IMAGING IN MEDICAL DIAGNOSIS AND THERAPY

MEDICAL IMAGE SYNTHESIS

Methods and Clinical Applications

Edited by

Xiaofeng Yang

CRC Press
Taylor & Francis Group

Medical Image Synthesis

Image synthesis across and within medical imaging modalities is an active area of research with broad applications in radiology and radiation oncology. This book covers the principles and methods of medical image synthesis, along with state-of-the-art research.

First, various traditional non-learning-based, traditional machine-learning-based, and recent deep-learning-based medical image synthesis methods are reviewed. Second, specific applications of different inter- and intra-modality image synthesis tasks and of synthetic image-aided segmentation and registration are introduced and summarized, listing and highlighting the proposed methods, study designs, and reported performances with the related clinical applications of representative studies. Third, the clinical usages of medical image synthesis, such as treatment planning and image-guided adaptive radiotherapy, are discussed. Last, the limitations and current challenges of various medical synthesis applications are explored, along with future trends and potential solutions to solve these difficulties.

The benefits of medical image synthesis have sparked growing interest in a number of advanced clinical applications, such as magnetic resonance imaging (MRI)-only radiation therapy treatment planning and positron emission tomography (PET)/MRI scanning. This book will be a comprehensive and exciting resource for undergraduates, graduates, researchers, and practitioners.

Xiaofeng Yang is an Associate Professor and Vice Chair in Medical Physics Research in the Department of Radiation Oncology at Emory University, where he completed his postdoctoral and medical physics residency training. He holds B.S., M.S., and Ph.D. degrees in Biomedical Engineering from Xi'an Jiaotong University and completed his Ph.D. training and thesis at Emory University. Dr. Yang is also an adjunct faculty member in the medical physics department at Georgia Institute of Technology, as well as in the biomedical informatics department and the Wallace H. Coulter Department of Biomedical Engineering at Emory University and Georgia Institute of Technology. As a board-certified medical physicist with expertise in image-guided radiotherapy, artificial intelligence (AI), multimodality medical imaging, and medical image analysis, Dr. Yang leads the Deep Biomedical Imaging Laboratory. His lab is focused on developing novel AI-aided analytical and computational tools to enhance the role of quantitative imaging in cancer treatment and to improve the accuracy and precision of radiation therapy. His research has been funded by the NIH, DOD, and industrial funding agencies. Dr. Yang has published over 180 peer-reviewed journal papers and has received numerous scientific awards, including awards from SPIE Medical Imaging, AAPM, ASTRO, and SNMMI. In 2020, he received the John Laughlin Young Scientist Award from the American Association of Physicists in Medicine. Dr. Yang currently serves as an Associate Editor for Medical Physics and the Journal of Applied Clinical Medical Physics.

Imaging in Medical Diagnosis and Therapy
Series Editors: Bruce R. Thomadsen, David W. Jordan

Beam's Eye View Imaging in Radiation Oncology
Ross I. Berbeco, Ph.D.

Principles and Practice of Image-Guided Radiation Therapy of Lung Cancer
Jing Cai, Joe Y. Chang, Fang-Fang Yin

Radiochromic Film: Role and Applications in Radiation Dosimetry
Indra J. Das

Clinical 3D Dosimetry in Modern Radiation Therapy
Ben Mijnheer

Hybrid Imaging in Cardiovascular Medicine
Yi-Hwa Liu, Albert J. Sinusas

Observer Performance Methods for Diagnostic Imaging: Foundations, Modeling, and Applications with R-Based Examples
Dev P. Chakraborty

Ultrasound Imaging and Therapy
Aaron Fenster, James C. Lacefield

Dose, Benefit, and Risk in Medical Imaging
Lawrence T. Dauer, Bae P. Chu, Pat B. Zanzonico

Big Data in Radiation Oncology
Jun Deng, Lei Xing

Monte Carlo Techniques in Radiation Therapy: Introduction, Source Modelling, and Patient Dose Calculations, Second Edition
Frank Verhaegen, Joao Seco

Monte Carlo Techniques in Radiation Therapy: Applications to Dosimetry, Imaging, and Preclinical Radiotherapy, Second Edition
Joao Seco, Frank Verhaegen

Introductory Biomedical Imaging: Principles and Practice from Microscopy to MRI
Bethe A. Scalettar, James R. Abney

Medical Image Synthesis: Methods and Clinical Applications
Xiaofeng Yang

For more information about this series, please visit: https://www.routledge.com/Imaging-in-Medical-Diagnosis-and-Therapy/book-series/CRCIMAINMED

Medical Image Synthesis
Methods and Clinical Applications

Edited by
Xiaofeng Yang

CRC Press
Taylor & Francis Group
Boca Raton New York London

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

Front cover image: vs148/Shutterstock.

First edition published 2024
by CRC Press
2385 NW Executive Center Drive, Suite 320, Boca Raton FL 33431

and by CRC Press
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

CRC Press is an imprint of Taylor & Francis Group, LLC

ISBN: 978-1-032-13388-1 (hbk)
ISBN: 978-1-032-15284-4 (pbk)
ISBN: 978-1-003-24345-8 (ebk)

DOI: 10.1201/9781003243458

Typeset in Palatino
by SPi Technologies India Pvt Ltd (Straive)

Contents

Contributors

Jing Cai
Hong Kong Polytechnic University
Kowloon, Hong Kong

Jin Chen
University of Kentucky
Lexington, Kentucky

Ying Chen
Faculty of Information Technology
Beijing University of Technology
Beijing, China

Zhiming Cheng
School of Automation
Hangzhou Dianzi University
Hangzhou, China

Xianjin Dai
Department of Radiation Oncology
Stanford University
California, USA

Ruogu Fang
J. Crayton Pruitt Family Department of Biomedical Engineering
and
Center for Cognitive Aging and Memory
McKnight Brain Institute
and
Department of Electrical and Computer Engineering
Herbert Wertheim College of Engineering
University of Florida
Florida, USA

Yabo Fu
Memorial Sloan Kettering Cancer Center
New York, USA

Kuang Gong
Gordon Center for Medical Imaging
Massachusetts General Hospital
Harvard Medical School
Massachusetts, USA

Kai Huang
Department of Biomedical Sciences
The University of Texas Health Science Center at Houston
Texas, USA

Yu-Hua Huang
Hong Kong Polytechnic University
Kowloon, Hong Kong

Se-In Jang
Gordon Center for Medical Imaging
Massachusetts General Hospital
Harvard Medical School
Massachusetts, USA

Yang Lei
Department of Radiation Oncology
and
Winship Cancer Institute
Emory University
Georgia, USA

Taoran Li
Department of Radiation Oncology
University of Pennsylvania
Pennsylvania, USA

Wen Li
Hong Kong Polytechnic University
Hong Kong

Hely Lin
J. Crayton Pruitt Family Department of Biomedical Engineering
University of Florida
Florida, USA

Dan Nguyen
Medical Artificial Intelligence and Automation (MAIA) Laboratory
Department of Radiation Oncology
University of Texas Southwestern Medical Center
Texas, USA

Richard L.J. Qiu
Department of Radiation Oncology
and
Winship Cancer Institute
Emory University
Georgia, USA

Liangqiong Qu
Department of Statistics and Actuarial Science
and
The Institute of Data Science
The University of Hong Kong
Hong Kong

Ge Ren
Hong Kong Polytechnic University
Kowloon, Hong Kong

Yang Sheng
Department of Radiation Oncology
Duke University Medical Center
North Carolina, USA

Zhen Tian
Department of Radiation Oncology
The University of Chicago
Illinois, USA

Jing Wang
Department of Radiation Oncology
and
Winship Cancer Institute
Emory University
Georgia, USA

Tonghe Wang
Department of Medical Physics
Memorial Sloan Kettering
Cancer Center
New York, USA

Jackie Wu
Department of Radiation Oncology
Duke University Medical Center
North Carolina, USA

Yao Xiao
The University of Texas MD Anderson Cancer Center
Houston, TX, USA

Lei Xing
Department of Radiation Oncology
Stanford University
California, USA

Xiaofeng Yang
Department of Radiation Oncology
and
Winship Cancer Institute
Emory University
Georgia, USA

Lequan Yu
Department of Statistics and Actuarial Science
The University of Hong Kong
Hong Kong

Hao Zhang
Department of Medical Physics
Memorial Sloan Kettering Cancer Center
New York, USA

Jie Zhang
University of Kentucky
Lexington, Kentucky

Xiaodan Zhang
Faculty of Information Technology
Beijing University of Technology
Beijing, China

Yongqin Zhang
School of Information Science and Technology
Northwest University
Xi'an, China

Zhicheng Zhang
JancsiTech
Hangzhou, China

Wei Zhao
Beihang University
Beijing, China

Yuyin Zhou
Department of Computer Science and Engineering
University of California
California, USA

Jiarui Zhu
Hong Kong Polytechnic University
Kowloon, Hong Kong

Lingting Zhu
Department of Statistics and Actuarial Science
The University of Hong Kong
Hong Kong

Introduction

Recently, medical image synthesis became a vital aspect of both medical imaging research and clinic study in radiology and radiation oncology fields. Within the medical imaging and medical physics communities, promising work on inter-modality synthesis (e.g., MRI-based synthetic CT/dual energy (DE) CT, CBCT-based synthetic CT, CBCT/CT-based synthetic MRI, CBCT/CT-based stopping power map estimation, CT-based ventilation image, PET-based synthetic CT and image-based dose plan prediction) and intra-modality (cross-modality) synthesis (e.g., low-count PET-based full-count PET image estimation, PET/SPECT attenuation correction, intra-multiparametric MRI transformation, ultrasound/CT/MRI high-resolution image generation, 2D-based 3D volumetric image generation, MRI inhomogeneity correction, low-dose CT, MRI/CT/US denoising and metal artifact reduction) is being performed in PET/SPECT attenuation correction, MRI-based treatment planning, CBCT-guided adaptive radiotherapy, image segmentation, multimodality image registration, high-quality image generation, low-dose PET or CT generation, fast MRI imaging, real-time motion tracking, and numerous other areas. Therefore, we believe that it is time to combine the mostly used and well-known methods – to prepare a book dedicated to the topic of medical image synthesis. The aim of this book is to provide principles and methods, and state-of-the-art research, for undergraduates, graduates, researchers and practitioners. In this book, first, various traditional non-learning-based, traditional machine-learning-based and recent deep-learning-based medical image synthesis methods are reviewed. Second, specific applications of different inter-modality and intra-modality image synthesis tasks and of synthetic image-aided segmentation and registration are introduced and summarized. Third, the clinic usage of medical image synthesis, such as treatment planning and image-guided adaptive radiotherapy, is discussed. Last, the limitations and current challenges of various medical synthesis applications are explored. The future trends or potential solution to solve these difficulties are discussed.

Section I

Methods and Principles

1

Non-Deep-Learning-Based Medical Image Synthesis Methods

Jing Wang and Xiaofeng Yang
Emory University, Atlanta, GA, USA

CONTENTS

1.1 Introduction

Image synthesis across multimodality has been actively studied and widely adopted in radiology and radiation oncology. For diagnosis and treatment purposes, it is quite common to involve medical imaging of several modalities, e.g., computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET). However, high-quality image acquisition can be too expensive, time-consuming, or laborsome and thus become infeasible during clinic workflow. In some cases, the radiation constraints and image registration complexity will also restrict the direct use of some imaging modalities [1]. To overcome such obstacles, image synthesis techniques have been introduced to generate difficult images from easy-to-obtain images (intermodality synthesis) or synthesize high-quality images from low-quality inputs (intramodality synthesis).

Before the era of deep learning, image synthesis was conventionally performed with non-learning-based or traditional machine-learning-based methods, including single/multi-atlas-based, random forest, support vector machine (SVM), and other techniques [2].

DOI: 10.1201/9781003243458-2

Such models are considered more difficult when the two involved imaging modalities carry very different information, e.g., between CT and MRI where MRI intensity information is not uniquely related to electron density like the CT Hounsfield unit (HU) value does [3]. In fact, MRI-based synthetic CT (sCT) or sometimes called pseudo–CT (pCT), has been a major topic in such studies since if we can obtain sCTs from standard clinical MRI sequences for treatment planning and patient positioning verification, then an MRI-only radiotherapy (RT) workflow can be implemented to greatly simplify the treatment process and reduce received doses, with a better lesion target and organ at risk (OAR) delineation provided by MRI [2].

In this chapter, we will review the published articles focusing on utilizing traditional computer-aided image synthesis before the development of modern deep learning-based techniques, which mainly include non-learning-based and traditional machine-learning-based methods. The remainder of the chapter is organized as follows: Section 1.2 provides an overview of non-learning-based methods, mainly the atlas-based algorithms. Section 1.3 introduces works utilizing traditional machine-learning-based methods. In Section 1.4 we will discuss the achievements and limitations for these conventional image synthesis techniques.

1.2 Overview of Non-learning-based Methods

Traditionally, one of the dominating applications of image synthesis in medicine is to generate sCTs from MRI sequences since CTs provide accurate patient positioning information and electron density distribution for RT treatment planning but suffer from low soft tissue contrast and extra radiation exposure; while clinical MRIs provide high soft tissue contrast for lesion targeting but don't contain electron density information. Thus, in the early days bulk density override methods were utilized to convert MRIs to sCTs intuitively, e.g., overriding the whole MRI with water-equivalent electron density for RT treatment planning [4–10]. Such early methods simply assume a homogeneous electron density across the volume and the treatment planning with sCTs could lead to a dose discrepancy of greater than 2% compared with the ground truth heterogeneous CTs [2]. Later, more sophisticated overriding methods were developed by assigning different electron density numbers or HU values to different classes on the MRI [11–20]. Though the more specific overriding methods achieved better accuracies, they often require manual contouring to segment different zones (air, bone, soft tissues, etc.) on MRIs.

Compared with the bulk density overriding methods, atlas-based methods are generally considered to achieve improved results with a fully automated workflow and are the most important non-learning-based type of algorithms for generating sCTs from MRIs, which can be used for treatment planning and PET attenuation correction for PET/MR systems. In atlas-based methods, numerous co-registered pre-acquired MRI-CT image pairs form a reference database. The incoming MR image will be registered or warped to the reference MRIs, and the warping deformation will be applied to corresponding reference CTs, which will be brought up to generate the final sCT of the incoming MRI. An example workflow of atlas-based sCT synthesis is shown in Figure 1.1 [21]. Considering the number of atlases CTs used to output the final sCT, there exist single-atlas-based methods and multiple-atlas-based methods.

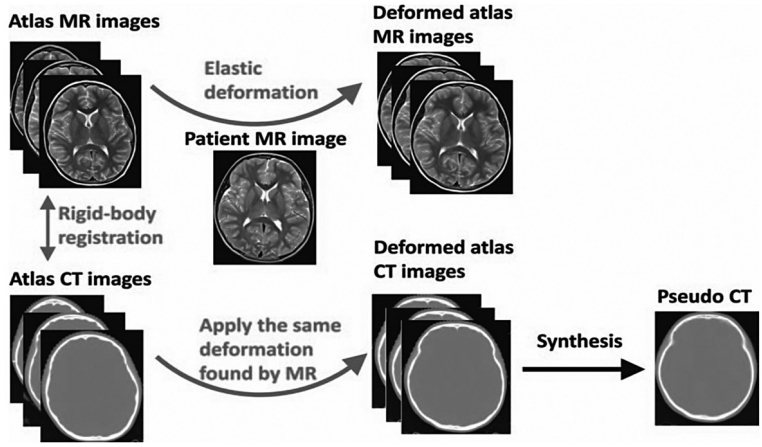

FIGURE 1.1
The workflow of generating sCT images from atlas images. The same warping deformation was applied to both atlas MRIs and the corresponding atlas CTs. The final sCT was a combination of the warped atlas CTs. (Reprint permission from [21].)

1.2.1 Single-Atlas-Based Method

Using a single atlas to create the sCTs is simple and straightforward. In a PET/MR study, Schreibmann et al. [22] used a multimodality optical flow deformable model to register a selected CT atlas template to the incoming patient MR images, by first performing a rigid registration to align the patient template with the patient MR images, and then using a BSpline deformable transform [23] to resolve the warping field between CT and MRIs, and finally applying a Hermosillo algorithm [24, 25] to refine the registration results. Thus, an sCT was created for PET attenuation correction. They compared the sCT attenuation with the true CT attenuation PET results for 17 clinical brain tumor cases, and a high agreement was found through the histogram analysis.

1.2.2 Multi-atlas-based Method

Compared to the single-atlas method, the multi-atlas-based method is more commonly used and considered to perform superior. Burgos et al. [26, 27] used a multi-atlas-based algorithm to synthesize CT from T1w MRI for PET attenuation correction. An atlas database was composed of affinely aligned CT and MRI pairs. To synthesize CT images for the target MRI, the atlas MRIs in the database were registered to the target MRI, and the same transformation was applied to corresponding atlas CTs. The convolution-based fast local normalized correlation coefficient (LNCC) [28] and the local normalized sum of square differences (NSSD), were used as two local similarity measures between the target MRI and transformed atlas MRIs. The ranks of local similarity measures were then converted to spatially varying weights. Then the sCT was generated by averaging the transformed atlas CTs with those weights. Finally, the HU values in sCT were converted to linear attenuation coefficients in cm^{-1} for PET image correction.

Uh et al. compared several atlas-based methods for MRI-CT synthesis of 14 pediatric patients with brain tumors. They used three types of atlas groups containing a single atlas,

six atlases, or twelve atlases. For the six atlases group, they employed a simple arithmetic mean process or pattern recognition with Gaussian process (PRGP) [29] to combine the warped CT atlases into the final output, while for the 12 atlases group, they only used the PRGP merging process. Their results suggest the average root-mean-square (RMSD) between pseudo and real CTs improves with more atlases, and the PRGP combination is better than the simple arithmetic mean process.

Sjölund et al. [30] proposed a novel method to combine the multiple registered CT atlases to form a final sCT, which iteratively estimates the final single sCT from the voxel-wise mean of warped reference CTs. They claimed their iterative mean registration estimation made significant improvement compared to a simple mean-estimator of the multiple CT atlases or the single best-warped atlas. Degen et al. [12] addressed the challenge of multimodal image registration with the self-similarity context as a matching criterion, to match CT atlases to patient anatomy. Multimodal local atlas fusion is then performed to jointly combine the registered atlases. The method compared the dissimilarity between the two modalities at each location and only fuse those atlases morphologically similar enough to the incoming MRI. Such a method could extend the MR-CT synthesis technique from brain studies with strict alignment of CT-MRI pairs to whole-body scans without well-aligned CT-MRI pairs. Burgos et al. [14] tried to jointly solve organ segmentation and pCT synthesis by a multi-atlas method. The mean absolute error (MAE) and mean error (ME) were 45.7 ± 4.6 HU and -1.6 ± 7.7 HU, respectively, between sCTs and true CTs.

1.2.3 Patch-based Atlas Method

Though it is intuitive to treat the MRI and CT images as a whole to extract the anatomy information, some studies extend to the patch-based atlas method in the sCT fusion/generation phase. Hofmann et al. [29] built an atlas database of co-registered CT-MR pairs from 17 patients, and then warped the atlas CTs by MRI registration between the atlas and patient MR images. After the prior atlas registration, they used a pattern recognition technique that searched neighboring MR patches in the warped atlases and then generate sCT patches by Gaussian process regression, integrating with the mean value of prior registered CT atlases. By testing the method on three independent patients, the mean percentage difference between the MRI-based attenuation correction and the CT-based PET attenuation is $3.2\% +/- 2.5\%$.

Besides synthesizing CTs from MRI, Roy et al. proposed a patch-based atlas method for generating T2-weighted (T2w) MRI from T1-weighted (T1w) MR images [31]. As an atlas-based method, they also created atlases with co-registered T1w and T2w images. From the atlas image pairs, then multiple subsampled patches were obtained from these image pairs to form numerous atlas patches. For each patch on the targeting T2w, a few similar-looking atlas patches were used to generate the synthetic T2w patches. Similar T1w atlas patches of the neighborhood of that targeting patch with similarity metrics were collected. Then the warped T2w reference atlases corresponding to these picked T1w atlases were combined to obtain the synthetic T2w patches, which were merged to form the final output T2w image. The peak signal-to-noise ratio (PSNR) is 28.68 vs. 25.62, and median cerebral spinal fluid (CSF) intensities are 2.17 vs. 1.91 for high vs. low resolution images, respectively, indicating a better performance on high-resolution datasets.

1.3 Overview of Traditional Machine-Learning-based Methods – Voxel-based Techniques

Though the atlas-based method is probably the most popular technique for image synthesis before the emergence of deep learning, there are other trials using traditional machine learning models such as random forest (RF) to investigate the image synthesis tasks at the voxel level.

Zhao et al. [32] proposed a bi-directional MR-CT synthesis framework using machine learning. The multiple co-registered MR-CT atlas pairs first performed a supervoxel over-segmentation process using the simple linear iterative clustering (SLIC) method, followed by a k-means clustering algorithm for tissue characterization into z-fields. Then for each field (tissue class), there are two RF regressors for sCTs from T1w or for T1w synthesis from CTs. Besides, two RF classifiers were trained to extract features and generate the z labels on both modalities, so that the image synthesis could be feasible by applying the bi-directional RF regressor of each tissue class. In an experiment involving six MR-CT atlas pairs, the structural similarity (SSIM) for sCT is 0.63–0.73, and the SSIM for synthetic MR is 0.7–0.8; while the PSNR for sCT is 27–30, and the PSNR for synthetic MR is 20–28 (Figure 1.2).

Besides the commonly discussed MR-CT synthesis problem, Jog et al. proposed a computationally efficient machine learning method to generate MR images of a desired modality from other input MR modalities, namely, the Regression Ensembles with Patch Learning for Image Contrast Agreement or REPLICA [33]. They extract multi-resolution features

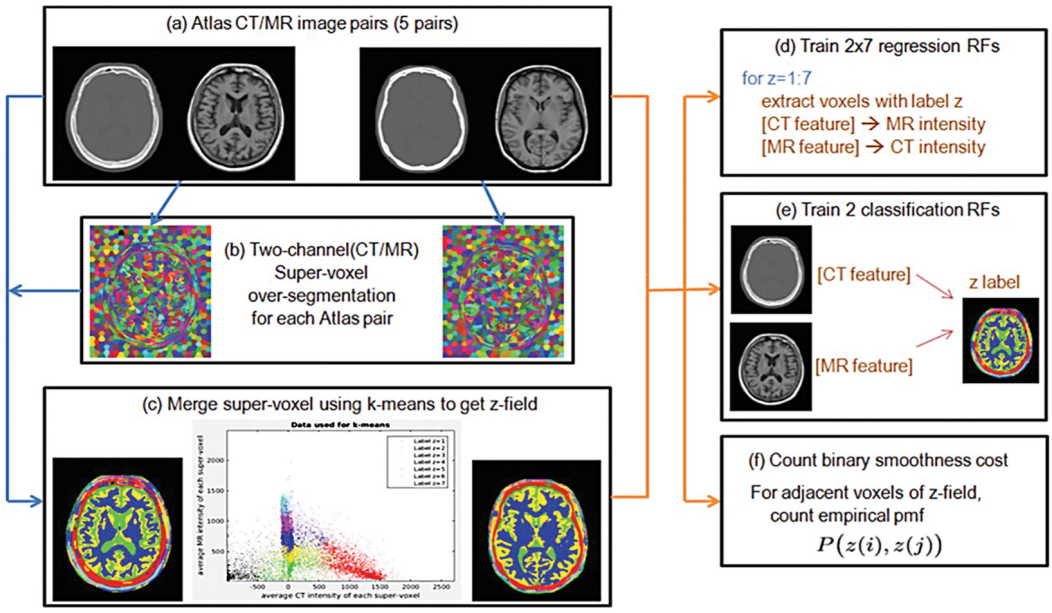

FIGURE 1.2
(a) example of CT-MR atlas pairs; (b) result of SLIC over-segmentation; (c) examples of z-fields after k-means clustering; (d) training of RF regressors; (e) RF classifiers trained to estimate z-fields from single modalities; and (f) computation of pairwise potentials for a Markov Random Field (MRF). (Reprint permission from [32].)

based on voxel intensities and high-resolution context descriptors to feed into an RF training model. The model learns to convert the input MR sequences to the target MR modality images at the voxel level. In an experiment of generating T2w MR images from 32 magnetization-prepared gradient echo (MPRAGE) scans, their method outperformed FUSION (a multi-atlas registration and intensity fusion method [34]) or MIMECS [34, 35] with a PSNR of 50.73, universal quality index (UQI) of 0.89, and an SSIM of 0.87. In another experiment of synthesizing 125 FLuid Attenuated Inversion Recovery (FLAIR) images of 84 patients from T1w, T2w, and P_D-weighted (PDw) sequences, their method achieved a PSNR of 21.73, UQI of 0.84, and SSIM of 0.81. They also pointed out that FUSION is an atlas-based model so the lesion intensity cannot be precisely synthesized if those intensities were not available in the atlases.

Bowles et al. [36] proposed an interesting application of synthesizing pseudo-healthy FLAIR images from an incoming T1w MRI. By subtracting this pseudo-healthy FLAIR from the patient's true FLAIR, they could obtain the difference voxel intensities corresponding to the probability of a pathological lesion, and train a support vector machine (SVM) to generate a probability map. The critical component is again the pseudo-healthy FLAIR synthesis from the input T1w images. They used a regression model to learn the voxel level relationship between healthy FLAIR and T1w, which directly mapped the T1w voxels to synthetic FLAIR images. The lesion detection of the proposed method is satisfactory, with a dice similarity coefficient (DSC) of 0.703, average symmetric surface distance (ASSD) of 1.23, Hausdorff distance (HD) of 38.6, and intraclass correlation (ICC) of 0.985.

We summarize the reviewed work for image synthesis in Table 1.1, including the study details, utilized methods, and obtained results.

1.4 Discussion

Medical image synthesis is an increasingly important topic in today's clinical improvement, since it could greatly save the time, economical cost, and other physical and psychological burden of patients by reducing the imaging modality involvement and simplifying the clinical workflow. Traditional image synthesis focuses on MR-CT synthesis or MR-MR synthesis, to extract anatomy or functional information from existing images, without the need of further medical scans. Most studies are preliminary and included small cohorts of patients, investigating mostly brain/head, and prostate or whole-body images.

1.4.1 Achievements

For MR-CT synthesis, the atlas-based methods seem to be the most commonly adopted technique [21, 30, 39–43]. In general, atlas-based methods perform better than the simple bulk density overriding techniques, and the atlas-based models can be carried out in a fully automated workflow while the density overriding often requires human contouring or segmentation of the MR images. Moreover, using multiple atlases outperforms the single-atlas method since more information can be tailored from the multi-atlas database. The major steps of the atlas-based method are the registration to align atlas MR-CT pairs to form co-registered atlas pairs, the registration of incoming MRI to atlas MRIs, and the fusion of warped CT atlases to generate final sCTs. The first two steps are highly dependent on the accuracy of rigid or non-rigid registration between multimodal images, and that is beyond the scope of this chapter.

TABLE 1.1

Summary of the Reviewed Work for Image Synthesis in Sections 2.1–2.3 and 3

Author	Year	Tasks	Location	Method	Dataset	Results
Hofmann et al. [29]	2008	sCT for PET/MRI correction	Brain	Patch-based atlas	20 patients	PET quantification with a mean error of 3.2% +/− 2.5%
Schreibmann et al. [22]	2010	sCT for PET/MRI correction	Brain	Single-atlas-based	17 brain tumor cases	mean voxel-to-voxel difference < 2 HU
Burgos et al. [26]	2013	sCT for PET/MRI correction	Brain	Multi-atlas-based	28 subjects	Average mean absolute residual (MAR) between sCTs and ground truth is 73–108 in HU for different CT modalities
Burgos et al. [27]	2014	sCT for PET/MRI correction	Brain	Multi-atlas-based	18 + 41 brain cases for parameter optimization and validation	The average MAE is 113 HU for full head and 240 HU for skull
Uh et al. [21]	2014	sCT from MRI	Brain	Single/multi-atlas-based	14 pediatric patients	More atlases and novel combination methods yielded better results
Sjölund et al. [30]	2015	sCT from MRI	Whole head	Multi-atlas-based	10 patients	Best mean absolute errors (HU) is 114.5 ± 20.5
Roy et al. [31]	2016	MRI synthesis from other sequences	Whole head	Patch-based atlas	44 patients	PSNR is 28.68 vs. 25.62 dB, and median CSF intensities are 2.17 vs. 1.91 for high- vs. low-resolution images
Degen et al. [37]	2016	sCT from MRI	Whole body	Multi-atlas-based	18 3D CT atlases	The dice overlap of segmentation labels between MRI and CT is greatly enhance after fusion
Bowles et al. [36]	2017	MRI synthesis from other sequences	Brain	SVM, Regression model	127 subjects	Lesion detection: DSC of 0.703, ASSD of 1.23, HD of 38.6, and ICC of 0.985
Burgos et al. [38]	2017	sCT from MRI	Prostate	Multi-atlas-based	15 subjects	MAE of 45.7 ± 4.6 HU, and ME of −1.6 ± 7.7 HU
Zhao et al. [32]	2017	Bi-directional MR/CT synthesis	Whole head	RF	6 MR/CT atlas pairs	SSIM for sCT is 0.63–0.73, and the SSIM for synthetic MR is 0.7–0.8; while the PSNR for sCT is 27–30, and the PSNR for synthetic MR is 20–28
Jog et al. [33]	2017	MRI synthesis from other sequences	Brain	RF	21 and 84 patients for T2w and FLAIR synthesis	A PSNR of 50.73, UQI of 0.89, and a SSIM of 0.87 for T2w MRI synthesis; a PSNR of 21.73, UQI of 0.84, and SSIM of 0.81 for FLAIR synthesis

Mean absolute error (MAE), mean absolute residual (MAR), mean error (ME), peak signal-to-noise ratio (PSNR), structural similarity (SSIM), universal quality index (UQI), mean surface distance (MSD), Hausdorff distance (HD), dice similarity coefficient (DSC), root-mean-square distance (RMSD), averaged symmetric surface distance (ASSD), intraclass correlation (ICC), and cerebral spinal fluid (CSF).

The last step, however, is unique in image synthesis and the novelty of many sCT studies focused on proposing devised fusion strategy. Common practices are taking the weighted average of warped atlas CTs as the sCTs or employing a patch-based method to fuse the final output. Similar principles also apply to atlas-based MR-MR synthesis.

Another commonly used type of implementation for CT-MR or MR-MR synthesis is a voxel-based method [15, 44–46]. The main idea is to obtain the voxel or supervoxel correlation between multimodal images via regression or machine learning models. Such techniques are claimed to outperform the atlas-based fusion methods [33]. In the scope of synthesizing CTs from MRIs, apart from the standard MRI sequences alone [47–53], there exist techniques utilizing ultra-short echo time (UTE) sequences in which the bone structures are more easily differentiated from air. However, taking these UTE sequences is costly in time and resources and not included in standard clinical workflow, so we didn't elaborate on those UTE work in this chapter.

Besides the most frequently seen CT and MR images, the CT synthesis from other modalities such as transrectal ultrasound was also investigated, which could potentially be useful in brachytherapy [54]. In the future, image synthesis tasks involving more imaging modalities can be expected, especially with the fast-developing deep-learning techniques.

1.4.2 Limitations

Though the atlas-based methods have achieved promising accuracy in image synthesis for both CT and MR images and have been considered as a robust method especially on unexpected or unusual cases [1], some limitations exist in such techniques. The first limitation is inherent within the method itself that the final synthetic images come from the co-registered atlas database. That explains why a single atlas can lead to noticeable deviance [40] due to its insufficiency of representing patient variations. Increasing the number of atlases in the database may relieve the problem, but could still perform poorly for atypical patients, e.g., patients with large tumors or surgical cavities [21]. Besides, increasing the atlases numbers can burden the cost of the already complex registration process in both the atlas MR-CT co-registration phase and the incoming MRI registration phase. Moreover, the effect of adding atlases is still not fully investigated. The optimal number of atlases needed for various tasks can depend on the similarity between atlases and the incoming images [42], while some studies reported limited benefits of adding atlases beyond 15 patients [55]. For machine learning techniques, they are often used with patch-based voxel level methods involving non-standard MRI sequences such as UTE that elongated patient stays, while for those used with standard MRI sequences, the need to manually contour bone and air may limit their use [2].

1.5 Conclusion

We have overviewed the state-of-the-art progress of conventional medical image synthesis before the blossom of deep learning. The main topics are MR-CT synthesis or MR-MR synthesis, and there are three major pathways to implement the tasks: the bulk density overriding, the atlas-based synthesis, and the traditional machine-learning-based methods. Though these techniques are relatively simple compared to the latest deep learning models, they have proven to be successful in generating desired images without the need for additional modality scans, and thus help reduce patient stays and lower radiation exposure.

References

1. Wang, T., et al., A review on medical imaging synthesis using deep learning and its clinical applications. *Journal of Applied Clinical Medical Physics*, 2021. **22**(1): pp. 11–36.
2. Johnstone, E., et al., Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging–only radiation therapy. *International Journal of Radiation Oncology* Biology* Physics*, 2018. **100**(1): pp. 199–217.
3. Karlsson, M., et al., Dedicated magnetic resonance imaging in the radiotherapy clinic. *International Journal of Radiation Oncology* Biology* Physics*, 2009. **74**(2): pp. 644–651.
4. Schad, L.R., et al., Radiosurgical treatment planning of brain metastases based on a fast, three-dimensional MR imaging technique. *Magnetic Resonance Imaging*, 1994. **12**(5): pp. 811–819.
5. Ramsey, C.R. and A.L. Oliver, Magnetic resonance imaging based digitally reconstructed radiographs, virtual simulation, and three-dimensional treatment planning for brain neoplasms. *Medical Physics*, 1998. **25**(10): pp. 1928–1934.
6. Prabhakar, R., et al., Feasibility of using MRI alone for 3D radiation treatment planning in brain tumors. *Japanese Journal of Clinical Oncology*, 2007. **37**(6): pp. 405–411.
7. Wang, C., et al., MRI-based treatment planning with electron density information mapped from CT images: a preliminary study. *Technology in Cancer Research & Treatment*, 2008. **7**(5): pp. 341–347.
8. Weber, D., et al., Open low-field magnetic resonance imaging for target definition, dose calculations and set-up verification during three-dimensional CRT for glioblastoma multiforme. *Clinical Oncology*, 2008. **20**(2): pp. 157–167.
9. Chen, L., et al., MRI-based treatment planning for radiotherapy: dosimetric verification for prostate IMRT. *International Journal of Radiation Oncology* Biology* Physics*, 2004. **60**(2): pp. 636–647.
10. Chen, L., et al., Dosimetric evaluation of MRI-based treatment planning for prostate cancer. *Physics in Medicine & Biology*, 2004. **49**(22): p. 5157.
11. Eilertsen, K., et al., A simulation of MRI based dose calculations on the basis of radiotherapy planning CT images. *Acta Oncologica*, 2008. **47**(7): pp. 1294–1302.
12. Karotki, A., et al., Comparison of bulk electron density and voxel-based electron density treatment planning. *Journal of Applied Clinical Medical Physics*, 2011. **12**(4): pp. 97–104.
13. Korsholm, M.E., L.W. Waring, and J.M. Edmund, A criterion for the reliable use of MRI-only radiotherapy. *Radiation Oncology*, 2014. **9**(1): pp. 1–7.
14. Chin, A.L., et al., Feasibility and limitations of bulk density assignment in MRI for head and neck IMRT treatment planning. *Journal of Applied Clinical Medical Physics*, 2014. **15**(5): pp. 100–111.
15. Doemer, A., et al., Evaluating organ delineation, dose calculation and daily localization in an open-MRI simulation workflow for prostate cancer patients. *Radiation Oncology*, 2015. **10**(1): pp. 1–9.
16. Jonsson, J.H., et al., Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions. *Radiation Oncology*, 2010. **5**(1): pp. 1–8.
17. Kristensen, B.H., et al., Dosimetric and geometric evaluation of an open low-field magnetic resonance simulator for radiotherapy treatment planning of brain tumours. *Radiotherapy and Oncology*, 2008. **87**(1): pp. 100–109.
18. Lambert, J., et al., MRI-guided prostate radiation therapy planning: Investigation of dosimetric accuracy of MRI-based dose planning. *Radiotherapy and Oncology*, 2011. **98**(3): pp. 330–334.
19. Stanescu, T., et al., A study on the magnetic resonance imaging (MRI)-based radiation treatment planning of intracranial lesions. *Physics in Medicine & Biology*, 2008. **53**(13): p. 3579.
20. Stanescu, T., et al., 3T MR-based treatment planning for radiotherapy of brain lesions. *Radiology and Oncology*, 2006. **40**(2): pp. 125–132.
21. Uh, J., et al., MRI-based treatment planning with pseudo CT generated through atlas registration. *Medical Physics*, 2014. **41**(5): p. 051711.

22. Schreibmann, E., et al., MR-based attenuation correction for hybrid PET-MR brain imaging systems using deformable image registration. *Medical Physics*, 2010. **37**(5): pp. 2101–2109.
23. Mattes, D., et al., PET-CT image registration in the chest using free-form deformations. *IEEE Transactions on Medical Imaging*, 2003. **22**(1): pp. 120–128.
24. De Craene, M., et al., Incorporating metric flows and sparse Jacobian transformations in ITK. *Insight Journal*, 2006. http://hdl.handle.net/1926/183
25. Cuadra, M.B., et al., Atlas-based segmentation of pathological MR brain images using a model of lesion growth. *IEEE Transactions on Medical Imaging*, 2004. **23**(10): pp. 1301–1314.
26. Burgos, N., et al., Attenuation correction synthesis for hybrid PET-MR scanners. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2013. Springer.
27. Burgos, N., et al., Attenuation correction synthesis for hybrid PET-MR scanners: application to brain studies. *IEEE Transactions on Medical Imaging*, 2014. **33**(12): pp. 2332–2341.
28. Cachier, P., et al., Iconic feature based nonrigid registration: the PASHA algorithm. *Computer Vision and Image Understanding*, 2003. **89**(2–3): pp. 272–298.
29. Hofmann, M., et al., MRI-based attenuation correction for PET/MRI: a novel approach combining pattern recognition and atlas registration. *Journal of Nuclear Medicine*, 2008. **49**(11): pp. 1875–1883.
30. Sjölund, J., et al., Generating patient specific pseudo-CT of the head from MR using atlas-based regression. *Physics in Medicine & Biology*, 2015. **60**(2): p. 825.
31. Roy, S., et al., Patch based synthesis of whole head MR images: Application to EPI distortion correction. In *International Workshop on Simulation and Synthesis in Medical Imaging*. 2016. Springer.
32. Zhao, C., et al., A supervoxel based random forest synthesis framework for bidirectional MR/CT synthesis. In *International Workshop on Simulation and Synthesis in Medical Imaging*. 2017. Springer.
33. Jog, A., et al., Random forest regression for magnetic resonance image synthesis. *Medical Image Analysis*, 2017. **35**: pp. 475–488.
34. Roy, S., A. Carass, and J.L. Prince, Magnetic resonance image example-based contrast synthesis. *IEEE Transactions on Medical Imaging*, 2013. **32**(12): pp. 2348–2363.
35. Roy, S., A. Carass, and J. Prince, A compressed sensing approach for MR tissue contrast synthesis. In *Biennial International Conference on Information Processing in Medical Imaging*. 2011. Springer.
36. Bowles, C., et al., Brain lesion segmentation through image synthesis and outlier detection. *NeuroImage: Clinical*, 2017. **16**: pp. 643–658.
37. Degen, J., and M.P. Heinrich, Multi-atlas based pseudo-CT synthesis using multimodal image registration and local atlas fusion strategies. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2016.
38. Burgos, N., et al., Iterative framework for the joint segmentation and CT synthesis of MR images: application to MRI-only radiotherapy treatment planning. *Physics in Medicine & Biology*, 2017. **62**(11): p. 4237.
39. Dowling, J.A., et al., An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy. *International Journal of Radiation Oncology* Biology* Physics*, 2012. **83**(1): pp. e5–e11.
40. Demol, B., et al., Dosimetric characterization of MRI-only treatment planning for brain tumors in atlas-based pseudo-CT images generated from standard T1-weighted MR images. *Medical Physics*, 2016. **43**(12): pp. 6557–6568.
41. Andreasen, D., et al., Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain. *Medical Physics*, 2015. **42**(4): pp. 1596–1605.
42. Andreasen, D., K. Van Leemput, and J.M. Edmund, A patch-based pseudo-CT approach for MRI-only radiotherapy in the pelvis. *Medical Physics*, 2016. **43**(8Part1): pp. 4742–4752.
43. Edmund, J.M., et al., Cone beam computed tomography guided treatment delivery and planning verification for magnetic resonance imaging only radiotherapy of the brain. *Acta Oncologica*, 2015. **54**(9): pp. 1496–1500.

44. Dowling, J.A., et al., Automatic substitute computed tomography generation and contouring for magnetic resonance imaging (MRI)-alone external beam radiation therapy from standard MRI sequences. *International Journal of Radiation Oncology* Biology* Physics*, 2015. **93**(5): pp. 1144–1153.
45. Jonsson, J.H., et al., Treatment planning of intracranial targets on MRI derived substitute CT data. *Radiotherapy and Oncology*, 2013. **108**(1): pp. 118–122.
46. Johansson, A., et al., Voxel-wise uncertainty in CT substitute derived from MRI. *Medical Physics*, 2012. **39**(6Part1): pp. 3283–3290.
47. Kapanen, M., and M. Tenhunen, T1/T2*-weighted MRI provides clinically relevant pseudo-CT density data for the pelvic bones in MRI-only based radiotherapy treatment planning. *Acta Oncologica*, 2013. **52**(3): pp. 612–618.
48. Korhonen, J., et al., A dual model HU conversion from MRI intensity values within and outside of bone segment for MRI-based radiotherapy treatment planning of prostate cancer. *Medical Physics*, 2014. **41**(1): p. 011704.
49. Korhonen, J., et al., Absorbed doses behind bones with MR image-based dose calculations for radiotherapy treatment planning. *Medical Physics*, 2013. **40**(1): p. 011701.
50. Korhonen, J., et al., Influence of MRI-based bone outline definition errors on external radiotherapy dose calculation accuracy in heterogeneous pseudo-CT images of prostate cancer patients. *Acta Oncologica*, 2014. **53**(8): pp. 1100–1106.
51. Koivula, L., L. Wee, and J. Korhonen, Feasibility of MRI-only treatment planning for proton therapy in brain and prostate cancers: dose calculation accuracy in substitute CT images. *Medical Physics*, 2016. **43**(8Part1): pp. 4634–4642.
52. Kim, J., et al., Implementation of a novel algorithm for generating synthetic CT images from magnetic resonance imaging data sets for prostate cancer radiation therapy. *International Journal of Radiation Oncology* Biology* Physics*, 2015. **91**(1): pp. 39–47.
53. Yu, H., et al., Toward magnetic resonance–only simulation: segmentation of bone in MR for radiation therapy verification of the head. *International Journal of Radiation Oncology* Biology* Physics*, 2014. **89**(3): pp. 649–657.
54. Satheesh, B.A., et al., Pseudo-CT image synthesis from ultrasound images for potential use in Brachytherapy treatment planning-initial results. In *TENCON 2021–2021 IEEE Region 10 Conference (TENCON)*. 2021. IEEE.
55. Siversson, C., et al., MRI only prostate radiotherapy planning using the statistical decomposition algorithm. *Medical Physics*, 2015. **42**(10): pp. 6090–6097.

2

Deep-Learning-Based Medical Image Synthesis Methods

Yang Lei and Richard L.J. Qiu
Emory University, Atlanta, GA, USA

Tonghe Wang
Memorial Sloan Kettering Cancer Center, New York, NY, USA

Xiaofeng Yang
Emory University, Atlanta, GA, USA

CONTENTS

DOI: 10.1201/9781003243458-3

2.1 Introduction

Image synthesis involves the generation of artificial or simulated images in a specific target image modality or target domain using input images from a different image modality [1]. The objective of a synthesis task is to use synthetic images to replace the physical patient during the imaging procedure. This is motivated by a variety of factors, such as the infeasibility of a specific image acquisition, the added cost and labor of the imaging procedure, the ionizing radiation exposure to patients by some of those imaging procedures, or the introduction of uncertainties via performing image registration across different modalities. Over the past decade, image synthesis research has been introduced considerable attention in radiation oncology, radiology, and biology fields [2]. Potential clinical applications of image synthesis include radiation treatment using magnetic resonance imaging (MRI) only [3–10], positron emission tomography (PET) attenuation correction (AC) and image quality enhancement [11, 12], stopping power estimation[13–15], medical image quality improvement [23–32], medical image reconstruction [33, 34], synthetic image-aided auto-delineation [16–22], super-resolution image estimation [35–37], and others [38–42].

Image synthesis technology has been a hot topic of investigation for the past decade. Conventional machine learning-based image synthesis methods often rely on prior knowledge that is explicitly and manually designed for converting images from one modality to another [43–50]. The robustness and accuracy of these methods are based on the way of prior knowledge definition and investigation [51–53]. In recent years, deep learning-based methods have dominated the field of image synthesis [54]. Contrary to standard machine learning, deep learning does not depend on manually extracted features ruled by prior knowledge [55–57]. It employs neural networks (NN) or convolutional neural networks (CNN) with hidden layers of neurons or convolutional kernels to learn features automatically. These methods mostly follow a general workflow that adopts a data-driven methodology to map image intensities.

The typical work includes a learning/training stage, where the network-based model learns the correlation between the arrival one and the object one, and an estimation stage, where the learned model generates the synthetic target based on an input. Deep learning-based methods have several advantages over conventional machine learning-based methods. They exhibit higher generalizability since the same network architecture can be applied to several sets of image modalities with the smallest adaptations. This makes it possible to fast map different imagery modalities which is clinically meaningful. Despite the requirement for substantial attempts in data gathering and mining during the training stage, the prediction of an image usually only takes a few seconds. Due to these advantages, there has been a significant increase in the clinical research interest of medical images used in radiotherapy.

In this chapter, we aim to:

- Summarize the latest network architecture designs of deep learning-based medical image synthesis.
- Summarize the latest deep learning-based medical image synthesis applications.
- Emphasize significant involvements and recognize current contests.

2.2 Literature Searching

The focus of this review is confined to deep learning-based methods designed for synthetic image estimation. Medical image synthesis applications include studies on multimodality MRI synthesis, proton stopping power estimation, image quality improvement/enhancement, super/high-resolution visualization, MRI-only-based radiotherapy, inter-modality image registration, segmentation, PET AC, and data augmentation for image analysis.

Searches of peer-reviewed journal articles and conference proceedings references were conducted on PubMed by means of the conditions in title or abstract as of December 2022: ("estimate*" OR "reconstruct*" OR "trans*" OR "syn*" OR "restore*" OR "corr*" OR "generate*") AND "deep learning" AND "generative" OR "adversarial" OR "convolution*" OR "neural" ("MR" OR "MRI" OR "CT" OR "PET" OR "Medical" OR "Biomedical" etc.).

2.3 Network Architecture

The frameworks of the reviewed investigations are roughly categorized into four categories: NN, CNN, FCN, and GAN. These categories of methods are not entirely distinct from one another; instead, they demonstrate incremental increases in architectural complexity. After introducing the four categories of network design, we will then summarize feasible loss functions that can be used to train these networks.

2.3.1 NN

A NN is a network or circuit of biological neurons, or, in a modern sense, an artificial NN, composed of artificial neurons or nodes. In medical image synthesis, a most popular NN design is the autoencoder (AE) network.

AE and its variants have been extensively explored in the literature and remain widely employed in the field of analysis of medical images [58]. AE typically involves a NN encoding layer, which represents the image via latent layers with minimum representation error. The aim of this layer is to re-establish the arrival data within a low-dimensional latent space. AE is able to discover relevant patterns in the data by restricting the dimension of latent space vector representation.

To avoid potential overfitting issues caused by AE learning an identity function, recently, improvement of AE has been studied. A widely-known enhancement of AE is the stacked autoencoder (SAE), which is constructed by utilizing stacking operators. SAEs are composed of multiple AEs arranged in stacked layers. The yield of an individual layer is coupled with the input of the subsequent layers [59]. SAE serves a greedy layer-wise learning to obtain optimal parameters. The advantage of SAE is its ability to capture the significant hierarchical features of the input data [59].

Another variant of the AE is the denoizing autoencoder (DAEs), which enables the creation of improved higher-level feature representations [60]. DAEs avoid learning insignificant solutions for the trained model, such as estimating a denoised output from its noisy input [61]. The stacked denoising autoencoder (SDAE) is a learning model that uses the capabilities of DAE [62].

By incorporating a sparsity constraint, it can to overcome the weakness of AEs potentially, which typically exhibit a modest number of neurons. By introducing many neurons to exhibit a low average output via the sparse AE, the network ensures that many neurons remain inactive for most of the time [63].

SAE requires layer-wise learning, which can be time-consuming and tedious due to the fully connected layers used in its construction. Li et al. did the earliest attempt to train a convolutional AE (CAE) in an end-to-end manner without pre-learning [64]. Guo et al. recommended that CAEs are beneficial for learning features for images, preserving local structures, and avoiding alteration of feature space [65]. Wang et al. recommended a hybrid model of transfer learning and CAE for automated chest screening [66].

2.3.2 CNN

CNN is a variant of NNs. The core of CNN is the convolution layer, which is aimed at information mining [67]. The convolution layer generates information maps based on its optimized convolutional kernels. A pooling layer conducts a summarizing of the information by taking either the maximum or average value within the specified neighboring region, thereby reducing the structural resolution of arrival information. The rectified linear unit (ReLU) is the commonly employed activation function in CNNs [68]. The fully connected layer establishes connections between each neuron in the preceding layer and every neuron in the succeeding layer such that the information can be transferred from the beginning to the end layer. Then, the end layer predicts the probability from the previous information.

During training, the CNN architecture's learnable parameters are employed to predict the objective categorization of the training data. The loss function, typically cross-entropy (CE) loss, is then computed, and the weights are updated via the gradient descent (GD) method through back-propagation. Adam gradient descent (AGD) and Stochastic gradient descent (SGD) are among the greatest widespread methods used for optimization.

Lecun et al. earlier developed a CNN model, called LeNet, for hand-written numeral identification [69]. LeNet is compiled of a set of convolutional layers, fully connected layers, and pooling layers. The progress in the computer unit and the sufficient data made CNN training possible. AlexNet was introduced by Krizhevsky et al. They won the ILSVRC-2012 Image Classification Competition [70] with a greatly lesser misclassification rate than the second rank [71]. Subsequently, CNNs have attracted prevalent consideration, and their alternatives have been established, leading to state-of-the-art performance in various image-processing tasks. ZFNet was developed by Zeiler and Fergus. It is developed as an enhancement to increase the execution of AlexNet. This study [72] showed that shallow networks could realize the boundary, intensity, and structural information of an image. They also demonstrated that deeper networks can achieve better performance. The primary enhancement in ZFNet is the utilization of a deconvolution network for visualizing the information maps.

Simonyan and Zisserman introduced the Visual Geometry Group (VGG) to investigate the capabilities of network models with intense layers [73]. The focal advance of VGG is an utter assessment of the network's performance when increasing its layers' depth. The study showed that a significant improvement on the prior-art configurations can be achieved by increasing the depth to 16–19 layers. GoogLeNet was introduced to expand the network structure [74]. Through the integration of the proposed inception module, GoogLeNet emerged victorious in the ImageNet Large-Scale Visual Recognition Challenge

2014 (ILSVRC14), an image classification and detection competition. The inception module aids the CNN model in providing more descriptive representations of the entered information, all the while augmenting the size of the network model.

These new technologies allow the network to have increased size which improves the performance. Nevertheless, the merely deeper network would result in overfitting issues. In order to reduce the complexity of optimizing a deep model and solve the overfitting consequence produced by adding more layers, He et al. introduced a residual network (ResNet) [75]. ResNet, primarily comprised of residual blocks, has been proven capable of surpassing a 100-layer obstacle and even reaching 1,000 layers.

Taking inspiration from the ResNet, Huang et al. later introduced a densely connected convolutional network (DenseNet) by establishing connections between individual layer to every other layer [76]. In contrast to ResNet, which focuses on learning the fundamental change between the input data and the output data, DenseNet [77] intended to merge multiple frequencies of information from the preceding and flow layers using dense block [78].

2.3.3 FCN

Most CNNs take an image or patch as input and then estimate a voxel-wise value, which is correlated to the central of the patch. The first proposal for a CNN was introduced by Shelhamer et al., in which the fully connected layers of previous CNN are substituted with convolution layers. The new network is referred to as a fully convolutional network (FCN) since all the layers of the network are convolutional layers. Due to significant enhancements in deconvolution kernels used for up-sampling the feature map, an FCN enables the model to achieve dense voxel-wise prediction for the entire full-size volume, rather than patch-wise classification, as seen in traditional CNNs [79]. FCN allows segmenting the entire image in a single forward pass, combining high-resolution information with low-resolution information, and then passing them to the remaining layers. FCN can improve localization performance and produce more accurate output.

The U-Net architecture, which was originally developed by Ronneberger [80], is a widely-known FCN model used in auto-delineation. The U-Net comprises an encoding set of FCN layers and a decoding set of FCN layers. In between, long skip connections exist between the two sets. These operators supply high-frequency information to the decoding path, allowing the network to address the potential imbalance between the localization accuracy of a delineated organ and the organ boundary specification. This issue arises because using bulky-sized data requires extra downsized operators, which in turn can lower the localization precision. In contrast, minor-sized data are not able to capture unlimited texture information to represent the organ boundary. Later on, V-Net [81] appears to be an improved network based on the U-Net architecture, with a similar structure.

In place of relying solely on single optimization, as done in traditional FCNs, the main objective behind the deep supervision model [82, 83] is to offer intended optimization for middle sets of FCN layers and transmit learned information to the remaining layers. The approach extends optimization to the hidden and shallower layers of FCN, which increases the judicial capability of learned information in distinguishing multiorgans in auto-delineating projects. Lately, attention gates and transformers have been integrated into FCN to advance the execution of recognition [84]. The attention gate has the capability of learning how to diminish unrelated information and emphasize outstanding information that is beneficial for users' tasks.

2.3.4 GAN

GANs have garnered significant attention in medical imaging because of their capacity to generate data without the need for explicit modeling of the probability density function. As compared to CNN, GANs include an additional sub-network, called discriminator network. This approach offers an innovative means of integrating samples with no ground truth data into learning procedures and enforcing sophisticated constancy. This approach demonstrated its utility in numerous cases, including but not limited to image reconstruction [85], image enhancement [86–89], segmentation [90–92], diagnosis [93], data generation [94], and multimodality image estimation [95].

Characteristic GANs are composed of two opposing models, called a generator model and a discriminator model [96]. The generator model is used to produce synthetic samples that are closely approximated to target distribution from a feature representation field. The discriminator model is used to differentiate between the generated sample and the definite sample. The discriminator model forces the generator model to obtain more faithful samples by imposing penalties on idealistic generated samples during the optimization process. The two networks compete. As is summarized by Yi et al., [97], Multiple GAN variants can be categorized into three groups: 1) variants of the discriminator's objective, 2) variants of the generator's objective, and 3) variants of the architecture.

2.3.4.1 Conditional GAN

Traditional GAN was trained without limitations on data generation. Subsequently, it was enhanced by incorporating restricted restrictions to generate samples with required assets, known as conditional generative adversarial network (cGAN). The generator model of cGAN is built via FCN, either an end-to-end manner or a non-end-to-end manner. The former manner can produce equal-sized samples as input, while the latter manner cannot. The end-to-end FCN is commonly structured with a set of encoding layers and a set of decoding layers. The combination of these two sets of layers is able to derive equal-sized output. The encoding set uses convolutional layers with different settings of stride sizes and w/o pooling based on the users' purpose. On the other hand, the decoding set comprises deconvolution layers to achieve end-to-end mapping, several convolution layers, and a final layer for regression. Short skip connections, such as residual blocks [87] or dense blocks [95], can be employed by connecting the two sets of layers. The benefit of a residual block is mining the disagreement between the input and output. The aim of the dense block is the aggregation of the low- and high-frequency information from input. The long residual block creates a skip connection that directs information between long-distance layers, guiding the entire layers to concentrate on the optimization purpose [98]. Certain studies have incorporated attention gates into the long skip connection to trap crucial information instead of expanding the receptive field [17]. The kernels updated in hidden layers can be used to remove irrelevant features. The non-end-to-end FCN, on the other hand, typically only consists of a set of encoding layers, possibly followed by NN layers. The discriminator of the GAN model is implemented similarly to non-end-to-end FCN, expect the last layer culminating in a sigmoid or softmax operator to facilitate recognition or assessment.

Many different variations of the cGAN model have been suggested in bumping into specific anticipated requirements. This review work explores several cGAN models that can be either designed for or adaptable to medical image synthesis, including pix2pix, deep convolutional GAN (DCGAN), and Information Maximizing GAN (InfoGAN).

2.3.4.1.1 DCGAN

DCGAN can yield better and more steady training results when a fully convolutional layer is used as opposed to a fully convolutional layer. This architecture is illustrated in the work of [99]. In the core of the framework, pooling layers were replaced with fractional-stride convolutions, which allowed it to learn from a random input noise vector by using its own spatial up-sampling to generate an image from it. Two important changes were adopted to modify the architecture of early cGAN: batch normalization and leaky ReLU. Batch normalization [100] was employed to address inadequate initialization and avoid model collapse. The model collapse was a significant obstacle for GAN frameworks in earlier research. In all the layers of a discriminator, Leaky ReLU [101] activation was introduced as a replacement for maxout activation [96], leading to an improvement in the resolution of image output [38].

2.3.4.1.2 Pix2pix

The Pix2pix model, which is trained in a supervised manner, is an image-to-image translation model introduced by Isola et al. [102]. The network aims to learn a translation between the arrival data and the learning target and requires the generated data to be tightly close to the distribution of the target. To achieve the goal, the discriminator loss, which assesses the truthfulness of generated data, and the image-wise error (e.g., mean absolute error (MAE)) are employed. The use of image-wise error trains the model under supervised means, making it more fit for imaging synthesis tasks when the ground truth data are given. An illustration of this is the task of MRI-only radiation therapy, where the planning CT and co-registered MRI are provided for learning the model. After training, the synthetic CT (sCT) generated from the input MRI has accurate Hounsfield values (HUs) like a real CT, which are critical for accurate dose calculation in radiotherapy planning.

2.3.4.1.3 InfoGAN

Recently, there are groups of synthesis tasks that show that the trained model should not only rely on image intensity loss but also on histogram similarity: for example, the cone beam CT (CBCT) scatter correction [103, 104] and PET AC [105–108]. Under these requirements, a model may not be properly learned by just utilizing image-wise error as the loss function [109, 110]. InfoGAN is a variation of cGAN that was developed to learn disentangled representations via counting information-theoretic extensions [111]. It does this by maximizing the mutual information, which is an assessment of distribution similarity [112]. InfoGAN has been successful in disentangling written characters from the digit structures on the Modified National Institute of Standards and Technology public dataset [112]. It has also been used for generating volumetric CT from 2D kilovoltage image [113].

2.3.4.2 CycleGAN

Numerous diverse variations of CycleGAN models have been proposed for synthesis tasks recently. This work summarizes the use of these models in medical imaging [87, 95, 114–116]. As mentioned, in cGANs, there are two sub-networks: a discriminator and a generator. These two sub-networks are trained alternatively and competitively. Mismatches might persist between the input and output for specific clinic use, even with successful image registration, leading to potential issues. To tackle this concern, CycleGAN proposes