five-fold cross-validation scheme for 1000 epochs, and the best-performing models were automatically identified and ensembled for the final GTV autocontouring of the test sets. The models were implemented with the PyTorch and trained on an NVIDIA Tesla V100 GPUs with 32 GB VRAM. The loss function was a combination of Dice similarity coefficient (DSC) loss and cross-entropy loss.

### 4.3.4 Stage 1: GTV localization

In stage 1, the PET/CT image pairs in Dataset I were used to train an auto-segmentation model for GTV localization and initial segmentation. The nnU-Net architecture was employed for auto-segmentation, with dual inputs of PET and CT images.

After the model was trained, it was utilized to generate the initial GTV contours for the 128 HN patients in Dataset II. To accomplish this, a global rigid registration was performed between planning CT and the CT in the PET/CT pairs, as shown in Figure 1. The resulted spatial transformation was then used to resample the PET images to achieve the alignment with the planning CT. The global rigid registration was implemented using SimpleITK[175] with the Mattes mutual information[176] as the similarity metrics. Once aligned, the planning CT and PET images were input into the trained model to create the initial GTV.

### 4.3.5 Stage 2: GTV segmentation

After the initial GTVs were generated for the patients in Dataset II, regions of interest (ROIs) were defined by selecting bounding boxes with the size of 128×128×48 to cover the initial GTVs. A local rigid registration using the same algorithm in stage 1 was then applied to register PET and MR images to planning CT with a focus on the defined ROIs. This local rigid registration aims to improve the local alignment within the GTV region so that the setup variations in different scans have a minimum impact on the GTV segmentation. In routine clinic, local registration is usually easy to achieve with a manual defined local region or ROI in most treatment planning systems. To achieve the local registration automatically, we will need to first localize the tumor in stage 1 before the local registration algorithm can be applied automatically.

Following the local rigid registration, the 128×128×48 patches of planning CT, PET, MR images, and initial GTV generated in stage 1 were all used to train another nnU-Net model for the GTV segmentation. This model was designed to leverage information from all image modalities to improve the accuracy of the initial GTV, simulating the clinical process. The network architecture was modified to accommodate multiple inputs from the various image modalities in this stage.

### 4.3.5 Evaluations and implementation

To assess the quality of the GTV segmentation from both stage 1 and stage 2 models, the auto-generated contours are compared with the manual contours both qualitatively and quantitatively. The performance of the stage 1 model was evaluated on all patients in Dataset II by generating the GTV(CT/PET). In stage 2, 25 out of the 128 patients in Dataset II were randomly selected as test patients, and the proposed model was trained on the remaining patients to generate GTV auto-segmentations. This process was repeated for three times, resulting in the GTV(CT/PET/MR) for 75 patients.

Dice similarity coefficient (DSC), mean surface distance (MSD), and 95% of Hausdorff distance (HD95) were used to quantitatively evaluate the performance of the models. Additionally, we utilized Dice's decompositions, precision, and recall in our study to further analyze the segmentation performance and identify over-segmentation and under-segmentation. To assess whether there was a significant difference in performance between the coarse and fine GTV auto-segmentation approaches, a two-sided Wilcoxon Rank-Sum test was conducted, with the significance level set at $p < 0.05$.

Clinical evaluation was also performed on the auto-generated GTVs in stage 1 [GTV(CT/PET)] and stage 2 [GTV(CT/PET/MR)], of ten patients in Dataset II. This evaluation was conducted by three HN radiation oncologists from our institution through visual inspection and focused on two aspects: (1) the clinical acceptability of GTV (CT/PET) and GTV (CT/PET/MR), and (2) the impact of MR images on the accuracy of GTV delineation. For the first aim, GTVs were scored by two HN radiation oncologists using the five-point Likert scale as shown in Table 3. For the second aim, we mainly wanted to evaluate if MR images can benefit the GTV delineation process. Thus, the two oncologists evaluate the test patients and scored using the two-point Likert scale as shown in Table 4.

Table 3. Five-point scale for evaluating the quality of auto-generated GTVs.

| Score | Acceptability | Description |
|---|---|---|
| 5 | **Acceptable:** Use as-is | Clinically acceptable, could be used for treatment without change |
| 4 | **Acceptable:** Minor edits that are not necessary | Stylistic differences, but not clinically important; the current contours/plans are acceptable |
| 3 | **Unacceptable:** Minor edits that are necessary | Edits that are clinically important, but it is more efficient to edit the automatically generated contours/plans than to start from scratch |
| 2 | **Unacceptable:** Major edits | Edits that are required to ensure appropriate treatment and sufficiently significant that the user would prefer to start from scratch |
| 1 | **Unacceptable:** Unusable | Automatically generated contours/plans are so bad that they are unusable (i.e., wrong body area, outside confines of body, etc.) |

Table 4**.** Two-point scale for evaluating the impact of MR images on the accuracy of GTV delineation.

| Score | Acceptability | Description |
|---|---|---|
| 1 | **Impact** | Improve the accuracy of GTV delineation |
| 0 | **No impact** | Do not affect the accuracy of GTV delineation |

## 4.4 Results

The performance of the proposed two-staged auto-segmentation framework was evaluated on Dataset II, and GTV(CT/PET) and GTV(CT/PET/MR) were successfully generated by the models in both stage 1 and 2 for each patient.

The primary objective of stage 1 was to localize the GTV region. Our model identified all GTVs (CT/PET) correctly, without any false positive and false negative detection. These findings demonstrated the effectiveness and reliability of our proposed framework for accurately localizing GTV regions using CT and PET images in stage 1.

### 4.4.1 Quantitative results

The quantitative results measured by DSC, MSD, HD95, precision and recall were summarized in Table 5. The results showed that the GTV auto-segmentation performance was significantly improved with the incorporation of MR images in stage 2. This improvement was supported by the observed increase in DSC (0.71±0.11) and decrease in MSD (2.3±1.35 mm) and HD95 (8.78±6.7 mm) scores for GTV (CT/PET/MR) compared to the GTV(CT/PET) (DSC=0.60±0.14, MSD=2.9±1.6 mm, HD95=10.4±6.5 mm). The statistical significance of the performance improvement was confirmed using a two-sided Wilcoxon Rank-Sum test ($p<0.05$), with the exception of precision.

Upon comparing the performance of GTV(CT/PET) and GTV(CT/PET/MR), we found that the precision of GTV(CT/PET) achieved higher value of 0.89±0.13 when comparing to GTV(CT/PET/MR). In the contrast, GTV (CT/PET/MR) showed higher recall result. Precision and recall are both important metrics used in evaluating the effectiveness of the auto-segmentation models. Precision measures the accuracy of positive predictions, while recall measures the completeness of positive predictions. The higher precision and lower recall of GTV (CT/PET) indicated that the use of CT and PET images in stage 1 can precisely localize GTV, but may result in underestimation of the actual GTV volume.

Furthermore, the impact of including MR images in the training data was also explored. To achieve this, we conducted a separate experiment, where the model in stage 2 was trained with only planning CT and PET images. We then tested the model on Dataset II, and compared it to the GTV (CT/PET/MR) from stage 2. The comparison was made using the same evaluation metrics, as summarized in Table 3. We denoted the auto-segmentation generated based on CT/PET images in stage 2 as GTV (CT/PET*). The model trained with all planning CT, PET, and MR images achieved superior performance, with an average DSC of 0.71 for GTV (CT/PET/MR) and 0.67 for GTV (CT/PET*). These results suggested that the inclusion of MR images in the training data could improve the performance of the model, although the improvement was not statistically significant ($p$=0.16).

Table 5**.** Evaluation results of test set for GTVs auto-generated in stage 1: GTV (CT/PET), and in stage 2: GTV(CT/PET$^*$) and GTV(CT/PET/MR) using DSC, MSD, HD95, precision and recall. GTV(CT/PET$^*$) represents the GTV generated based on CT and PET images in stage 2.

| | GTV(CT/PET) | GTV(CT/PET$^*$) | GTV(CT/PET/MR) |
|---|---|---|---|
| DSC | 0.60±0.14 | 0.67±0.12 | **0.71±0.11** |
| MSD (mm) | 2.9±1.6 | 2.8±1.4 | **2.3±1.4** |
| HD95 (mm) | 10.4±6.5 | 10.0±6.5 | **8.8±6.7** |
| Precision | **0.89±0.13** | 0.76±0.15 | 0.78±0.14 |
| Recall | 0.47±0.17 | 0.66±0.19 | **0.69±0.18** |

The best scores are in **bolded** in the table. DSC: Dice similarity coefficient; MSD: mean surface distance; HD95: 95% Hausdorff distance.

Furthermore, the boxplot in Figure 12 demonstrated the distribution of the quantitative scores for GTV(CT/PET), GTV (CT/PET*), and GTV(CT/PET/MR). The median scores for GTV(CT/PET/MR) demonstrated a significant improvement over those for GTV(CT/PET), which suggests an overall better performance. Moreover, the interquartile range for GTV(CT/PET/MR) was smaller than that for GTV(CT/PET) and GTV (CT/PET*), indicating less variability in the scores and a more consistent performance. However, the presence of outliers for both methods suggested certain cases where the performance of auto-segmentation models was unstable.

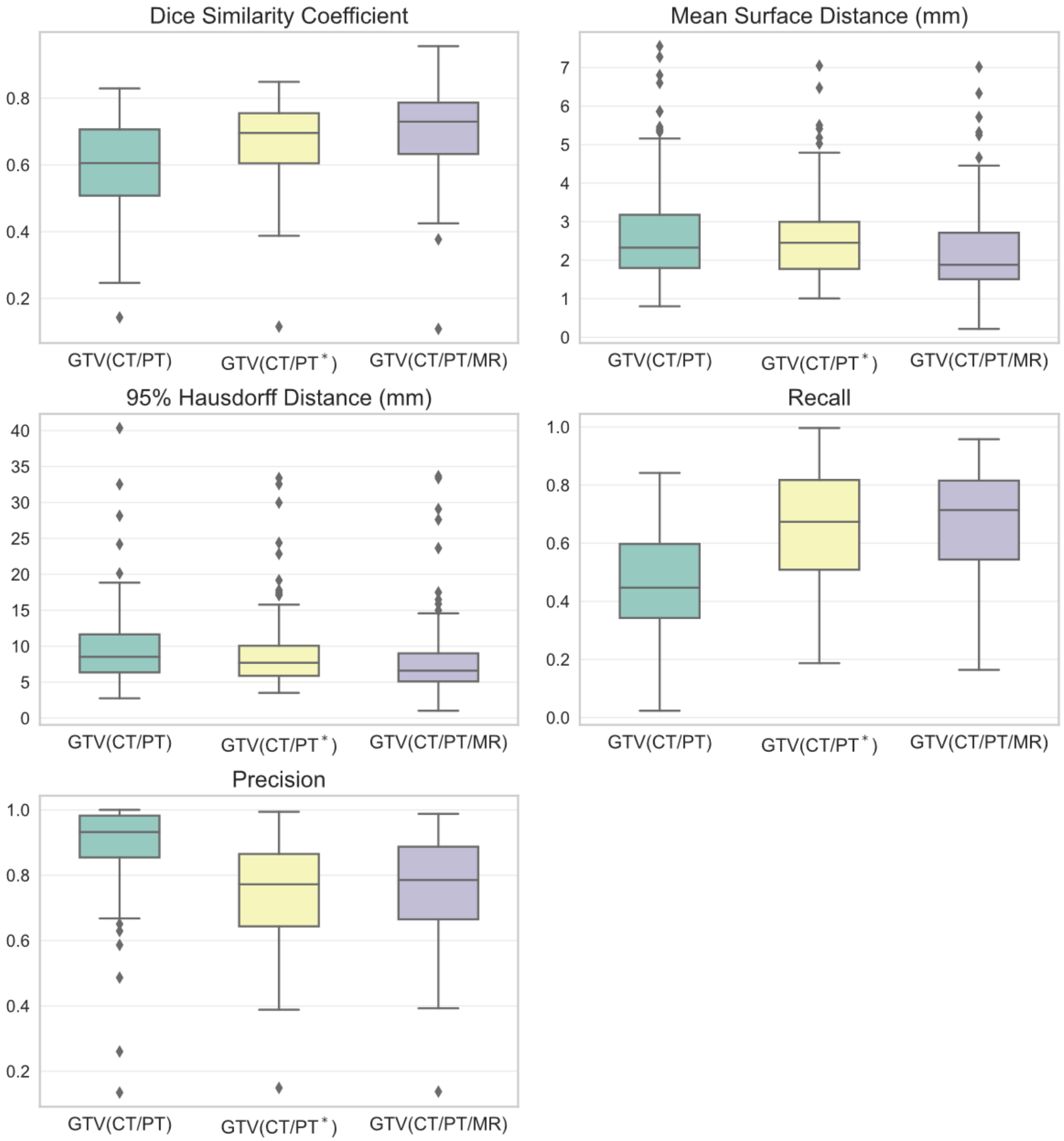


Figure 12. Quantitative evaluation results. Boxplots show the Dice similarity coefficient, mean surface distance, 95% Hausdorff Distance, precision, and recall results of GTV(CT/PET) and GTV(CT/PET/MR) compared with the ground-truth manual contours.

### 4.4.2 Qualitative results

The visual assessment also demonstrated consistent results with the quantitative analysis. Two selected cases are presented in Figure 13 for comparative analysis, with manual contours as ground-truth shown in green. The auto-segmentation of GTVs (CT/PET/MR), as represented by the blue contours, achieved the DSC scores of 0.81 and 0.88, while GTVs (CT/PET) in red have lower DSC scores of 0.58 and 0.62, respectively. Both cases in Figure 3 show high degree of agreement between the GTVs (CT/PET/MR) and ground-truth contours. In contrast, the GTVs (CT/PET) in stage 1 demonstrated under-segmentation of the tumors due to the missing of MR information.

High standardized uptake values on PET images is best suitable for localizing the tumor. The strong physiological signal on PET is easy to identify the tumor location in the body. However, due to the poor resolution of PET, the tumor detailed extension cannot be identified accurately. MR with high spatial resolution and tumor contrast is best severed to identify the detailed tumor extents. The proposed framework in stage 2 incorporated all CT, PET, and MR images to accurately delineate the detailed boundary of the tumor by using the information from MR images. In contrast, stage 1 using only CT and PET images lacked the detailed tumor extents. This comparison example demonstrated that the auto-segmentation framework can effectively leverage anatomical and functional information from all CT, PET, and MR images in stage 2, resulting in superior performance compared to the model in stage 1.

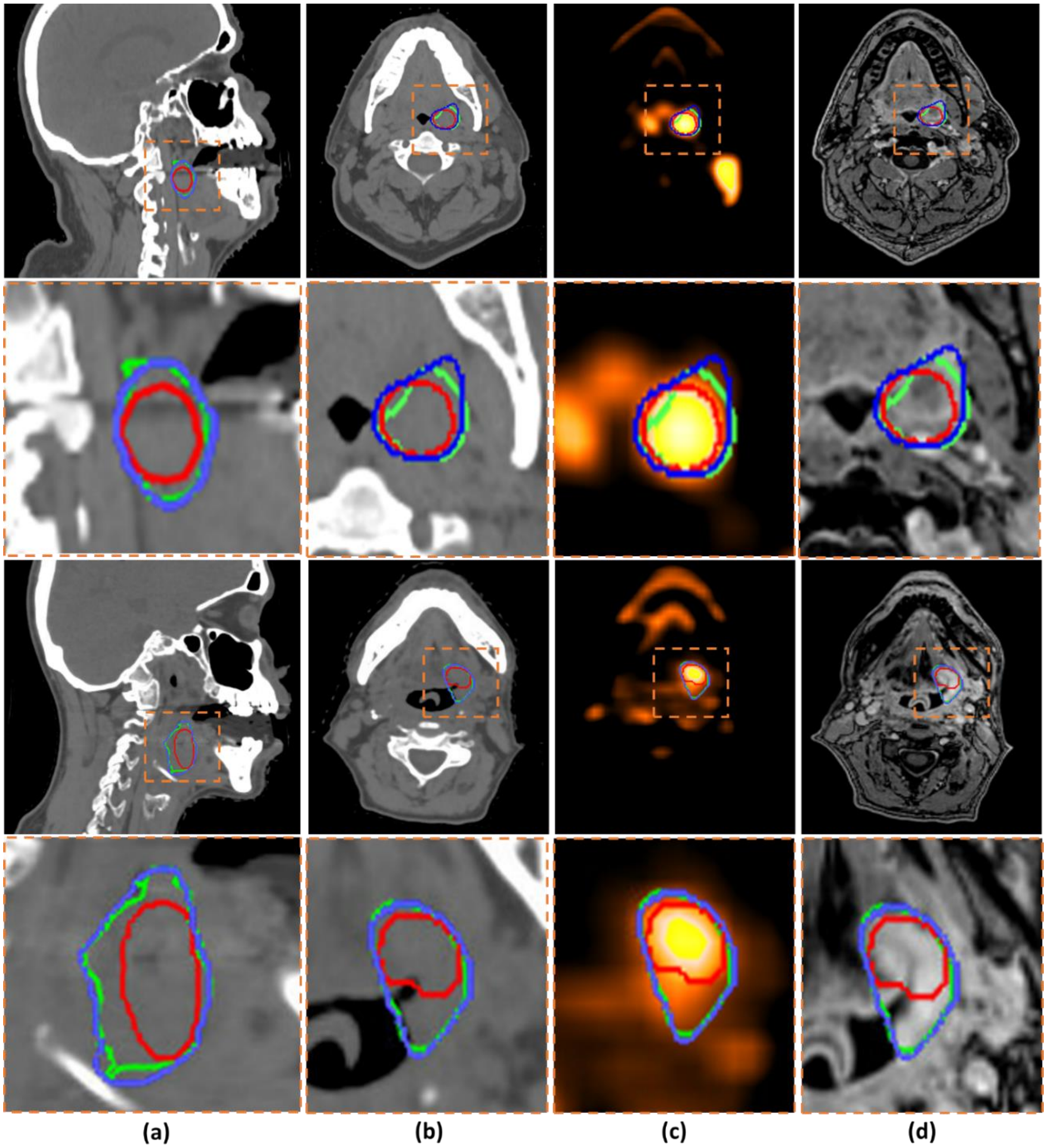


Figure 13. The auto-generated GTVs compared to the physician contours. (a) are the sagittal views of two patients. (b-d) are the axial view of the GTVs, shown on planning CT, PET, and MR images, respectively. Green is the GTV delineated by physician; Red is the GTV contour auto-generated based on CT/PET (in Stage 1); Blue is the GTV contour auto-

generated based on CT/PET/MR (in Stage 2). Second row are the zoom-in figures of the orange boxes shown in the first row.

### 4.4.3 Clinical evaluation

The clinical acceptability of GTV (CT/PET) and GTV (CT/PET/MR) evaluated by radiation oncologists are summarized in Table 6. Based on the results, it can be observed that both GTV (CT/PET) and GTV (CT/PET/MR) for each patient received a clinical acceptability score of 3 or greater. This suggested that all the auto-generated GTVs in stage 1 and stage 2 were acceptable by radiation oncologists for use in treatment planning, with only minor modifications in some cases. However, when examining the scores for higher levels of acceptability (≥4), there was a notable improvement in GTV (CT/PET/MR). Specifically, 80% of the GTVs (CT/PET/MR) achieved a score of 4 or higher, indicating that these auto-generated GTVs can be directly used for treatment planning without any modification. Overall, the high scores for both GTV (CT/PET) and GTV (CT/PET/MR) suggested that our auto-segmentation framework can generate clinically acceptable GTVs, potentially reducing the need for manual editing and saving time in treatment planning.

Table 6**.** The evaluation scores of the clinical acceptability of GTV (CT/PET) and GTV (CT/PET/MR) of 10 randomly selected patents.

| | **≥3** | **≥4** | **5** |
|---|---|---|---|
| **GTV(CT/PET)** | 100% | 60% | 40% |
| **GTV(CT/PET/MR)** | 100% | 80% | 40% |

Besides the evaluation of the clinical acceptability, the radiation oncologists evaluated the impact of using MR for GTV delineation for each patient. The objective was to evaluate whether the introduction of MR images in stage 2 improved the accuracy of the GTV contours and whether the improvement was clinically meaningful. The evaluation results demonstrated that 70% of the cases can benefit from the incorporation of MR images during the GTV delineation process.

### 4.4.4 Tumor size impact

We investigated the relationship between the segmentation performance and the size of the GTV volume. We separated the tumors in Dataset II into three groups based on their volume: (1) small volume size < 15 $cm^3$, (2) medium volume size between 15-30 $cm^3$, and (3) large volume size $>$30 $cm^3$. The resulting mean and standard deviation of DSC scores for the auto-generated GTV (CT/PET/MR) for each group are plotted in Figure 14.

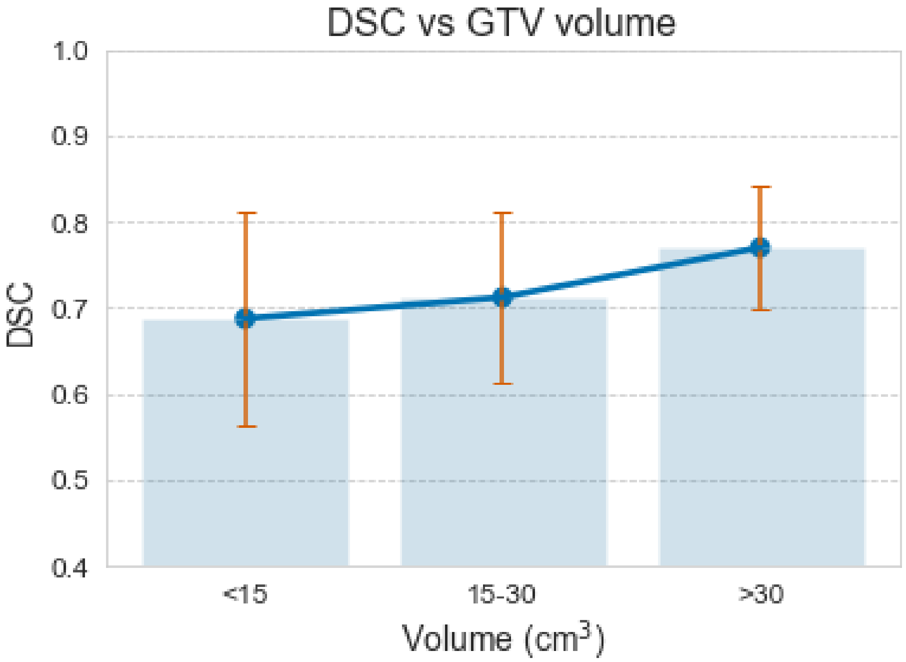


Figure 14. Relationship between DSC and GTV volume size. The mean and standard deviation of DSC results are shown for small volume, medium volume, and large volume GTVs, respectively.

Our findings demonstrated that there is a difference in DSC scores between the different GTV groups, with small volume group achieving a DSC of 0.68±0.12, the medium volume group achieving a DSC of 0.71±0.10, and the large volume group achieving a DSC of 0.77±0.07. It is known that DSC value is correlated to tumor volume. A small volume usually results to a lower DSC score; however, it does not necessary mean that the segmentation is worse compared with a large volume with a higher DSC score. Figure 15 shows the segmentation for a tumor volume of 6.6 $cm^3$ with a DSC score of 0.65, compared with a tumor volume of 57.3 $cm^3$ with a DSC score of 0.78. Compare the two segmentation visually, their performance did not show much difference. In Figure 5, the small volume group exhibited a large variance in DSC scores indicating the DSC is more sensitive for small tumor volumes. In contrast, the large volume group has a smaller standard deviation, indicating DSC is less sensitive to segmentation performance difference. This study emphasized the need of using a combination of segmentation evaluation metrics, particularly for tumor segmentation with a large size variation.

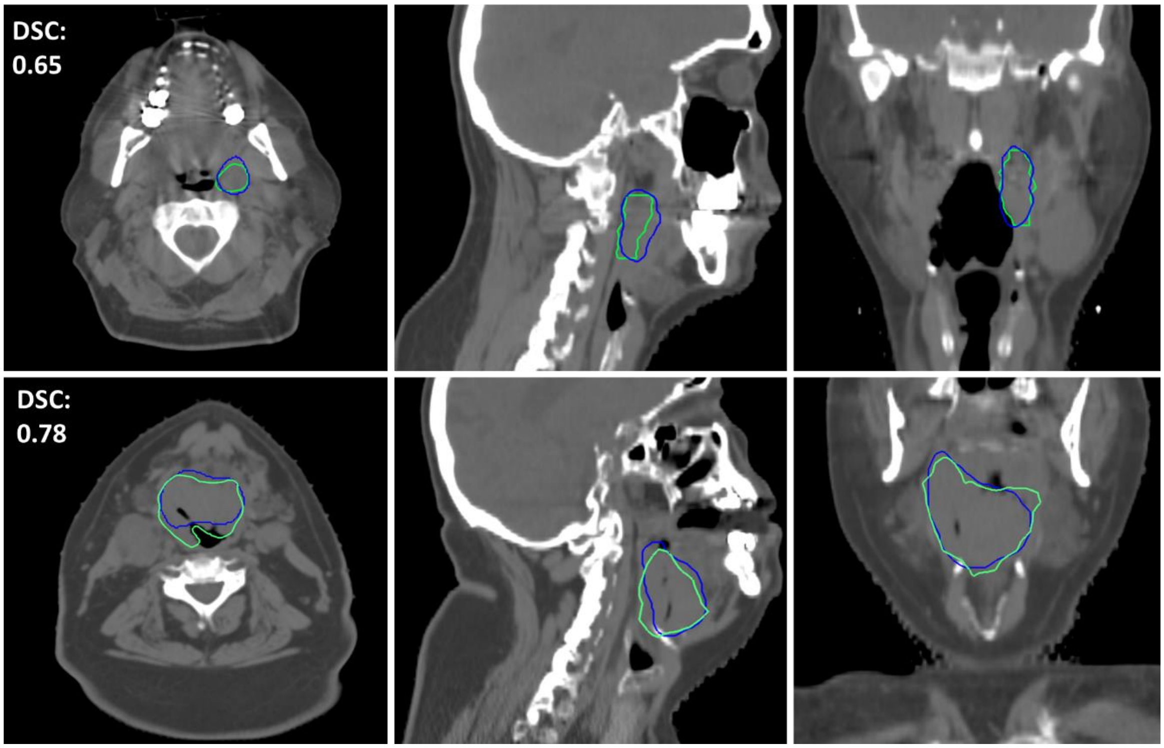


Figure 15. Examples of two cases with large and small tumor sizes. The first row shows the axial, sagittal and coronal views of a patient with tumor volume of 6.6 cm3, and the DSC for the auto-generated GTV is 0.65. The second row shows a patient with tumor volume of 57.3 cm3 with DSC of 0.78. Green is the GTV delineated by physician; Blue is the GTV contour auto-generated based on CT/PET/MR (in Stage 2).

### 4.4.5 Auto-segmentation with missing modality

In many GTV delineation cases, not all image modalities are available for contouring. To address the missing image modality issue, we developed our models to allow flexible data input with missing a channel of image modality. We assume that the planning CT is always available, but allow PET or MR image is not available. The results of missing MR images have been presented and discussed in section 4.4.1. In this section, we focus on discussing the missing of PET images.

Our stage 1 model was trained to generate an initial GTV with or without PET images. If a PET image is not available for a patient, we generated a blank image automatically and use it as the replacement of PET image. During training process, the model was randomly presented with either real PET/CT image pair or the blank/CT image pair. During prediction, if PET image is not available, a blank image is generated automatically and fed into the PET channel. The effectiveness of our proposed model (Model 1) was evaluated by testing it with only CT images and CT/PET image pairs in Dataset II, and compared with two other models: Model 2 trained with CT images only, and Model 3 trained with real PET/CT images.

Table 7. Quantitative results for our model trained to handle missing PET images. Model 1 was trained with PET/CT and blank/CT images, Model 2 was trained with CT images only, and Model 3 was trained with real PET/CT images.

| Model | Test image | DSC | MSD (mm) | HD95 (mm) |
|---|---|---|---|---|
| **Model 1** | **CT only** | 0.56±0.17 | 3.4±2.5 | 14.8±8.6 |
| **Model 2** | | 0.56±0.17 | 3.3±2.6 | 14.4±8.5 |
| **Model 1** | **PET/CT** | **0.61±0.14** | **2.8±1.6** | **10.2±6.6** |
| **Model 3** | | 0.60±0.14 | 2.9±1.6 | 10.4±6.5 |

The best scores are in **bolded** in the table. DSC: Dice similarity coefficient; MSD: mean surface distance; HD95: 95% Hausdorff distance.

Out of 75 test cases, there were 6 cases where both Model 1 and 2 cannot identify the GTV using CT image only. All of these 6 cases have a small tumor size ($<10$ cm$^3$) with low contrast to surrounding normal tissues. They are difficult to identify without PET images. All GTVs were successfully localized by Model 1 and 3 when using both CT and PET images. After excluding the 6 cases, the quantitative results of these models were summarized in Table 7. The results indicated that Model 1, which was developed to handle missing PET images performed similarly to the other models that were trained for specific modality segmentations. Specifically, Model 1 achieved comparable results to Model 2 when tested with only CT images, and comparable results to Model 3 when tested with both CT and PET images.

In stage 2, we developed two separate models for GTV auto-segmentation to address the potential challenge of missing PET images. The first model was trained using a combination of CT/PET/MR images, while the second model utilized only CT/MR images. We evaluated the performance of the two models on Dataset II by predicting GTV (CT/PET/MR) and GTV (CT/MR), respectively. The quantitative comparison of the two models was summarized in Table 8. The table showed that both models achieved comparable results, with a DSC of 0.71 indicating a high level of agreement between the predicted and ground truth GTV contours. This suggests that even in the absence of PET images, the models in stage 2 can still accurately segment the GTV with CT and MR images.

Table 8. Quantitative comparison of two models for GTV auto-segmentation using CT/PET/MR and CT/MR images.

| | **DSC** | **MSD (mm)** | **HD95 (mm)** |
|---|---|---|---|
| **GTV (CT/PET/MR)** | 0.71±0.11 | 2.3±1.4 | 8.8±6.7 |
| **GTV (CT/MR)** | 0.71±0.12 | 2.4±1.6 | 9.0±7.2 |

DSC: Dice similarity coefficient; MSD: mean surface distance; HD95: 95% Hausdorff distance.

## 4.5 Discussion

The development of an accurate and efficient auto-segmentation framework for GTVs from multi-modality images is a critical challenge in radiotherapy. In this study, we introduced a novel two-stage auto-segmentation framework for delineating GTVs from multi-modality images (CT/PET/MR), providing an automatic, fast, and consistent solution for GTV segmentation. This is the first auto-segmentation framework to simulate the clinical workflow of manual GTV delineation. To achieve this, the framework was developed based on the clinical procedures for GTV delineation[177], where physicians typically start with PET to identify the GTV region and then contour the GTVs separately using PET and MRI information. The final $GTV_m$ (Composite) used for treatment planning is a combination of $GTV_m$ (PET) and $GTV_m$ (MR), as shown in Figure 16.

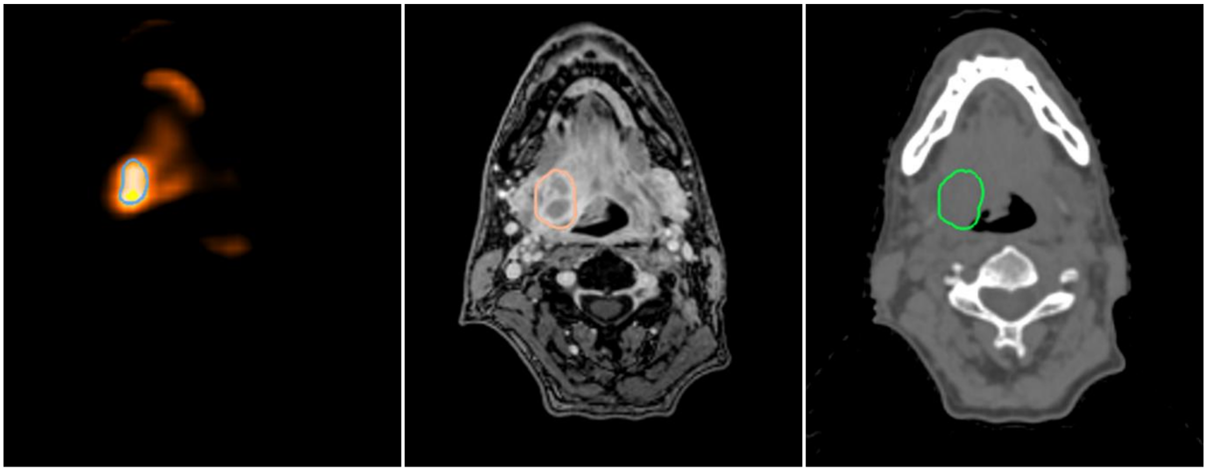


Figure 16. An example of the $GTV_m$ (PET), $GTV_m$ (MR), and $GTV_m$ (Composite) used in clinic, which are shown on PET, MR and planning CT images, respectively.

Our proposed approach does not show superior segmentation performance purely based on mean DSC values by comparing with similar studies in literature.[38]–[40], [172], [173] However, it is worth of noting the impact of tumor volume to DSC values when DSC is used to evaluate the segmentation performance. According to Guo et al.[74], the size of tumor has a substantial impact on the segmentation performance. Their study on GTV auto-segmentation based CT and PET images demonstrated a significant difference in the DSC scores between large GTV group (volume size > 30 $cm^3$, mean DSC: 0.75) and small GTV group (volume size < 30 $cm^3$, mean DSC: 0.6x). In our study, the average size of GTVs in test dataset is 16 $cm^3$, with 43% of the GTVs having a volume < 10 $cm^3$, 72% having a volume < 20 $cm^3$, and only 12% having a volume > 30 $cm^3$. For tumor volumes > $cm^3$, our method achieved a mean DSC value of 0.77, which outperforms most reported results in literature.[38]–[40], [172], [173] On the other hand, our study also signifies the importance of using a combination of multiple quantitative metrics to evaluate segmentation performance of GTV due to the large variation of the tumor volumes. In general, a combination of DSC, MSD, and HD95 is necessary to fairly evaluate GTV segmentation. By examining our MSD results, our method outperformed all other reported results [38]–[40], [172], [173].

In our evaluation of GTV(PET/CT) or GTV (PET/CT*), the ground-truth used for comparison is the final GTV contour manually delineated by radiation oncologists using a combination of planning CT, PET, and MR images. It is known that the detailed boundary of the tumor is not visible on either CT or PET. Therefore, the segmentation results of GTV(PET/CT) or GTV(PET/CT*) could miss the tumor detailed boundaries that are only available on MR images. The relatively low DSC values may reflect the importance of incorporating MR images for GTV segmentation. On the other hand, some reported studies using PET/CT only usually use manual GTV contours delineated from PET/CT only as ground-truth. Since both auto-segmented contours and ground-truth did not use any MR information, a better agreement may be expected. We used a clinical case to demonstrate this point. In this case, the manual GTV contour was created in the following way in clinic. The GTV contours were first delineated separately on PET image [$GTV_m$(PET)], MR image [$GTV_m$(MR)] and the final $GTV_m$(Composite) was determined through a fusion of $GTV_m$(PET) and $GTV_m$(MR) by leveraging all available information from planning CT, PET, and MR images. Our auto-segmented GTV (CT/PET) was compared with the manual contour $GTV_m$

(PET) and $GTV_m$ (Composite) shown in Figure 17. The figure shows that the auto-segmented GTV (CT/PET) closely resembles the manual $GTV_m$ (PET) when only PET/CT information was used for the segmentation [DSC = 0.89]. It also showed that by incorporating the MR image, the $GTV_m$ (Composite) greatly deviated from $GTV_m$ (PET) by including the tumor boundary details. Our auto-segmented GTV(PET/CT) has a DSC of 0.64 by comparing with the $GTV_m$ (Composite).

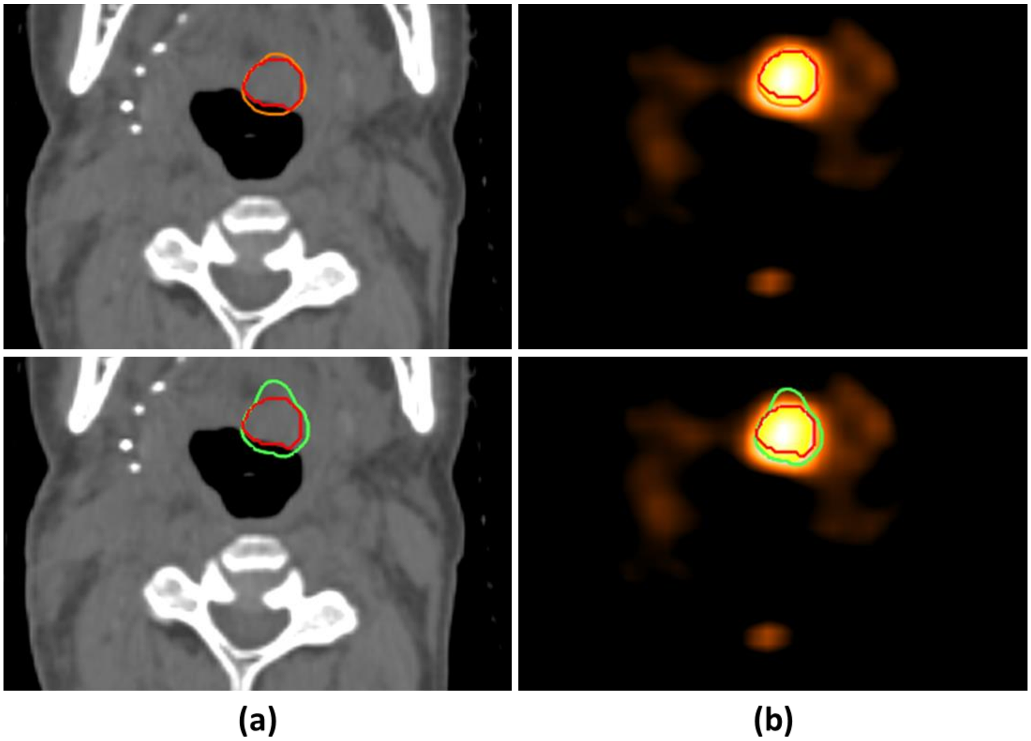


Figure 17. The auto-generated GTV (PET/CT) compared to the physician contours. Auto-generated GTV (PET/CT), the clinical $GTV_m$ (PET) delineated by physician and the clinical $GTV_m$ (Composite) are shown as red, orange, and green contours, respectively. The first row is the comparison between GTV (PET/CT) and clinical $GTV_m$ (PET) shown in planning CT (a) and PET (b) images. The second row is the comparison between GTV (PET/CT) and corresponding $GTV_m$ (Composite) shown in planning CT (a) and PET (b) images.

Auto-segmentation from multi-modality images is also challenged by accurate alignment of planning CT, PET, and MR images. Despite short intervals between image acquisitions, variations may exist due to factors such as patient discomfort, changes in tumor size, or differences in positioning instruction. The degree of flexion or extension of the neck can vary significantly between different sessions, making it difficult to accurately register the multi-modality images. For example, Ren et al.[39] observed poor GTV autosegmentation quality using a combination of CT and MR images due to poor registered CT/MRI data when MR images were deformably registered to the planning CT images. To address the tumor alignment issue, we applied both global and local rigid registration to minimize the misalignment. In stage 1, a global rigid registration was implemented to align planning CT and PET images, with the main goal for localizing the tumor and generating an initial GTV. A high-accurate registration was not required for the localization purpose. In stage 2, we introduced a local rigid registration based on the initial GTV generated in stage 1. This local registration focusses on the GTV region to improve the alignment accuracy for tumor. To ensure sufficient registration accuracy, the results of registration in each stage were manually inspected. It is worth of noting that deformable registration is not suitable for pre-alignment of multimodal images for segmentation because deformable registration could potentially change the morphology of tumor causing inaccurate segmentation. The hierarchical rigid registration is best suited for this purpose. The two-stage segmentation is partially due to the implementation of the hierarchical rigid registration automatically.

Besides quantitative evaluation, our clinical evaluation of the contours scored by three radiation oncologists demonstrates the clinical usability of this tool. As shown in table 4, all the auto-generated GTVs in stage 2 received a score greater than 3, of which, 80% having a score 4 or 5 does not need any editing and 20% having a score of 3 needs minor editing.

The use of multi-modal imaging techniques, such as PET/CT and MRI, has been shown to improve the accuracy and consistency of GTV delineation in radiation treatment planning.

[31]–[33] However, these modalities may not always be available in clinical settings. To address the data availability issue, we proposed the two-staged auto-segmentation framework that is adaptable to different scenarios. Specifically, our proposed two-stage auto-segmentation framework can generate GTV contours using different image combinations (CT, CT/PET, CT/MR, or CT/PET/MR), depending on the available imaging modalities. This flexibility makes our framework a valuable tool in clinical practice where access to all modalities may not always be possible. The experiments conducted in Section 3.6 demonstrate the robustness of our framework in generating accurate and consistent GTV segmentations with missing certain modalities.

In our study, the inclusion of MRI was demonstrated to improve the performance of GTV delineation. Our results are consistent with the findings of Bollen et al. but in opposition to those of Ren et al. As previously discussed, the differences in outcomes may have been affected by inherent challenges, including the poor MR-CT registration and non-standardized intensity of MR images. It is worth noting that the advantages of using MRI in GTV delineation have been extensively reported in previous studies [35], [38], [73]. Moreover, the development of MRI-guided radiotherapy for HNC enables the acquisition of MR images in the treatment position with an immobilization mask. This allows for improved registration accuracy between MR and planning CT images, and thus enhances the potential benefits of automated delineation based on the combination of CT and MR images. The incorporation of MRI in the radiotherapy workflow can facilitate more accurate and consistent GTV delineation.

Despite the promising results and clinical potential of our proposed GTV auto-segmenting framework, our study still has several limitations. First, the performance of our models heavily relied on the quality of manual GTV contours used for model training. As the ground-truth GTVs were delineated by different HN radiation oncologists, there existed inter-observer variability. The inconsistencies in the manual GTVs may carry over to segmentation and lower segmentation performance. To address this issue, more training data with variations are needed. Second, our model was trained and evaluated only for the auto-segmentation of GTV in the oral region, including base of tongue, oropharynx, and tonsil disease. This is rather limited, however, due to the anatomical difference, it cannot be extended directly to other HN subsites. In addition, our model only segments primary tumors while ignores nodal GTVs. To autocontour nodal GTVs or GTVs in other subsites, a new set of data need to be curated to train a new model. Future studies should aim to validate our segmentation framework on larger and more diverse datasets (e.g. multi-institutional datasets) to better understand its potential impact on clinical outcomes. Last, we used only contrast-enhanced T1-weighted out-of-phase Dixon MR sequences for our model development. It would be beneficial to further investigate other more general MR sequences for segmentation, (e.g. T2-weighted MRI).

### 4.6 Conclusion

In this study, we successfully developed an automated tool for accurate GTV delineation in HN cancer patients using multimodal images from PET/CT/MR. Our tool allows for flexible inputs with missing PET and/or MR images, making it a valuable tool in clinical settings where access to all imaging modalities may not always be possible. Our two-stage auto-segmentation framework simulates the clinical workflow of GTV delineation by radiation oncologists, which is a significant advantage over previous studies. The efficacy of our proposed auto-segmentation tool could potentially reduce the physician workload and improve the accuracy and consistency of GTV delineation from multimodal images, which ultimately will improve radiation treatment outcomes for HNC patients.

# Chapter 5: Automated Deformable Image Registration for MR-guided Adaptive Radiotherapy

This chapter is based upon the following article:

Yao Zhao, Xinru Chen, Brigid McDonald, Cenji Yu, Laurence E. Court, Tinsu Pan, He Wang, Xin Wang, Jack Phan, and Jinzhong Yang "Patch-RegNet: a hierarchical deformable registration framework for inter-/intra-modality head-and-neck image registration with ViT-Morph", Proc. SPIE 12464, Medical Imaging 2023: Image Processing, 1246403 (3 April 2023). doi: 10.1117/12.2653352

## 5.1 Abstract

**Background**: Deformable image registration (DIR) between daily and reference images is fundamentally important for adaptive radiotherapy. In the last decade, deep learning-based image registration methods have been developed with faster computation time and improved robustness compared to traditional methods. However, the registration performance is often degraded in extra-cranial sites with large volume containing multiple anatomic regions, such as Computed Tomography (CT)/Magnetic Resonance (MR) images used in head and neck (HN) radiotherapy.

**Purpose**: To improve the accuracy and speed of CT-MR and MR-MR registration for head-and-neck MR-Linac treatments through an automated deformable image registration (DIR) framework.

**Methods**: We developed a hierarchical registration framework. Following a whole-volume rigid registration, the input images were divided into overlapping patches. Then a patch-based rigid registration was applied to achieve accurate local alignment for subsequent DIR. We developed a ViT-Morph model, a combination of a convolutional neural network (CNN) and the Vision Transformer (ViT), for the patch-based DIR. A modality independent neighborhood descriptor was adopted in our model as the similarity metric to account for both inter-modality and intra-modality registration. The CT-MR and MR-MR DIR models were trained with 242 CT-MR and 213 MR-MR image pairs from 36 patients, respectively, and both tested with 48 image pairs (24 CT-MR and 24 MR-MR) from 6 other patients. The registration performance was evaluated with 7 manually contoured organs (brainstem, spinal cord, mandible, left/right parotids, left/right submandibular glands) using Dice similarity coefficient (DSC) and mean surface distance (MSD). These indexes were compared with the traditional registration methods in Monaco treatment planning system and the popular deep learning-based DIR framework, Voxelmorph.

**Results**: The average DSC and MSD calculated over all organs between the deformed contours and the reference manual contours were 0.76±0.05/1.9±0.5mm for CT-MR registration and 0.87±0.06/0.9±0.3mm for MR-MR registration, respectively. Our method outperformed VoxelMorph by 6% for CT-CT registration, and 4% for MR-MR registration based on DSC measurements. Of the organs, right parotid achieved the best results (0.81±0.03/2.0±0.5mm) while mandible had the worst (0.71±0.04/1.9±0.3mm) in CT-MR registration; brainstem had the best results (0.93±0.02/0.6±0.2mm) while right submandibular gland had the worst (0.84±0.05/1.0±0.4mm) in MR-MR registration.

**Conclusions**: We developed a fully automated hierarchical registration framework that achieved significantly improved DIR accuracy of both CT-MR and MR-MR registration for head-and-neck MR-guided adaptive radiotherapy.

## 5.2 Introduction

Recent technological advancement in radiotherapy (RT) has enabled online adaptive radiotherapy (ART) to optimize radiation treatment plans on-the-fly based on daily anatomy changes to improve treatment accuracy[178]. ART is attractive in the treatment of head and neck cancer (HNC) because potential significant tumor shrinkage, large anatomical deformation of organs at risk (OARs), and substantial weight loss are commonly observed for HNC patients during treatment[179]–[181]. Without ART, the anatomical changes can lead to a compromised plan for dose delivery, with either unnecessary damage to critical organs or undertreatment near the tumor boundary. Cone beam computed tomography (CBCT) has been in use for decades for patient alignment, it enables the observation of anatomical changes and assessing the necessity of conducting offline ART[182]. However, the inferior image quality and poor soft tissue contrast of CBCT can limit its performance in clinical implementation. The advent of magnetic resonance (MR) imaging-guided linear accelerators (MR-Linacs) has provided an excellent platform for online ART, enabling MR-guided radiotherapy (MRgRT)[95], [183], [184]. Integrated MR-Linacs allow the acquisition of on-board MR images just before radiation delivery, thereby providing superior visualization of the soft tissue for daily online plan re-optimization.

In MRgRT, daily MR images are used for daily set up verification and plan adaptation. The key enabling technology for online ART is deformable image registration (DIR), which establishes spatial relationship between two images by transforming a moving image to a fixed image space. DIR establishes the relationship between the planning computed tomography (CT) and MRI images so that the original treatment plan on CT can be adapted and re-optimized based on daily anatomy on MRI to achieve optimal treatment delivery[184]–[186]. The current ART workflow relies on the DIR process to deform contours delineated on planning CT onto the daily MR image, followed by a full plan re-optimization using the established spatial relationship between CT and MR images. However, DIR between CT and MR images has long been a difficult task due to significant contrast differences and anatomical changes over time, particularly for extracranial anatomical sites. For example, the neck flexion is often unavoidable from fraction to fraction no matter how carefully the patient is set up for HNC treatment. In addition, the treatment volume of HNC often spans a large region from top of nasopharynx to lower neck or upper lung, further complicating the DIR task because of locoregional anatomical variations in the HN region, as shown in Figure 18. Existing registration tools in clinic could not ensure accurate and robust contour propagation, especially for flexible structures such as the neck and spine[95]. They always require the physician to review and revise the deformed contours on daily MR images, which is tedious and time-consuming. Therefore, to facilitate the online ART workflow of MRgRT of HNC, it is crucial to develop a fast and efficient DIR method for CT-to-MR registration.

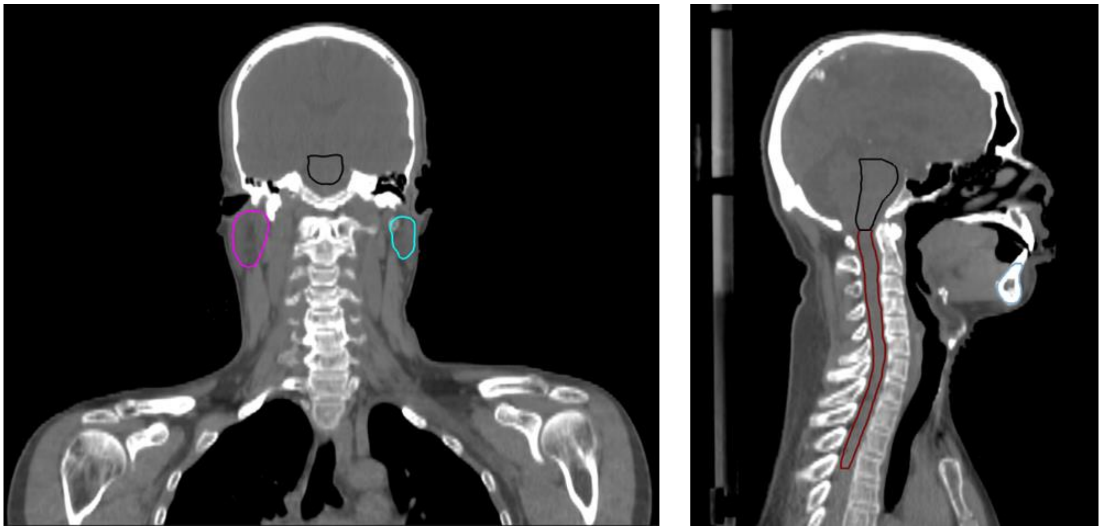


Figure 18: Illustration of the anatomical coverage of head and neck cancer patient

Many approaches have been developed to address the challenges in inter-/intra modality image registration. For decades, traditional registration algorithms based on maximizing image similarity between two images have been successively implemented in many applications[98]. However, these iterative optimization-based methods usually require high computational cost and are not robust or accurate for cases with large deformation. Recently, deep learning (DL)-based approaches have been demonstrated to have superior performance and speed compared to the traditional methods [187]–[190]. The DL-based DIR methods are broadly categorized into (i) supervised and (ii) unsupervised learning methods[187]–[189]. In supervised learning methods, the networks are trained with ground-truth deformation vector fields (DVFs) that are usually generated from traditional methods or synthetic data. The registration performance is limited by the quality of ground-truth DVFs which may be different from the actual anatomical deformation. On the other hand, unsupervised learning methods have been developed to overcome these limitations by training the networks to optimize similarity metrics between deformed and fixed images like traditional registration methods. VoxelMorph, proposed by Balakrishanan et al.[191], was an example of unsupervised learning methods that utilized a spatial transformer network (STN)[192] to generate the deformed image during training process. However, the application of the unsupervised registration methods for CT-to-MR comes with many challenges due to the inherent limitation of available similarity metrics. While mutual information (MI)[193], [194] has been commonly used in CT-to-MR registration, its performance for DIR may be diminished due to its intrinsic global measurement and limited capacity in distinguishing tissue types with similar intensities [195]. To overcome this problem, translation methods using Generative Adversarial Network (GAN)[196] have been proposed to generate synthetic CT (sCT) from MR images, which are used as a bridge for CT-to-MR registration. McKenzie et al.[197] reduced the multi-modal registration problem to a mono-modal one by using a cycle-consistency GAN model to generate sCT from MR images for HNC patients. Xu et al.[198] proposed to further combine the sCT-to-CT registration together with MR-to-CT registration, which could leverage the deformation predictions from both multi-modal and mono-modal registrations. However, the performance of this method is restricted by the reliability of the anatomical features in synthetic images[198]. Furthermore, the convolutional neural network (CNN)-based registration methods with a limited size effective receptive field may suffer from loss of long-range spatial relations in moving and fixed images during registration. Recently, there has been an increase of applying transformer to computer vision tasks, such as segmentation[199]–[201], image reconstruction[202], [203], and image recognition[204], [205]. This technique has been demonstrated to have superior performance in image registration because of its self-attention mechanism and ability to build associations between distant parts of images[190], [206]–[208].

In this study, we attempt to achieve accurate and rapid DIR between CT/MR or MR/MR for online MR-guided ART for HNC. We propose a novel hierarchical registration framework named Patch-RegNet, where the patch-based registration is introduced to improve the local alignment. One difficulty in training a DL-based DIR network for HNC is to achieve a good initial position between moving and fixed images owing to their extensive coverage over multiple anatomic sites, including the head, neck, shoulders, and upper lungs. Our Patch-RegNet addresses this issue using a three-stage workflow: a whole volume rigid registration, a patch-based rigid registration, and a patch-based deformable registration. The patch-based rigid registration ensures an improved local pre-alignment between two images for subsequent DIR. Our patch-based DIR network, ViT-Morph, is built upon the combination of Vision Transformer (ViT)[204] and VoxelMorph[191] to take advantage of both CNN features and the long-range spatial relationships from the transformer. In addition, the modality independent neighborhood descriptor (MIND)[195] is used as the similarity metric in Patch-RegNet to account for both inter- (CT-to-MR) and intra- (MR-to-MR) modality DIR. Our proposed Patch-RegNet is evaluated on the clinical HNC CT and MR images acquired for radiotherapy both qualitatively and quantitatively.

## 5.3 Methods and Materials

### 5.3.1 Overview

Patch-RegNet has a hierarchical registration framework to complete registration tasks robustly and automatically in a fully unsupervised manner. The overall framework is shown in Figure 19.

The hierarchical registration framework starts with a whole-volume rigid registration. The moving and fixed images are first rigidly registered based on whole-image volumes. Overlapping patches are then extracted from the pre-aligned images for the subsequent patch-based registration model. Within patch-based registration, a rigid registration is applied to further refine the local alignment of patch pairs for following DIR, which is achieved by training a patch-based DIR network, ViT-Morph, for registration. The ViT-Morph model is trained to capture the deformation field between moving and fixed patches. A deformed patch is generated by warping the moving patch with the deformation field using a spatial transformer network (STN). Finally, the deformed patches are fused together to obtain the entire deformed image.

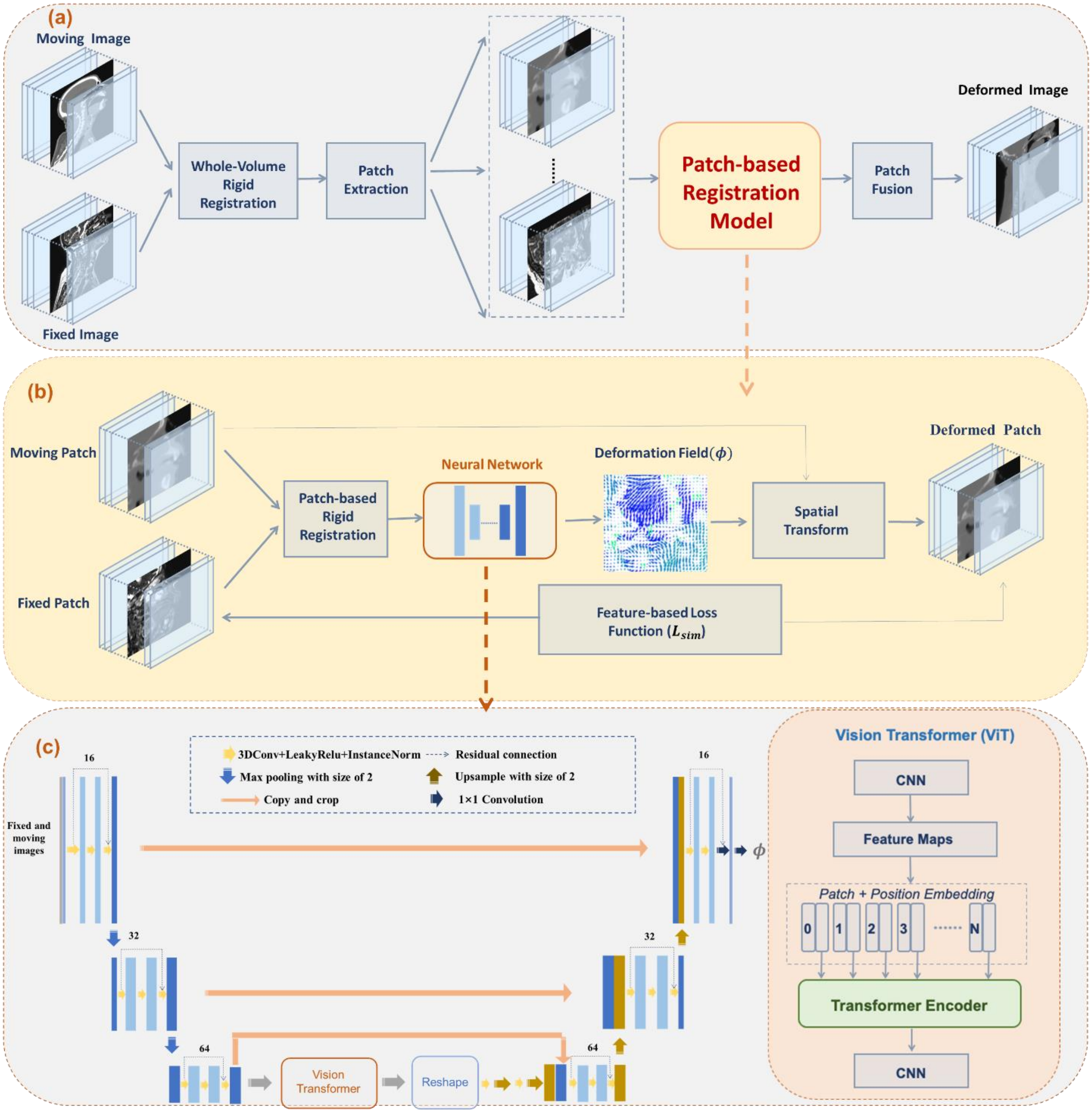


Figure 19: Overall framework of Patch-RegNet consisting of three-stage registrations: a whole volume rigid registration, a patch-based rigid registration, and a patch-based deformable registration. The patch-based registration model that includes stages 2 and 3 is shown in (b) and (c). (b) The schematic illustration of ViT-Morph: a hybrid network of vision transformer (ViT) and VoxelMorph; (c) the details of the convolutional neural network (CNN): a combination of a modified residual-UNet and ViT. The transformer encoder consists of 12

alternating layers of Multihead Self-Attention (MSA) and Multi-Layer Perceptron (MLP) blocks.

### 5.3.2 Whole volume global registration

Let $V_M, V_F \in \mathbb{R}^{D \times H \times W}$ denote the whole volumes of moving and fixed images. $V_M$ and $V_F$ are first rigidly aligned to account for different anatomical scanning range as an initial step to facilitate the subsequent DL-based DIR process. In Patch-RegNet, we use SimpleITK[209] to implement a rigid registration method that minimizes the Mattes MI[210] between moving and fixed images with a gradient optimization algorithm. Instead of using an affine transform, a similarity transform is employed for the linear registration. We observe that a similarity transform could often give more robust results than an affine transform for the sake of slightly inferior local alignment accuracy. After the whole volume linear registration, $V_F$ and pre-aligned $V_M$ are cropped to the overlapped region, represented by $V_F'$ and $V_M'$, for the following registration.

### 5.3.3 Patch-based local registration

The whole volume registration techniques generally do not provide sufficient accuracy in some local regions for subsequent DIR, especially for images covering a large area with multiple anatomical sites like HN images.

To address this problem, we propose the patch-based rigid registration to provide the DL-based registration with improved regional guidance. Let $P_m, P_f \in \mathbb{R}^{d \times h \times w}$ represent the corresponding patches extracted from the fixed and pre-aligned moving images. The parameters $d, h, w$ are chosen to enable the patch to cover the complete anatomic site/sites and keep it with the flexibility of locoregional alignment. Body masks are generated for $V_F'$ to assist the patch extraction. Based on the body mask, the patch centers are limited to be within the patient body to minimize the air region in each patch. As shown in Figure 20, two tunable parameters, stride $s$ and cropping $c$, allow the flexibility of patch extraction and border exclusion during patch fusion. As registration for each image voxel needs the surrounding spatial information, the patches are extracted with large overlapping regions which are controlled by the user specified strides $s$. However, registration in the peripheral regions of patches might be inaccurate due to the lack of sufficient surrounding spatial information. Thus, the patches are cropped to abandon the peripheral regions in the patch-fusion process.

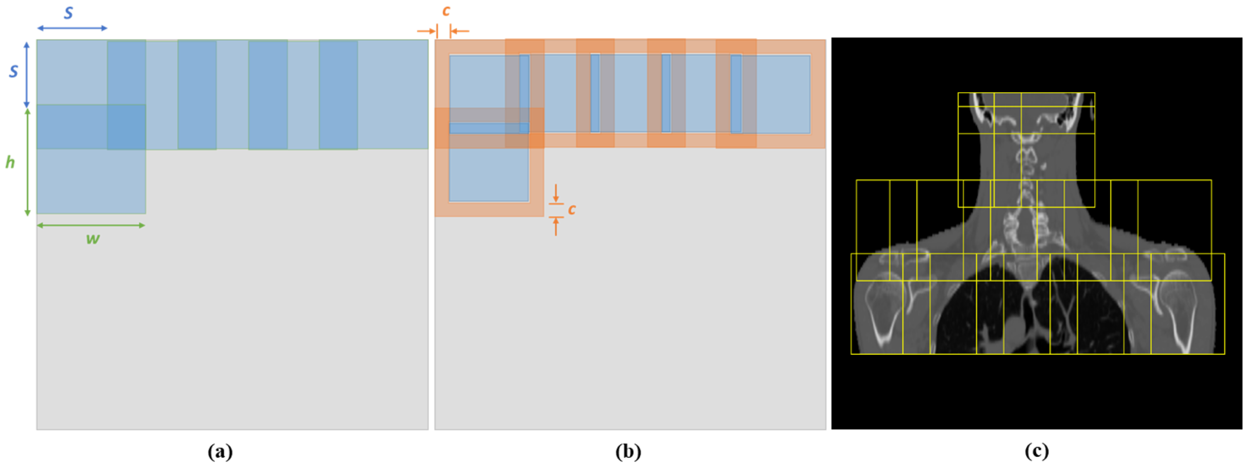


Figure 20: Illustration of patch processing with adjustable parameters stride $s$ and cropping $c$. (a) Patch extraction. (b) Patch fusion. (c) Patch extraction is restricted within the patient body. $h \times w$ is the patch size. The diagram is shown in 2D, but the actual implementation is in 3D.

After patch extraction, $P_m$ and $P_f$ are registered using the same algorithm used in whole volume linear registration for further local alignment. The pre-aligned patch $P_m'$, generated from $P_m$, is then deformably registered to $P_f$ using the patch-based ViT-Morph network. During patch fusion, each patch is cropped by cropping parameter $c$ to keep the central part of the patch for the final deformed image.

### 5.3.4 Patch-based deformable registration: ViT-Morph

The objective of the final stage is to predict a deformation vector field $\phi \in \mathbb{R}^{3\times d\times h\times w}$ that can warp the moving patch $P_m'$ to match the fixed patch $P_f$. Based on the framework of VoxelMorph, we adapted the UNet architecture and combine it with Vision Transformer (ViT) to construct the ViT-Morph network. The framework is shown in Figure 19 (b) and (c).

Specifically, a residual-UNet is adopted in the netork. It can take advantage of long skip connections like UNet and alleviate the gradient vanishing issue with residual connections. The overview of our architecture is described in Figure 19 (b). The implementation of ViT has been detailed described in previous work[204], [199], [206]. ViT is integrated as an encoder with the residual-UNet in feature map level. To be more specific, after the images $(P_f, P_m')$ are encoded into feature maps through residual-UNet decoders, the high-level feature maps are fed into ViT and extracted into vectorized patches $\{x_p^i \in \mathbb{R}^{P^3.C} | i = 1, \dots, N\}$, where $x_p^i$ denotes the $i^{th}$ vectorized patch, $P$ represents the patch size, $N = \frac{dhw}{P^3}$ is the number of patches, and $C$ denotes the channel size. The patches are then mapped into a latent D-dimensional space using a trainable linear projection. The patch positional information is preserved by adding specific position embeddings to the patch embeddings:

$$\boldsymbol{z_0} = \left[x_p^1\boldsymbol{E}; x_p^2\boldsymbol{E}; \dots . x_p^N\boldsymbol{E}\right] + \boldsymbol{E_{pos}}$$

where $\boldsymbol{E} \in \mathbb{R}^{(P^3.C)\times D}$ is the patch embedding projection and $\boldsymbol{E_{pos}} \in \mathbb{R}^{N\times D}$ is the position embedding.

Following the patch and position embeddings, the embedded patches are fed into the Transformer encoder, which consists of 12 layers of Multihead Self-Attention (MSA) and Multi-Layer Perceptron (MLP) blocks[211]. The output of the $\ell^{th}$ Transformer encoder can be expressed as following:

$$\boldsymbol{z_\ell'} = MSA\big(LN(\boldsymbol{z_{\ell-1}})\big) + \boldsymbol{z_{\ell-1}}$$

$$\boldsymbol{z}_{\boldsymbol{\ell}} = MLP\big(LN(\boldsymbol{z}'_{\boldsymbol{\ell}})\big) + \boldsymbol{z}'_{\boldsymbol{\ell}}$$

where $\boldsymbol{z}_{\boldsymbol{\ell}}$ represents encoded image representation and $LN(.)$ denotes the layer normalization, which is added before each MSA and MLP block.

Finally, the resulting sequence of hidden feature from ViT is reshaped and decoded with residual-UNet decoders to output the final deformation field $\phi$. STN is implemented to warp the moving patch $P'_m$ with $\phi$, and the dissimilarity between $P_f$ and $P'_m \circ \phi$ is then calculated as the similarity loss function for network training.

### 5.3.5 Loss function and training strategy

The simple intensity-based similarity loss in VoxelMorph could not be used effectively for inter-modality (CT-to-MR) registration. MI was introduced for the rigid alignment of multi-modal images. However, its application to deformable CT-to-MR registration comes with many difficulties[195]. Because MI is intrinsically a global measurement, it would result in inaccurate image registration with the loss of local accuracy. In our network, we employed MIND[195], a feature of 12 channels extracting the distinctive structure similarity from a local neighborhood, as the similarity metric for DIR, described further below. It has been demonstrated to have superior performance in MR-to-CT DIR tasks using conventional registration algorithms[212]. Following the similar implementation, the MIND features are calculated for fixed $P_f$ and deformed patches $P'_m \circ \phi$ during the training process. The MIND-based unsupervised loss $\mathcal{L}_{MIND}$ penalizes the differences between their MIND features and is defined as follows:

$$\mathcal{L}_{MIND}\big(P_f\, , P'_m \circ \phi, x\big) = \frac{1}{|R|} \sum_{s \in R} \left\| MIND\big(P_f, x, s\big) - MIND(P'_m \circ \phi, x, s) \right\|^2$$

where $\mathcal{L}_{MIND}\big(P_f\, , P'_m \circ \phi, x\big)$ represent the similarity loss between fixed $P_f$ and deformed patches $P'_m \circ \phi$ at voxel $x$, and $MIND(\cdot,\cdot,\cdot)$ denotes the MIND feature. Specifically, for voxel x in image patch $I$, its MIND feature in a local search region $R$ is defined as:

$$MIND(I,x,s) = \frac{1}{M}\exp\left(-\frac{D_p(I,x,x+s)}{V(I,x)}\right) \qquad s \in R$$

where M is a normalization constant to ensure the maximum value of $MIND(I,x,s)$ is 1, $D_p(I,x,x+s)$ denotes the patch distance, $V(I,x)$ represents the local variance estimation, and the spatial search region is set to be $|R| = 6$. To be more specific, $D_p(I,x,x+s)$ is the $L_2$ distance between two patches $P$ with the size of $(2p+1)^3$ centered at voxel $x$ and voxel $x+s$, defined as:

$$D_p(I,x,x+s) = \sum_{\boldsymbol{p}\in P}\big(I(x+\boldsymbol{p}) - I(x+s+\boldsymbol{p})\big)^2$$

And the variance estimate $V(I,x)$ is calculated based on the mean of the patch distances within a six-neighborhood $n \in N$:

$$V(I,x) = \frac{1}{6}\sum_{n\in N} D_p(I,x,x+n)$$

In order to ensure the generation of a reasonable deformation vector field $\phi$, a smoothness constraint is needed for regularization. In our study, the $l_2$-norm of first order gradient of $\phi$ was applied as the diffusion regularizer $\mathcal{L}_{smooth}$ as:

$$\mathcal{L}_{smooth}(\phi) = \sum_{\boldsymbol{p}\in\omega} \|\nabla\boldsymbol{u}\,(\boldsymbol{p})\|^2$$

where $\omega$ is the set of all voxels and $\boldsymbol{u}$ denotes the displacement field. Thus, the final loss function used for training the network is:

$$\mathcal{L}\big(P_f, P_m', \phi\big) = \mathcal{L}_{MIND}\big(P_f\,, P_m' \circ \phi\big) + \lambda\, \mathcal{L}_{smooth}(\phi)$$

where $\lambda$ is a tunable regularization parameter.

### 5.3.6 Data acquisition and preprocessing

MR and CT images of 42 head and neck patients who were treated on a 1.5 T MR-Linac (Unity; Elekta AB; Stockholm, Sweden) at The University of Texas MD Anderson Cancer Center were included in this study. Each patient had a pre-treatment CT simulation scan and daily T2-weighted MR scans. The number of treatment fractions ranged from 2 to 35, giving a total of 266 pairs of CT-MR and 237 pairs of MR-MR scans. The voxel intensities of MR images can vary significantly among different patients, scans, and machines. To reduce the intensity variations in MR images, the Z-score normalization[213] within the patient body was applied. The body contour for each MR image was generated in RayStation (version 11B; RaySearch Laboratories; Stockholm, Sweden), and the Z-score normalization was conducted only within the body contours.

After the whole volume registration in the first stage, all fixed and moving images were preprocessed to have an isotropic resolution of $1mm^3$ and spatial dimensions (420×420×300). In the following patch-based registration step, the stride was chosen to be s =96, 96, 80, and the overlapping patches were extracted with a size of (160×160×128), which can sufficiently cover one complete anatomical structure in HN region. 40 to 80 patches were extracted from each scan, depending on the patient's body size. In the patch fusion, we used the cropping parameter c = 20, 20, 16 to exclude the boundary region in the fusion process.

### 5.3.7 Evaluations and implementation

We selected 48 pairs of images (24 CT-MR and 24 MR-MR) from 6 patients as a testing dataset, and the remaining data were used as a training dataset. Seven manually labelled organs at risk (OARs) were used for evaluation, including brain stem, spinal cord, mandible, right and left parotid glands, right and left submandibular glands. These OARs were contoured separately on CT and T2-weigthed MR images by experienced radiation oncologists. After the DIR by our model, the deformed contours of the moving images were compared with those on the fixed images by calculating Dice similarity coefficient (DSC) and mean surface distance (MSD). Those labelled structures were only used for quantitative evaluation and not used for model training. Visual assessment was also performed by overlaying the deformed contours on those ground-truth contours in the fixed images.

To demonstrate the effectiveness of our method, we compared our Patch-RegNet with VoxelMorph-MIND and a clinical DIR tool in Monaco Treatment Planning System (version 5.4 Elekta AB; Stockholm, Sweden). VoxelMorph-MIND is the VoxelMorph network using MIND as the similarity metric and whole-volume global registration for pre-alignment. Input images were cropped and resized to 256×256×288 to fit the network. In addition, we also trained a ViT-Morph-MIND mode without using hierarchical registration framework for comparison. The ViT-Morph-MIND model was trained with whole-volume images, which have the same image size of 256×256×288 as that in VoxelMorph-MIND. The paired two-tailed t-test was included in our study to evaluate the statistical difference between our proposed method and other methods at a significance level of 0.05 (defined by a $p < 0.05$).

Our Patch-RegNet was implemented with the TensorFlow and trained on an NVIDIA QUADRO RTX 8000 for 120 epochs. We applied Adam optimizer[214] to train our model with a learning rate of 0.0001. The regularization parameter λ is set to 0.1 to achieve the best network performance based on our investigation. To improve the model performance during training process, we adopted rotation, random flipping, and non-linear transformation for data augmentation.

## 5.4 Results

### 5.4.1 Registration accuracy: inter-modality (CT-to-MR) registration

The DIR results between CT and daily MR images were quantitatively and qualitatively analyzed to evaluate the performance of our Patch-RegNet for inter-modality registration.

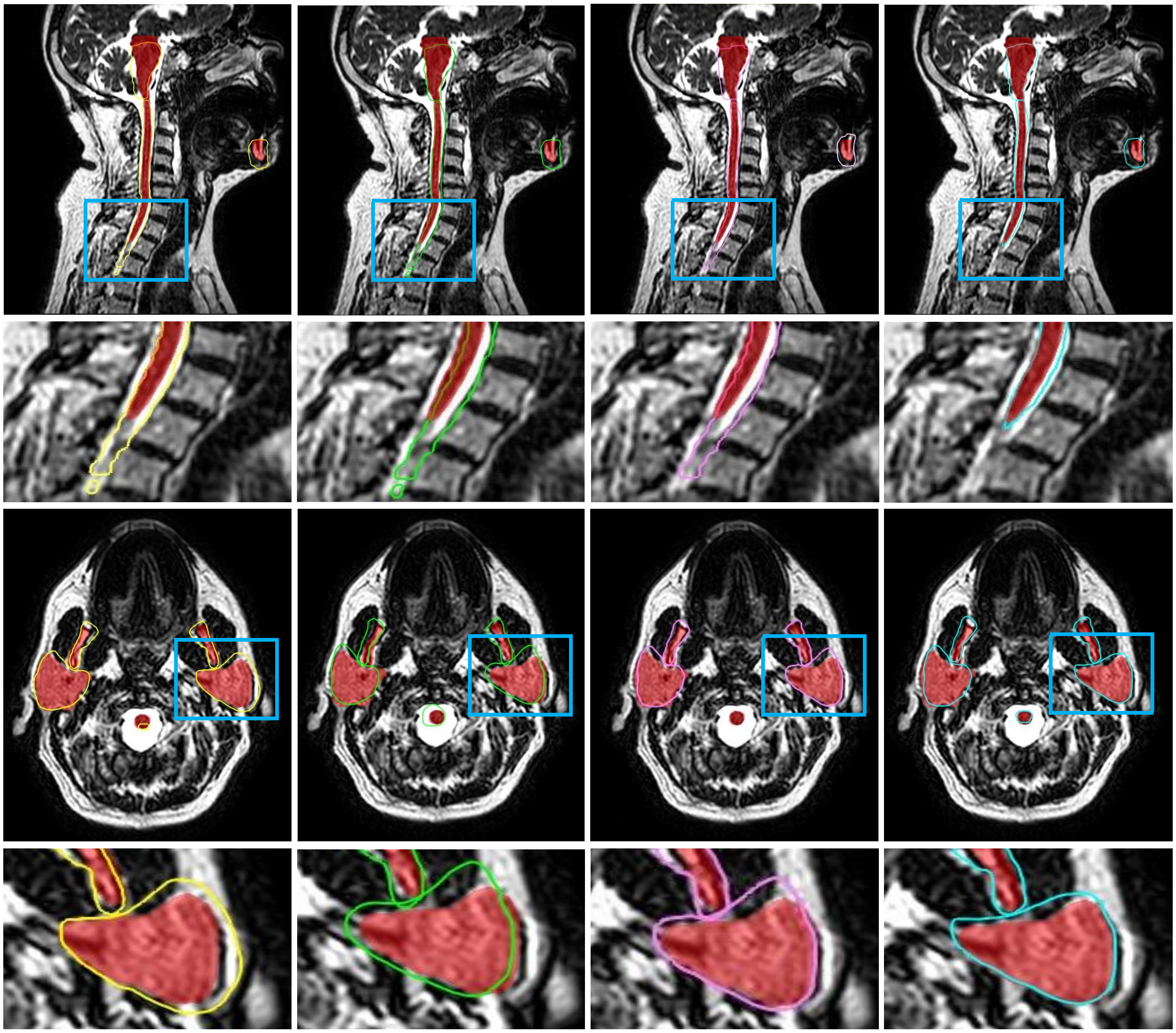


Figure 21: Qualitative evaluation results of different registration methods. The manual contours (red color-wash) are compared with the deformed contours from Monaco (yellow), VoxelMorph-MIND (green), ViT-Morph-MIND (purple), and our Patch-RegNet (blue) methods.

Examples of the CT-to-MR registration results are shown in Figure 21. The contours on the simulation CT images were propagated onto the daily MR images using different registration methods, including the traditional DIR tool in Monaco (a1, c1), VoxelMorph-MIND (a2, c2), ViT-Morph-MIND (a3, c3), and Patch-RegNet (a4, c4). The deformed contours were overlaid on the MR images and compared to the ground-truth contours shown in red color-wash for the evaluation of DIR accuracy. For better visualization, the insets (b1-b4, d1-d4) show the zoomed-in images of the deformed contours of specific structures (a1-a4, c1-c4) indicated by the dashed-line boxes.

As can be seen in Figure 21, the deformed contours by Monaco have a considerable amount of disagreement with the manual contours, while our Patch-RegNet achieves the most accurate registration results among the four approaches. As can be seen from the zoomed-in images (a1-a4, c1-c4), there are clear distinctions between our Patch-RegNet and other methods, especially for the alignment of spinal cord and parotid glands. In addition, the ViT-Morph-MIND achieves much-improved accuracy when compared with VoxelMorph-MIND, which can be shown from a2 vs. a3 and c2 vs. c3 in Figure 21.

Table 9. Quantitative comparison of different methods for inter-modality (CT-to-MR) registration. The average Dice similarity coefficient (DSC) (%) and mean surface distance (MSD) (mm) and their standard deviations are calculated over all 7 organs for all test patients. The bolded numbers denote the highest scores.

| **Methods** | **Rigid** | **Monaco** | **VoxelMorph** | **ViT-Morph** | **Patch-RegNet** |
|---|---|---|---|---|---|
| DSC | 0.61±0.16 | 0.69±0.11 | 0.72±0.10 | 0.73±0.09 | **0.76±0.05** |
| MSD (mm) | 3.1±1.8 | 2.6±1.1 | 2.2±0.7 | 2.1±0.7 | **1.9±0.5** |
| Times (s) | - | 60 | **0.03** | **0.03** | 5.6 |

DSC: Dice similarity coefficient; MSD: mean surface distance

Table 9 reports the average DSC and MSD over all organs calculated for different registration methods. Among them, our Path-RegNet achieved the best performance with the highest average DSC of 0.76 and lowest average MSD of 1.9 mm. Based on the DSC measurements, Our Patch-RegNet outperforms the VIT-Morph-MIND, VoxelMorph-MIND and the traditional DIR in Monaco by 4%, 6%, and 10%, respectively. The results demonstrate that our proposed Patch-RegNet achieves a statistically significant ($p<0.05$ using a two-tailed t-test) improvement over other methods for inter-modality (CT-to-MR) registration.

The boxplots of organ-specific results are shown in Figure 22 for detailed comparison. These results further demonstrate that our Patch-RegNet achieves improved registration performance for each organ under consideration. The improvement is more noticeable in organs/structures that are more likely to move from day to day due to setup, e.g. the mandible, because the local alignment in the hierarchical registration framework of Patch-RegNet can effectively correct this local displacement. In addition, Figure 22 shows that our Patch-RegNet has smaller variance and improved median performance compared to other methods, indicating improved robustness and reliability of Patch-RegNet. ViT-Morph-MIND also achieves superior performance than the VoxelMorph-MIND for each specific organ, demonstrating the effectiveness of the combination of Transformer and CNN for image registration. The average run-time for DIR process of each method is listed in Table 9 as well. All three deep-learning methods show much faster registration than the traditional method in Monaco.

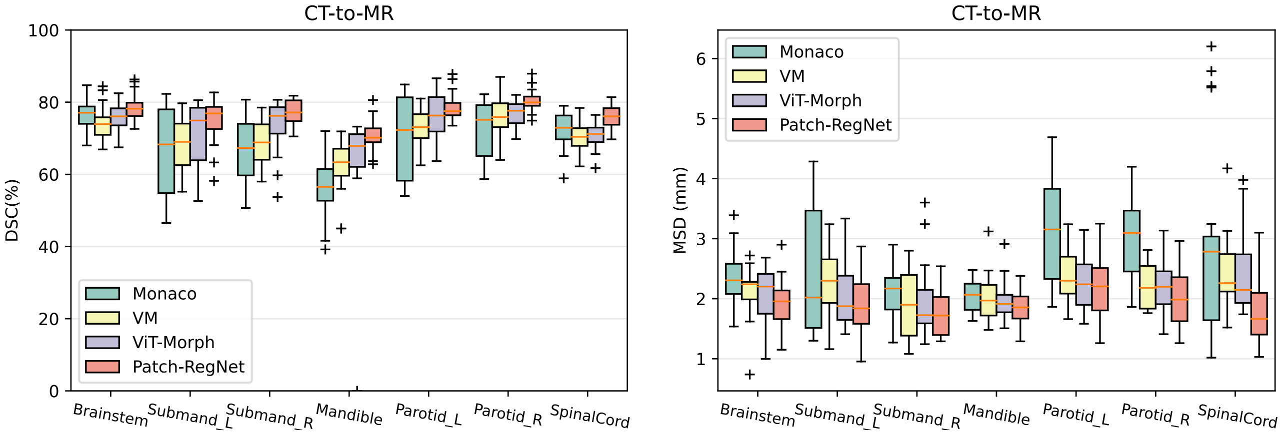


Figure 22: Quantitative evaluation results of different methods for inter-modality (CT-to-MR) registration. Boxplots showing DSC and MSD results for various anatomical structures using Monaco, VM (VoxelMorph-MIND), ViT-Morph (ViT-Morph-MIND), and our Patch-RegNet methods. DSC: Dice similarity coefficient. MSD: mean surface distance.

### 5.4.2 Registration accuracy: intra-modality (MR-to-MR) registration

The performance of our Patch-RegeNet for intra-modality (MR-to-MR) registration was also evaluated and compared with the traditional DIR tool in Monaco and VoxelMorph-MIND. Table 10 summarizes the quantitative results of the average DSC and MSD over all organs calculated for different registration methods. Similar to the results of inter-modality registration, our Patch-RegNet achieves the best average DSC of 0.86 and average MSD of 0.9mm. Compared with VoxelMorph-MIND and the traditional DIR tool in Monaco, Patch-RegNet improves the DSC accuracy by 4% and 7%, respectively. As shown in the quantitative results, our proposed Patch-RegNet achieves a statistically significant ($p<0.05$ using a two-tailed t-test) improvement over all other methods for intra-modality (MR-to-MR) registration. It indicates our Patch-RegNet method can work effectively not only for inter-modality (CT-to-MR) but also for intra-modality (MR-to-MR) registration.

Table 10. Quantitative comparison of different methods for intra-modality (MR-to-MR) registration. The average Dice similarity coefficient (DSC) (%) and mean surface distance (MSD) (mm)) and their standard deviations are calculated over all 7 organs for all test patients. The bolded numbers denote the best scores.

| **Methods** | **Rigid** | **Monaco** | **VoxelMorph** | **Patch-RegNet** |
|---|---|---|---|---|
| DSC (%) | 0.76±0.16 | 0.80±0.11 | 0.83±0.7 | **0.86±0.06** |
| MSD (mm) | 2.0±1.5 | 1.5±1.5 | 1.1±0.6 | **0.9±0.3** |
| Times (s) | - | 60 | **0.03** | 5.2 |

DSC: Dice similarity coefficient; MSD: mean surface distance

The detailed DSC and MSD results for each structure are presented as boxplots in Figure 23**.** Although Monaco is observed to achieve better registration results for some specific patients, its registration performance is unstable among different patients. On the contrary, our Patch-RegNet is much more robust for different cases, demonstrated by the smaller variance and improved median values in the boxplots. Therefore, overall performance of our Patch-RegNet is better than both the traditional DIR tool in Monaco and VoxelMorph-MIND. In addition, as shown in Table 9 and Table 10, the average run-times for DIR process of CT-to-MR and MR-to-MR are comparable, which are more than 10 times faster than the traditional method.

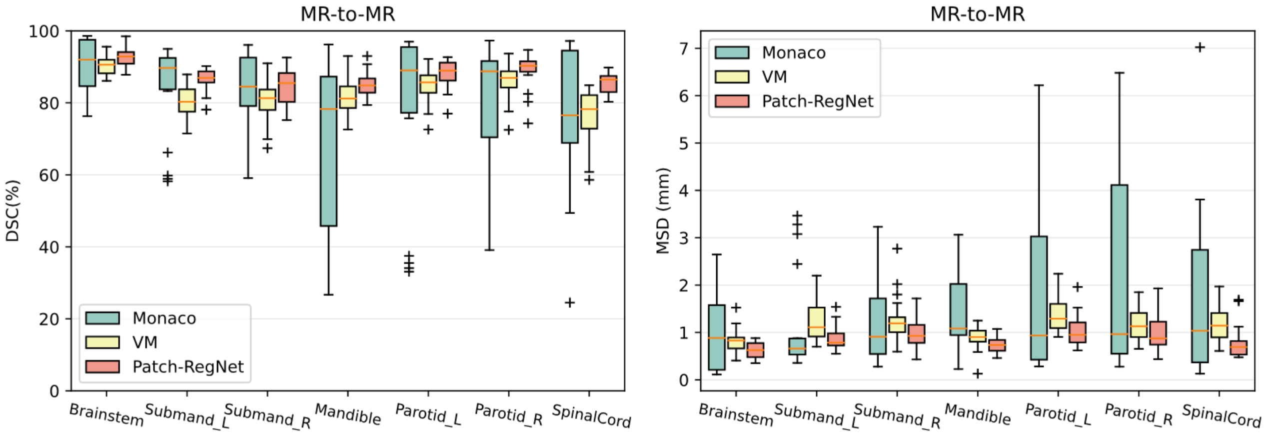


Figure 23: Quantitative evaluation results of different methods for intra-modality (MR-to-MR) registration. Boxplots showing DSC and MSD results for various anatomical structures using Monaco, VM (VoxelMorph-MIND), and our Patch-RegNet methods. DSC: Dice similarity coefficient. MSD: mean surface distance.

### 5.4.3 Regularization Analysis

We investigated the influence of the hyperparameter $\lambda$, which regularizes the smoothness of the generated DVFs, to our Patch-RegNet. The average DSC results for the test dataset for different values of the regularization parameter $\lambda$ are plotted in Figure 24. The results for both CT-MR and MR-MR registration show that our method is relatively robust to the choice of $\lambda$ values. The optimal value for the regularization parameter is demonstrated to be $\lambda$=1.0 for both inter- and intra-modality registration.

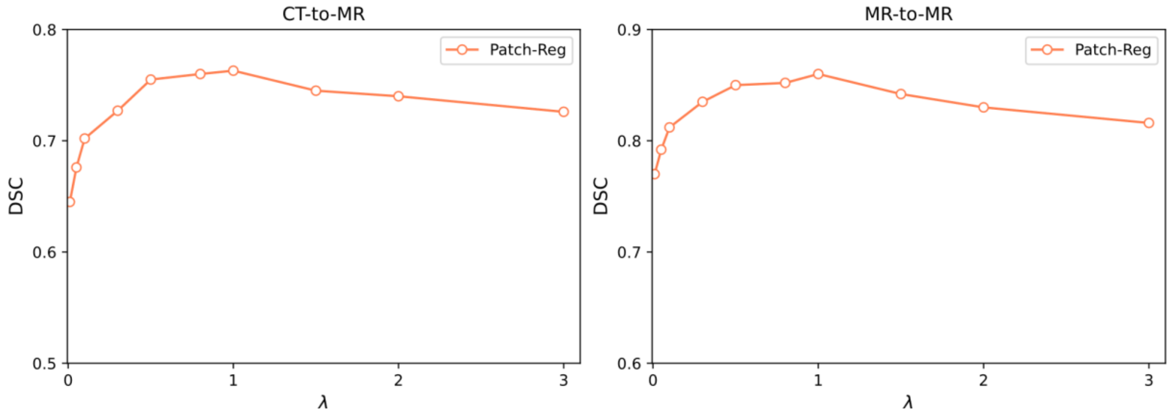


Figure 24: Average DSC results of test data for Patch-RegNet with varied regularization parameter λ

### 5.5 Discussion

In last two decades, ART has gained widespread popularity due to its ability to deliver more personalized and accurate treatment. Meanwhile, the advent of MR-Linacs has significantly advanced the field of ART by enabling the integration of MR imaging into radiotherapy with online plan adaptation. DIR plays a crucial role in ART to ensure accurate contour prorogation with daily anatomical changes so that treatment plan can be reoptimized in real time. In this study, we introduced a novel hierarchical framework for the implementation of efficient and precise CT-to-MR and MR-to-MR DIR. This framework features the utilization of patch-based registration, which can improve the local registration over a large anatomical region. The results of this study indicate that our proposed method significantly outperforms current DIR techniques in CT-MR and MR-to-MR registration accuracy. Moreover, the study demonstrates that the proposed Patch-RegNet achieves faster registration speed compared to the traditional method in Monaco, as presented in Table 9 and Table 10. Although our Patch-RegNet is slower than VoxelMorph-MIND due to multiple patch-based registrations, a total run time of 5-6 seconds is acceptable for most clinical applications and the improved registration accuracy is worth the cost in registration time.

In this work, we introduced a novel deep learning-based model, named ViT-Morph, for DIR by incorporating Vision Transformer (ViT) into conventional Convolutional Neural Network (CNN) networks. ViT-Morph leverages both the local features provided by CNNs and the long-range image relationships obtained through the self-attention mechanism in ViT. We expected the combination of these two models would provide a more comprehensive representation of the medical images, which will provide superior image registration performance. The result of comparison between the performance of ViT-Morph and VoxelMorph in CT-to-MR DIR demonstrated a marked improvement in performance of ViT-Morph compared to the CNN-based network, VoxelMorph. These results proved that the combination of ViT and CNN indeed enhanced the performance of inter-modality DIR.

One of the main challenges for deep learning-based registration methods is the pre-alignment of the moving and fixed images prior to the DIR process. It is common practice to apply a global rigid or affine registration to the moving and fixed images as an initial step for the following more complicated DIR process. However, this approach, which involves the application of rigid or affine registration to the entire image volumes alone, may not provide adequate alignment accuracy in some local regions for subsequent DIR, particularly for images that cover a large area and contain more than one anatomical site, such as HN images. Our study demonstrated that the hierarchical framework in Patch-RegNet can fundamentally address the local pre-alignment issue. This framework involves the extraction of patches from the whole image volumes and the application of patch-based local registration, which provides improved regional guidance for DL-based DIR by focusing the registration on local regions and specific anatomical sites. The results indicate a substantial improvement in CT-to-MR and MR-to-MR DIR through the use of our hierarchical framework. This framework can greatly benefit organs that are particularly susceptible to daily setup uncertainty during treatment, such as the mandible and spinal cord. Our Patch-RegNet has been compared to the same DL network ViT-Morph that did not use a hierarchical framework in Section 5.4.1. The comparison results highlighted the effectiveness of the hierarchical framework, demonstrating that the improved performance of Patch-RegNet is not solely attributed to the introduction of a new DL network, but also to the combination of the new network and the hierarchical framework. Table 1 provides further evidence, showing the impact of our hierarchical framework on the DIR performance. Additionally, the flexibility of Patch-RegNet allows for the replacement of ViT-Morph with a more advanced networks, thus providing the potential for enhanced registration performance.

We used the HN patients to demonstrate the effectiveness of our Patch-RegNet for both inter-modality (CT-to-MR) and intra-modality (MR-to-MR) DIR. However, this registration method can be extended for the registration tasks of other body sites and other image modalities. To facilitate this expansion, a continuous process of data accumulation and thorough curation of large image datasets will be necessary. Additionally, the selection of appropriate patch size is crucial when extending the application of Patch-RegNet to other body sites. In this study, we employed overlapping patches with dimensions of 160×160×128, ensuring adequate coverage for complete anatomic structures in the HN region. It is possible that varying patch sizes may be required for applications to other body sites.

Besides the development of the framework, the MIND descriptor was introduced as the similarity metric in Patch-RegNet for inter- and intra-modality DIR. The MIND extracts the distinctive features within a local neighborhood to create descriptor vectors, allowing the transformation of images from different modalities into a common domain. This enables straightforward similarity measurement through the use of metrics such as the sum of squared differences. Consequently, the incorporation of the MIND descriptor in Patch-

RegNet enhances its robustness against image noise and non-linear intensity variations, making it suitable for the registration of images acquired from various modalities. Furthermore, the similarity metric used in Patch-RegNet can be easily changed to alternative metrics to accommodate other specific registration tasks.

In this study, we evaluated the registration accuracy of our proposed deep learning-based image registration model, Patch-RegNet, through both qualitative and quantitative methods. Despite the promising results of our study, there are some limitations that should be noted. First, we only evaluated the performance of our proposed method for CT-to-MR and MR-to-MR registration. Thus, the applicability of Patch-RegNet to other modalities, such as PET and SPECT, remains to be investigated. Second, the study only focuses on head and neck cancer patients, and the extension of this work to other anatomical sites will need to be further assessed. Furthermore, the limited data used in our study may affect the generalizability of the results to other patient populations. Therefore, future studies with larger and more diverse datasets are necessary to validate the performance of our proposed method. Lastly, the performance of Patch-RegNet was only evaluated using a limited number of metrics, and further investigations are necessary to assess its clinical impact.

### 5.6 Conclusions

We developed the Patch-RegNet, a fully automated hierarchical registration framework, for inter-modality (CT-to-MR) and intra-modality (MR-to-MR) DIR. The hierarchical framework enables our Patch-RegNet to achieve markedly improved registration for large-volume images containing multiple anatomic sites. The patch-based ViT-Morph in our Patch-RegNet takes advantage of both CNN and ViT features of long-range spatial relationships. Additionally, MIND is incorporated as similarity metric to effectively train the network for multi-modality registrations. The Patch-RegNet is validated using HN cancer patient images and demonstrated superior results compared to the traditional DIR and other DL-based DIR methods.

# Chapter 6: Automated Synthetic CT Generation from MRI for MR-based Treatment Planning

This chapter is based upon the following article:

Zhao, Yao, He Wang, Cenji Yu, Laurence E. Court, Xin Wang, Qianxia Wang, Tinsu Pan, Yao Ding, Jack Phan, and Jinzhong Yang. "Compensation cycle consistent generative adversarial networks (Comp-GAN) for synthetic CT generation from MR scans with truncated anatomy." *Medical physics* (2023). doi:10.1002/mp.16246

## 6.1 Abstract

**Background**: MR scans used in radiotherapy can be partially truncated due to the limited field of view, affecting dose calculation accuracy in MR-based radiation treatment planning.

**Purpose**: We proposed a novel Compensation-cycleGAN (Comp-cycleGAN) by modifying the cycle-consistent Generative Adversarial Network (cycleGAN), to simultaneously create synthetic CT (sCT) images and compensate the missing anatomy from the truncated MR images.

**Methods**: CT and T1 MR images with complete anatomy of 79 head-and-neck patients were used for this study. The original MR images were manually cropped 10-25 mm off at the posterior head to simulate clinically truncated MR images. Fifteen patients were randomly chosen for testing and the rest of the patients were used for model training and validation. Both the truncated and original MR images were used in the Comp-cycleGAN training stage, which enables the model to compensate for the missing anatomy by learning the relationship between the truncation and known structures. After the model was trained, sCT images with complete anatomy can be generated by feeding only the truncated MR images into the model. In addition, the external body contours acquired from the CT images with full anatomy could be an optional input for the proposed method to leverage the additional information of the actual body shape for each test patient. The mean absolute error (MAE)

of Hounsfield units (HU), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM) were calculated between sCT and real CT images to quantify the overall sCT performance. To further evaluate the shape accuracy, we generated the external body contours for sCT and original MR images with full anatomy. The Dice similarity coefficient (DSC) and mean surface distance (MSD) were calculated between the body contours of sCT and original MR images for the truncation region to assess the anatomy compensation accuracy.

**Results**: The average MAE, PSNR, and SSIM calculated over test patients were 93.1 HU/91.3 HU, 26.5 dB/27.4 dB, and 0.94/0.94 for the proposed Comp-cycleGAN models trained without/with body-contour information, respectively. These results were comparable with those obtained from the cycleGAN model which is trained and tested on full-anatomy MR images, indicating the high quality of the sCT generated from truncated MR images by the proposed method. Within the truncated region, the mean DSC and MSD were 0.85/0.89 and 1.3 mm/0.7 mm for the proposed Comp-cycleGAN models trained without/with body contour information, demonstrating good performance in compensating the truncated anatomy.

**Conclusions**: We developed a novel Comp-cycleGAN model that can effectively create synthetic CT with complete anatomy compensation from truncated MR images, which could potentially benefit the MRI-based treatment planning.

### 6.2 Introduction

Magnetic resonance (MR) imaging has been widely used in radiation therapy to accurately delineate the tumor target and organs at risk since it has superior soft-tissue contrast compared to computed tomography (CT) imaging.[215], [216] In radiation treatment of head-and-neck cancers, several studies have shown that the application of MR images can substantially reduce the inter-observer variability of tumor delineation and improve the treatment outcomes.[216]–[219] However, CT images are still required for treatment planning because MR images cannot provide electron density maps for dose calculation.[220] Usually the MR images are fused to simulate CT through inter-modality registration and physicians draw contours based on the fusion MR view. However, this process has systematic errors and will introduce geometrical uncertainties for the contours.[221]–[223] Therefore, it is desirable to develop a treatment planning workflow using MR images only. The development of MR-based treatment planning could essentially benefit radiotherapy since it eliminates the inherent CT/MR registration error, reduces unnecessary radiation exposure for patients, and advances the efficiency of clinical workflow.[224] Additionally, the advent of MR-guided linear accelerator (MR-Linac) further drives the need for MR-based treatment planning and workflow[225], [226].

Since the electron density information acquired in CT images is necessary for dose calculation, estimating Hounsfield Unit (HU) and generating synthetic CT (sCT) from MR images become a key step to enable the MR-based treatment planning.[220] So far, a lot of different methods have been proposed to address this issue, which can be mainly divided into three categories: segmentation-based[44], [227], atlas-based[228]–[233], and learning-based method[234]–[236]. The segmentation-based methods generate sCT images by assigning uniform bulk densities to structures delineated on the MR images. However, these methods heavily rely on the accuracy of organ segmentation and fail to account for heterogeneity within each structure.[237] Besides, Atlas-based methods have been widely used for sCT generation which depends on the deformable image registration to correlate Atlas CT images with the real MR images. One limitation of these methods is their heavy computational burden, making them impractical for clinical implementation. Also, their performance is limited by the underlying deformable image registration algorithms.[233]

In recent years, learning-based methods, including traditional machine learning and deep learning methods, have gained significant attention for synthetic image generation. These methods exploit self-learning and self-optimizing strategies to learn the MR-CT mapping for sCT generation. Among them, deep learning methods using convolutional neural networks (CNNs) have been demonstrated to have more promising performance in sCT generation without the need of extracting handcrafted features.[45] For the deep learning methods, generally a model is trained to establish a nonlinear mapping from MR to CT domain based on a large databases of MR and CT pairs. Once the deep learning model has been trained, sCT images can be easily generated in a short amount of time by feeding the new MR image into the model. Han *et al*.[236] first developed a UNet architecture[238] to successfully generate two-dimensional (2D) sCT images of brain patients from T1-weighted MR images. To fully utilize the image information in all dimensions, Nie *et al.*[239] proposed a three-dimensional (3D) fully convolution neural network (FCN) to learn the complex translation mapping between MR and CT images. While the CNN method[240], [241] improved the efficiency and quality of sCT generation, its performance is affected by the voxel-wise accuracy of MR-CT registration and might suffer from blurriness during image synthesis[242]. To generate high-quality sCT images with less blurriness, the generative adversarial network (GAN)[243] which consists of a generator and a discriminator has been proposed. An adversarial loss function was also introduced to simultaneously optimize the generator and discriminator to improve the sCT image quality. Isola *et al*.[244] further extended the GAN model and proposed the conditional GAN, in which the output sCT image is constrained by the input MR image. Although the GAN-based methods have achieved great success in generating synthetic images[245], [105], [107], [103], [246], [106], [247], training a GAN model usually requires perfectly co-registered image pairs[104], which is especially challenging for inter-modal (MR-CT) images. Nevertheless, the cycle-consistent generative adversarial network (cycleGAN) model proposed by Zhu *et al*.[248] could be trained to generate synthetic images without the requirement of spatially aligned image pairs. With the incorporation of cycle-consistency loss, cycleGAN models trained with unpaired CT/MR images could even outperform GAN models trained with paired images in the aspect of the image quality of generated sCT [249]. Due to this advantage, cycleGAN has been applied for generating sCT images in radiotherapy planning for a variety of anatomical sites[45], [246], [250]–[255].

However, in MR-guided radiotherapy, in order to avoid the geometric distortion in peripheral regions and optimize the sequence for acquisition time and image quality, the field of view (FOV) of MR scans is often limited[256]. Thus, the MR images might be partially truncated in the peripheral regions, e.g. the posterior area of the head for the head-and-neck patients, as shown in Figure 25. Moreover, the truncation might also be observed in some cases where the regions of interest are distant from the posterior head region. The truncation in MR images usually does not affect the tumor delineation or diagnosis. However, this truncation poses significant challenge to MR-only treatment planning, since it might cause significant dose calculation errors due to the missing anatomy in the generated sCT images. To the best of our knowledge, current existing methods cannot accurately compensate for the missing structures in the truncated region during sCT generation.[45]

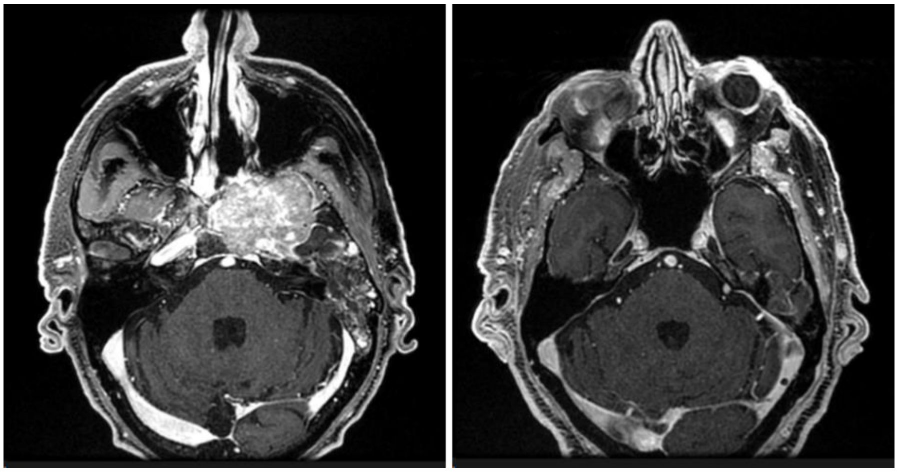


Figure 25: Illustration of the truncated MR images in the clinic. The truncation is observed in the posterior region of the head.

In this study, we aimed to generate sCT images with complete anatomy from truncated MR images to facilitate the MR-based radiotherapy workflow. We proposed a novel deep learning network named Compensation-cycleGAN (Comp-cycleGAN), which is based on cycleGAN, to enable anatomy compensation and sCT generation at the same time. Specifically, MR images with complete anatomy were collected and applied as training targets in our approach, making our Comp-cycleGAN model capture the complex relationship between the truncated regions and the given anatomical structures. The cycle-consistency loss in the traditional cycleGAN model was accordingly modified to constrain the process of sCT generation and anatomy compensation. We assessed the effectiveness of our proposed method with head-and-neck cancer patients with truncation at the posterior head on MR images.

## 6.3 Methods

### 6.3.1 Overview

The proposed novel algorithm, Comp-cycleGAN, which can generate sCT images with complete anatomy from truncated MRI images is composed of a training stage and a generation stage. The CT and MR images with complete anatomy of head-and-neck patients were collected. The MR images were then manually cropped 10-25mm off at the posterior head to simulate real truncated MR images in the clinic. For a given MR image, the manually truncated image, the original image, and paired CT image were all used for training. The CT image was used for learning a complex MR-to-CT mapping, while the original MR image was used as a target for the truncated image to learn the anatomy compensation. Furthermore, full anatomy is usually available on CT images. To leverage this additional information during sCT generation, the body contours with complete anatomy obtained from these CT images could be used to guide the anatomy compensation. Therefore, the body contours were applied as the optional inputs during our network development. The workflow schematic of our proposed method is illustrated in Figure 26.

In the training stage, the truncated MRI image and body contours(optional) were fed into the generator ($G_{MR-CT}$) to be translated into synthetic CT (sCT). The sCT was trained to be realistic and compensate for the missing anatomy in the truncated MR image. Then, another generator ($G_{CT-MR}$) was trained to translate the input sCT back into a synthetic MR image, which approximates the original MR image with complete anatomy. To improve training stability, the backward cycle was also trained, translating CT images into synthetic MR images and back into CT images. In addition, two discriminators ($D_{CT}, D_{MR}$) were trained to distinguish synthetic images from real images.

In the prediction stage, the sCT images with full anatomy can be generated by directly feeding the truncated MRI images and body contour (optional) into the trained generator ($G_{MR-CT}$).

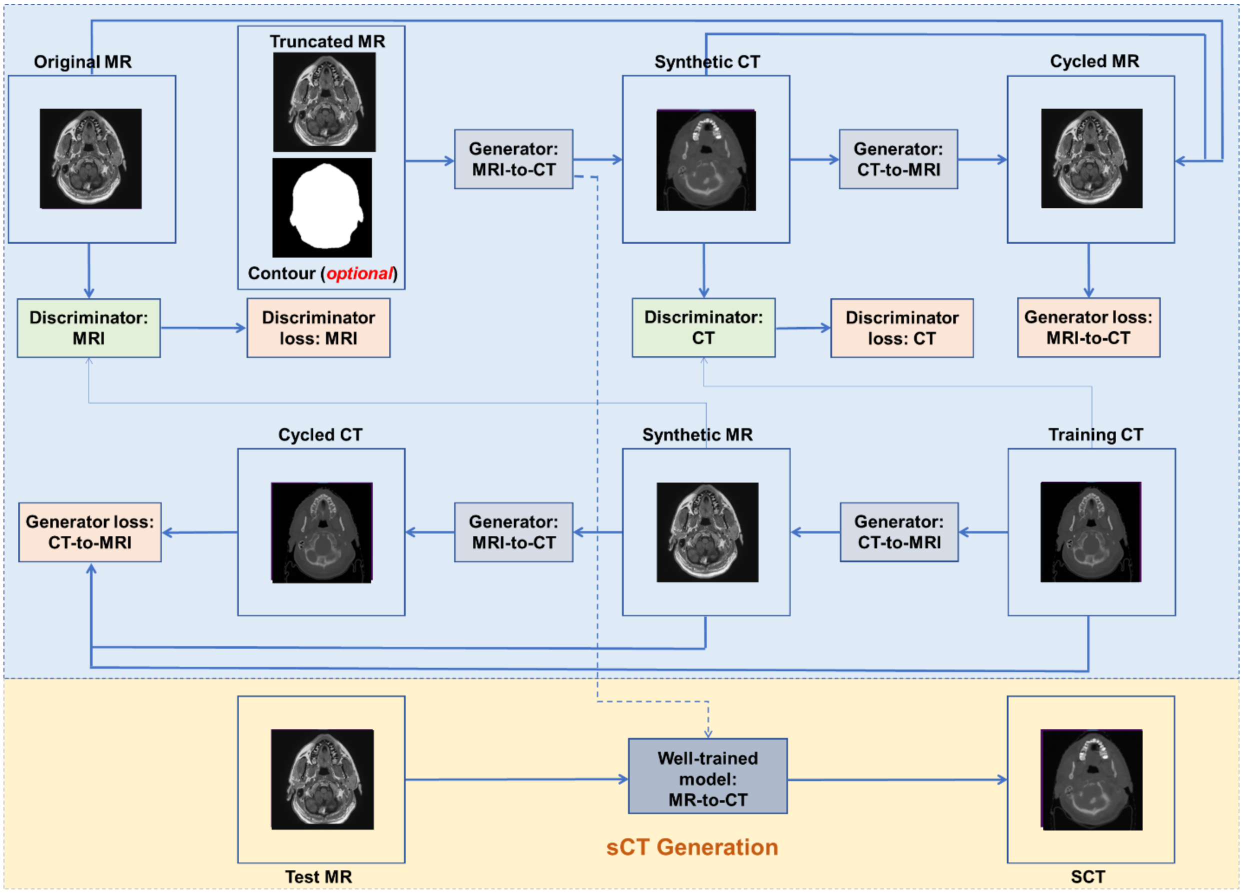


Figure 26: Schematic flow chart of the proposed model for truncated MR-based synthetic CT generation. **Blue part** shows the training stage of our proposed method, which consists of two generators and two discriminators. Both the truncated MR and original MR images are used for training. **Yellow part** shows the synthesizing stage, where a new testing MR image is fed into the well-trained generator to create the sCT image.

### 6.3.2 Data acquisition

MR and CT images of 79 head-and-neck patients who received external photon beam radiation treatment at The University of Texas MD Anderson Cancer Center were included in this study, under an institutional review board-approved protocol and waiver of informed consent (RCR03-0400). These patients were randomly selected without specific restrictions on sex, age, or histology type. The median patient age was 68 yr (range 30-84 yr) at the time of image acquisition; 60 patients (76%) were men and 19 patients (24%) were women. The treatment sites include oral cavity, pharynx, larynx, paranasal sinuses and nasal cavity, and salivary glands. The CT images were acquired on a Philips scanner (Big Bore) with 120 kVp, 434mA tube current, 887ms exposure time, $1.1719 \times 1.1719 \times 1.0\ mm^3$ resolution, and $512 \times 512$ reconstruction matrix. The corresponding MR images were acquired using a 1.5 T MR system (Magnetom Aera, Siemens Healthineers) with a pair of large flex 4 coils and built-in spine coils covering the head and neck region. The post-contrast T1-weighted MR imaging protocol included a 3D gradient dual-echo Dixon sequence with repetitive time = 7.11 ms, echo time1 = 2.39 ms, echo time2 = 4.77 ms, Pixel bandwidth = 405 Hz, flip angle = 10°, field of view = 256×256×240 mm3 and reconstructed voxel size = 1×1×1 mm3. The CT and MR images of each patient were acquired using the same setup with an interval of less than one week. Since both the original MR (with complete anatomy) and truncated MR images were required in the training stage, all the MR images were carefully selected to contain the complete anatomy.

### 6.3.3 Preprocessing

CT images were first rigidly registered to MR images for each patient using a commercial software Velocity AI v.3.0.1 (Varian Medical System, Atlanta, GA). A binary body mask was generated to remove the head mask and couch outside the patient body for each image. The voxel number of the region outside the body mask was set to 0 and -1024 HU for MR and CT images, respectively. The intensity of each MR image was normalized by Z-score standardization using only voxels in the body mask, and then scaled to a similar numeric range. All the MR and CT images were resampled to have the same voxel size of $1.1719 \times 1.1719 \times 1.0\ mm^3$. Then, each axial slice of the registered MR and CT pairs was cropped to a $256 \times 256$ 2D patch that kept the head and neck regions in the middle of the image. For the shoulder region, four $256 \times 256$ patches were cropped for each slice with 128 overlapping pixels.

In the clinic, the MR images might be partially truncated at the posterior head, with the depth ranging from 5-25mm. To simulate the truncation, the acquired MR images were manually cropped at the posterior head, making them close to the clinically truncated MR images. Furthermore, to acquire adequate truncated images for network training, all the slices within the head region (above the lower boundary of the mandible) were utilized. Each slice within the head region was randomly cropped 5-25 mm off at the posterior area based on the body mask. Specifically, the most posterior part of the head within each slice was identified by its corresponding body mask. Then, the cropping of 5-25 mm was randomly selected for each slice and all the pixels within that area were assigned a value of 0. The cropping can simulate all possible clinical MR truncations and can serve as data augmentation for network training.

### 6.3.4 Preprocessing

The overall architecture of the Comp-cycleGAN network is shown in Figure 26, consisting of four CNNs: two generators ($G_{MR-CT}, G_{CT-MR}$) and two discriminators ($D_{CT}, D_{MR}$). The model was trained based on 2D $256 \times 256$ patches of axial slices. Generators ($G_{MR-CT}, G_{CT-MR}$) were developed based on a hybrid of UNet architecture and residual blocks, residual-UNet. The residual-UNet comprises 9 residual units and its detailed architecture is demonstrated in Figure 27. Each residual unit contains two sets of one $3 \times 3$ convolutional layer followed by instance normalization and Rectified Linear Unit (ReLU) activation. Instead of using a pooling operation to downsample the feature map, a stride of 2 was used for the first convolutional layer in each residual unit during the decoding stage. On the other hand, in the decoding stage, an up-sampling operation was applied to recover the size of the feature map. Like original UNet architecture, the long skip connections were also used to copy low-level features to the corresponding high-level features. After the last level of the decoding stage, another $1 \times 1$ convolutional layer was used to output the generated synthetic image with the same size as the input image. In addition, the generators in our model were modified to take dual-channel inputs: the truncated MR images and the body contours (optional) which might be generated from corresponding CT images with complete anatomy.

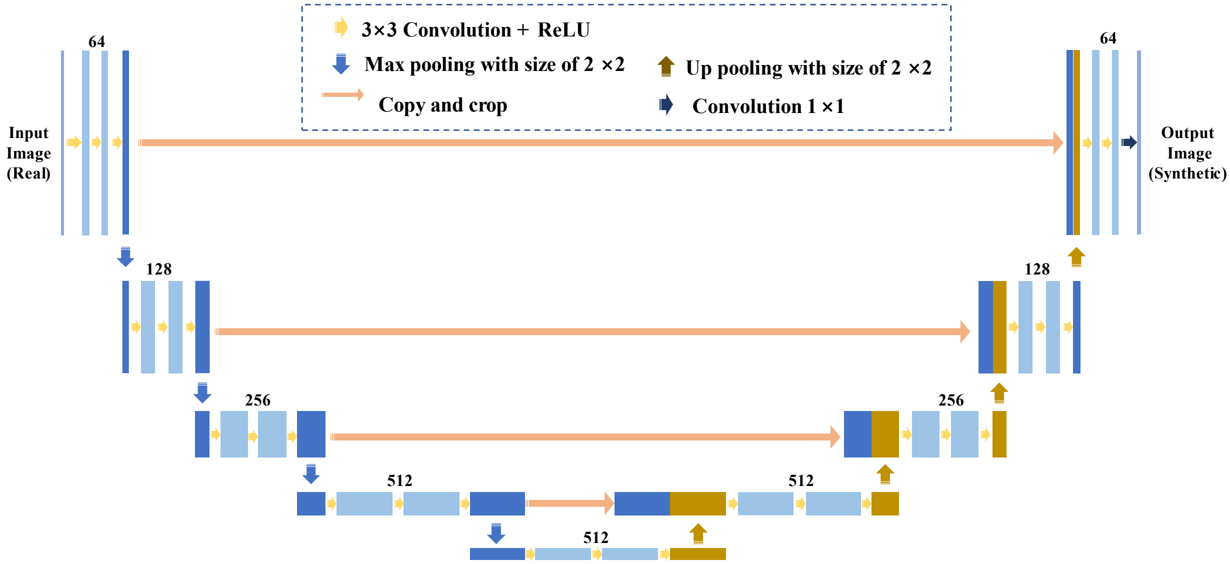


Figure 27: The details of the residual-UNet network: a combination of UNet architecture and residual blocks.

Discriminators ($D_{CT}, D_{MR}$) were trained to distinguish real images and synthetic images generated by generators ($G_{MR-CT}, G_{CT-MR}$), respectively. The discriminator is built by five successively $4 \times 4$ convolutional with the filter number of 64, 129, 256, 512, and 1 to generate a sub-regional estimation of the authenticity of images. All filters in the discriminator were followed by a Leak ReLU except for the last layer.

### 6.3.5 Network training

In the training stage, the generators ($G_{MR-CT}, G_{CT-MR}$) and discriminators ($D_{CT}, D_{MR}$) were trained simultaneously to achieve an optimal solution in an adversarial manner. Typically, the generators are trained to create realistic synthetic images to "fool" the discriminators, and the discriminators are trained to differentiate between synthetic and real images by decreasing the judging error of the discriminator network. In the original cycleGAN, the MR-CT texture translation is based on the constraints of adversarial loss and cycle-consistency loss. However, if the truncated MR images are used to train a cycleGAN, the model will not be able to generate stable solutions for volume changes due to the missing anatomy in MR images, which will deteriorate texture translation for sCT generation. To address these issues, we used three different images for each patient: the manually truncated MR image, the original MR image, and paired CT image, to develop our Comp-cycleGAN. By doing so, we expect to establish a stable solution for $G_{MR-CT}$ in anatomy compensation in the truncation region and achieve MR-CT translation at the same time.

Similar to traditional cycleGAN, our training loss also includes a cycle-consistency loss ($\mathcal{L}_{cycle}$), an adversarial loss ($\mathcal{L}_{adv}$), and an identity loss ($\mathcal{L}_{identity}$). In order to build the correlation between synthetic images and input images, the cycle-consistency loss is introduced in cycleGAN to constrain generators, satisfying: $G_{CT-MR}(G_{MR-CT}(I_{MR_{trunc}})) \approx I_{MR_{trunc}}$ and $G_{MR-CT}\,(G_{CT-MR}\,(I_{CT})) \approx I_{CT}$, where $I_{MR_{trunc}}$ and $I_{CT}$ represent the truncated MR image and the paired CT image, respectively. With this cycle-consistency loss, the cycleGAN model is expected to prevent generators from producing synthetic images that are irrelevant to the input images. However, as aforementioned, this cycle-consistency loss cannot ensure accurate compensation of the truncated regions during sCT generation when the truncated MR images are used for model training, as shown in Figure 28(a). Since the cycle-consistency loss forces the reconstructed synthetic MR images ($cycled\ MR$) to be identical to their inputs $I_{MR_{trunc}}$, the generated $sCT$ from $G_{MR-CT}(I_{MR_{trunc}})$ might be also truncated or randomly compensated.

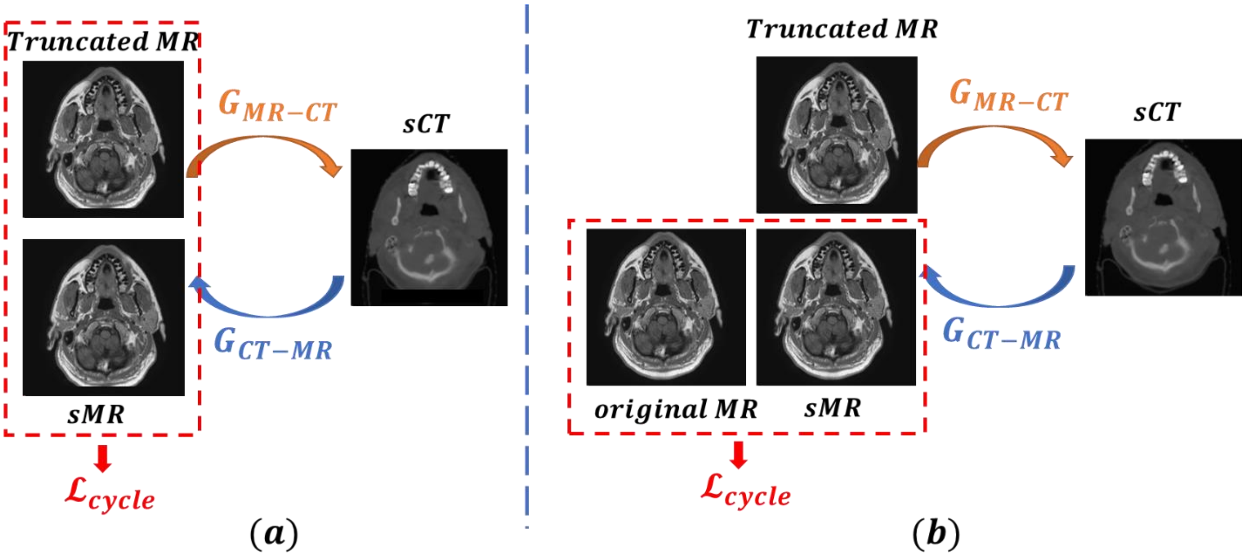


Figure 28: Illustration of the cycle-consistency loss calculation. (a) The cycle-consistency loss is calculated between synthetic MR (sMR) and truncated MR images. (b) The cycle-consistency loss is calculated between sMR and original MR images.

To overcome this issue, we modified the cycleGAN, especially the cycle-consistency loss, in our method. Instead of forcing $cycled\ MR$ to be identical to the input $I_{MR_{trunc}}$, the original MR image with full anatomy $I_{MR_{ori}}$ is utilized as the target for the reconstructed $cycled\ MR$ and penalized the cycle-consistency loss, as shown in Figure 28(b). The new cycle-consistency loss is defined below:

$$\mathcal{L}_{cycle}(G_{MR-CT,}\, G_{CT-MR})$$

$$= \left\| G_{CT-MR}\left(G_{MR-CT}(I_{MR_{trunc,}}\, C_{body})\right), I_{MR_{ori}} \right\|_1 + \left\| G_{MR-CT}\left(G_{CT-MR}(I_{CT})\right), I_{CT} \right\|_1$$

where $\|.\|_1$ denotes the calculation of $L_1$-norm distance, and the $C_{body}$ denotes the body contour of this patient, which might be acquired from the CT image and is an optional input for our model. By targeting the $cycled\ MR$ to the original MR image with full anatomy, the generator $G_{MR-CT}$ is trained to simultaneously learn MR to CT texture translation and anatomy compensation.

The adversarial loss $\mathcal{L}_{adv}$ is applied in cycleGAN to guarantee correct domain translation for synthetic images, which is defined as:

$$\mathcal{L}_{adv}(G_{MR-CT}\,, D_{CT}) = \left(1 - D_{CT}(G_{MR-CT}\left(I_{MR_{trunc}}, C_{body}\right))\right)^2 + \left(D_{CT}(I_{CT})\right)^2$$

$$\mathcal{L}_{adv}(G_{CT-MR}\,, D_{MR}) = (1 - D_{MR}(G_{CT-MR}(I_{CT})))^2 + \left(D_{MR}\left(I_{MR_{trunc}}\right)\right)^2$$

In the training process, the generators $(G_{MR-CT}, G_{CT-MR})$ are optimized to generate synthetic images close to the real ones, and discriminators$(D_{CT}, D_{MR})$ are trained to distinguish between the generated synthetic images and the real ones. However, if $I_{MR_{trunc}}$ is applied in the adversarial loss for $D_{MR}$, the generator $G_{CT-MR}$ will be optimized to generate truncated $cycled\ MR$ that is similar to $I_{MR_{trunc}}$ due to the adversarial relationship between the generator $G_{CT-MR}$ and discriminator $D_{MR}$. In this situation, $\mathcal{L}_{adv}$ will enforce the $cycled\ MR$ to be truncated, while $\mathcal{L}_{cycle}$ will regularize the $cycled\ MR$ to be identical to the original MR image with full anatomy. Consequently, the conflict between $\mathcal{L}_{adv}$ and $\mathcal{L}_{cycle}$ will impair the model training and image quality of sCT. In this study, we address this issue by modifying the discriminator loss in adversarial loss $\mathcal{L}_{adv}(G_{CT-MR}\,, D_{MR})$ as following:

$$\mathcal{L}'_{adv}(G_{CT-MR}\,, D_{MR}) = (1 - D_{MR}(G_{CT-MR}(I_{CT})))^2 + \left(D_{MR}\left(I_{MR_{ori}}\right)\right)^2$$

Thus, the objective of training discriminator $D_{MR}$ will be to decrease the judging error of the discriminator network and encourage generator $G_{CT-MR}$ to produce a synthetic image that has similar features with $I_{MR_{ori}}$.

An additional identity loss $\mathcal{L}_{identity}$ is also introduced to constrain the generator to an identity mapping if the input images are from the target domain:

$$\mathcal{L}_{adv}(G_{CT-MR}, G_{MR-CT}) = \|G_{MR-CT}(I_{CT}), I_{CT}\|_1 + \left\|G_{CT-MR}\left(I_{MR_{ori}}\right), I_{MR_{ori}}\right\|_1$$

Therefore, the final cost function to be optimized in our method is defined as:

$$\theta_{G,D}(G_{CT-MR}, G_{MR-CT}, D_{CT}, D_{MR})$$

$$= arg\min_{G}\max_{D}\,(\lambda_1\mathcal{L}_{cycle}\left(G_{MR-CT,}\ G_{CT-MR}\right) + \mathcal{L}_{adv}(G_{MR-CT}\ , D_{CT}) + \mathcal{L}'_{adv}(G_{CT-MR}\ , D_{MR})$$

$$+\lambda_2\mathcal{L}_{adv}(G_{CT-MR}, G_{MR-CT}))$$

where $\theta_{G,D}$ denotes the parameters for generators and discriminators. $\lambda_1, \lambda_2$ are the hyperparameters to control the relative weight of the losses, which were set as 10 and 5 based on our experiments to balance the variance uncertainty of each task.

### 6.3.6 Validation and evaluations

To evaluate the performance of our proposed model, 64 patients (80%) were randomly selected for training and validation, and the remaining 15 patients (20%) were used for model testing. We performed five-fold cross validation in our study, in which 52 patients and 13 patients were used for training and validation, respectively. The best model from all folds was selected to generate sCT for the independent test dataset. The 15 test patients were manually cropped 25 mm off to simulate the most severe truncation in clinical scenarios.

Note that the body contour is an optional input for our method. To investigate the impact of utilizing body contours, we separately trained two independent models: (1) modified cycleGAN without body contours (Comp-cycleGAN), and (2) modified cycleGAN with body contours (Comp-cycleGAN with contours). The body contours were created from the real CT images, and then rigidly registered to the corresponding MR images to be used as inputs for Comp-cycleGAN (contour) model. Therefore, for each tested patient, two sCT images were generated by feeding the truncated MR image into the two models (Comp-cycleGAN and Comp-cycleGAN (contour)). The paired CT image for each patient was deformably registered to the corresponding MR image using a commercial software Velocity AI v.3.0.1 (Varian Medical System, Atlanta, GA). The deformed CT image was visually checked in Velocity and then used as the ground-truth image to evaluate the image quality of sCT images in this study. The mean absolute error (MAE) of Hounsfield units (HU), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) between the sCT and ground-truth CT image were calculated within the patient body to quantify the comparison. Furthermore, to demonstrate the effectiveness of our method, we compare with two additional 2D models: (1) cycleGAN-trunc and (2) cycleGAN-full. The cycleGAN-turnc model is a traditional cycleGAN model trained and predicted using truncated MR images which were the same data used in our proposed method. The cycleGAN-full model is a cycleGAN model trained and predicted using the MR images with full anatomy, which were the original MR images without any truncation in the head region.

To further evaluate the model performance in terms of anatomy compensation, the external body contours were generated for sCT and original MR images with full anatomy. The Dice similarity coefficient (DSC) and mean surface distance (MSD) were calculated between body contours of sCT and original MR images only within the truncated region to assess the shape accuracy. The MAE of HU between the sCT and ground-truth CT image was also calculated only within the truncated region to evaluate the structural accuracy.

For the comparison, we also performed paired two-tailed t-tests between our proposed method and comparison methods at a significance level of 0.05 (defined by a $p < 0.05$) to evaluate the statistical difference.

## 6.4 Results

### 6.4.1 Imaging quality evaluation

The image quality of sCT images generated by different models was evaluated by comparing to the ground-truth CT images. Figure 29 shows the visual comparison, which contains the axial view of truncated MR (a1, b1), the original MR (a2, b2), the corresponding real CT (a3, b3), and sCT images generated from cycleGAN-trunc(a4, b4), the proposed Comp-cycleGAN (a5, b5), the proposed Comp-cycleGAN with body contours (a6, b6), and the cycleGAN-full (a7, b7). Figure 6 shows the visual comparison in the sagittal and coronal views. The external body contours of the original MR images (a2, b2) are shown on CT and sCT images as the green outlines in the Figure 30 for comparison. As shown in Figure 29 and Figure 30, the image quality of sCT images generated by both Comp-cycleGAN and Comp-cycleGAN (contour) is much better than cycleGAN-trunc with respect to anatomy compensation in the truncated regions and overall structural accuracy. Training with truncated MR images, sCT images generated by the cycleGAN-trunc model are inaccurate and deteriorated. Specifically, the shapes of these sCT images (a4, b4) are inconsistent with the actual patient shape (a2, b2) in the truncation region. Even in the untruncated region, the overall image quality is degraded and some generated structures in sCT are inconsistent with the input MR images. On the other hand, sCT images generated by both Comp-cycleGAN (a5, b5) and Comp-cycleGAN (contour) (a6, b6) demonstrate superior image quality and improved compensation of the truncated region. Additionally, the sCT images generated with truncated MR images by Comp-cycleGAN (a5, b5) and Comp-cycleGAN (contour) (a6, b6) are close to those generated with full-anatomy MR images by cycleGAN-full (a7, b7).

We also performed quantitative comparison of sCT images generated by different models. The detailed results of the average MAE, PSNR, and SSIM over 15 test patients are illustrated in Table 11. The results indicate high quality of the sCT images generated by our proposed method. Specifically, our Comp-cycleGAN models trained with/without body contours can achieve an average MAE of 93.3 and 95.1 HU, which is close to the cycleGAN-full model (MAE = 92.5 HU). Compared with cycleGAN-full, our Comp-cycleGAN and Comp-cycleGAN model (contour) show comparable image quality of generated sCT images, and there are no statistically significant differences $(p > 0.05)$ for all evaluation metrics except the PSNR for only Comp-cycleGAN. In contrast, cycleGAN-trunc has the worst quantitative results (MAE = 147.6), and our proposed Comp-cycleGAN models have statistically significant improvement ($p < 0.05$) over cycleGAN-trunc in all metrics.

Table 11. Quantitative comparison of image quality of sCT generated by different models.

| Methods | MAE(HU) | PSNR (dB) | SSIM |
|---|---|---|---|
| cycleGAN-trunc | 147.6±14.3 | 22.5±1.1 | 0.86±0.03 |
| Comp-cycleGAN | 93.1±11.4 | 26.5±1.0 | 0.94±0.02 |
| Comp-cycleGAN(contour) | 91.3±10.9 | 27.4±1.0 | 0.94±0.01 |
| cycleGAN-full | 90.5±10.7 | 27.9±0.9 | 0.95±0.01 |

MAE: mean absolute error; PSNR: peak signal-to-noise ratio; SSIM: structural similarity; HU: Hounsfield units; cycleGAN-trunc: the traditional cycleGAN model trained and tested with truncated MR images; Comp-cycleGAN: the modified cycleGAN model trained and tested with truncated MR images without body contours as input; Comp-cycleGAN(contour): the modified cycleGAN model trained and tested with truncated MR images with body contours as input; cycleGAN-full: the traditional cycleGAN model trained and predicted using the original MR images with full anatomy

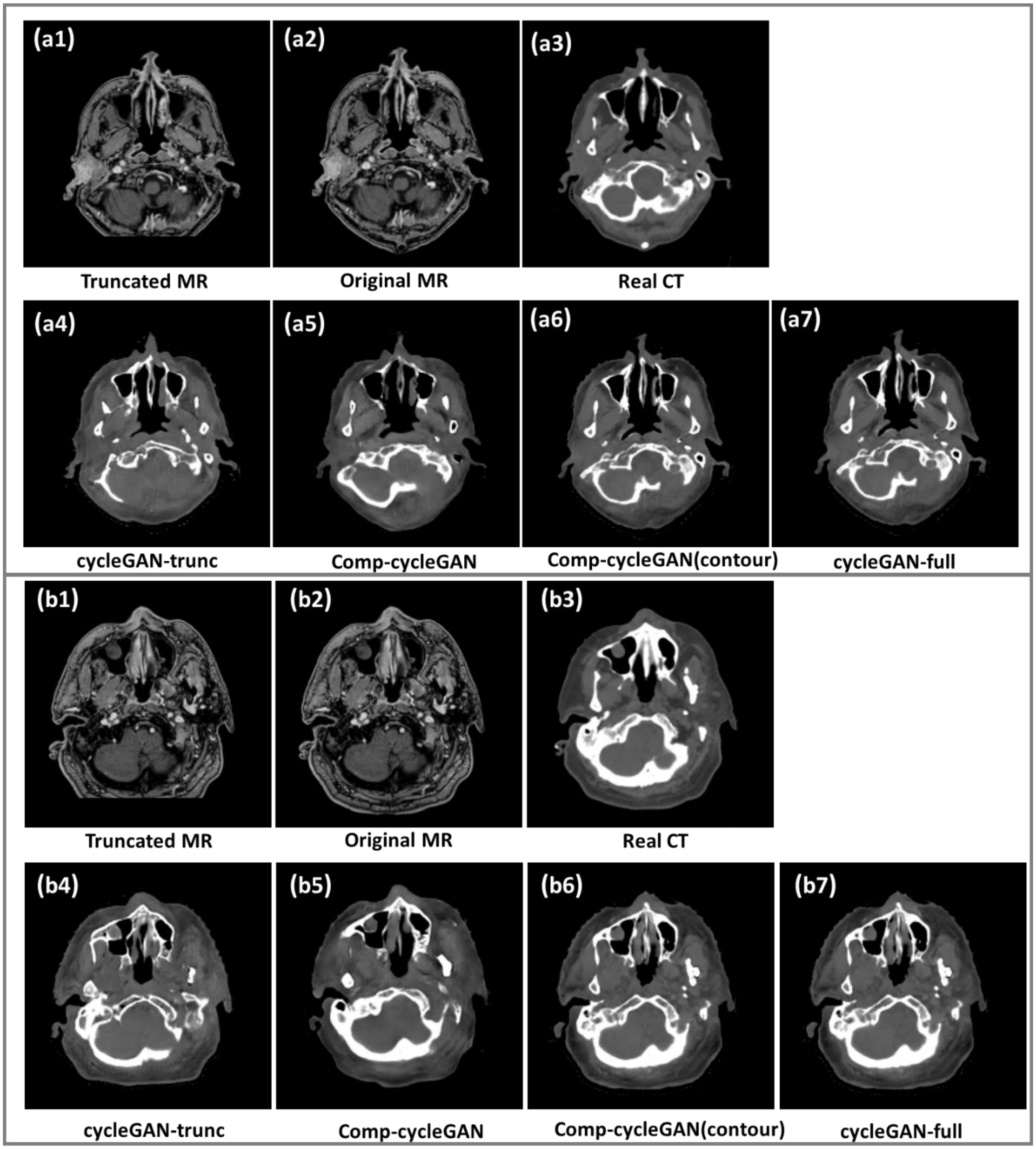


Figure 29: Comparison of actual computed tomography (CT) and synthetic CT (sCT) images generated by different models. Two examples (a, b) are included for illustration. (a1, b1) and (a2, b2) are the axial view of truncated MR images and original MR images with full anatomy. (a3, b3) are the corresponding real CT images for comparison. The sCT images

generated by different models are shown as (a4, b4) for cycleGAN-trunc model, (a5, b5) for Comp-cycleGAN model, (a6, b6) or Comp-cycleGAN (contour) model, and (a7, b7) for cycleGAN-full model.

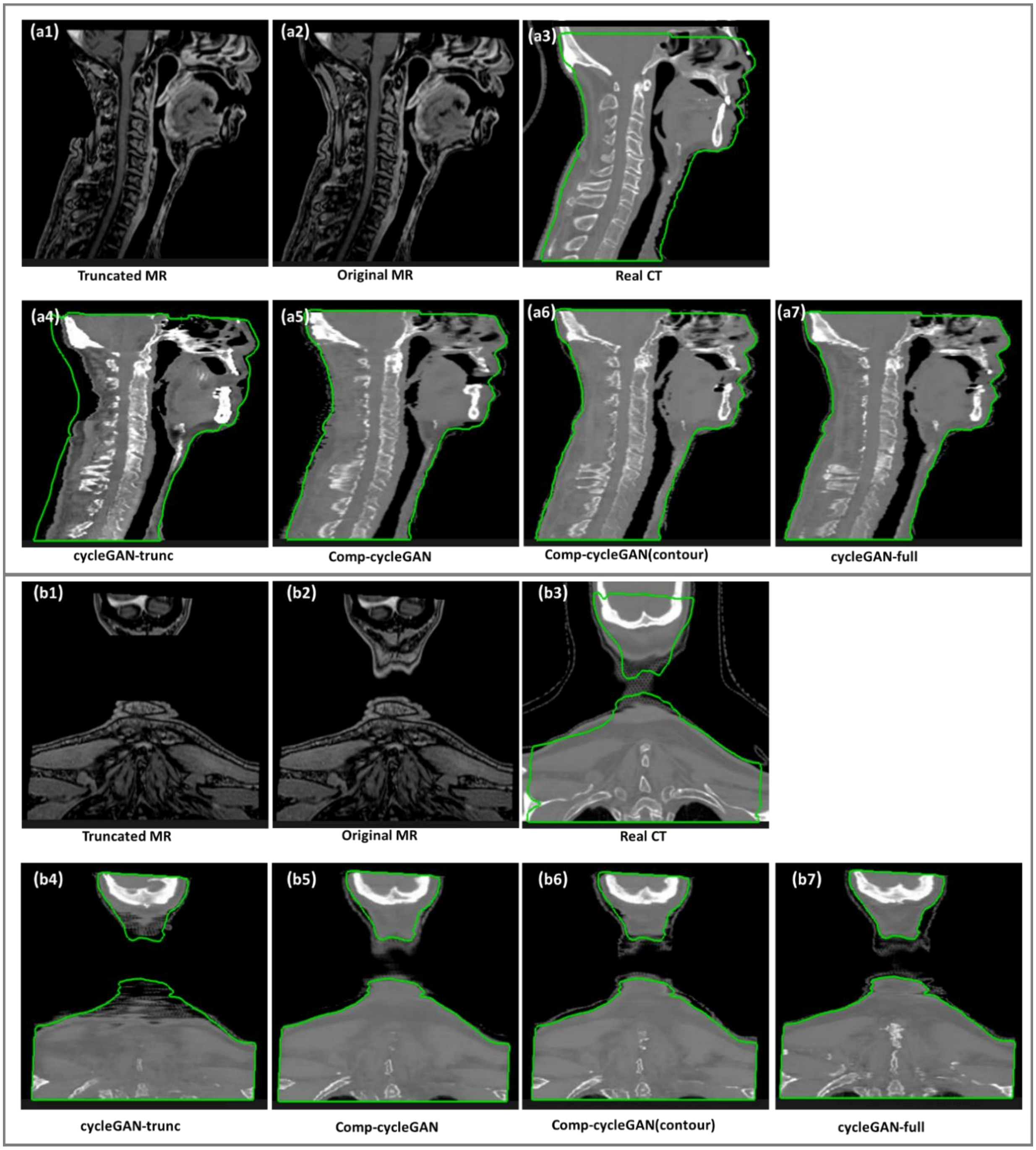


Figure 30: Comparison of actual computed tomography (CT) and synthetic CT (sCT) images generated by different models. Two examples (a, b) are included for illustration. (a1, b1) and

(a2, b2) are the sagittal and coronal views of truncated MR images and original MR images with full anatomy, respectively. (a3, b3) are the corresponding real CT images for comparison. The sCT images generated by different models are shown as (a4, b4) for cycleGAN-trunc model, (a5, b5) for Comp-cycleGAN model, (a6, b6) or Comp-cycleGAN (contour) model, and (a7, b7) for cycleGAN-full model. Green outlines are the external body contours of the original MR images and shown on all CT and sCT images. (This figure is best viewed in the online version)

### 6.4.2 Imaging quality evaluation

Figure 31 illustrates the visual inspection of the sCT images in the truncated region. The truncated MR images (a5, c5) and the original MR images (a6, c6) are shown together with the sCT images generated by cycleGAN-trunc models (a1, c1), Comp-cycleGAN (a2, c2), Comp-cycleGAN contour (a3, c3), and cycleGAN-full (a4, c4), respectively. The insets (b1-b4, d1-d4) show the zoomed-in images of the truncated regions of sCT images (a1-a4, c1-c4) indicated by the dashed-line boxes, to highlight the differences in anatomy compensation by using different methods. The external body contours are generated for the original MR images (a6, c6) and shown on all sCT images as the green outlines in the figure for comparison. As the cycleGAN-full was trained and tested using original MR images with full anatomy, the generated sCT images (a4, c4) have the same shape as the original MR images (a6, c6).

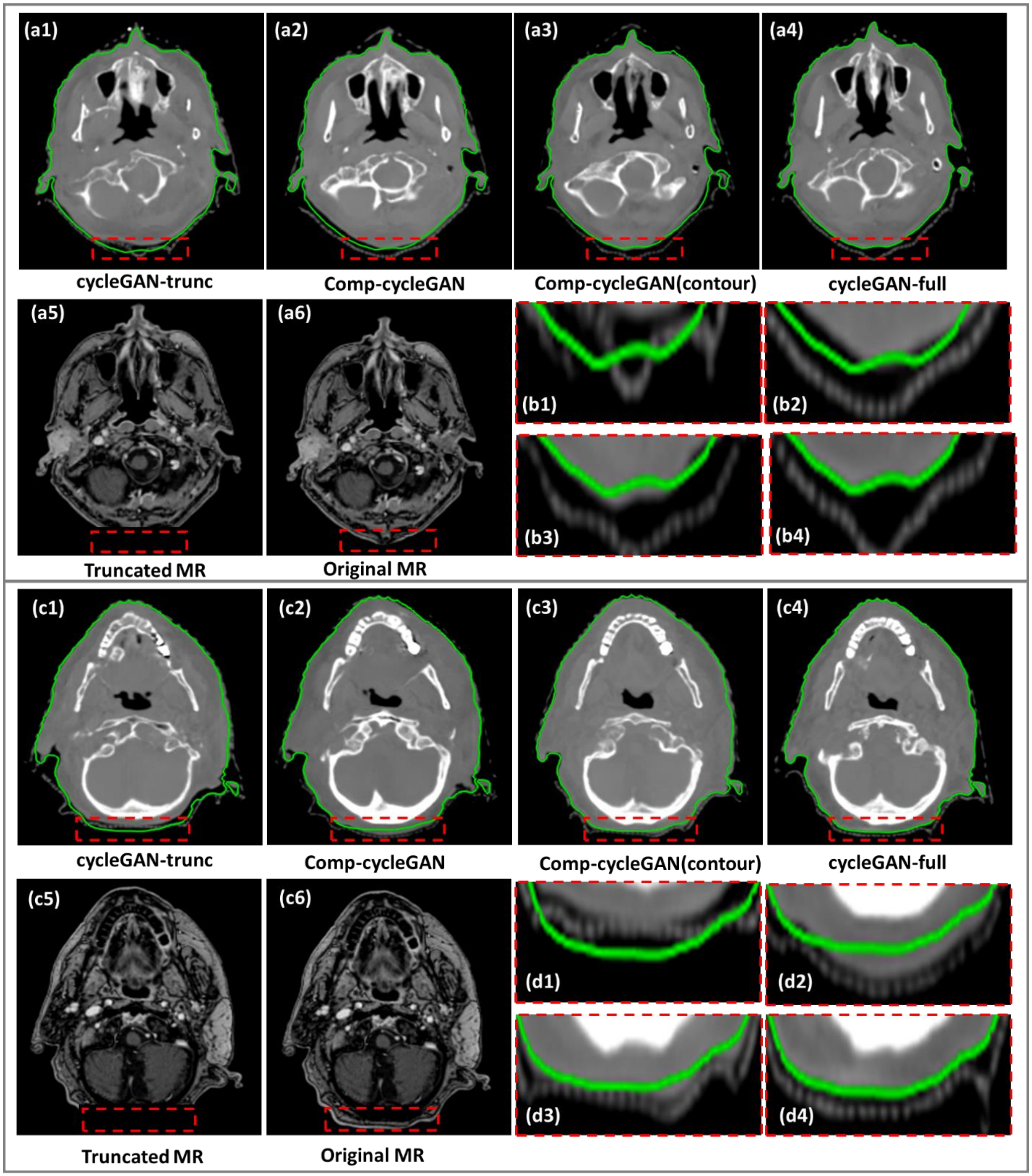


Figure 31: Comparison of anatomy compensation in the truncated regions of sCT images generated by different models. The axial view of truncated MR and original MR images are shown as (a5, c5) and (a6, c6), respectively. The first and second rows show the sCT images generated by cycleGAN-trunc (a1, c1), the Comp-cycleGAN model (a2, c2), the

Comp-cycleGAN(contour) model (a3, c3), and the cycleGAN-full model (a4, c4). The insets (b1-b4, d1-d4) show the zoomed-in images of truncation regions in sCT images (a1-a4, c1-c4), outlined by the red boxes. (This figure is best viewed in the online version)

By using the cycleGAN-trunc model, the generated sCT images cannot accurately compensate for the truncated regions (b1, d1) and there are notable differences when compared to the cyleGAN-full model (b4, d4). On the contrary, anatomy compensation quality in sCT images generated by Comp-cycleGAN (b2, d2) and Comp-cycleGAN contour (b3, d3) outperforms cycleGAN-full (b1, d1), and are similar to the structures in truncated regions of sCT images generated by cycleGAN-full (b4, d4). The sCT images in the truncated region (b2, b3, d2, d3) are close to the real body outlines in the original MR images. In addition, with the use of body contours, the shape and anatomical structures of sCT images in the truncated regions (b3, d3) are more similar to those generated by cycleGAN-full (b4, d4), and have a notable improvement over the Comp-cycleGAN without using contours (b2, d2).

The quantitative comparison of anatomy compensation in the truncated region has also been performed over 15 test patients by calculating MSD and DSC to quantify the shape accuracy and MAE to indicate texture accuracy. The quantitative results are summarized in Table 12. The DSC and MSD were not calculated for cycleGAN-full because the external body outlines of the generated sCTs are identical to those of the original MR images. Among these models, the sCT images generated by cycleGAN-trunc have the worst accuracy with an average MAE of 219.8 HU, MSD of 3.9 mm, and DSC of 0.62 in the truncated regions. Instead, both the proposed Comp-cycleGAN and Comp-cycleGAN (contour) achieve statistically significant improvement ($p < 0.05$) over cycleGAN-tunc in all evaluation metrics. The MAE comparison results show that sCT images generated by Comp-cycleGAN with/without contours have comparable image quality as those generated from cycleGAN-full in the truncated region, and the differences were not significant ($p > 0.05$). Furthermore, the Comp-cycleGAN (contour) shows superior performance to Comp-cycleGAN in anatomy compensation accuracy, which is also consistent with the visual comparison. The Comp-cycleGAN (contour) is significantly better than Comp-cycleGAN based on the evaluation metrics of MSD and DSC ($p < 0.05$).

Table 12. Quantitative comparison between different models for anatomy compensation. All the evaluation metrics are calculated within the truncated regions.

| Methods | MAE(HU) | MSD (mm) | DSC |
|---|---|---|---|
| cycleGAN-trunc | 219.8±34.3 | 3.9±1.1 | 0.62±0.08 |
| Comp-cycleGAN | 65.6±15.1 | 1.3±0.5 | 0.85±0.03 |
| Comp-cycleGAN(contour) | 62.1±13.7 | 0.7±0.3 | 0.89±0.02 |
| cycleGAN-full | 59.3±9.2 | - | - |

MAE: mean absolute error; MSD: mean surface distance, DSC: Dice similarity coefficient; HU: Hounsfield units.

### 6.4.3 Effect of the extent of truncation

The 15 test patients were manually cropped 25 mm off within the head region to simulate the most severely truncated cases during the model evaluation. To further evaluate the performance of our proposed method on truncated MR images with various extent, we also manually cropped the 15 test patients with 10 mm, 15 mm, 20 mm, and 25 mm off. Our proposed Comp-cycleGAN and Comp-cycleGAN (contour) were tested on those different cases and evaluated by calculating MAE, MSD, and DSC. Note that the MSD and DSC are calculated in the truncated region only. The quantitative results for Comp-cycleGAN and Comp-cycleGAN (contour) are listed in Table 13and Table 14, respectively. The results demonstrate that our proposed methods are robust for different truncation cases. For different truncation levels, the MAE results calculated within the whole volume of the patient body demonstrate the consistent image quality of the generated sCT images. With the decrease of the truncation level from 25mm to 10mm, the MSD results reduce from 1.3mm to 0.8 mm, 0.7 mm to 0.6 m, and DSC results increase from 0.85 to 0.90, 0.90 to 0.92 for Comp-cycleGAN and Comp-cycleGAN (contour) models. It indicates that our models have improved performance in anatomy compensation for cases with smaller truncation.

Table 13. Quantitative evaluation of Comp-cycleGAN for different truncation cases:10mm, 15mm, 20mm, 25mm. The metric, *MAE Whole*, is calculated within the whole volume of the patient body. The other evaluation metrics are only calculated within the corresponding truncated regions

| Truncation (mm) | MAE (HU) Whole | MAE (HU) Truncated | MSD (mm) | DSC |
|---|---|---|---|---|
| 10 | 93.0±10.6 | 61.1±12.4 | 0.8 ±0.2 | 0.90±0.02 |
| 15 | 92.7±10.8 | 61.6±13.2 | 0.9±0.4 | 0.86±0.03 |
| 20 | 93.2±11.3 | 63.7±14.0 | 1.0±0.5 | 0.86±0.03 |
| 25 | 93.1±11.4 | 65.6±15.1 | 1.3±0.5 | 0.85±0.03 |

MAE: mean absolute error; MSD: mean surface distance, DSC: Dice similarity coefficient; HU: Hounsfield units.

Table 14**.** Quantitative evaluation of Comp-cycleGAN for different truncation cases:10mm, 15mm, 20mm, 25mm. The metric, *MAE Whole*, is calculated within the whole volume of the patient body. The other evaluation metrics are only calculated within the corresponding truncated regions

| Truncation (mm) | MAE (HU) |  | MSD (mm) | DSC (%) |
|---|---|---|---|---|
|  | Whole | Truncated |  |  |
| 10 | 91.2±10.8 | 58.8±13.0 | 0.6±0.2 | 0.92±0.02 |
| 15 | 90.8±10.5 | 58.6±12.6 | 0.6±0.3 | 0.91±0.02 |
| 20 | 91.4±11.2 | 61.0±13.2 | 0.7±0.2 | 0.90±0.03 |
| 25 | 91.3±10.9 | 62.1±13.7 | 0.7±0.3 | 0.90±0.02 |

MAE: mean absolute error; MSD: mean surface distance, DSC: Dice similarity coefficient; HU: Hounsfield units.

### 6.4.4 Dosimetric evaluation

The dosimetric accuracy of using the sCT images for planning was evaluated by comparing them to clinical treatment plans. The study included sCT images of five HNC patients, contaning the tumor sites of larynx, base of tonuge, and neck. The prescribed dose for the primary tumor varied between 66 and 70 Gy.

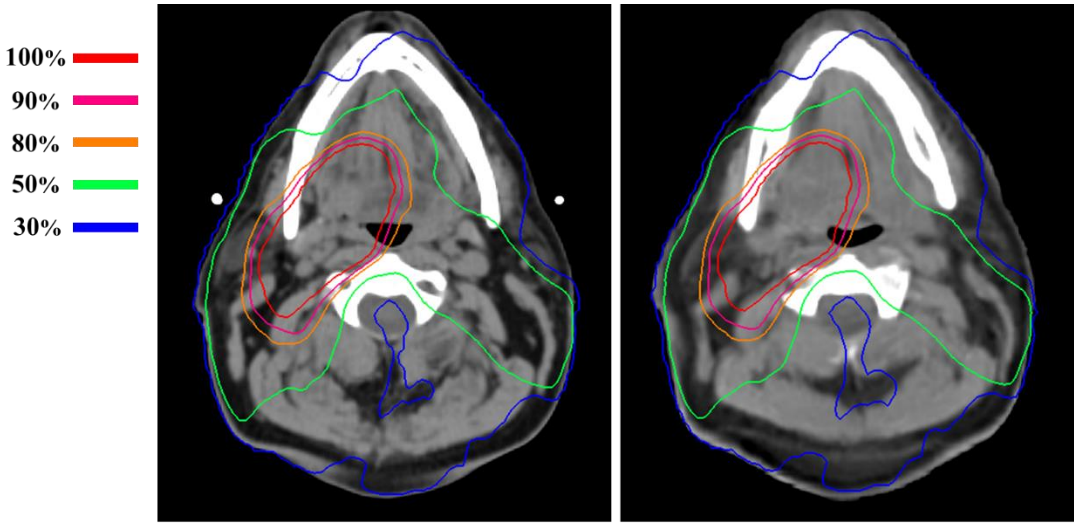
For each patient, the clinical VMAT plan on the planning CT image was transferred to its corresponding synthetic CT image, and the planning dose distribution was re-computed without any optimization. The synthetic CT used the same imaging system to covert the CT number to electron density as the original CT. The Dose-volume histograms (DVHs) for plan on original CT and synthetic CT were analyzed for target and OARs. Specifically, the DVH parameters of mean dose ($D_{mean}$), $D_{95\%}$, and $D_{5\%}$ were calculated for CTV (7000cGy), CTV (5016cGy), right submandibular gland, mandible, spinal cord, and brainstem. $D_{95\%}$, and $D_{5\%}$ refer to the minimum dose delivered to 95% and 5% of the regions of interest, respectively. The mean and standard deviation of the percent differences between DVH parameters in original CT and synthetic CT images in each target and OARs are summarized in Table 15. The average value for the DVH parameters of all targets and OARs is 0.4±0.2 %, demonstrating a high dosimetric accuracy of the synthetic CT images.

Table 15. The mean difference percentage between the DVH parameters calculated on planning CT and synthetic CT on all test patients. The parameters of $D_{mean}$, $D_{95\%}$, and $D_{5\%}$ are included for evaluation. SMG: submandibular gland.

| | **$D_{mean}$ (%)** | **$D_{95\%}$ (%)** | **$D_{5\%}$ (%)** |
|---|---|---|---|
| **CTV (7000cGy)** | 0.6±0.2 | 0.6±0.1 | 0.4±0.2 |
| **CTV (5016cGy)** | 0.4±0.1 | 0.5±0.2 | 0.3±0.1 |
| **Right SMG** | 0.4±0.2 | 0.6±0.3 | 0.4±0.1 |
| **Mandible** | 0.3±0.2 | 0.5±0.3 | 0.4±0.1 |
| **Spinal cord** | 0.3±0.2 | 0.4±0.1 | 0.3±0.1 |
| **Brainstem** | 0.2±0.1 | 0.4±0.1 | 0.2±0.1 |

For visual inspection, treatment plan of one patient was calculated using both original CT and synthetic CT and shown in Figure 32. The resulting isodose lines shown on both images were compared and found to be very close. The DVH plot comparison shown in Figure 33 also demonstrates the consistency of plans calculated from original CT and synthetic CT.

The results demonstrate that the synthetic CT images generated through our proposed DL model are of high quality and can be reliably used in place of real CT images for radiation therapy planning.



Planning CT

Synthetic CT

Figure 32: Dose comparison between planning CT and synthetic CT for a same treatment planning. The isodose lines are shown for comparison.

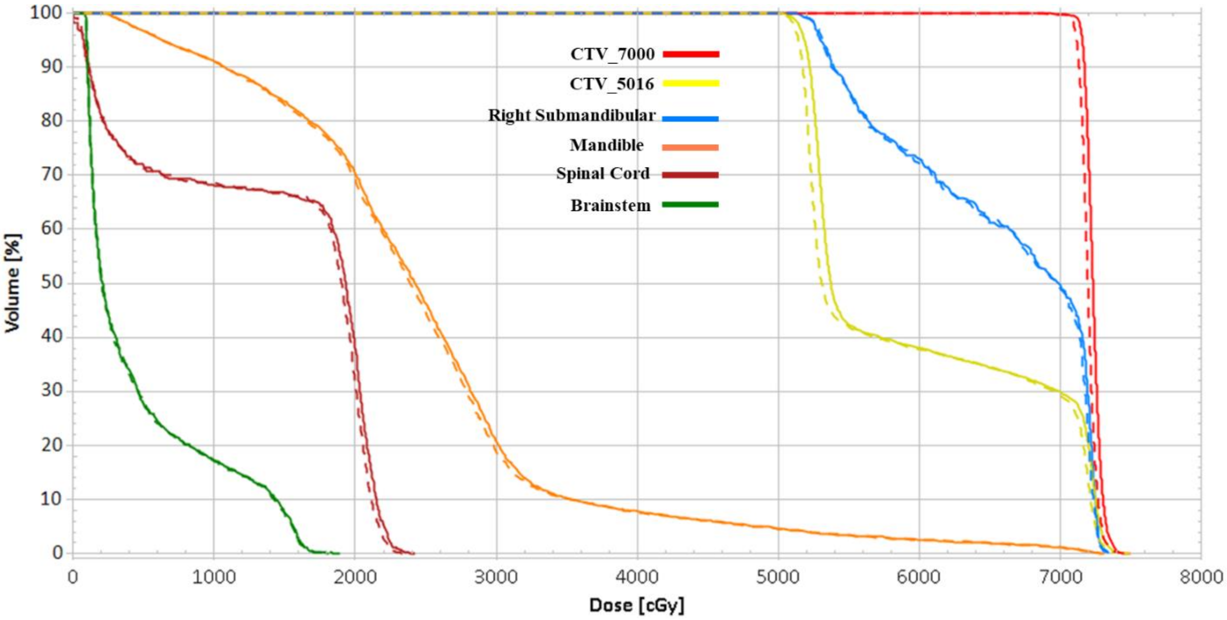


Figure 33: Comparison between DVH plots of planning CT (solid lines) and synthetic CT (dashed lines). The target and organs at risk include CTV (7000cGy), CTV (5016 cGy), right submandibular gland, mandible, spinal cord, and brainstem.

### 6.5 Discussion

In this study, we proposed a novel method by modifying the traditional cycleGAN model to effectively generate sCT images with complete anatomy from truncated MR images. We aimed at training our Comp-cycleGAN model to effectively capture the relationship between the truncated area and non-truncated anatomical structures during the sCT generation process. Based on this relationship, our Comp-cycleGAN can predict the truncation area for unseen cases. The proposed method innovatively incorporated both truncated and full-anatomy MR images into the model and adjusted the loss functions to address the truncation issue. Additionally, the residual-UNet was developed as the generator in our methods to better capture the connection between input MR and sCT images. With the integration of UNet with residual blocks, the network can take advantage of strengths from both UNet features and residual learning: (1) the UNet architecture enables the effective combination of low-level information and high-level features; (2) the residual connections facilitate information propagation and alleviate the degradation problem during network training. In this work, we evaluated our method on 15 independent test patients with great success in both sCT generation and missing anatomy compensation. All the deep learning models were implemented in TensorFlow (v2.2.0) and were trained on an NVIDIA Tesla GPU (V100) with 32GB of memory. Training took around 120-150 hours for each model. The sCT generation took about 5-10 seconds for each patient, making the model practical for clinical implementation.

From our experiment, the traditional cycleGAN model was not able to generate accurate sCT images with full anatomy from truncated MR images. This is because cycleGAN is developed based on cycle-consistency constraints, which means the generators are trained to ensure the reconstructed MR (cycled MR) to be identical to the input MR images. Since the input MR images are truncated, the generators in cycleGAN will perform unstably by applying random compensation and cropping to the truncated regions to guarantee the consistency between cycled MR and input MR images. The accurate anatomy compensation thus cannot be learned by the generator in the cycleGAN model. Moreover, due to the lack of constraints for compensation/cropping in the cycled process, the generators would perform unstably for sCT image generation. This might further substantially decrease the sCT image quality (shown in Table 11), and even lead to inconsistent anatomy structures in sCT compared to input MR images (shown in Figure 29 and Figure 30). Consequently, the decreased image quality (an average MAE of 147.6 HU) and missing anatomy in sCT images might remarkably affect dose calculation in MR-based radiotherapy planning.

By introducing the original MR images with full anatomy to the model training process, our method can significantly outperform the traditional cycleGAN model in terms of the sCT image quality and the accuracy of anatomy compensation in the truncated region. In our work, the MR images with full anatomy were collected and used together with the truncated MR images in the training process. The generator was trained to learn the anatomy compensation for the truncated region by targeting the cycled MR to the original MR images with full anatomy. As the datasets used in our study were the head-and-neck MR images, the truncation at the posterior head shares some similar shapes and anatomical structures (e.g. skin and skull). During the model training, the original MR images with full anatomy were used as the ground truth for the cycled MR images to constrain the anatomy compensation, one example being that the actual posterior head usually is rounded shape. Our Comp-cycleGAN not only can translate MR to CT domain, but also can compensate the truncated region based on the known anatomical information on the truncated MR image. However, as observed in Figure 31, the generated sCT image in the truncated region may not be perfectly identical to the actual anatomy in the sCT image from the cycleGAN-full model. This is mainly because the trained Comp-cycleGAN model would compensate the truncated region based on the anatomical features it learned from the training data. Thus, its capability to compensate the missing anatomy in sCT might be limited as the anatomical structures can substantially vary from patient to patient. Although the generated sCT images in the truncation region by our method might be imperfect, they were sufficiently close to the real anatomy of patients based on both quantitative and qualitative evaluation. In addition, one clear advantage of our method is that it does not require complete-anatomy MR images for sCT generation in the prediction stage. The original MR images with full anatomy were only used during the training stage. Once the model has been trained, it can directly generate sCT images from truncated MR images.

The body contours are optional input for our method, and we also trained the Comp-cycleGAN (contour) model to leverage the additional contour information in this work. In the clinic, CT images covering the complete anatomy might be acquired for patients from the treatment simulation process. The body contours of the CT images contain information about the actual shape of each patient. In MR-guided adaptive radiotherapy, daily MR images are acquired for adaptive planning and dose calculation. After rigid registration, the simulation CT and MR images for the same patient should have similar body outline with minor variations due to daily setup uncertainty. In this scenario, the acquired body contours can be used as guidance for anatomy compensation during sCT generation from daily MR images, and our Comp-cycleGAN (contour) model was trained to take advantage of the supplementary information from the body contours. The results in Table 11 and Table 12 show our Comp-cycleGAN (contour) has superior performance compared to Comp-cycleGAN, demonstrating the effectiveness of utilizing the contour information for anatomy compensation. However, if the patient setup is extremely different between the CT and MR images or CT images are unavailable for MR-only based treatment planning, there will be no body contours for the Comp-cycleGAN (contour) model. In this condition, the Comp-cycleGAN (contour) model cannot be used, but the Comp-cycleGAN model can still be used to generate sCT image and compensate the truncated region because it does not rely on the contour information during training and prediction.

To demonstrate the effectiveness of our method, we also compared it to the cycleGAN-full model, a regular cycleGAN model trained and tested using original MR images with complete anatomy. The sCT images generated from cycleGAN-full had an average MAE of 90.5 HU, PSNR of 27.9 dB, and SSIM of 0.95, which were comparable to the results published in recent deep learning studies of head-and-neck sCT generation.[45], [107], [240], [246], [252], [257] The quantitative results demonstrated that sCT images generated from truncated MR images by our methods were comparable to those generated from complete-anatomy MR images by the cycleGAN full model. Furthermore, a dosimetric evaluation of our synthetic CT generation was conducted and the results showed a similar level of dosimetric accuracy as reported in previous studies [107], [240], [246], [257]. This confirms the high quality of our synthetic CT images for use in treatment planning.

Furthermore, in this study, we used the truncated MR images of head-and-neck patients to demonstrate the effectiveness of our method for simultaneous anatomy compensation and sCT generation. However, this method can be used for any other anatomical sites where the MR images might be truncated. The other common case would be pelvis and prostate patients. Their MR images are usually truncated at the peripheral regions because of the limited field of view of MR scans to avoid geometric distortion and optimize the sequence for acquisition time and image quality. The major challenge for the pelvis and prostate sites is the non-rigid body shape that might result in the inferior performance of the Comp-cycleGAN model in compensating the missing anatomy. From our observation, the Comp-cycleGAN can be trained to predict missing data of a rigid body shape like head very well. Non-rigid body shape like the abdomen is less predictable in general, and more training data are needed for Comp-cycleGAN to achieve reasonable prediction.

All CT and MR images used to train and test the Comp-cycleGAN model were from our institution. The trained Comp-cycleGAN model might not be directly applied to MR images obtained from other institutions. The main reason lies in that the voxel intensities of MR images do not correspond to specific physical meaning because they depend on the combination of tissue properties and hardware-specific settings.[258] Although the Z-score standardization could be used to reduce the intensity variability[259], there is still a lack of consistency of voxel intensity, which prevents the direct application of the trained model across the different institutions. Transfer learning is usually needed to train a new model for new institution data.

One limitation of the proposed method is that the model might not perform robustly if the input MR images are severely truncated (e.g. larger than 3cm). This is mainly because the truncated regions might contain some critical structures which could not be inferred from the given anatomy in the truncated images. In this scenario, our proposed method could still generate sCT images and compensate the missing anatomy based on the learning of anatomical features from the patient population in the training dataset. However, the accuracy and robustness of our method would be considerably reduced. Another limitation of this work is that our evaluation metrics (MAE, PSNR, and SSIM) are affected by the accuracy of registration between CT and MR images. Even though training our model does not require registered MR-CT pairs, the misalignment between CT and MR images will affect our evaluation accuracy. While deformable image registration was applied to alleviate this issue, inter-modality image registration is still a challenging problem, especially for head-and-neck patients where the neck flexion can be quite different. Furthermore, the capability of our proposed method to compensate the missing anatomy is learned from the structural features in training data. In the future study, we plan to collect more patients to train and test our method. We plan to include more patients with increased variations in shapes and weights in our study, improving the diversity of structural features in our training database. This is expected to enhance the performance of our model.

### 6.6 Conclusion

We proposed a novel deep-learning method to generate sCT images with complete anatomy from truncated MR images. Based on cycleGAN, we modified the cycle-consistency loss and innovatively introduced original MR images with complete anatomy in the training process to facilitate anatomy compensation during sCT creation. Extensive experiments demonstrated that our method can generate sCT images with high image quality and reliable anatomy compensation. This technique has the great potential to be a useful tool to facilitate MR-based radiation treatment planning.

# Chapter 7: Discussion

## 7.1 Project Summary

The primary goal of this project is to enable the MR-guided online ART for SABR of HNC by addressing several challenges and gaps in the current clinical workflow. To achieve this goal, we first developed an auto-segmentation framework for GTV delineation from multimodality images in Aim 1. Next, a DIR framework was developed for CT-to-MR and MR-to-MR registration for online adaptive planning in Aim2. In addition, we developed the automated methods for creation of synthetic CT from MR images to improve dosimetric accuracy of MR-based planning in Aim 3.

## 7.2 Discussion of Specific Aims

### 7.2.1 Aim 1: Automate GTV delineation from fusion of multi-modal images for treatment planning

This aim was split into two sub-aims, the first of which was a development of data augmentation method for efficiently training auto-segmentation models (in Chapter 3), and the second of which focused on developing an auto-segmentation framework for GTV delineation based on the combination of planning CT, PET, and MR images (in Chapter 4).

For the first sub-aim, we successfully developed a simple but effective approach based on PCA method to generated synthetic images with contours for training auto-segmentation models. Our proposed method enabled the training of high-quality auto-segmentation models for HNC OARs from a severely limited number of well-contoured data (e.g. 10 cases). This makes DL-based auto-segmentation more applicable in clinical settings where data availability may be limited and can significantly reduce the effort needed for data curation.

Sub-aim 1.2, “*Gross Tumor Volume Auto-segmentation based on Multimodality Images”* was discussed in Chapter 4 of this dissertation. Accurate GTV delineation is essential for SABR treatment of HNC due to the extreme high dose delivered in each fraction and the presence of a lot of critical structures in HN region. Thus, we developed a two-staged auto-segmentation framework to automate the GTV contouring process based on the combination of CT, PET, and MR images. This framework is the first to simulate the GTV delineation process of radiation oncologists in clinic by dividing the process into two stages. The first stage was to localize and generate a coarse GTV based on CT and PET images. The second stage generated the final GTV by refining the coarse GTV using all multimodal images including CT, PET, and MR images. We trained our auto-segmentation models using data from 7 different institutions and evaluated the auto-contours on patient data from MD Anderson Cancer Center.

The results of our study demonstrated the high accuracy of our auto-contouring framework. Specifically, evaluating our approach on a dataset comprising small GTVs (average size:16 $cm^3$), our approach achieved an average DSC of 0.71, an MSD of 2.3 mm, and a HD95 of 8.8 mm using the combination of CT, PET, and MR images. These quantitative results provide strong evidence of the effectiveness of our segmentation approach. To further evaluate the clinical applicability of our framework, we conducted a clinical evaluation of the auto-generated contours. Two experienced HNC radiation oncologists independently reviewed the auto-generated contours and assessed their clinical acceptability. The results of the clinical evaluation showed that our auto-segmentation system was able to generate clinically acceptable contours for 90% of the GTVs, and the other 10% cases just needed minor edit for treatment planning. This finding is particularly significant as it highlights the potential clinical benefits of our proposed auto-segmentation framework. By improving the efficiency and accuracy of GTV segmentation based on multi-modality images, our approach has the potential to improve treatment planning for SABR and ultimately lead to better patient outcomes.

In addition, our auto-segmentation framework was designed to be capable of handling missing modality, which is a common challenge in clinical practice. To achieve this capability, we trained our models with different combinations of available modalities. This approach enabled our framework to generate accurate GTV contours in the absence of certain imaging modalities. Specifically, our two-stage auto-segmentation framework could localize the GTV in stage 1 based on the available CT/PET or CT images. In stage 2, the final GTV was predicted based on all available modalities. With missing PET or/and MR images, our model was able to produce a result not worse than a model trained with corresponding available image modalities only. The ability of our framework to handle missing modalities enhances the practical utility of our automated segmentation method, making it a valuable tool for radiation oncologists in real clinical settings where access to all imaging modalities may not be feasible.

### 7.2.2 Aim 2: Automate deformable registration of simulation CT and daily MRI for adaptive planning

Specific Aim 2 was the focus of Chapter 5 of this dissertation, in which we proposed a fully automated hierarchical registration framework for inter-modality (CT-to-MR) and intro-modality (MR-to-MR) DIR. The hierarchical registration framework is designed to handle large deformations between images by dividing the registration process into multiple stages, including whole volume-based registration, patch-based local registration, and patch-based deformable registration. The models were trained and evaluated based on the data of patients treated with MR-Linac at MD Anderson Cancer Center. The proposed method demonstrated superior performance in registration accuracy and computational efficiency compared to other deep learning methods and traditional DIR tool in Monaco treatment planning. Specifically, the average DSC and MSD calculated over all organs between the deformed contours by our method and the reference manual contours were

0.76±0.05/1.9±0.5mm for CT-MR registration and 0.87±0.06/0.9±0.3mm for MR-MR registration, respectively.

For MR-guided ART, DIR is crucial to ensure accurate contour prorogation with daily anatomical changes so that treatment plan can be reoptimized in real time. However, current registration tools in the clinic often fall short in achieving sufficient accuracy for CT-to-MR DIR, leading to the need for manual intervention and time-consuming contour editing. Our proposed method addresses these limitations and has the potential to significantly improve the accuracy of MR-guided ART. Furthermore, the computational efficiency of our method makes it more feasible for clinical use, allowing for fast registration and real-time adaptive planning.

### 7.2.3 Aim 3: Automate synthetic CT generation from MR images for MR-based adaptive planning

In Aim 3, we focused on the automation of high-quality synthetic CT generation from MR images for MR-based adaptive planning, which was included in Chapter 6. We proposed a cycleGAN-based method that was specifically designed to generate sCT images while compensating for the missing anatomy from truncated MR images. To evaluate the effectiveness of our approach, we compared the generated sCT images with real CT images. Our results indicated that the average MAE, PSNR, and SSIM calculated over test patients were 91.3 HU, 27.4 dB, and 0.94 for the sCT generated by our proposed DL models, demonstrating high-quality sCT generation.

Given that the goal of our work was to use sCT for dose calculation in MR-guided ART workflow, we further evaluated the dosimetric performance of sCT by comparing the dose distribution of the same treatment planning calculated on sCT and planning CT. The DVHs parameters of mean dose ($D_{mean}$), $D_{95\%}$, and $D_{5\%}$ were calculated for tumor targets and OARS for evaluation. We found that the absolute dose difference within the targets and

OARs was less than 0.5 %, which is superior to synthetic CT generation using bulk density override in the Monaco system [260]. In addition to the dosimetric accuracy, our method does not rely on the accuracy of contour propagation, which is a common source of error in sCT based on bulk density assignment. Instead, it uses a cycleGAN-based method to learn the mapping between MR images and sCT images, ensuring that the generated sCT images are consistent and accurate. Our method has the potential to significantly improve the efficiency and accuracy of MR-guided ART workflow for HNC patients, by providing high-quality sCT images that can be used for treatment planning and dose calculation.

### 7.3 Study Limitations and Future Direction

Although we successfully developed an auto-contouring system for GTV delineation in the oral region (including base of tongue, oropharynx, and tonsil disease) in Specific Aim 1, our model's coverage is currently limited to this specific HNC subsite. Expanding our model to other subsites is crucial for wider clinical applications. This requires collecting more patient data to include in the model training process or training a new model for each anatomical site. Furthermore, while our model performs well in segmenting primary GTVs, it currently does not account for nodal GTVs. This is a significant limitation, as nodal GTVs are an important aspect of HNC treatment planning. To address this limitation, we plan to include nodal GTVs in our future model development with existing curated data. This will enable our model to provide more comprehensive and accurate GTV delineation, covering both primary and nodal GTVs. In addition, while our auto-segmentation framework has demonstrated high clinical acceptability with a 100% success rate, it is essential to develop a quality assurance method to automatically detect and flag any inaccurate auto-contours. Another important aspect to consider in future studies is to improve the generalizability and robustness of our auto-contouring system. We aim to validate our segmentation framework on larger and more diverse datasets, including data from multiple institutions. This will enable us to better understand the potential impact of our contouring system on clinical outcomes, as well as identify potential challenges or limitations in real-world clinical settings.

In Specific Aim 2, we proposed and validated our DL-based registration framework for MR-guided adaptive planning, with a focus on its application for contour propagation. While contour propagation is a crucial step in MR-guided ART to account for daily anatomical changes, DIR has several other important applications in radiation therapy, including dose mapping and accumulation [261], [262]. Specifically, in MR-guided ART, DIR plays an important role in registering daily MR images to the planning CT and enabling the accumulation of dose for plan evaluation and re-planning. This approach allows for a more precise assessment of the accumulated dose distribution in the target volume and surrounding OARs over the course of treatment. Furthermore, in the case of re-treatment, dose mapping is crucial in transforming the previous dose distribution to the new planning images to assess the dose tolerances of OARs and ensure that dose is delivered accurately to the target while minimizing the risk of side effects to surrounding healthy tissue [263], [264]. However, there are several challenges in implementing our DL-based DIR framework for dose deformation, especially for the inter-modality CT-MR cases. One challenge is the limited accuracy of DVFs generated during the deformable registration process, which in turn affects the accuracy of DIR-based dose accumulation. Despite outperforming traditional DIR tools in Monaco and other DL methods, our proposed method still faces limitations in achieving high accuracy for inter-modality (CT-to-MR) registration, with an average DSC of only 0.76. Thus, improving the accuracy of inter-modality DIR remains a crucial area for future research in MR-guided ART. In addition, even with precise deformable registration, accurate dose accumulation is inherently difficult. As the volumes of tumor targets and normal tissues can vary from fraction to faction, the deformation process involves volumetric changes, including the merging of adjacent voxels or the creation of new voxels. These changes may lead to the violation of the energy conservation principle, since a constant energy density that is deformed to a greater or smaller mass can lead to the creation or elimination of energy. The voxelized nature of the moving and reference images during computation further challenged the dose mapping task. To address this problem, many dose accumulation methods have been proposed based traditional DIR methods, including center of mass method [265], interpolation methods [266], [267], direct voxel tracking method [268], and energy transfer method [269]. In our future studies, we plan to implement the interpolation method with our DL-based registration framework, due to their fast speed and no requirement of Mote Carlo simulation. As illustrated in Figure 34, the interpolation method involves deforming the center of each voxel in the reference image to locations on the dose grid of the moving image. The corresponding point of dose in the moving images is estimated by trilinear interpolation, and then mapped back to the original dose grid center on the reference image. However, the limitation of this method is that it can lead to accuracy degradation in regions of steep dose gradients, and it does not account for the physical aspects of dose deposition.

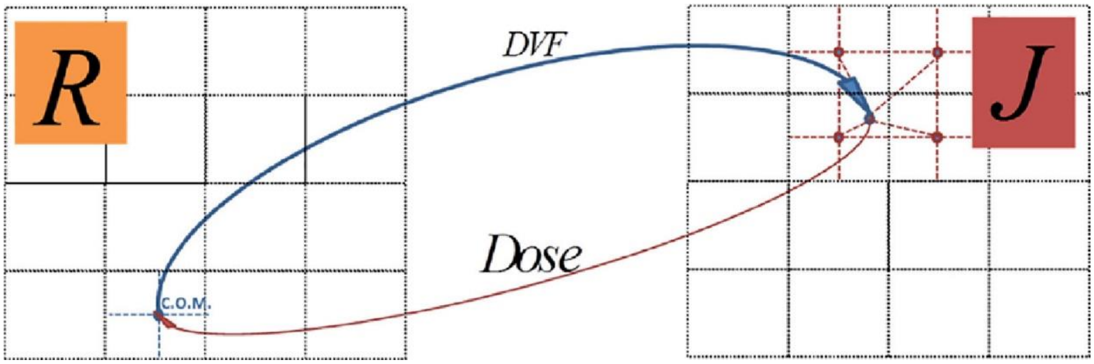


Figure 34: The illustration of trilinear interpolation method for dose accumulation. The left is the reference image, and the right is the moving image. Image courtesy of Chetty el al. [263]

In Specific Aim 3, a DL model was developed to generate synthetic CT from daily MR images with high dosimetric accuracy for MR-guided adaptive planning. The future work is to establish an MR-only treatment workflow using the MR-linac technology. In the current clinical practice, the treatment planning and dose calculation rely on simulation CT images, while both simulation CT and simulation MR images are acquired before treatment [94]. However, acquiring both simulation CT and MR images increases resource usage and cause additional challenges in terms of registration and fusion of the two image sets. An MR-only workflow would eliminate the need for simulation CT imaging and streamline the treatment planning process, resulting in reduced radiation exposure for the patient and increased efficiency for the clinical team. By generating synthetic CT from simulation MR images, we can use it for initial treatment planning and avoid the need for additional CT images. However, it is important to note that the use of undistorted MR images for synthetic CT generation is crucial to ensure high-precision MR-based planning. MR images are prone to distortion due to magnetic field inhomogeneities and susceptibility artifacts, which can result in geometric inaccuracies in the generated synthetic CT images. Therefore, future research should focus on developing and evaluating methods for distortion correction in MR images used for synthetic CT generation, as well as investigating the impact of residual distortion on treatment planning accuracy.

The DL models developed in this project have shown promising results in facilitating MR-guided ART for HNC, with automated GTV delineation, improved online contour propagation, and accurate synthetic CT generation. These models can be used collectively or separately to assist MR-guided ART for SABR of HNC. An end-to-end evaluation will be conducted to assess the collective performance of these DL models in achieving an automated MR-guided ART workflow for HNC. This evaluation will involve implementing all the DL models together to streamline the treatment planning process and improve treatment accuracy.

### 7.4 Conclusion

In this study, we successfully developed DL-based models to facilitate the MR-guided online adaptive RT workflow for HNC SABR. The automation not only facilitates GTV delineation in treatment planning stage, but also reduces the time needing for contouring review/editing during online adaptive planning. Also, the developed synthetic CT generation is expected to improve dose calculation accuracy for MR-based adaptive planning. This outcome is expected to have an important positive impact on improving treatment efficiency and improving the overall treatment quality of HNC SABR.

## References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, “Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries,” *CA: a cancer journal for clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2018,” *CA: a cancer journal for clinicians*, vol. 68, no. 1, pp. 7–30, 2018.

[3] L. Q. Chow, “Head and neck cancer,” *New England Journal of Medicine*, vol. 382, no. 1, pp. 60–72, 2020.

[4] M. Mourad, T. Jetmore, A. A. Jategaonkar, S. Moubayed, E. Moshier, and M. L. Urken, “Epidemiological trends of head and neck cancer in the United States: a SEER population study,” *Journal of Oral and Maxillofacial Surgery*, vol. 75, no. 12, pp. 2562–2572, 2017.

[5] A. K. Chaturvedi, E. A. Engels, R. M. Pfeiffer, B. Y. Hernandez, W. Xiao, E. Kim, B. Jiang, M. T. Goodman, M. Sibug-Saber, and W. Cozen, “Human papillomavirus and rising oropharyngeal cancer incidence in the United States,” *Journal of clinical oncology*, vol. 29, no. 32, p. 4294, 2011.

[6] P.-S. Ho, Y.-C. Ko, Y.-H. C. Yang, T.-Y. Shieh, and C.-C. Tsai, “The incidence of oropharyngeal cancer in Taiwan: an endemic betel quid chewing area,” *Journal of oral pathology & medicine*, vol. 31, no. 4, pp. 213–219, 2002.

[7] C. Y. Mimi and J.-M. Yuan, “Epidemiology of nasopharyngeal carcinoma,” in *Seminars in cancer biology*, Elsevier, 2002, pp. 421–429.

[8] P. L. Horn-Ross, B.-M. Ljung, and M. Morrow, “Environmental factors and the risk of salivary gland cancer,” *Epidemiology*, pp. 414–419, 1997.

[9] S. Preston-Martin, D. C. Thomas, S. C. White, and D. Cohen, “Prior exposure to medical and dental x-rays related to tumors of the parotid gland1,” *JNCI: Journal of the National Cancer Institute*, vol. 80, no. 12, pp. 943–949, 1988.

[10] S. J. Hamilton-Dutoit, M. H. Therkildsen, N. H. Nielsen, H. Jensen, J. H. Hansen, and G. Pallesen, “Undifferentiated carcinoma of the salivary gland in Greenlandic Eskimos: demonstration of Epstein-Barr virus DNA by in situ nucleic acid hybridization,” *Human pathology*, vol. 22, no. 8, pp. 811–815, 1991.

[11] A. Beddok, S. Krieger, L. Castera, D. Stoppa-Lyonnet, and J. Thariat, “Management of Fanconi Anemia patients with head and neck carcinoma: Diagnosis and treatment adaptation,” *Oral Oncology*, vol. 108, p. 104816, 2020.

[12] A. Sanabria, A. L. Carvalho, R. L. Melo, J. Magrin, M. K. Ikeda, J. G. Vartanian, and L. P. Kowalski, “Predictive factors for complications in elderly patients who underwent head and neck oncologic surgery,” *Head & Neck: Journal for the Sciences and Specialties of the Head and Neck*, vol. 30, no. 2, pp. 170–177, 2008.

[13] M. E. Daly, D. H. Lau, D. G. Farwell, Q. Luu, P. J. Donald, and A. M. Chen, “Feasibility and toxicity of concurrent chemoradiation for elderly patients with head and neck cancer,” *American Journal of Otolaryngology*, vol. 34, no. 6, pp. 631–635, 2013.

[14] D. Alterio, G. Marvaso, A. Ferrari, S. Volpe, R. Orecchia, and B. A. Jereczek-Fossa, “Modern radiotherapy for head and neck cancer,” in *Seminars in oncology*, Elsevier, 2019, pp. 233–245.

[15] G. N. Marta, V. Silva, H. de Andrade Carvalho, F. F. de Arruda, S. A. Hanna, R. Gadia, J. L. F. da Silva, S. F. M. Correa, C. E. C. V. Abreu, and R. Riera, “Intensity-modulated radiation therapy for head and neck cancer: systematic review and meta-analysis,” *Radiotherapy and oncology*, vol. 110, no. 1, pp. 9–15, 2014.

[16] T. A. Ratko, G. W. Douglas, J. A. De Souza, S. E. Belinson, and N. Aronson, “Radiotherapy treatments for head and neck cancer update,” 2015.

[17] C. M. Lim, D. A. Clump, D. E. Heron, and R. L. Ferris, “Stereotactic Body Radiotherapy (SBRT) for primary and recurrent head and neck tumors,” *Oral Oncol*, vol. 49, no. 5, pp. 401–406, May 2013, doi: 10.1016/j.oraloncology.2012.12.009.

[18] L. Stanisce, Y. Koshkareva, Q. Xu, A. Patel, C. Squillante, N. Ahmad, K. Rajagopalan, and G. J. Kubicek, “Stereotactic Body Radiotherapy Treatment for Recurrent, Previously Irradiated Head and Neck Cancer,” *Technol Cancer Res Treat*, vol. 17, p. 1533033818780086, Jun. 2018, doi: 10.1177/1533033818780086.

[19] J. A. Vargo, M. C. Ward, J. J. Caudell, N. Riaz, N. E. Dunlap, D. Isrow, S. J. Zakem, J. Dault, M. J. Awan, K. A. Higgins, C. Hassanadeh, J. J. Beitler, C. A. Reddy, S. Marcrom, D. H. Boggs, J. A. Bonner, M. Yao, M. Machtay, F. Siddiqui, A. M. Trotti, N. Y. Lee, S. A. Koyfman, R. L. Ferris, and D. E. Heron, “A Multi-institutional Comparison of SBRT and IMRT for Definitive Reirradiation of Recurrent or Second Primary Head and Neck Cancer,” *International Journal of Radiation Oncology*Biology*Physics*, vol. 100, no. 3, pp. 595–605, Mar. 2018, doi: 10.1016/j.ijrobp.2017.04.017.

[20] M. Swain and S. Ghosh-Laskar, “Stereotactic body radiotherapy (SBRT) for primary non-metastatic head and neck cancer: When less is enough,” *Oral Oncology*, vol. 116, p. 105265, May 2021, doi: 10.1016/j.oraloncology.2021.105265.

[21] S. Boeke, D. Mönnich, J. E. van Timmeren, and P. Balermpas, “MR-Guided Radiotherapy for Head and Neck Cancer: Current Developments, Perspectives, and Challenges,” *Frontiers in Oncology*, vol. 11, 2021, Accessed: May 28, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fonc.2021.616156

[22] C. Kurz, G. Buizza, G. Landry, F. Kamp, M. Rabe, C. Paganelli, G. Baroni, M. Reiner, P. J. Keall, C. A. T. van den Berg, and M. Riboldi, “Medical physics challenges in

clinical MR-guided radiotherapy," *Radiat Oncol*, vol. 15, no. 1, p. 93, Dec. 2020, doi: 10.1186/s13014-020-01524-4.

[23] H. Ashamalla, A. Guirgius, E. Bieniek, S. Rafla, A. Evola, G. Goswami, R. Oldroyd, B. Mokhtar, and K. Parikh, "The impact of positron emission tomography/computed tomography in edge delineation of gross tumor volume for head and neck cancers," *Int J Radiat Oncol Biol Phys*, vol. 68, no. 2, pp. 388–395, Jun. 2007, doi: 10.1016/j.ijrobp.2006.12.029.

[24] A. M. Berson, N. F. Stein, A. C. Riegel, S. Destian, T. Ng, L. B. Tena, R. J. Mitnick, and S. Heiba, "Variability of gross tumor volume delineation in head-and-neck cancer using PET/CT fusion, Part II: the impact of a contouring protocol," *Med Dosim*, vol. 34, no. 1, pp. 30–35, 2009, doi: 10.1016/j.meddos.2007.08.003.

[25] S. L. Breen, J. Publicover, S. De Silva, G. Pond, K. Brock, B. O'Sullivan, B. Cummings, L. Dawson, A. Keller, J. Kim, J. Ringash, E. Yu, A. Hendler, and J. Waldron, "Intraobserver and interobserver variability in GTV delineation on FDG-PET-CT images of head and neck cancers," *Int J Radiat Oncol Biol Phys*, vol. 68, no. 3, pp. 763–770, Jul. 2007, doi: 10.1016/j.ijrobp.2006.12.039.

[26] M. Mukesh, R. Benson, R. Jena, A. Hoole, T. Roques, C. Scrase, C. Martin, G. A. Whitfield, J. Gemmill, and S. Jefferies, "Interobserver variation in clinical target volume and organs at risk segmentation in post-parotidectomy radiotherapy: can segmentation protocols help?," *Br J Radiol*, vol. 85, no. 1016, pp. e530–e536, Aug. 2012, doi: 10.1259/bjr/66693547.

[27] A. C. Riegel, A. M. Berson, S. Destian, T. Ng, L. B. Tena, R. J. Mitnick, and P. S. Wong, "Variability of gross tumor volume delineation in head-and-neck cancer using CT and PET/CT fusion," *Int J Radiat Oncol Biol Phys*, vol. 65, no. 3, pp. 726–732, Jul. 2006, doi: 10.1016/j.ijrobp.2006.01.014.

[28] C. F. Njeh, “Tumor delineation: The weakest link in the search for accuracy in radiotherapy,” *J Med Phys*, vol. 33, no. 4, pp. 136–140, 2008, doi: 10.4103/0971-6203.44472.

[29] C. F. Njeh, L. Dong, and C. G. Orton, “Point/Counterpoint. IGRT has limited clinical value due to lack of accurate tumor delineation,” *Med Phys*, vol. 40, no. 4, p. 040601, Apr. 2013, doi: 10.1118/1.4789492.

[30] E. Weiss and C. F. Hess, “The impact of gross tumor volume (GTV) and clinical target volume (CTV) definition on the total accuracy in radiotherapy theoretical aspects and practical experiences,” *Strahlenther Onkol*, vol. 179, no. 1, pp. 21–30, Jan. 2003, doi: 10.1007/s00066-003-0976-5.

[31] H. Ashamalla, A. Guirgius, E. Bieniek, S. Rafla, A. Evola, G. Goswami, R. Oldroyd, B. Mokhtar, and K. Parikh, “The impact of positron emission tomography/computed tomography in edge delineation of gross tumor volume for head and neck cancers,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 68, no. 2, pp. 388–395, 2007.

[32] A. C. Riegel, A. M. Berson, S. Destian, T. Ng, L. B. Tena, R. J. Mitnick, and P. S. Wong, “Variability of gross tumor volume delineation in head-and-neck cancer using CT and PET/CT fusion,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 65, no. 3, pp. 726–732, 2006.

[33] X. Geets, J.-F. Daisne, S. Arcangeli, E. Coche, M. De Poel, T. Duprez, G. Nardella, and V. Grégoire, “Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: comparison between CT-scan and MRI,” *Radiotherapy and oncology*, vol. 77, no. 1, pp. 25–31, 2005.

[34] A. C. Paulino, M. Koshy, R. Howell, D. Schuster, and L. W. Davis, “Comparison of CT-and FDG-PET-defined gross tumor volume in intensity-modulated radiotherapy for

head-and-neck cancer," *International Journal of Radiation Oncology* Biology* Physics*, vol. 61, no. 5, pp. 1385–1392, 2005.

[35] T. Nishioka, T. Shiga, H. Shirato, E. Tsukamoto, K. Tsuchiya, T. Kato, K. Ohmori, A. Yamazaki, H. Aoyama, and S. Hashimoto, "Image fusion between 18FDG-PET and MRI/CT for radiotherapy planning of oropharyngeal and nasopharyngeal carcinomas," *International Journal of Radiation Oncology* Biology* Physics*, vol. 53, no. 4, pp. 1051–1057, 2002.

[36] D. J. Rhee, C. E. Cardenas, H. Elhalawani, R. McCarroll, L. Zhang, J. Yang, A. S. Garden, C. B. Peterson, B. M. Beadle, and L. E. Court, "Automatic detection of contouring errors using convolutional neural networks," *Med Phys*, vol. 46, no. 11, pp. 5086–5097, Nov. 2019, doi: 10.1002/mp.13814.

[37] Y. Zhao, D. J. Rhee, C. Cardenas, L. E. Court, and J. Yang, "Training deep-learning segmentation models from severely limited data," *Med Phys*, vol. 48, no. 4, pp. 1697–1706, Apr. 2021, doi: 10.1002/mp.14728.

[38] H. Bollen, S. Willems, M. Wegge, M. Frederik, and S. Nuyts, "Benefits of automated gross tumor volume segmentation in head and neck cancer using multi-modality information," *Radiotherapy and Oncology*, p. 109574, 2023.

[39] J. Ren, J. G. Eriksen, J. Nijkamp, and S. S. Korreman, "Comparing different CT, PET and MRI multi-modality image combinations for deep learning-based head and neck tumor segmentation," *Acta Oncologica*, vol. 60, no. 11, pp. 1399–1406, 2021.

[40] Y. M. Moe, A. R. Groendahl, O. Tomic, E. Dale, E. Malinen, and C. M. Futsaether, "Deep learning-based auto-delineation of gross tumour volumes and involved nodes in PET/CT images of head and neck cancer patients," *European journal of nuclear medicine and molecular imaging*, vol. 48, pp. 2782–2792, 2021.

[41] B. A. McDonald, S. Vedam, J. Yang, J. Wang, P. Castillo, B. Lee, A. Sobremonte, S. Ahmed, Y. Ding, A. S. R. Mohamed, P. Balter, N. Hughes, D. Thorwarth, M. Nachbar,

M. E. P. Philippens, C. H. J. Terhaard, D. Zips, S. Böke, M. J. Awan, J. Christodouleas, C. D. Fuller, and MR-Linac Consortium Head and Neck Tumor Site Group, “Initial Feasibility and Clinical Implementation of Daily MR-Guided Adaptive Head and Neck Cancer Radiation Therapy on a 1.5T MR-Linac System: Prospective R-IDEAL 2a/2b Systematic Clinical Evaluation of Technical Innovation,” *Int J Radiat Oncol Biol Phys*, vol. 109, no. 5, pp. 1606–1618, Apr. 2021, doi: 10.1016/j.ijrobp.2020.12.015.

[42] C. Kurz, G. Buizza, G. Landry, F. Kamp, M. Rabe, C. Paganelli, G. Baroni, M. Reiner, P. J. Keall, C. A. T. van den Berg, and M. Riboldi, “Medical physics challenges in clinical MR-guided radiotherapy,” *Radiation Oncology*, vol. 15, no. 1, p. 93, May 2020, doi: 10.1186/s13014-020-01524-4.

[43] U. A. van der Heide, A. C. Houweling, G. Groenendaal, R. G. H. Beets-Tan, and P. Lambin, “Functional MRI for radiotherapy dose painting,” *Magn Reson Imaging*, vol. 30, no. 9, pp. 1216–1223, Nov. 2012, doi: 10.1016/j.mri.2012.04.010.

[44] S.-H. Hsu, Y. Cao, K. Huang, M. Feng, and J. M. Balter, “Investigation of a method for generating synthetic CT models from MRI scans of the head and neck for radiation therapy,” *Phys Med Biol*, vol. 58, no. 23, pp. 8419–8435, Dec. 2013, doi: 10.1088/0031-9155/58/23/8419.

[45] M. Boulanger, J.-C. Nunes, H. Chourak, A. Largent, S. Tahri, O. Acosta, R. De Crevoisier, C. Lafond, and A. Barateau, “Deep learning methods to generate synthetic CT from MRI in radiotherapy: A literature review,” *Phys Med*, vol. 89, pp. 265–281, Sep. 2021, doi: 10.1016/j.ejmp.2021.07.027.

[46] N. H. Malik, M. S. Kim, H. Chen, I. Poon, Z. Husain, A. Eskander, G. Boldt, A. V. Louie, and I. Karam, “Stereotactic radiation therapy for de novo head and neck cancers: a systematic review and meta-analysis,” *Advances in Radiation Oncology*, vol. 6, no. 1, p. 100628, 2021.

[47] A. Amini, J. D. McDermott, G. Gan, S. Bhatia, W. Sumner, C. M. Fisher, A. Jimeno, D. W. Bowles, D. Raben, and S. D. Karam, “Stereotactic body radiotherapy as primary therapy for head and neck cancer in the elderly or patients with poor performance,” *Frontiers in Oncology*, vol. 4, p. 274, 2014.

[48] J. G. Douglas, R. Goodkin, and G. E. Laramore, “Gamma knife stereotactic radiosurgery for salivary gland neoplasms with base of skull invasion following neutron radiotherapy,” *Head & Neck: Journal for the Sciences and Specialties of the Head and Neck*, vol. 30, no. 4, pp. 492–496, 2008.

[49] E. F. Lartigau, E. Tresch, J. Thariat, P. Graff, B. Coche-Dequeant, K. Benezery, L. Schiappacasse, M. Degardin, P.-Y. Bondiau, and D. Peiffert, “Multi institutional phase II study of concomitant stereotactic reirradiation and cetuximab for recurrent head and neck cancer,” *Radiotherapy and Oncology*, vol. 109, no. 2, pp. 281–285, 2013.

[50] K. R. Unger, C. E. Lominska, J. F. Deeken, B. J. Davidson, K. A. Newkirk, G. J. Gagnon, J. Hwang, R. S. Slack, A.-M. Noone, and K. W. Harter, “Fractionated stereotactic radiosurgery for reirradiation of head-and-neck cancer,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 77, no. 5, pp. 1411–1419, 2010.

[51] F. Siddiqui, M. Patel, M. Khan, S. McLean, J. Dragovic, J.-Y. Jin, B. Movsas, and S. Ryu, “Stereotactic body radiation therapy for primary, recurrent, and metastatic tumors in the head-and-neck region,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 74, no. 4, pp. 1047–1053, 2009.

[52] D. Owen, F. Iqbal, B. E. Pollock, M. J. Link, K. Stien, Y. I. Garces, P. D. Brown, and R. L. Foote, “Long-term follow-up of stereotactic radiosurgery for head and neck malignancies,” *Head & Neck*, vol. 37, no. 11, pp. 1557–1562, 2015.

[53] G. Voynov, D. E. Heron, S. Burton, J. Grandis, A. Quinn, R. Ferris, C. Ozhasoglu, W. Vogel, and J. Johnson, “Frameless stereotactic radiosurgery for recurrent head and

neck carcinoma," *Technology in cancer research & treatment*, vol. 5, no. 5, pp. 529–535, 2006.

[54] G. Ozyigit, M. Cengiz, G. Yazici, F. Yildiz, M. Gurkaynak, F. Zorlu, D. Yildiz, S. Hosal, I. Gullu, and F. Akyol, "A retrospective comparison of robotic stereotactic body radiotherapy and three-dimensional conformal radiotherapy for the reirradiation of locally recurrent nasopharyngeal carcinoma," *International Journal of Radiation Oncology* Biology* Physics*, vol. 81, no. 4, pp. e263–e268, 2011.

[55] N. G. Burnet, S. J. Thomas, K. E. Burton, and S. J. Jefferies, "Defining the tumour and target volumes for radiotherapy," *Cancer Imaging*, vol. 4, no. 2, p. 153, 2004.

[56] J. C. Ho and J. Phan, "Reirradiation of skull base tumors with advanced highly conformal techniques," *Current oncology reports*, vol. 19, pp. 1–12, 2017.

[57] M. M. i Garau, "Radiobiology of stereotactic body radiation therapy (SBRT)," *Reports of Practical Oncology and Radiotherapy*, vol. 22, no. 2, pp. 86–95, 2017.

[58] K. K. Fu, T. F. Pajak, A. Trotti, C. U. Jones, S. A. Spencer, T. L. Phillips, A. S. Garden, J. A. Ridge, J. S. Cooper, and K. K. Ang, "A Radiation Therapy Oncology Group (RTOG) phase III randomized study to compare hyperfractionation and two variants of accelerated fractionation to standard fractionation radiotherapy for head and neck squamous cell carcinomas: first report of RTOG 9003," *International Journal of Radiation Oncology* Biology* Physics*, vol. 48, no. 1, pp. 7–16, 2000.

[59] J. M. Brown, D. J. Carlson, and D. J. Brenner, "The tumor radiobiology of SRS and SBRT: are more than the 5 Rs involved?," *International Journal of Radiation Oncology* Biology* Physics*, vol. 88, no. 2, pp. 254–262, 2014.

[60] D. Winkel, G. H. Bol, P. S. Kroon, B. van Asselen, S. S. Hackett, A. M. Werensteijn-Honingh, M. P. Intven, W. S. Eppinga, R. H. Tijssen, and L. G. Kerkmeijer, "Adaptive radiotherapy: the Elekta Unity MR-linac concept," *Clinical and translational radiation oncology*, vol. 18, pp. 54–59, 2019.

[61] C. E. Cardenas, J. Yang, B. M. Anderson, L. E. Court, and K. B. Brock, “Advances in auto-segmentation,” in *Seminars in radiation oncology*, Elsevier, 2019, pp. 185–197.

[62] Y. Zhong, Y. Yang, Y. Fang, J. Wang, and W. Hu, “A preliminary experience of implementing deep-learning based auto-segmentation in head and neck cancer: a study on real-world clinical cases,” *Frontiers in oncology*, vol. 11, p. 638197, 2021.

[63] M. Kosmin, J. Ledsam, B. Romera-Paredes, R. Mendes, S. Moinuddin, D. de Souza, L. Gunn, C. Kelly, C. O. Hughes, and A. Karthikesalingam, “Rapid advances in auto-segmentation of organs at risk and target volumes in head and neck cancer,” *Radiotherapy and Oncology*, vol. 135, pp. 130–140, 2019.

[64] J. Y. Lim and M. Leech, “Use of auto-segmentation in the delineation of target volumes and organs at risk in head and neck,” *Acta Oncologica*, vol. 55, no. 7, pp. 799–806, 2016.

[65] T. Vrtovec, D. Močnik, P. Strojan, F. Pernuš, and B. Ibragimov, “Auto-segmentation of organs at risk for head and neck radiotherapy planning: from atlas-based to deep learning methods,” *Medical physics*, vol. 47, no. 9, pp. e929–e950, 2020.

[66] D. J. Rhee, C. E. Cardenas, H. Elhalawani, R. McCarroll, L. Zhang, J. Yang, A. S. Garden, C. B. Peterson, B. M. Beadle, and L. E. Court, “Automatic detection of contouring errors using convolutional neural networks,” *Medical physics*, vol. 46, no. 11, pp. 5086–5097, 2019.

[67] Y. Liu, Y. Lei, Y. Fu, T. Wang, J. Zhou, X. Jiang, M. McDonald, J. J. Beitler, W. J. Curran, and T. Liu, “Head and neck multi-organ auto-segmentation on CT images aided by synthetic MRI,” *Medical physics*, vol. 47, no. 9, pp. 4294–4302, 2020.

[68] K. Men, H. Geng, C. Cheng, H. Zhong, M. Huang, Y. Fan, J. P. Plastaras, A. Lin, and Y. Xiao, “More accurate and efficient segmentation of organs-at-risk in radiotherapy with convolutional neural networks cascades,” *Medical physics*, vol. 46, no. 1, pp. 286–292, 2019.

[69] K. Kisling, R. McCarroll, L. Zhang, J. Yang, H. Simonds, M. Du Toit, C. Trauernicht, H. Burger, J. Parkes, and M. Mejia, “Radiation planning assistant-a streamlined, fully automated radiotherapy treatment planning system,” *JoVE (Journal of Visualized Experiments)*, no. 134, p. e57411, 2018.

[70] R. E. McCarroll, B. M. Beadle, P. A. Balter, H. Burger, C. E. Cardenas, S. Dalvie, D. S. Followill, K. D. Kisling, M. Mejia, and K. Naidoo, “Retrospective validation and clinical implementation of automated contouring of organs at risk in the head and neck: a step toward automated radiation treatment planning for low-and middle-income countries,” *Journal of Global Oncology*, vol. 4, pp. 1–11, 2018.

[71] K. L. Moore, “Automated radiotherapy treatment planning,” in *Seminars in radiation oncology*, Elsevier, 2019, pp. 209–218.

[72] K. Kisling, L. Zhang, H. Simonds, N. Fakie, J. Yang, R. McCarroll, P. Balter, H. Burger, O. Bogler, and R. Howell, “Fully automatic treatment planning for external-beam radiation therapy of locally advanced cervical cancer: a tool for low-resource clinics,” *Journal of Global Oncology*, vol. 5, pp. 1–9, 2019.

[73] J. Yang, B. M. Beadle, A. S. Garden, D. L. Schwartz, and M. Aristophanous, “A multimodality segmentation framework for automatic target delineation in head and neck radiotherapy,” *Medical physics*, vol. 42, no. 9, pp. 5310–5320, 2015.

[74] Z. Guo, N. Guo, K. Gong, and Q. Li, “Gross tumor volume segmentation for head and neck cancer radiotherapy using deep dense multi-modality network,” *Physics in Medicine & Biology*, vol. 64, no. 20, p. 205015, 2019.

[75] V. Andrearczyk, V. Oreiller, M. Abobakr, A. Akhavanallaf, P. Balermpas, S. Boughdad, L. Capriotti, J. Castelli, C. Cheze Le Rest, and P. Decazes, “Overview of the HECKTOR challenge at MICCAI 2022: automatic head and neck tumor segmentation and outcome prediction in PET/CT,” in *Head and Neck Tumor Segmentation and Outcome Prediction: Third Challenge, HECKTOR 2022, Held in*

*Conjunction with MICCAI 2022, Singapore, September 22, 2022, Proceedings*, Springer, 2023, pp. 1–30.

[76] Y. Zhao, D. J. Rhee, C. Cardenas, L. E. Court, and J. Yang, “Training deep-learning segmentation models from severely limited data,” *Medical physics*, vol. 48, no. 4, pp. 1697–1706, 2021.

[77] D. L. Schwartz, A. S. Garden, S. J. Shah, G. Chronowski, S. Sejpal, D. I. Rosenthal, Y. Chen, Y. Zhang, L. Zhang, and P.-F. Wong, “Adaptive radiotherapy for head and neck cancer—dosimetric results from a prospective clinical trial,” *Radiotherapy and Oncology*, vol. 106, no. 1, pp. 80–84, 2013.

[78] D. L. Schwartz, A. S. Garden, J. Thomas, Y. Chen, Y. Zhang, J. Lewin, M. S. Chambers, and L. Dong, “Adaptive radiotherapy for head-and-neck cancer: initial clinical outcomes from a prospective trial,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 83, no. 3, pp. 986–993, 2012.

[79] J. Castelli, A. Simon, G. Louvel, O. Henry, E. Chajon, M. Nassef, P. Haigron, G. Cazoulat, J. D. Ospina, and F. Jegoux, “Impact of head and neck cancer adaptive radiotherapy to spare the parotid glands and decrease the risk of xerostomia,” *Radiation Oncology*, vol. 10, pp. 1–10, 2015.

[80] P. Kupelian and J.-J. Sonke, “Magnetic resonance–guided adaptive radiotherapy: a solution to the future,” in *Seminars in radiation oncology*, Elsevier, 2014, pp. 227–232.

[81] K. P. Nesteruk, M. Bobić, A. Lalonde, B. A. Winey, A. J. Lomax, and H. Paganetti, “Ct-on-rails versus in-room cbct for online daily adaptive proton therapy of head-and-neck cancers,” *Cancers*, vol. 13, no. 23, p. 5991, 2021.

[82] R. J. D. Prestwich, J. Sykes, B. Carey, M. Sen, K. E. Dyker, and A. F. Scarsbrook, “Improving target definition for head and neck radiotherapy: a place for magnetic resonance imaging and 18-fluoride fluorodeoxyglucose positron emission tomography?,” *Clinical Oncology*, vol. 24, no. 8, pp. 577–589, 2012.

[83] M. Ahmed, M. Schmidt, A. Sohaib, C. Kong, K. Burke, C. Richardson, M. Usher, S. Brennan, A. Riddell, and M. Davies, “The value of magnetic resonance imaging in target volume delineation of base of tongue tumours–a study using flexible surface coils,” *Radiotherapy and Oncology*, vol. 94, no. 2, pp. 161–167, 2010.

[84] C. Kontaxis, G. H. Bol, J. J. W. Lagendijk, and B. W. Raaymakers, “Towards adaptive IMRT sequencing for the MR-linac,” *Physics in Medicine & Biology*, vol. 60, no. 6, p. 2493, 2015.

[85] B. A. McDonald, S. Vedam, J. Yang, J. Wang, P. Castillo, B. Lee, A. Sobremonte, S. Ahmed, Y. Ding, A. S. R. Mohamed, P. Balter, N. Hughes, D. Thorwarth, M. Nachbar, M. E. P. Philippens, C. H. J. Terhaard, D. Zips, S. Böke, M. J. Awan, J. Christodouleas, and C. D. Fuller, “Initial Feasibility and Clinical Implementation of Daily MR-Guided Adaptive Head and Neck Cancer Radiation Therapy on a 1.5T MR-Linac System: Prospective R-IDEAL 2a/2b Systematic Clinical Evaluation of Technical Innovation,” *International Journal of Radiation Oncology*Biology*Physics*, vol. 109, no. 5, pp. 1606–1618, Apr. 2021, doi: 10.1016/j.ijrobp.2020.12.015.

[86] S. Fu, Y. Li, Y. Han, H. Wang, Y. Chen, O. Yan, Q. He, H. Ma, L. Liu, and F. Liu, “Diffusion-weighted MRI-guided dose painting in patients with locoregionally advanced nasopharyngeal carcinoma treated with induction chemotherapy plus concurrent chemoradiotherapy: a randomized, controlled clinical trial.,” *International Journal of Radiation Oncology, Biology, Physics*, 2022.

[87] S. Klüter, “Technical design and concept of a 0.35 T MR-Linac,” *Clinical and translational radiation oncology*, vol. 18, pp. 98–101, 2019.

[88] B. W. Raaymakers, J. J. W. Lagendijk, J. Overweg, J. G. M. Kok, A. J. E. Raaijmakers, E. M. Kerkhof, R. W. Van Der Put, I. Meijsing, S. P. M. Crijns, and F. Benedosso, “Integrating a 1.5 T MRI scanner with a 6 MV accelerator: proof of concept,” *Physics in Medicine & Biology*, vol. 54, no. 12, p. N229, 2009.

[89] A. J. Raaijmakers, B. W. Raaymakers, and J. J. Lagendijk, “Integrating a MRI scanner with a 6 MV radiotherapy accelerator: dose increase at tissue–air interfaces in a lateral magnetic field due to returning electrons,” *Physics in Medicine & Biology*, vol. 50, no. 7, p. 1363, 2005.

[90] K. Wachowicz, T. Stanescu, S. D. Thomas, and B. G. Fallone, “Implications of tissue magnetic susceptibility-related distortion on the rotating magnet in an MR-linac design,” *Medical physics*, vol. 37, no. 4, pp. 1714–1721, 2010.

[91] B. G. Fallone, “The rotating biplanar linac–magnetic resonance imaging system,” in *Seminars in radiation oncology*, Elsevier, 2014, pp. 200–202.

[92] P. J. Keall, M. Barton, and S. Crozier, “The Australian magnetic resonance imaging–linac program,” in *Seminars in radiation oncology*, Elsevier, 2014, pp. 203–206.

[93] G. P. Liney, U. Jelen, H. Byrne, B. Dong, T. L. Roberts, Z. Kuncic, and P. Keall, “the first live treatment on a 1.0 Tesla inline MRI-linac,” *Medical Physics*, vol. 46, no. 7, pp. 3254–3258, 2019.

[94] C.-Y. Huang, B. Yang, W. W. Lam, H. Geng, K. Y. Cheung, and S. K. Yu, “Magnetic field induced dose effects in radiation therapy using MR-linacs,” *Medical Physics*, 2023.

[95] B. A. McDonald, S. Vedam, J. Yang, J. Wang, P. Castillo, B. Lee, A. Sobremonte, S. Ahmed, Y. Ding, and A. S. Mohamed, “Initial feasibility and clinical implementation of daily mr-guided adaptive head and neck cancer radiation therapy on a 1.5 t mr-linac system: Prospective r-ideal 2a/2b systematic clinical evaluation of technical innovation,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 109, no. 5, pp. 1606–1618, 2021.

[96] H. Bahig, Y. Yuan, A. S. Mohamed, K. K. Brock, S. P. Ng, J. Wang, Y. Ding, K. Hutcheson, M. McCulloch, and P. A. Balter, “Magnetic resonance-based response assessment and dose adaptation in human papilloma virus positive tumors of the

oropharynx treated with radiotherapy (MR-ADAPTOR): an R-IDEAL stage 2a-2b/Bayesian phase II trial," *Clinical and translational radiation oncology*, vol. 13, pp. 19–23, 2018.

[97] P. J. Keall, C. K. Glide-Hurst, M. Cao, P. Lee, B. Murray, B. W. Raaymakers, A. Tree, and U. A. van der Heide, "ICRU REPORT 97: MRI-Guided Radiation Therapy Using MRI-Linear Accelerators," *Journal of the ICRU*, vol. 22, no. 1, pp. 1–100, 2022.

[98] K. K. Brock, S. Mutic, T. R. McNutt, H. Li, and M. L. Kessler, "Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132," *Medical physics*, vol. 44, no. 7, pp. e43–e76, 2017.

[99] S.-H. Hsu, Z. Han, J. E. Leeman, Y.-H. Hu, R. H. Mak, and A. Sudhyadhom, "Synthetic CT generation for MRI-guided adaptive radiotherapy in prostate cancer," *Frontiers in Oncology*, vol. 12, 2022.

[100] A. M. Owrangi, P. B. Greer, and C. K. Glide-Hurst, "MRI-only treatment planning: benefits and challenges," *Physics in Medicine & Biology*, vol. 63, no. 5, p. 05TR01, 2018.

[101] A. L. Chin, A. Lin, S. Anamalayil, and B.-K. K. Teo, "Feasibility and limitations of bulk density assignment in MRI for head and neck IMRT treatment planning," *Journal of Applied Clinical Medical Physics*, vol. 15, no. 5, pp. 100–111, 2014.

[102] J. Kim, K. Garbarino, L. Schultz, K. Levin, B. Movsas, M. S. Siddiqui, I. J. Chetty, and C. Glide-Hurst, "Dosimetric evaluation of synthetic CT relative to bulk density assignment-based magnetic resonance-only approaches for prostate radiotherapy," *Radiation Oncology*, vol. 10, pp. 1–9, 2015.

[103] A. Largent, A. Barateau, J.-C. Nunes, E. Mylona, J. Castelli, C. Lafond, P. B. Greer, J. A. Dowling, J. Baxter, H. Saint-Jalmes, O. Acosta, and R. de Crevoisier, "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*, vol. 105, no. 5, pp. 1137–1150, Dec. 2019, doi: 10.1016/j.ijrobp.2019.08.049.

[104] S. Kazemifar, A. M. Barragán Montero, K. Souris, S. T. Rivas, R. Timmerman, Y. K. Park, S. Jiang, X. Geets, E. Sterpin, and A. Owrangi, “Dosimetric evaluation of synthetic CT generated with GANs for MRI-only proton therapy treatment planning of brain tumors,” *J Appl Clin Med Phys*, vol. 21, no. 5, pp. 76–86, Mar. 2020, doi: 10.1002/acm2.12856.

[105] H. Emami, M. Dong, S. P. Nejad-Davarani, and C. Glide-Hurst, “Generating Synthetic CTs from Magnetic Resonance Images using Generative Adversarial Networks,” *Med Phys*, p. 10.1002/mp.13047, Jun. 2018, doi: 10.1002/mp.13047.

[106] S. Kazemifar, S. McGuire, R. Timmerman, Z. Wardak, D. Nguyen, Y. Park, S. Jiang, and A. Owrangi, “MRI-only brain radiotherapy: Assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach,” *Radiother Oncol*, vol. 136, pp. 56–63, Jul. 2019, doi: 10.1016/j.radonc.2019.03.026.

[107] P. Klages, I. Benslimane, S. Riyahi, J. Jiang, M. Hunt, J. O. Deasy, H. Veeraraghavan, and N. Tyagi, “Patch-Based Generative Adversarial Neural Network Models for Head and Neck MR-Only Planning,” *Med Phys*, vol. 47, no. 2, pp. 626–642, Feb. 2020, doi: 10.1002/mp.13927.

[108] S. Regnery, C. Buchele, F. Weykamp, M. Pohl, P. Hoegen, T. Eichkorn, T. Held, J. Ristau, C. Rippke, and L. König, “Adaptive MR-guided stereotactic radiotherapy is beneficial for ablative treatment of lung tumors in high-risk locations,” *Frontiers in Oncology*, vol. 11, p. 5431, 2022.

[109] I. Hadi, C. Eze, S. Schönecker, R. von Bestenbostel, P. Rogowski, L. Nierer, R. Bodensohn, M. Reiner, G. Landry, and C. Belka, “MR-guided SBRT boost for patients with locally advanced or recurrent gynecological cancers ineligible for brachytherapy:

feasibility and early clinical experience,” *Radiation Oncology*, vol. 17, no. 1, pp. 1–9, 2022.

[110] L. A. Gharzai, B. S. Rosen, B. Mittal, M. L. Mierzwa, and P. Yadav, “Magnetic Resonance Guided Radiotherapy for Head and Neck Cancers,” *Journal of Clinical Medicine*, vol. 11, no. 5, p. 1388, 2022.

[111] T. M. Ma, J. M. Lamb, M. Casado, X. Wang, T. V. Basehart, Y. Yang, D. Low, K. Sheng, N. Agazaryan, and N. G. Nickols, “Magnetic resonance imaging-guided stereotactic body radiotherapy for prostate cancer (mirage): a phase iii randomized trial,” *BMC cancer*, vol. 21, no. 1, p. 538, 2021.

[112] D. lin, A. V. Vasilakos, Y. Tang, and Y. Yao, “Neural networks for computer-aided diagnosis in medicine: A review,” *Neurocomputing*, vol. 216, pp. 700–708, Dec. 2016, doi: 10.1016/j.neucom.2016.08.039.

[113] C. E. Cardenas, R. E. McCarroll, L. E. Court, B. A. Elgohari, H. Elhalawani, C. D. Fuller, M. J. Kamal, M. A. M. Meheissen, A. S. R. Mohamed, A. Rao, B. Williams, A. Wong, J. Yang, and M. Aristophanous, “Deep Learning Algorithm for Auto-Delineation of High-Risk Oropharyngeal Clinical Target Volumes With Built-In Dice Similarity Coefficient Parameter Optimization Function,” *International Journal of Radiation Oncology*Biology*Physics*, vol. 101, no. 2, pp. 468–478, Jun. 2018, doi: 10.1016/j.ijrobp.2018.01.114.

[114] S. Nikolov, S. Blackwell, R. Mendes, J. De Fauw, C. Meyer, C. Hughes, H. Askham, B. Romera-Paredes, A. Karthikesalingam, C. Chu, D. Carnell, C. Boon, D. D’Souza, S. A. Moinuddin, K. Sullivan, D. R. Consortium, H. Montgomery, G. Rees, R. Sharma, M. Suleyman, T. Back, J. R. Ledsam, and O. Ronneberger, “Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy,” *arXiv:1809.04430 [physics, stat]*, Sep. 2018, Accessed: Jun. 25, 2020. [Online]. Available: http://arxiv.org/abs/1809.04430

[115] B. Ibragimov and L. Xing, “Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks,” *Med. Phys.*, vol. 44, no. 2, pp. 547–557, Feb. 2017, doi: 10.1002/mp.12045.

[116] C. E. Cardenas, J. Yang, B. M. Anderson, L. E. Court, and K. B. Brock, “Advances in Auto-Segmentation,” *Seminars in Radiation Oncology*, vol. 29, no. 3, pp. 185–197, Jul. 2019, doi: 10.1016/j.semradonc.2019.02.001.

[117] A. Halevy, P. Norvig, and F. Pereira, “The Unreasonable Effectiveness of Data,” *IEEE Intell. Syst.*, vol. 24, no. 2, pp. 8–12, Mar. 2009, doi: 10.1109/MIS.2009.36.

[118] D. N. Teguh, P. C. Levendag, P. W. J. Voet, A. Al-Mamgani, X. Han, T. K. Wolf, L. S. Hibbard, P. Nowak, H. Akhiat, M. L. P. Dirkx, B. J. M. Heijmen, and M. S. Hoogeman, “Clinical Validation of Atlas-Based Auto-Segmentation of Multiple Target Volumes and Normal Tissue (Swallowing/Mastication) Structures in the Head and Neck,” *International Journal of Radiation Oncology*Biology*Physics*, vol. 81, no. 4, pp. 950–957, Nov. 2011, doi: 10.1016/j.ijrobp.2010.07.009.

[119] T. S. Hong, W. A. Tomé, and P. M. Harari, “Heterogeneity in head and neck IMRT target design and clinical practice,” *Radiotherapy and Oncology*, vol. 103, no. 1, pp. 92–98, Apr. 2012, doi: 10.1016/j.radonc.2012.02.010.

[120] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *J Big Data*, vol. 6, no. 1, p. 60, Dec. 2019, doi: 10.1186/s40537-019-0197-0.

[121] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[122] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, “Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images,” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1240–1251, May 2016, doi: 10.1109/TMI.2016.2538465.

[123] A. Oliveira, S. Pereira, and C. A. Silva, “Augmenting data when training a CNN for retinal vessel segmentation: How to warp?,” in *2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG)*, Coimbra, Portugal: IEEE, 2017, pp. 1–4. doi: 10.1109/ENBENG.2017.7889443.

[124] J. Shijie, W. Ping, J. Peiyi, and H. Siping, “Research on data augmentation for image classification based on convolution neural networks,” in *2017 Chinese Automation Congress (CAC)*, Jinan: IEEE, Oct. 2017, pp. 4165–4170. doi: 10.1109/CAC.2017.8243510.

[125] L. Taylor and G. Nitschke, “Improving Deep Learning using Generic Data Augmentation,” *arXiv:1708.06020 [cs, stat]*, Aug. 2017, Accessed: Jun. 25, 2020. [Online]. Available: http://arxiv.org/abs/1708.06020

[126] Z. Eaton-Rosen, F. Bragman, S. Ourselin, and M. J. Cardoso, “Improving Data Augmentation for Medical Image Segmentation,” Apr. 2018, Accessed: Jun. 25, 2020. [Online]. Available: https://openreview.net/forum?id=rkBBChjiG

[127] A. Dosovitskiy, P. Fischer, J. T. Springenberg, M. Riedmiller, and T. Brox, “Discriminative Unsupervised Feature Learning with Exemplar Convolutional Neural Networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 9, pp. 1734–1747, Sep. 2016, doi: 10.1109/TPAMI.2015.2496141.

[128] S. Hauberg, O. Freifeld, A. B. L. Larsen, J. W. Fisher III, and L. K. Hansen, “Dreaming More Data: Class-dependent Distributions over Diffeomorphisms for Learned Data Augmentation,” *arXiv:1510.02795 [cs]*, Jun. 2016, Accessed: Jun. 03, 2020. [Online]. Available: http://arxiv.org/abs/1510.02795

[129] A. J. Ratner, H. R. Ehrenberg, Z. Hussain, J. Dunnmon, and C. Ré, “Learning to Compose Domain-Specific Transformations for Data Augmentation,” *arXiv:1709.01643 [cs, stat]*, Sep. 2017, Accessed: Jun. 03, 2020. [Online]. Available: http://arxiv.org/abs/1709.01643

[130] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, “AutoAugment: Learning Augmentation Strategies From Data,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA: IEEE, Jun. 2019, pp. 113–123. doi: 10.1109/CVPR.2019.00020.

[131] A. Zhao, G. Balakrishnan, F. Durand, J. V. Guttag, and A. V. Dalca, “Data augmentation using learned transformations for one-shot medical image segmentation,” *arXiv:1902.09383 [cs]*, Apr. 2019, Accessed: May 21, 2020. [Online]. Available: http://arxiv.org/abs/1902.09383

[132] X. Yi, E. Walia, and P. Babyn, “Generative adversarial network in medical imaging: A review,” *Medical Image Analysis*, vol. 58, p. 101552, Dec. 2019, doi: 10.1016/j.media.2019.101552.

[133] M. J. M. Chuquicusma, S. Hussein, J. Burt, and U. Bagci, “How to fool radiologists with generative adversarial networks? A visual turing test for lung cancer diagnosis,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Washington, DC: IEEE, Apr. 2018, pp. 240–244. doi: 10.1109/ISBI.2018.8363564.

[134] M. Moradi, A. Madani, A. Karargyris, and T. F. Syeda-Mahmood, “Chest x-ray generation and data augmentation for cardiovascular abnormality classification,” in *Medical Imaging 2018: Image Processing*, E. D. Angelini and B. A. Landman, Eds., Houston, United States: SPIE, Mar. 2018, p. 57. doi: 10.1117/12.2293971.

[135] F. Calimeri, A. Marzullo, C. Stamile, and G. Terracina, “Biomedical Data Augmentation Using Generative Adversarial Neural Networks,” in *Artificial Neural Networks and Machine Learning – ICANN 2017*, A. Lintas, S. Rovetta, P. F. M. J.

Verschure, and A. E. P. Villa, Eds., in Lecture Notes in Computer Science, vol. 10614. Cham: Springer International Publishing, 2017, pp. 626–634. doi: 10.1007/978-3-319-68612-7_71.

[136] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, “GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification,” *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018, doi: 10.1016/j.neucom.2018.09.013.

[137] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, and X. Chen, “Improved Techniques for Training GANs,” in *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds., Curran Associates, Inc., 2016, pp. 2234–2242. [Online]. Available: http://papers.nips.cc/paper/6125-improved-techniques-for-training-gans.pdf

[138] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1–3, pp. 37–52, Aug. 1987, doi: 10.1016/0169-7439(87)80084-9.

[139] Z. H. Yu, R. Kudchadker, L. Dong, Y. Zhang, L. E. Court, F. Mourtada, A. Yock, S. L. Tucker, and J. Yang, “Learning anatomy changes from patient populations to create artificial CT images for voxel-level validation of deformable image registration,” *Journal of Applied Clinical Medical Physics*, vol. 17, no. 1, pp. 246–258, Jan. 2016, doi: 10.1120/jacmp.v17i1.5888.

[140] M. Söhn, M. Birkner, D. Yan, and M. Alber, “Modelling individual geometric variation based on dominant eigenmodes of organ deformation: implementation and evaluation,” *Phys. Med. Biol.*, vol. 50, no. 24, pp. 5893–5908, Dec. 2005, doi: 10.1088/0031-9155/50/24/009.

[141] A. M. Badawi, E. Weiss, W. C. Sleeman, C. Yan, and G. D. Hugo, “Optimizing principal component models for representing interfraction variation in lung cancer

radiotherapy: Optimizing interfraction lung models," *Med. Phys.*, vol. 37, no. 9, pp. 5080–5091, Aug. 2010, doi: 10.1118/1.3481506.

[142] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active Shape Models-Their Training and Application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, Jan. 1995, doi: 10.1006/cviu.1995.1004.

[143] T. Cootes, A. Hill, C. Taylor, and J. Haslam, "Use of active shape models for locating structures in medical images," *Image and Vision Computing*, vol. 12, no. 6, pp. 355–365, Jul. 1994, doi: 10.1016/0262-8856(94)90060-4.

[144] H. Wang, L. Dong, M. F. Lii, A. L. Lee, R. de Crevoisier, R. Mohan, J. D. Cox, D. A. Kuban, and R. Cheung, "Implementation and validation of a three-dimensional deformable registration algorithm for targeted prostate cancer radiotherapy," *International Journal of Radiation Oncology*Biology*Physics*, vol. 61, no. 3, pp. 725–735, Mar. 2005, doi: 10.1016/j.ijrobp.2004.07.677.

[145] C. Yan, H. Zhong, M. Murphy, E. Weiss, and J. V. Siebers, "A pseudoinverse deformation vector field generator and its applications," *Med Phys*, vol. 37, no. 3, pp. 1117–1128, Mar. 2010, doi: 10.1118/1.3301594.

[146] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," *arXiv:1606.04797 [cs]*, Jun. 2016, Accessed: Jun. 25, 2020. [Online]. Available: http://arxiv.org/abs/1606.04797

[147] P. Dost, "Ultrasonographic biometry in normal salivary glands," *Eur Arch Otorhinolaryngol*, vol. 254, no. S1, pp. S18–S19, Jan. 1997, doi: 10.1007/BF02439713.

[148] J. Yang, A. Amini, R. Williamson, L. Zhang, Y. Zhang, R. Komaki, Z. Liao, J. Cox, J. Welsh, L. Court, and L. Dong, "Automatic contouring of brachial plexus using a multi-atlas approach for lung cancer radiation therapy," *Practical Radiation Oncology*, vol. 3, no. 4, pp. e139–e147, Oct. 2013, doi: 10.1016/j.prro.2013.01.002.

[149] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv:1412.6980 [cs]*, Jan. 2017, Accessed: Jun. 25, 2020. [Online]. Available: http://arxiv.org/abs/1412.6980

[150] D. J. Rhee, C. E. Cardenas, H. Elhalawani, R. McCarroll, L. Zhang, J. Yang, A. S. Garden, C. B. Peterson, B. M. Beadle, and L. E. Court, “Automatic detection of contouring errors using convolutional neural networks,” *Med. Phys.*, vol. 46, no. 11, pp. 5086–5097, Nov. 2019, doi: 10.1002/mp.13814.

[151] R. Zhou, Z. Liao, T. Pan, S. A. Milgrom, C. C. Pinnix, A. Shi, L. Tang, J. Yang, Y. Liu, D. Gomez, Q.-N. Nguyen, B. S. Dabaja, L. Court, and J. Yang, “Cardiac atlas development and validation for automatic segmentation of cardiac substructures,” *Radiotherapy and Oncology*, vol. 122, no. 1, pp. 66–71, Jan. 2017, doi: 10.1016/j.radonc.2016.11.016.

[152] E. Shelhamer, J. Long, and T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017, doi: 10.1109/TPAMI.2016.2572683.

[153] J. Yang, B. Haas, R. Fang, B. M. Beadle, A. S. Garden, Z. Liao, L. Zhang, P. Balter, and L. Court, “Atlas ranking and selection for automatic segmentation of the esophagus from CT scans,” *Phys. Med. Biol.*, vol. 62, no. 23, pp. 9140–9158, Nov. 2017, doi: 10.1088/1361-6560/aa94ba.

[154] D. Winkel, G. H. Bol, P. S. Kroon, B. van Asselen, S. S. Hackett, A. M. Werensteijn-Honingh, M. P. W. Intven, W. S. C. Eppinga, R. H. N. Tijssen, L. G. W. Kerkmeijer, H. C. J. de Boer, S. Mook, G. J. Meijer, J. Hes, M. Willemsen-Bosman, E. N. de Groot-van Breugel, I. M. Jürgenliemk-Schulz, and B. W. Raaymakers, “Adaptive radiotherapy: The Elekta Unity MR-linac concept,” *Clin Transl Radiat Oncol*, vol. 18, pp. 54–59, Apr. 2019, doi: 10.1016/j.ctro.2019.04.001.

[155] R. B. Ger, J. Yang, Y. Ding, M. C. Jacobsen, C. D. Fuller, R. M. Howell, H. Li, R. Jason Stafford, S. Zhou, and L. E. Court, “Accuracy of deformable image registration on magnetic resonance images in digital and physical phantoms,” *Med. Phys.*, vol. 44, no. 10, pp. 5153–5161, Oct. 2017, doi: 10.1002/mp.12406.

[156] R. B. Ger, J. Yang, Y. Ding, M. C. Jacobsen, C. E. Cardenas, C. D. Fuller, R. M. Howell, H. Li, R. J. Stafford, S. Zhou, and L. E. Court, “Synthetic head and neck and phantom images for determining deformable image registration accuracy in magnetic resonance imaging,” *Med. Phys.*, vol. 45, no. 9, pp. 4315–4321, Sep. 2018, doi: 10.1002/mp.13090.

[157] J. Yang, B. M. Beadle, A. S. Garden, B. Gunn, D. Rosenthal, K. Ang, S. Frank, R. Williamson, P. Balter, L. Court, and L. Dong, “Auto-segmentation of low-risk clinical target volume for head and neck radiation therapy,” *Practical Radiation Oncology*, vol. 4, no. 1, pp. e31–e37, Jan. 2014, doi: 10.1016/j.prro.2013.03.003.

[158] J. Yang, Y. Zhang, L. Zhang, and L. Dong, “Automatic Segmentation of Parotids from CT Scans Using Multiple Atlases,” *Medical Image Analysis for the Clinic: A Grand Challenge*, Jan. 2010.

[159] H. Wang, A. S. Garden, L. Zhang, X. Wei, A. Ahamad, D. A. Kuban, R. Komaki, J. O’Daniel, Y. Zhang, R. Mohan, and L. Dong, “Performance Evaluation of Automatic Anatomy Segmentation Algorithm on Repeat or Four-Dimensional Computed Tomography Images Using Deformable Image Registration Method,” *International Journal of Radiation Oncology*Biology*Physics*, vol. 72, no. 1, pp. 210–219, Sep. 2008, doi: 10.1016/j.ijrobp.2008.05.008.

[160] R. E. McCarroll, B. M. Beadle, P. A. Balter, H. Burger, C. E. Cardenas, S. Dalvie, D. S. Followill, K. D. Kisling, M. Mejia, K. Naidoo, C. L. Nelson, C. B. Peterson, K. Vorster, J. Wetter, L. Zhang, L. E. Court, and J. Yang, “Retrospective Validation and Clinical Implementation of Automated Contouring of Organs at Risk in the Head and

Neck: A Step Toward Automated Radiation Treatment Planning for Low- and Middle-Income Countries,” *JGO*, no. 4, pp. 1–11, Aug. 2018, doi: 10.1200/JGO.18.00055.

[161] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *J Big Data*, vol. 6, no. 1, p. 60, Jul. 2019, doi: 10.1186/s40537-019-0197-0.

[162] C. Bowles, L. Chen, R. Guerrero, P. Bentley, R. Gunn, A. Hammers, D. A. Dickie, M. V. Hernández, J. Wardlaw, and D. Rueckert, “GAN Augmentation: Augmenting Training Data using Generative Adversarial Networks,” *arXiv:1810.10863 [cs]*, Oct. 2018, Accessed: Nov. 11, 2020. [Online]. Available: http://arxiv.org/abs/1810.10863

[163] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, Jun. 2001, doi: 10.1109/34.927467.

[164] C. Fitzmaurice, C. Allen, R. M. Barber, L. Barregard, Z. A. Bhutta, H. Brenner, D. J. Dicker, O. Chimed-Orchir, R. Dandona, and L. Dandona, “Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study,” *JAMA oncology*, vol. 3, no. 4, pp. 524–548, 2017.

[165] I. J. Boero, A. J. Paravati, B. Xu, E. E. Cohen, L. K. Mell, Q.-T. Le, and J. D. Murphy, “Importance of radiation oncologist experience among patients with head-and-neck cancer treated with intensity-modulated radiation therapy,” *Journal of Clinical Oncology*, vol. 34, no. 7, p. 684, 2016.

[166] M. Mukesh, R. Benson, R. Jena, A. Hoole, T. Roques, C. Scrase, C. Martin, G. A. Whitfield, J. Gemmill, and S. Jefferies, “Interobserver variation in clinical target volume and organs at risk segmentation in post-parotidectomy radiotherapy: can segmentation protocols help?,” *The British journal of radiology*, vol. 85, no. 1016, pp. e530–e536, 2012.

[167] S. L. Breen, J. Publicover, S. De Silva, G. Pond, K. Brock, B. O'Sullivan, B. Cummings, L. Dawson, A. Keller, and J. Kim, "Intraobserver and interobserver variability in GTV delineation on FDG-PET-CT images of head and neck cancers," *International Journal of Radiation Oncology* Biology* Physics*, vol. 68, no. 3, pp. 763–770, 2007.

[168] K. Wang, D. E. Heron, J. C. Flickinger, J.-C. M. Rwigema, R. L. Ferris, G. J. Kubicek, J. P. Ohr, A. E. Quinn, C. Ozhasoglu, and B. F. Branstetter, "A retrospective, deformable registration analysis of the impact of PET-CT planning on patterns of failure in stereotactic body radiation therapy for recurrent head and neck cancer," *Head & neck oncology*, vol. 4, pp. 1–10, 2012.

[169] A. M. Chen, D. G. Farwell, Q. Luu, L. M. Chen, S. Vijayakumar, and J. A. Purdy, "Misses and near-misses after postoperative radiation therapy for head and neck cancer: Comparison of IMRT and non-IMRT techniques in the CT-simulation era," *Head & neck*, vol. 32, no. 11, pp. 1452–1459, 2010.

[170] A. M. Chen, R. Chin, P. Beron, T. Yoshizaki, A. G. Mikaeilian, and M. Cao, "Inadequate target volume delineation and local–regional recurrence after intensity-modulated radiotherapy for human papillomavirus-positive oropharynx cancer," *Radiotherapy and Oncology*, vol. 123, no. 3, pp. 412–418, 2017.

[171] N.-N. Chung, L.-L. Ting, W.-C. Hsu, L. T. Lui, and P.-M. Wang, "Impact of magnetic resonance imaging versus CT on nasopharyngeal carcinoma: primary tumor target delineation for radiotherapy," *Head & Neck: Journal for the Sciences and Specialties of the Head and Neck*, vol. 26, no. 3, pp. 241–246, 2004.

[172] V. Andrearczyk, V. Oreiller, M. Vallières, J. Castelli, H. Elhalawani, M. Jreige, S. Boughdad, J. O. Prior, and A. Depeursinge, "Automatic segmentation of head and neck tumors and nodal metastases in PET-CT scans," in *Medical imaging with deep learning*, PMLR, 2020, pp. 33–43.

[173] A. Myronenko, M. M. R. Siddiquee, D. Yang, Y. He, and D. Xu, “Automated Head and Neck Tumor Segmentation from 3D PET/CT HECKTOR 2022 Challenge Report,” in *Head and Neck Tumor Segmentation and Outcome Prediction: Third Challenge, HECKTOR 2022, Held in Conjunction with MICCAI 2022, Singapore, September 22, 2022, Proceedings*, Springer, 2023, pp. 31–37.

[174] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation,” *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.

[175] R. Beare, B. Lowekamp, and Z. Yaniv, “Image segmentation, registration and characterization in R with SimpleITK,” *Journal of statistical software*, vol. 86, 2018.

[176] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellyn, and W. Eubank, “Nonrigid multimodality image registration,” in *Medical imaging 2001: image processing*, Spie, 2001, pp. 1609–1620.

[177] C. E. Cardenas, A. S. Mohamed, R. Tao, A. J. Wong, M. J. Awan, S. Kuruvila, M. Aristophanous, G. B. Gunn, J. Phan, and B. M. Beadle, “Prospective qualitative and quantitative analysis of real-time peer review quality assurance rounds incorporating direct physical examination for head and neck cancer radiation therapy,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 98, no. 3, pp. 532–540, 2017.

[178] Q. J. Wu, T. Li, Q. Wu, and F.-F. Yin, “Adaptive Radiation Therapy: Technical Components and Clinical Applications,” *The Cancer Journal*, vol. 17, no. 3, pp. 182–189, May 2011, doi: 10.1097/PPO.0b013e31821da9d8.

[179] J. Castelli, A. Simon, C. Lafond, N. Perichon, B. Rigaud, E. Chajon, B. De Bari, M. Ozsahin, J. Bourhis, and R. de Crevoisier, “Adaptive radiotherapy for head and neck cancer,” *Acta Oncologica*, vol. 57, no. 10, pp. 1284–1292, Oct. 2018, doi: 10.1080/0284186X.2018.1505053.

[180] A. Bahl, A. Elangovan, C. B. Dracham, S. Kaur, A. S. Oinam, G. Trivedi, R. Verma, N. K. Panda, and S. Ghoshal, “Analysis of volumetric and dosimetric changes in mid treatment CT scan in carcinoma nasopharynx: implications for adaptive radiotherapy,” *J Exp Ther Oncol*, vol. 13, no. 1, pp. 33–39, Jan. 2019.

[181] N. Burela, T. P. Soni, N. Patni, and T. Natarajan, “Adaptive intensity-modulated radiotherapy in head-and-neck cancer: A volumetric and dosimetric study,” *J Cancer Res Ther*, vol. 15, no. 3, pp. 533–538, Sep. 2019, doi: 10.4103/jcrt.JCRT_594_17.

[182] L. Belshaw, C. E. Agnew, D. M. Irvine, K. P. Rooney, and C. K. McGarry, “Adaptive radiotherapy for head and neck cancer reduces the requirement for rescans during treatment due to spinal cord dose,” *Radiation Oncology*, vol. 14, pp. 1–7, 2019.

[183] P. Kupelian and J.-J. Sonke, “Magnetic resonance–guided adaptive radiotherapy: a solution to the future,” in *Seminars in radiation oncology*, Elsevier, 2014, pp. 227–232.

[184] B. W. Raaymakers, J. J. W. Lagendijk, J. Overweg, J. G. M. Kok, A. J. E. Raaijmakers, E. M. Kerkhof, R. W. Van Der Put, I. Meijsing, S. P. M. Crijns, and F. Benedosso, “Integrating a 1.5 T MRI scanner with a 6 MV accelerator: proof of concept,” *Physics in Medicine & Biology*, vol. 54, no. 12, p. N229, 2009.

[185] A. M. Owrangi, P. B. Greer, and C. K. Glide-Hurst, “MRI-only treatment planning: benefits and challenges,” *Physics in Medicine & Biology*, vol. 63, no. 5, p. 05TR01, 2018.

[186] M. A. Schmidt and G. S. Payne, “Radiotherapy planning using MRI,” *Physics in Medicine & Biology*, vol. 60, no. 22, p. R323, 2015.

[187] X. Chen, A. Diaz-Pinto, N. Ravikumar, and A. F. Frangi, “Deep learning in medical image registration,” *Progress in Biomedical Engineering*, vol. 3, no. 1, p. 012003, 2021.

[188] Y. Fu, Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang, “Deep learning in medical image registration: a review,” *Physics in Medicine & Biology*, vol. 65, no. 20, p. 20TR01, 2020.

[189] G. Haskins, U. Kruger, and P. Yan, “Deep learning in medical image registration: a survey,” *Machine Vision and Applications*, vol. 31, pp. 1–18, 2020.

[190] J. Chen, E. C. Frey, Y. He, W. P. Segars, Y. Li, and Y. Du, “Transmorph: Transformer for unsupervised medical image registration,” *Medical image analysis*, vol. 82, p. 102615, 2022.

[191] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, “VoxelMorph: a learning framework for deformable medical image registration,” *IEEE transactions on medical imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[192] M. Jaderberg, K. Simonyan, and A. Zisserman, “Spatial transformer networks,” *Advances in neural information processing systems*, vol. 28, 2015.

[193] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, “Multimodality image registration by maximization of mutual information,” *IEEE transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997.

[194] P. Viola and W. M. Wells, “Alignment by maximization of mutual information,” in *Proceedings of IEEE International Conference on Computer Vision*, IEEE, 1995, pp. 16–23.

[195] M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, S. M. Brady, and J. A. Schnabel, “MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration,” *Medical Image Analysis*, vol. 16, no. 7, pp. 1423–1435, Oct. 2012, doi: 10.1016/j.media.2012.05.008.

[196] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[197] E. M. McKenzie, A. Santhanam, D. Ruan, D. O'Connor, M. Cao, and K. Sheng, "Multimodality image registration in the head-and-neck using a deep learning-derived synthetic CT as a bridge," *Medical physics*, vol. 47, no. 3, pp. 1094–1104, 2020.

[198] Z. Xu, J. Luo, J. Yan, R. Pulya, X. Li, W. Wells, and J. Jagadeesan, "Adversarial Uni- and Multi-modal Stream Networks for Multimodal Image Registration," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*, A. L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M. A. Zuluaga, S. K. Zhou, D. Racoceanu, and L. Joskowicz, Eds., Cham: Springer International Publishing, 2020, pp. 222–232.

[199] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:2102.04306*, 2021.

[200] Y. Zhang, H. Liu, and Q. Hu, "Transfuse: Fusing transformers and cnns for medical image segmentation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*, Springer, 2021, pp. 14–24.

[201] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part III*, Springer, 2023, pp. 205–218.

[202] A. Güngör, B. Askin, D. A. Soydan, E. U. Saritas, C. B. Top, and T. Çukur, "TranSMS: Transformers for super-resolution calibration in magnetic particle imaging," *IEEE Transactions on Medical Imaging*, vol. 41, no. 12, pp. 3562–3574, 2022.

[203] P. Guo, Y. Mei, J. Zhou, S. Jiang, and V. M. Patel, "ReconFormer: Accelerated MRI reconstruction using recurrent transformer," *arXiv preprint arXiv:2201.09376*, 2022.

[204] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, and S. Gelly, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.

[205] J. Wu, R. Hu, Z. Xiao, J. Chen, and J. Liu, “Vision Transformer-based recognition of diabetic retinopathy grade,” *Medical Physics*, vol. 48, no. 12, pp. 7850–7863, 2021.

[206] J. Chen, Y. He, E. C. Frey, Y. Li, and Y. Du, “Vit-v-net: Vision transformer for unsupervised volumetric medical image registration,” *arXiv preprint arXiv:2104.06468*, 2021.

[207] Y. Zhang, Y. Pei, and H. Zha, “Learning dual transformer network for diffeomorphic registration,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part IV 24*, Springer, 2021, pp. 129–138.

[208] L. Liu, Z. Huang, P. Liò, C.-B. Schönlieb, and A. I. Aviles-Rivero, “Pc-swinmorph: patch representation for unsupervised medical image registration and segmentation,” *arXiv preprint arXiv:2203.05684*, 2022.

[209] R. Beare, B. Lowekamp, and Z. Yaniv, “Image segmentation, registration and characterization in R with SimpleITK,” *Journal of statistical software*, vol. 86, 2018.

[210] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellyn, and W. Eubank, “Nonrigid multimodality image registration,” in *Medical imaging 2001: image processing*, Spie, 2001, pp. 1609–1620.

[211] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.

[212] M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, M. Brady, and J. A. Schnabel, “MIND: Modality independent neighbourhood descriptor for multi-modal deformable registration,” *Medical image analysis*, vol. 16, no. 7, pp. 1423–1435, 2012.

[213] K. A. Wahid, R. He, B. A. McDonald, B. M. Anderson, T. Salzillo, S. Mulder, J. Wang, C. S. Sharafi, L. A. McCoy, M. A. Naser, S. Ahmed, K. L. Sanders, A. S. R. Mohamed, Y. Ding, J. Wang, K. Hutcheson, S. Y. Lai, C. D. Fuller, and L. V. van Dijk, "Intensity standardization methods in magnetic resonance imaging of head and neck cancer," *Physics and Imaging in Radiation Oncology*, vol. 20, pp. 88–93, 2021, doi: https://doi.org/10.1016/j.phro.2021.11.001.

[214] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[215] C. F. Njeh, "Tumor delineation: The weakest link in the search for accuracy in radiotherapy," *J Med Phys*, vol. 33, no. 4, pp. 136–140, 2008, doi: 10.4103/0971-6203.44472.

[216] M. A. Schmidt and G. S. Payne, "Radiotherapy Planning using MRI," *Phys Med Biol*, vol. 60, no. 22, pp. R323–R361, Nov. 2015, doi: 10.1088/0031-9155/60/22/R323.

[217] N.-N. Chung, L.-L. Ting, W.-C. Hsu, L. T. Lui, and P.-M. Wang, "Impact of magnetic resonance imaging versus CT on nasopharyngeal carcinoma: primary tumor target delineation for radiotherapy," *Head Neck*, vol. 26, no. 3, pp. 241–246, Mar. 2004, doi: 10.1002/hed.10378.

[218] C. R. Rasch, R. J. Steenbakkers, I. Fitton, J. C. Duppen, P. J. Nowak, F. A. Pameijer, A. Eisbruch, J. H. Kaanders, F. Paulsen, and M. van Herk, "Decreased 3D observer variation with matched CT-MRI, for target delineation in Nasopharynx cancer," *Radiat Oncol*, vol. 5, no. 1, p. 21, Dec. 2010, doi: 10.1186/1748-717X-5-21.

[219] Y. L. Dai and A. D. King, "State of the art MRI in head and neck cancer," *Clinical radiology*, vol. 73, no. 1, pp. 45–59, 2018.

[220] J. Edmund and T. Nyholm, "A review of substitute CT generation for MRI-only radiation therapy," *Radiation Oncology*, vol. 12, Jan. 2017, doi: 10.1186/s13014-016-0747-y.

[221] N. Burgos, F. Guerreiro, J. McClelland, B. Presles, M. Modat, S. Nill, D. Dearnaley, N. deSouza, U. Oelfke, A.-C. Knopf, S. Ourselin, and M. J. Cardoso, “Iterative framework for the joint segmentation and CT synthesis of MR images: application to MRI-only radiotherapy treatment planning,” *Phys. Med. Biol.*, vol. 62, no. 11, pp. 4237–4253, 2017, doi: 10.1088/1361-6560/aa66bf.

[222] T. Nyholm and J. Jonsson, “Counterpoint: Opportunities and challenges of a magnetic resonance imaging-only radiotherapy work flow,” *Semin Radiat Oncol*, vol. 24, no. 3, pp. 175–180, Jul. 2014, doi: 10.1016/j.semradonc.2014.02.005.

[223] K. Ulin, M. M. Urie, and J. M. Cherlow, “Results of a Multi-Institutional Benchmark Test for Cranial CT/MR Image Registration,” *Int J Radiat Oncol Biol Phys*, vol. 77, no. 5, pp. 1584–1589, Aug. 2010, doi: 10.1016/j.ijrobp.2009.10.017.

[224] V. S. Khoo and D. L. Joon, “New developments in MRI for target volume delineation in radiotherapy,” *Br J Radiol*, vol. 79 Spec No 1, pp. S2-15, Sep. 2006, doi: 10.1259/bjr/41321492.

[225] S. Boeke, D. Mönnich, J. E. van Timmeren, and P. Balermpas, “MR-Guided Radiotherapy for Head and Neck Cancer: Current Developments, Perspectives, and Challenges,” *Front Oncol*, vol. 11, p. 616156, Mar. 2021, doi: 10.3389/fonc.2021.616156.

[226] D. Winkel, G. H. Bol, P. S. Kroon, B. van Asselen, S. S. Hackett, A. M. Werensteijn-Honingh, M. P. W. Intven, W. S. C. Eppinga, R. H. N. Tijssen, L. G. W. Kerkmeijer, H. C. J. de Boer, S. Mook, G. J. Meijer, J. Hes, M. Willemsen-Bosman, E. N. de Groot-van Breugel, I. M. Jürgenliemk-Schulz, and B. W. Raaymakers, “Adaptive radiotherapy: The Elekta Unity MR-linac concept,” *Clin Transl Radiat Oncol*, vol. 18, pp. 54–59, Sep. 2019, doi: 10.1016/j.ctro.2019.04.001.

[227] W. Zheng, J. P. Kim, M. Kadbi, B. Movsas, I. J. Chetty, and C. K. Glide-Hurst, “Magnetic Resonance-Based Automatic Air Segmentation for Generation of Synthetic

Computed Tomography Scans in the Head Region," *Int J Radiat Oncol Biol Phys*, vol. 93, no. 3, pp. 497–506, Nov. 2015, doi: 10.1016/j.ijrobp.2015.07.001.

[228] D. Andreasen, K. Van Leemput, and J. M. Edmund, "A patch-based pseudo-CT approach for MRI-only radiotherapy in the pelvis," *Med Phys*, vol. 43, no. 8, p. 4742, Aug. 2016, doi: 10.1118/1.4958676.

[229] H. Arabi and H. Zaidi, "One registration multi-atlas-based pseudo-CT generation for attenuation correction in PET/MRI," *Eur J Nucl Med Mol Imaging*, vol. 43, no. 11, pp. 2021–2035, Oct. 2016, doi: 10.1007/s00259-016-3422-5.

[230] J. A. Dowling, N. Burdett, P. B. Greer, J. Sun, J. Parker, P. Pichler, P. Stanwell, S. Chandra, D. Rivest-Hénault, S. Ghose, O. Salvado, and J. Fripp, "Automatic Atlas Based Electron Density and Structure Contouring for MRI-based Prostate Radiation Therapy on the Cloud," *J. Phys.: Conf. Ser.*, vol. 489, p. 012048, 2014, doi: 10.1088/1742-6596/489/1/012048.

[231] J. A. Dowling, J. Sun, P. Pichler, D. Rivest-Hénault, S. Ghose, H. Richardson, C. Wratten, J. Martin, J. Arm, L. Best, S. S. Chandra, J. Fripp, F. W. Menk, and P. B. Greer, "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*, vol. 93, no. 5, pp. 1144–1153, Dec. 2015, doi: 10.1016/j.ijrobp.2015.08.045.

[232] M. S. R. Gudur, W. Hara, Q.-T. Le, L. Wang, L. Xing, and R. Li, "A unifying probabilistic Bayesian approach to derive electron density from MRI for radiation therapy treatment planning," *Phys Med Biol*, vol. 59, no. 21, pp. 6595–6606, Nov. 2014, doi: 10.1088/0031-9155/59/21/6595.

[233] J. Uh, T. E. Merchant, Y. Li, X. Li, and C. Hua, "MRI-based treatment planning with pseudo CT generated through atlas registration," *Med Phys*, vol. 41, no. 5, p. 051711, May 2014, doi: 10.1118/1.4873315.

[234] T. Huynh, Y. Gao, J. Kang, L. Wang, P. Zhang, J. Lian, D. Shen, and Alzheimer's Disease Neuroimaging Initiative, "Estimating CT Image From MRI Data Using Structured Random Forest and Auto-Context Model," *IEEE Trans Med Imaging*, vol. 35, no. 1, pp. 174–183, Jan. 2016, doi: 10.1109/TMI.2015.2461533.

[235] X. Cao, J. Yang, Y. Gao, Y. Guo, G. Wu, and D. Shen, "Dual-core steered non-rigid registration for multi-modal images via bi-directional image synthesis," *Med Image Anal*, vol. 41, pp. 18–31, Oct. 2017, doi: 10.1016/j.media.2017.05.004.

[236] X. Han, "MR-based synthetic CT generation using a deep convolutional neural network method," *Med Phys*, vol. 44, no. 4, pp. 1408–1419, Apr. 2017, doi: 10.1002/mp.12155.

[237] A. L. Chin, A. Lin, S. Anamalayil, and B.-K. K. Teo, "Feasibility and limitations of bulk density assignment in MRI for head and neck IMRT treatment planning," *J Appl Clin Med Phys*, vol. 15, no. 5, p. 4851, Sep. 2014, doi: 10.1120/jacmp.v15i5.4851.

[238] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[239] D. Nie, X. Cao, Y. Gao, L. Wang, and D. Shen, "Estimating CT Image from MRI Data Using 3D Fully Convolutional Networks," *Deep Learn Data Label Med Appl (2016)*, vol. 2016, pp. 170–178, 2016, doi: 10.1007/978-3-319-46976-8_18.

[240] A. M. Dinkla, M. C. Florkow, M. Maspero, M. H. F. Savenije, F. Zijlstra, P. A. H. Doornaert, M. van Stralen, M. E. P. Philippens, C. A. T. van den Berg, and P. R. Seevinck, "Dosimetric evaluation of synthetic CT for head and neck radiotherapy generated by a patch-based three-dimensional convolutional neural network," *Med Phys*, vol. 46, no. 9, pp. 4095–4104, Sep. 2019, doi: 10.1002/mp.13663.

[241] A. M. Dinkla, J. M. Wolterink, M. Maspero, M. H. F. Savenije, J. J. C. Verhoeff, E. Seravalli, I. Išgum, P. R. Seevinck, and C. A. T. van den Berg, “MR-Only Brain Radiation Therapy: Dosimetric Evaluation of Synthetic CTs Generated by a Dilated Convolutional Neural Network,” *Int J Radiat Oncol Biol Phys*, vol. 102, no. 4, pp. 801–812, Nov. 2018, doi: 10.1016/j.ijrobp.2018.05.058.

[242] J. Wolterink, A. Dinkla, M. Savenije, P. Seevinck, C. Berg, and I. Išgum, “Deep MR to CT Synthesis Using Unpaired Data,” in *SASHIMI@MICCAI*, 2017. doi: 10.1007/978-3-319-68127-6_2.

[243] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2014. Accessed: May 26, 2022. [Online]. Available: https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html

[244] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[245] D. Cusumano, J. Lenkowicz, C. Votta, L. Boldrini, L. Placidi, F. Catucci, N. Dinapoli, M. V. Antonelli, A. Romano, V. De Luca, G. Chiloiro, L. Indovina, and V. Valentini, “A deep learning approach to generate synthetic CT in low field MR-guided adaptive radiotherapy for abdominal and pelvic cases,” *Radiother Oncol*, vol. 153, pp. 205–212, Dec. 2020, doi: 10.1016/j.radonc.2020.10.018.

[246] Y. Peng, S. Chen, A. Qin, M. Chen, X. Gao, Y. Liu, J. Miao, H. Gu, C. Zhao, X. Deng, and Z. Qi, “Magnetic resonance-based synthetic computed tomography images generated using generative adversarial networks for nasopharyngeal carcinoma

radiotherapy treatment planning," *Radiotherapy and Oncology*, vol. 150, pp. 217–224, Sep. 2020, doi: 10.1016/j.radonc.2020.06.049.

[247] M. Maspero, L. G. Bentvelzen, M. H. Savenije, F. Guerreiro, E. Seravalli, G. O. Janssens, C. A. van den Berg, and M. E. Philippens, "Deep learning-based synthetic CT generation for paediatric brain MR-only photon and proton radiotherapy," *Radiotherapy and Oncology*, vol. 153, pp. 197–204, 2020.

[248] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks." arXiv, Aug. 24, 2020. Accessed: May 26, 2022. [Online]. Available: http://arxiv.org/abs/1703.10593

[249] J. M. Wolterink, A. M. Dinkla, M. H. F. Savenije, P. R. Seevinck, C. A. T. van den Berg, and I. Isgum, "Deep MR to CT Synthesis using Unpaired Data," *arXiv:1708.01155 [cs]*, Aug. 2017, Accessed: Sep. 28, 2020. [Online]. Available: http://arxiv.org/abs/1708.01155

[250] Y. Lei, J. Harms, T. Wang, Y. Liu, H.-K. Shu, A. B. Jani, W. J. Curran, H. Mao, T. Liu, and X. Yang, "MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks," *Med Phys*, vol. 46, no. 8, pp. 3565–3581, Aug. 2019, doi: 10.1002/mp.13617.

[251] W. Li, Y. Li, W. Qin, X. Liang, J. Xu, J. Xiong, and Y. Xie, "Magnetic resonance image (MRI) synthesis from brain computed tomography (CT) images based on deep learning methods for magnetic resonance (MR)-guided radiotherapy," *Quant Imaging Med Surg*, vol. 10, no. 6, pp. 1223–1236, Jun. 2020, doi: 10.21037/qims-19-885.

[252] Y. Liu, A. Chen, H. Shi, S. Huang, W. Zheng, Z. Liu, Q. Zhang, and X. Yang, "CT synthesis from MRI using multi-cycle GAN for head-and-neck radiation therapy," *Computerized Medical Imaging and Graphics*, vol. 91, p. 101953, 2021, doi: https://doi.org/10.1016/j.compmedimag.2021.101953.

[253] Y. Liu, Y. Lei, T. Wang, O. Kayode, S. Tian, T. Liu, P. Patel, W. J. Curran, L. Ren, and X. Yang, "MRI-based treatment planning for liver stereotactic body radiotherapy: validation of a deep learning-based synthetic CT generation method," *Br J Radiol*, vol. 92, no. 1100, p. 20190067, Aug. 2019, doi: 10.1259/bjr.20190067.

[254] G. Shafai-Erfani, Y. Lei, Y. Liu, Y. Wang, T. Wang, J. Zhong, T. Liu, M. McDonald, W. J. Curran, J. Zhou, H.-K. Shu, and X. Yang, "MRI-Based Proton Treatment Planning for Base of Skull Tumors," *Int J Part Ther*, vol. 6, no. 2, pp. 12–25, 2019, doi: 10.14338/IJPT-19-00062.1.

[255] H. Yang, J. Sun, A. Carass, C. Zhao, J. Lee, J. L. Prince, and Z. Xu, "Unsupervised MR-to-CT Synthesis Using Structure-Constrained CycleGAN," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 4249–4261, 2020, doi: 10.1109/TMI.2020.3015379.

[256] C. Hong, D.-H. Lee, and B. S. Han, "Characteristics of geometric distortion correction with increasing field-of-view in open-configuration MRI," *Magn Reson Imaging*, vol. 32, no. 6, pp. 786–790, Jul. 2014, doi: 10.1016/j.mri.2014.02.007.

[257] M. Qi, Y. Li, A. Wu, Q. Jia, B. Li, W. Sun, Z. Dai, X. Lu, L. Zhou, X. Deng, and T. Song, "Multi-sequence MR image-based synthetic CT generation using a generative adversarial network for head and neck MRI-only radiotherapy," *Medical Physics*, vol. 47, no. 4, pp. 1880–1894, 2020, doi: https://doi.org/10.1002/mp.14075.

[258] J. L. Bloem, M. Reijnierse, T. W. Huizinga, and A. H. van der Helm-van, "MR signal intensity: staying on the bright side in MR image interpretation," *RMD open*, vol. 4, no. 1, p. e000728, 2018.

[259] K. A. Wahid, R. He, B. A. McDonald, B. M. Anderson, T. Salzillo, S. Mulder, J. Wang, C. S. Sharafi, L. A. McCoy, and M. A. Naser, "MRI Intensity Standardization Evaluation Design for Head and Neck Quantitative Imaging Applications," *MedRxiv*, 2021.

[260] B. Tang, M. Liu, B. Wang, P. Diao, J. Li, X. Feng, F. Wu, X. Yao, X. Liao, and Q. Hou, “Improving the clinical workflow of a MR-Linac by dosimetric evaluation of synthetic CT,” *Frontiers in Oncology*, vol. 12, 2022.

[261] L. A. Olteanu, I. Madani, W. De Neve, T. Vercauteren, and W. De Gersem, “Evaluation of deformable image coregistration in adaptive dose painting by numbers for head-and-neck cancer,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 83, no. 2, pp. 696–703, 2012.

[262] Y. J. Graves, A.-A. Smith, D. Mcilvena, Z. Manilay, Y. K. Lai, R. Rice, L. Mell, X. Jia, S. B. Jiang, and L. Cerviño, “A deformable head and neck phantom with in-vivo dosimetry for adaptive radiotherapy quality assurance,” *Medical physics*, vol. 42, no. 4, pp. 1490–1497, 2015.

[263] I. J. Chetty and M. Rosu-Bubulac, “Deformable registration for dose accumulation,” in *Seminars in radiation oncology*, Elsevier, 2019, pp. 198–208.

[264] E. Boman, M. Kapanen, L. Pickup, and S.-L. Lahtela, “Importance of deformable image registration and biological dose summation in planning of radiotherapy retreatments,” *Medical Dosimetry*, vol. 42, no. 4, pp. 296–303, 2017.

[265] H. Paganetti, H. Jiang, J. A. Adams, G. T. Chen, and E. Rietzel, “Monte Carlo simulations with time-dependent geometries to investigate effects of organ motion with high temporal resolution,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 60, no. 3, pp. 942–950, 2004.

[266] M. Rosu, I. J. Chetty, J. M. Balter, M. L. Kessler, D. L. McShan, and R. K. Ten Haken, “Dose reconstruction in deforming lung anatomy: dose grid size effects and clinical implications,” *Medical physics*, vol. 32, no. 8, pp. 2487–2495, 2005.

[267] G. D. Hugo and M. Rosu, “Advances in 4D radiation therapy for managing respiration: part I–4D imaging,” *Zeitschrift für Medizinische Physik*, vol. 22, no. 4, pp. 258–271, 2012.

[268] E. Heath and J. Seuntjens, “A direct voxel tracking method for four-dimensional Monte Carlo dose calculations in deforming anatomy,” *Medical physics*, vol. 33, no. 2, pp. 434–445, 2006.

[269] J. V. Siebers and H. Zhong, “An energy transfer method for 4D Monte Carlo dose calculation,” *Medical physics*, vol. 35, no. 9, pp. 4096–4105, 2008.

# Vita

Yao Zhao was born in China. After graduating from Jiangsu Xishan Senior High School, he entered Lanzhou University in China He obtained his Bachelor of Science degree with a major in nuclear physics from Lanzhou University in 2016. Following this, he continued his education at Duke Kunshan University, where he obtained a Master of Science degree majoring in Medical Physics in 2019. In Fall 2019, he began the pursuit of a Doctor of Philosophy in Medical Physics at The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences. After his PhD, he will continue his training at MD Anderson Cancer Center through the Radiation Physics Medical Physics Residency Program.