Medical Image Analysis 92 (2024) 103046

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Survey paper

Deep learning based synthesis of MRI, CT and PET: Review and analysis

Sanuwani Dayarathna [a,*], Kh Tohidul Islam [b], Sergio Uribe [c], Guang Yang [d], Munawar Hayat [a], Zhaolin Chen [a,b]

[a] Department of Data Science and AI, Faculty of Information Technology, Monash University, Clayton VIC 3800, Australia
[b] Monash Biomedical Imaging, Clayton VIC 3800, Australia
[c] Department of Medical Imaging and Radiation Sciences, Faculty of Medicine, Monash University, Clayton VIC 3800, Australia
[d] Bioengineering Department and Imperial-X, Imperial College London, W12 7SL, United Kingdom

ARTICLE INFO

Keywords:
Medical image synthesis
Generative deep-learning models
Pseudo-CT
Synthetic MR
Synthetic PET

ABSTRACT

Medical image synthesis represents a critical area of research in clinical decision-making, aiming to overcome the challenges associated with acquiring multiple image modalities for an accurate clinical workflow. This approach proves beneficial in estimating an image of a desired modality from a given source modality among the most common medical imaging contrasts, such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Positron Emission Tomography (PET). However, translating between two image modalities presents difficulties due to the complex and non-linear domain mappings. Deep learning-based generative modelling has exhibited superior performance in synthetic image contrast applications compared to conventional image synthesis methods. This survey comprehensively reviews deep learning-based medical imaging translation from 2018 to 2023 on pseudo-CT, synthetic MR, and synthetic PET. We provide an overview of synthetic contrasts in medical imaging and the most frequently employed deep learning networks for medical image synthesis. Additionally, we conduct a detailed analysis of each synthesis method, focusing on their diverse model designs based on input domains and network architectures. We also analyse novel network architectures, ranging from conventional CNNs to the recent Transformer and Diffusion models. This analysis includes comparing loss functions, available datasets and anatomical regions, and image quality assessments and performance in other downstream tasks. Finally, we discuss the challenges and identify solutions within the literature, suggesting possible future directions. We hope that the insights offered in this survey paper will serve as a valuable roadmap for researchers in the field of medical image synthesis.

1. Introduction

Medical imaging is crucial in clinical diagnosis and treatment monitoring as it provides specific information about the human body. Imaging modalities, including Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and Positron Emission Tomography (PET), are commonly used in clinical workflow, each providing unique structural, functional, and metabolic information that supports comprehensive clinical decisions. However, specific imaging modalities, such as PET and CT, pose a risk of radiation exposure, particularly for paediatric patients (Armanious et al., 2020). Furthermore, the acquisition process of comprehensive multi-modal images is costly, and longer scanner time also introduces artefacts (Zhan et al., 2021). Consequently, acquiring precise imaging in a safer manner is challenging in practical applications (e.g. MRI-only radiation therapy treatment planning).

Medical image synthesis offers a potential solution to these challenges by mapping from a given source image modality to a target modality, enabling the translation of images from one modality to another. Synthesizing medical images can maximize the utility of acquired images and reduce scanner time and operation costs (e.g. radiotracers) (Wang et al., 2021b). Consequently, medical image synthesis has gained significant traction in clinical applications such as MRI-only radiation therapy treatment planning, PET/MRI scanning, image segmentation, and image super-resolution (Armanious et al., 2020).

Mapping between image modalities poses a significant challenge in medical image synthesis due to high dimensionality, non-linear differences and the ill-posed nature of the problem (Nie et al., 2018). Conventional methods were initially the most common approaches used in medical image synthesis, which highly rely on handcrafted features selected by trained professionals (Yu et al., 2020). However, these methods have limited capability to represent complex image details and adversely affect the performance of the synthesis task. Consequently, conventional methods have demonstrated limited applications

[*] Corresponding author.
E-mail address: sanuwani.hewamunasinghe@monash.edu (S. Dayarathna).

https://doi.org/10.1016/j.media.2023.103046
Received 5 June 2023; Received in revised form 14 November 2023; Accepted 29 November 2023
Available online 1 December 2023

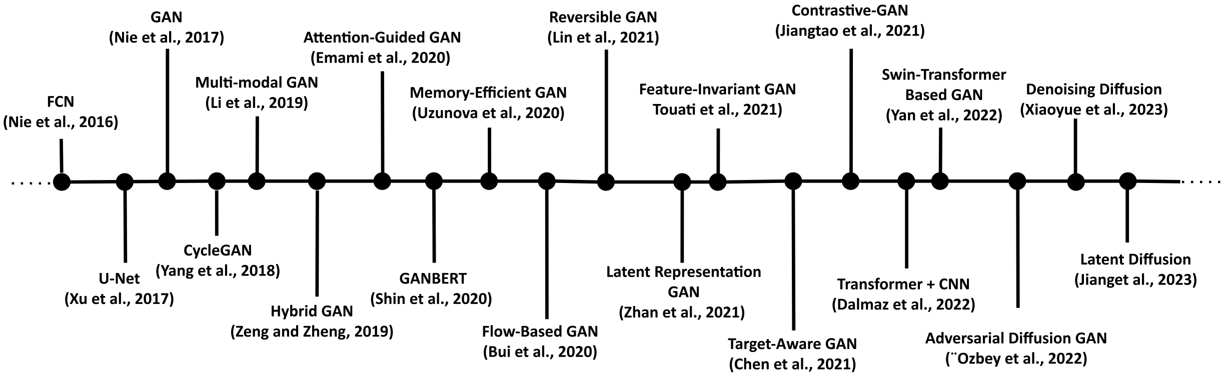

Fig. 1. A brief chronology of deep learning networks for medical image synthesis and generation in the literature.

and their reliability and generalizability are limited (Bahrami et al., 2020).

Recently, deep learning has become a prominent approach in computer vision by showing significantly improved performance in medical image synthesis for learning the complex non-linear relationship between image modalities (Wang et al., 2021). This has proven that deep learning-based data-driven methods can accurately model domain-specific characteristics across imaging modalities and synthesize images from different modalities. Further, these deep learning-based methods involve leveraging knowledge adopted from one specific task to improve the performance of another related task with minimal modification (Spadea et al., 2019). These benefits have facilitated the clinical applications of medical image synthesis with considerably enhanced performances.

Motivation and Contributions: Medical image synthesis has achieved promising results in the recent literature, owing to the expeditious growth of advanced deep learning frameworks. These methods have allowed accurate translation between diverse imaging modalities with the synthesis of clinically useful results (Lenkowicz et al., 2022). Inspired by this rapidly evolving field, in this paper, we present a comprehensive study of the applications of deep learning-based methods in cross-modality medical image synthesis, providing a reference and analysis of the literature. Fig. 1 provides a brief chronology of deep learning-based networks for medical image synthesis in the literature. The main contributions of this survey paper are as follows:

- This survey provides a comprehensive review of the cross-modality medical image synthesis methods with an analysis of network structures, loss functions, performance, and an overview of data availability.
- This review further identifies challenges and limitations in the current methods and discusses the possible future directions in medical image synthesis.

Search Criteria: We conducted a literature search using Scopus and PubMed databases to find relevant articles published from 2018 to 2023 July with the search term **("synth*" OR "pseudo" OR "translat*") AND "deep learning" AND ("medical imag*" OR "CT" OR "MRI" OR "MR" OR "PET" OR "low dose" OR "low count" OR "low field")** in the title, abstract or keywords. Google Scholar was also used to identify possible omissions and excluded the review articles. We first screened titles and abstracts, considering inclusion criteria such as peer-reviewed journals and conference papers and excluding irrelevant topics such as image translation for super-resolution and reconstruction. After a full document review and cross-referencing,173 articles were included. Fig. 2 illustrates the methodological process followed in the literature review search flow.

While there have been several literature surveys on deep learning-based medical image synthesis (Wang et al., 2020; Fard et al., 2022), our review presents a unique and more comprehensive scope. We not only delve deeper into the analysis of model performance and compare datasets but also set ourselves apart by including the most recent advancements in Diffusion and Transformer-related models. In contrast, recent surveys such as those by Yi et al. (2019), Skandarani et al. (2023) primarily concentrate on Generative Adversarial Network (GAN) and U-Net models. Our broader perspective allows for an exhaustive comparison of various synthesis tasks, offering a holistic understanding of deep learning-based architectures and their efficiencies. Furthermore, we shed light on datasets that have been overlooked in prior works. We also present a detailed comparative analysis of outcomes across diverse network architectures and synthesis tasks, an aspect often neglected in many contemporary studies. Finally, our review identifies challenges in the domain and collates solutions presented in the existing literature.

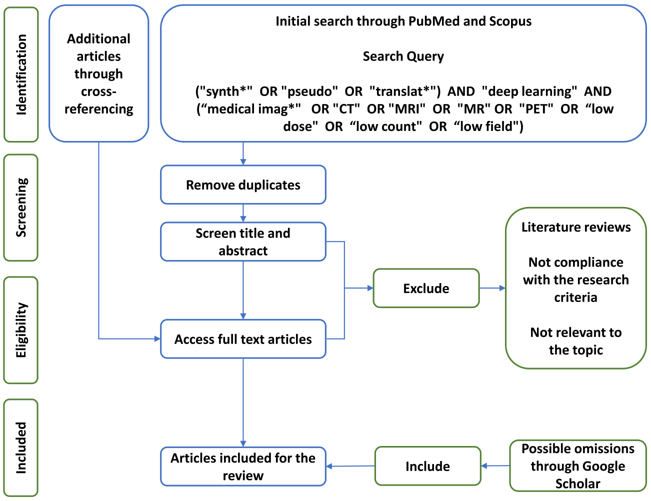
Identification
Screening
Eligibility
Included
Additional articles through cross-referencing
Initial search through PubMed and Scopus
Search Query
("synth*" OR "pseudo" OR "translat*") AND "deep learning" AND ("medical imag*" OR "CT" OR "MRI" OR "MR" OR "PET" OR "low dose" OR "low count" OR "low field")
Remove duplicates
Screen title and abstract
Exclude
Literature reviews
Not compliance with the research criteria
Not relevant to the topic
Access full text articles
Articles included for the review
Include
Possible omissions through Google Scholar

Fig. 2. An overview of the methodological process followed in the literature search flow.

Paper Organization: The schematic representation of this manuscript's layout can be found in Fig. 3. Section 2 provides an overview of synthetic image contrasts and deep learning-based networks in medical image synthesis. Section 3 presents a review of various deep learning-based medical image synthesis methods. Section 4 analyses the network architectures and their applications in image synthesis, including utilizing different loss functions and model training techniques. Sections 5 and 6 offer an overview of the other datasets used in the literature and the performance evaluation, including a description of the various evaluation methods and metrics used in the proposed methods. Section 7 discusses the challenges and potential future directions of the literature reviewed, and Section 8 concludes the review.

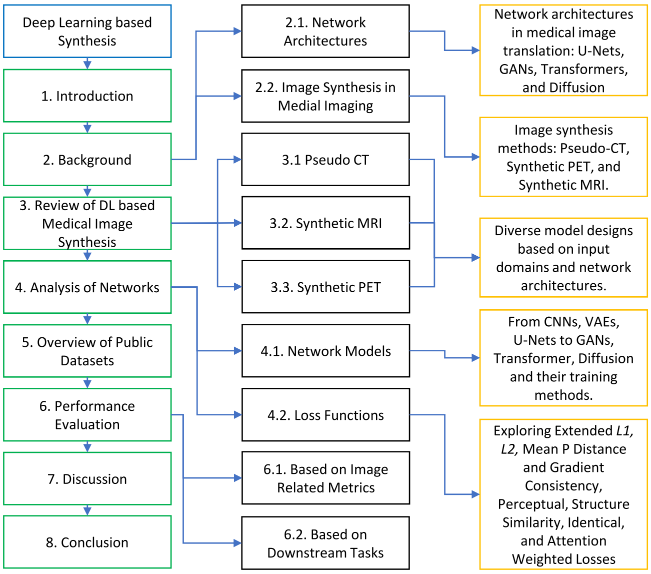
Deep Learning based Synthesis
1. Introduction
2. Background
3. Review of DL based Medical Image Synthesis
4. Analysis of Networks
5. Overview of Public Datasets
6. Performance Evaluation
7. Discussion
8. Conclusion
2.1. Network Architectures
2.2. Image Synthesis in Medial Imaging
3.1 Pseudo CT
3.2. Synthetic MRI
3.3. Synthetic PET
4.1. Network Models
4.2. Loss Functions
6.1. Based on Image Related Metrics
6.2. Based on Downstream Tasks
Network architectures in medical image translation: U-Nets, GANs, Transformers, and Diffusion
Image synthesis methods: Pseudo-CT, Synthetic PET, and Synthetic MRI.
Diverse model designs based on input domains and network architectures.
From CNNs, VAEs, U-Nets to GANs, Transformer, Diffusion and their training methods.
Exploring Extended L1, L2, Mean P Distance and Gradient Consistency, Perceptual, Structure Similarity, Identical, and Attention Weighted Losses

Fig. 3. An overview of the paper organization.

2. Background

2.1. Background of network architectures in medical image translation

In the early stages of medical image synthesis, Autoencoders and Fully Convolutional Networks (FCNs) were the architectures of choice, as evidenced by studies such as Nie et al. (2016), Xiang et al. (2017). As the field matured, more advanced architectures, namely U-Net and Generative Adversarial Networks (GANs), became prominent, as delineated by works like Nie et al. (2017). Recent times have witnessed the emergence of avant-garde models, specifically Vision-Transformer and Diffusion-based architectures, which exemplify the next wave of image-generative networks. This evolutionary trajectory illustrates a progression from rudimentary autoencoder-based designs to more sophisticated Diffusion-based models, as highlighted in Khader et al. (2023), Kazerouni et al. (2022). Our exhaustive review of contemporary medical image synthesis literature shows that U-Nets, GANs, Transformers, and Diffusion-based architectures dominate the discourse. Thus, the following sections of this review will elucidate upon these predominant model architectures.

In this paper, we adopt the following notational conventions. The generator network within GAN architectures is symbolized by G, whereas D stands for the discriminator network. For image-to-image translation between two distinct modalities, we designate the source modality as x and the target modality as y. The notation z is reserved to represent a random low-dimensional noise vector.

2.1.1. U-Nets

Ronneberger et al. (2015) initially proposed U-Net for image segmentation, featuring an encoder (contraction path), a bottleneck layer, and a decoder (expansion path) that form a U-shaped structure. The symmetrical architecture between the encoder and decoder allows for the extraction and concatenation of feature maps. Meanwhile, encoded features are transmitted to the decoder via skip connections and a bottleneck layer. U-Net is an efficient Convolutional Neural Network (CNN) architecture for image translation designed to capture the images' high-level and low-level features.

As depicted in Fig. 4, the U-Net encoder architecture comprises a series of convolutional layers followed by Rectified Linear Unit (ReLU) activation (Spadea et al., 2019). The encoder module extracts features, which are downsampled using max-pooling operations, simultaneously increasing feature channels to produce spatially contracted feature maps. These features subsequently pass through a bottleneck layer containing cascaded convolutional layers and are sent to the decoder. The decoder block consists of up-convolutions that concatenate high-resolution features originating from the encoder (Bahrami et al., 2020). The most commonly utilized loss functions for U-Net based image to image translation includes intensity based pixel-wise loss functions such as $\mathcal{L}_1 := \|\cdot\|_1$ and $\mathcal{L}_2 := \|\cdot\|_2$. In practice, the loss functions are calculated by using Mean Absolute Error (MAE) and Mean Squared Error (MSE) by averaging the absolute differences between synthesized and ground truth image's intensity values (Sikka et al., 2018).

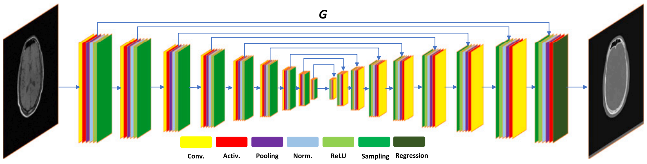

Fig. 4. Overview of UNet-based CT synthesis from MRI.

2.1.2. Generative adversarial networks

Generative Adversarial Networks (GANs) involve two competing networks — the generator and the discriminator optimized through a minimax-based method (Li et al., 2021b). The generator takes z as input, originating from a distribution $p(z)$ that is either uniform or Gaussian and learns to map z from low-dimensional space to high-dimensional real space. The discriminator receives the generated fake sample ($G(z)$) from the generator and the real sample from the training set. The discriminator's task is to classify generated data as fake or real. At the same time, the generator aims to create data as similar as possible to real samples, making it difficult for the discriminator to distinguish between fake and real data. In this manner, the generator and discriminator enhance each other's performance during the training process by optimizing the same objective function (Li et al., 2021b), given as (Eq. (1))

$$\mathcal{L}_{GAN} = \mathbb{E}_{x\sim p(x)}\left[\log D(x)\right] + \mathbb{E}_{z\sim p(z)}\left[\log(1 - D(G(z)))\right] \tag{1}$$

where $\mathcal{L}_{GAN}$ denotes the loss functions of the generator and the discriminator, the optimization happens while the discriminator maximizes and the generator minimizes the same objective function. $\mathbb{E}_x$ represents the expected values over the x, and $\mathbb{E}_z$ represents the expected values over z to the generator network.

Various GAN networks have been developed for specific generative tasks following the original GAN architecture. Deep Convolutional GAN (DCGAN) was created by integrating GANs with CNNs, making the architecture more suitable for image generation (Radford et al., 2015). Conditional-GAN (cGAN), commonly used in medical image synthesis, enhances the model's controllability by adding a constraint to the objective function, which could be labelled data in another mode or even an image (Li et al., 2021b). Fig. 5 (A) presents an overview of a cGAN for synthesizing CT images from MRI data. During the image generation process, MRI serves as a conditional input to the GAN, which then generates a CT image using the generator G. The discriminator D receives both the generated and the real CT images, aiming to distinguish between real and generated images.

CycleGAN uses two generator and discriminator models to perform image translation between two domains (A and B) without paired images (Zhu et al., 2017). G_A and G_B execute image translation from $B \rightarrow A$ and $A \rightarrow B$, respectively, where x belongs to domain A and y belongs to domain B. The CycleGAN objective function comprises the regular GAN generator loss $\mathcal{L}_{GAN}(G_A, D_A)$ and $\mathcal{L}_{GAN}(G_B, D_B)$, as well as an additional cycle consistency loss ($\mathcal{L}_{cyc}$) that calculates the pixel-wise loss or $\mathcal{L}_1$ loss between real and cyclic data, as shown below ((2)

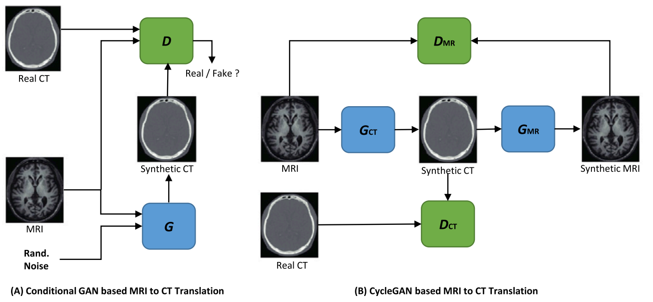

Fig. 5. Overview of cGAN and CycleGAN-based CT synthesis from MRI.

and (3))

$$\mathcal{L}_{cycGAN}(G_A, G_B, D_A, D_B) = \mathcal{L}_{GAN}(G_A, D_A) + \mathcal{L}_{GAN}(G_B, D_B) + \mathcal{L}_{cyc}(G_A, G_B) \tag{2}$$

where

$$\mathcal{L}_{cyc}(G_A, G_B) = \mathbb{E}_{x\sim A}[\|G_A(G_B(x)) - x\|_1] + \mathbb{E}_{y\sim B}[\|G_B(G_A(y)) - y\|_1] \tag{3}$$

CycleGAN's key feature is its cycle consistency, which connects cGANs through inverse mapping. One cGAN receives an input image as the conditioning image and generates a new image, which then serves as the conditioning variable for the second cGAN network. Fig. 5 (B) presents an overview of CycleGAN-based CT synthesis from MRI data. The generator G_{CT} takes a real MRI image as a conditional input and produces a CT image. The generated CT image is then input for the generator G_{MR} to reconstruct the original MR image. Discriminator D_{MR} differentiates between the real and reconstructed MR images from the cyclic generation, while D_{CT} distinguishes between the generated and real CT images.

2.1.3. Vision transformers

Originally conceptualized for sequence inferencing tasks in Natural Language Processing, the Transformer architecture was pioneered by Vaswani et al. (2017). What sets Transformers apart from other sequence data handling architectures is their exceptional performance, underpinned by a self-attention mechanism adept at encapsulating long-range relations within data. In the domain of computer vision, Transformer-based architectures have obtained significant attention due to their capability to grasp the global context of the images. The most prominent model in Vision is Vision Transformers (ViT) which adapts the foundational structure of vanilla Transformer networks. Since then, ViTs have been extensively used in most common vision-based applications such as object detection, image classification, and image segmentation. Recently, these vision Transformer-based network architectures have seen extensive use in medical imaging (Shamshad et al., 2022).

In ViTs, images are delineated as sequences of non-overlapping image patches. These patches undergo processing via an encoder, followed by task-specific decoder modules. Crucially, alongside the image patches, the associated positional information is integrated into the encoder block. This block is structured with multi-head self-attention (MHSA), normalization, and multi-layer perceptron (MLP) layers, as depicted in Fig. 6. Through the MHSA layer, attention maps are formulated for the embedded image tokens, thereby enabling the network to selectively prioritize the most salient regions within the image.

2.1.4. Denoising diffusion probabilistic models

Denoising Diffusion Probabilistic Models (DDPMs), firstly introduced by Ho et al. (2020), is a novel approach that has demonstrated an excellent ability to model a generative process. As shown in Fig. 7, DDPMs are a parameterized Markov chain which is trained to map from pure noise to actual data through a gradual process over finite time steps T.

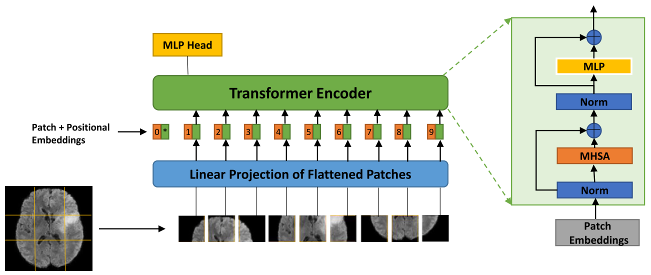

Fig. 6. Overview of ViT architecture (Shamshad et al., 2022).

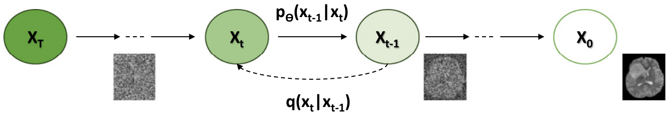

Fig. 7. Overview of Diffusion-based MRI synthesis.

The process of learning by denoising mainly consists of two processes as forward and reverse process. In the forward process (Eq. (4)), random Gaussian noise is added to the input image x_0 in a series of sufficiently large time steps T to obtain noisy image x_T from an isotropic Gaussian distribution. This forms a Markov chain by which the mean distribution of the current step x_t is conditioned on the sample from the previous step by following a noise variance schedule β_t as in Eq. (5).

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \tag{4}$$

$$x_t = \sqrt{1-\beta_t}x_{t-1} + \epsilon\sqrt{\beta_t}; \epsilon \sim \mathcal{N}(0, I) \tag{5}$$

where ϵ is the added noise, $\mathcal{N}$ is the Gaussian distribution and I is the identity covariance matrix. Due to the Markov property of this process, the marginal distribution x_t can be directly obtained with a given input sample x_0.

The reverse diffusion process (Eq. (6)), $p_\theta(x_{t-1}|x_t)$, starts with the pure noise distribution and denoises the sample in each time step by forming a Markov chain from x_T to x_0 as shown in Fig. 7 . With a smaller β_t, the transition between x_t and x_{t-1} can be approximated as Gaussian distribution as both forward and reverse processes have the same functional form,

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sum_\theta(x_t, t)) \tag{6}$$

Each reverse step of the Diffusion models maps through a neural network p_θ and trains by optimizing the variational lower bound on simplified log-likelihood.

The denoising network, p_θ, estimates the added noise ϵ by minimizing the loss between actual noise and the network estimate ϵ_θ as follows (Eq. (7)),

$$\mathcal{L}_{error} = \mathbb{E}_{t,x_0,\epsilon}[\|\epsilon - \epsilon_\theta(x_t)\|_2^2] \tag{7}$$

In each reverse step $t \in \{1, T\}$ the mean distribution μ_θ is derived using ϵ_θ and sample x_{t-1} as in eq. (6)

Although Diffusion-based medical image synthesis has demonstrated promising results, its efficacy is hindered by the computational burden of image sampling as a likelihood-based model, which necessitates substantial computational resources for modelling. Many recent work focus on improving their computational efficiency, e.g. Rombach et al. (2021).

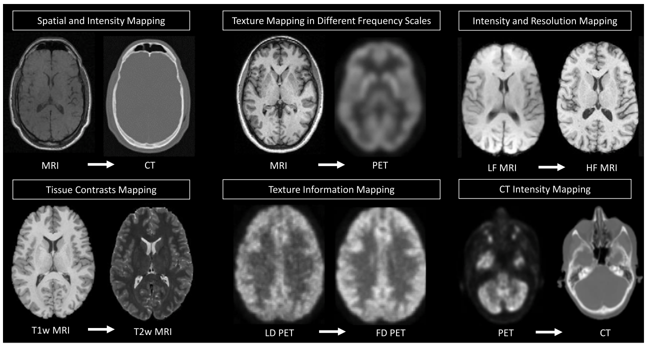
Spatial and Intensity Mapping
MRI
CT
Texture Mapping in Different Frequency Scales
MRI
PET
Intensity and Resolution Mapping
LF MRI
HF MRI
Tissue Contrasts Mapping
T1w MRI
T2w MRI
Texture Information Mapping
LD PET
FD PET
CT Intensity Mapping
PET
CT

Fig. 8. An overview of non-linear mapping between image modalities in Deep-learning-based medical image synthesis.

2.2. Background of image synthesis in medical imaging

Medical image synthesis offers an effective approach for the realistic simulation of various pathological conditions. This provides a robust solution to avoid repeated scans with high radiation exposure, especially for paediatric patients, which increase the risk of brain cancers and leukaemia (Boroojeni et al., 2022). The extensive utilization of medical image synthesis in clinical pathology is particularly prominent in the context of MRI-only radiotherapy treatment planning for cancer patients (Fu et al., 2020; Touati et al., 2021; Mendes et al., 2023). Additionally, it plays a vital role in enhancing computer-aided systems for clinical diagnosis by data augmentation, as they heavily rely on the availability of sufficient training data (Salem et al., 2019). Furthermore, this offers a solution for acquiring data from the normative population, which is challenging due to the predominance of medical images available from patients with pathological conditions. Moreover, medical image synthesis is of great value in acquiring imaging data pertaining to high-risk scenarios such as contrast-enhanced MRI, which is crucial for patients with liver tumours and examining their impact on diagnostic accuracy with non-contrast images (Zhao et al., 2020a).

Medical image synthesis methods encompass a range of modalities, including Pseudo-CT or synthetic CT, PET, and MR. These methods can be categorized based on their applications as image translation between different modalities or the translation between two different contrasts of the same modality. Fig. 8 presents an example of the synthetic MR, PET, and CT modalities, highlighting their underlying characteristics.

MR is naturally a multi-contrast imaging modality, and MRI data frequently comprise multiple imaging contrasts with complementary information (Zhao et al., 2021). However, acquiring multiple image contrasts is not always possible due to limited scanning time and the increasing cost of MRI. In addition, the acquisition of multiple contrasts is prone to unintended image artefacts with random noise, resulting in poor image quality (Chen et al., 2022). Therefore, synthesizing missing or corrupted MRI contrasts from other successfully obtained contrasts is vital for a reliable clinical diagnosis and to assist in comprehensive image analysis tasks such as image registration and segmentation (Zhan et al., 2022). Low-field (LF) MRI also offers a remarkable solution by providing more affordable equipment and reducing costs in medical imaging. However, these low-field MR scans yield images with lower signal-to-noise ratios due to the reduced magnetic field strength. Therefore, the synthesis of high-field (HF) MR images proves valuable in generating images with higher spatial resolution and enhanced contrast derived from low-field MR images.

MRI is important for radiotherapy treatment planning applications (Fu et al., 2020). However, standard MR-guided radiotherapy (MRgRT) still requires CT images for dose calculation, which may be limited due to interscan differences (Boni et al., 2021). To address this limitation, synthetic (pseudo) CT can be generated from MRI to delineate bone and muscle boundaries (Armanious et al., 2019). For PET/MR examinations, CT images can also be synthesized from MRI for attenuation correction (AC) (Baran et al., 2018; Pozaruk et al., 2020). Medical image synthesis allows for mapping anatomical, functional, and metabolic information, such as spatial resolutions, pixel intensities, and texture-wise features (Chen et al., 2018b).

Synthetic PET is potentially useful for diagnosing degenerative disorders, such as Alzheimer's disease, where grey matter atrophy, ventricular enlargement observed in MRI, and cerebral distribution of fluorodeoxyglucose (FDG) in PET serve as crucial differentiating factors (Sudarshan et al., 2021). Synthetic PET can potentially aid the diagnosis of cerebrovascular diseases using MRI-derived cerebral blood flow (CBF) maps (Hussein et al., 2022) and detect abnormalities of various anatomies and lesions. It can also contribute to generating diverse data for developing and evaluating PET reconstruction algorithms (Rajagopal et al., 2023). Further, low-dose (LD) PET images exhibit more complex spatial variation and statistical noise than high-dose images, rendering them less reliable for diagnostic purposes. Consequently, the full-dose (FD) PET synthesis process aims to recover the missing high-frequency details from low-dose images to achieve superior image quality (Sikka et al., 2021; Pain et al., 2022).

3. Review of deep learning based medical image synthesis

3.1. Pseudo CT

Deep learning-based techniques for CT synthesis have demonstrated remarkable results in capturing the highly complex, underlying non-linear mapping between CT and source modalities to generate realistic synthesized images (Kläser et al., 2018; Armanious et al., 2019; Qian et al., 2020; Shi et al., 2021; Koh et al., 2022). These methods have further proven their applicability in clinical and non-clinical applications by outperforming conventional CT synthesis methods, such as atlas-based and voxel-based (Xiang et al., 2018). CNNs and GAN-based approaches, in particular, have shown great potential for accurately estimating CT Hounsfield Units (HU) from source image modalities (Xiang et al., 2018; Emami et al., 2018). In addition, novel transformer and Diffusion-based networks have shown superior performance with high-fidelity synthetic images (Li et al., 2023a,b,c; Lyu and Wang, 2022). This section reviews the deep learning-based approaches for CT synthesis.

3.1.1. Model design based on input domain

- Preserving 3D information

Deep 3D networks have also demonstrated better results in encoding complex mapping between MRI and CT to obtain more accurate HU predictions. Fu et al. (2019) showcased this by marking the first end-to-end 3D CNN application for MRI-to-CT translation of pelvis images. Zimmermann et al. (2022) proposed a 3D U-Net approach that uses multiple input MRI sequences to generate synthetic CT images. Additionally, 3D cGAN-based approaches were widely adopted to address the discontinuity between image slices in 2D networks (Kläser et al., 2018; Lei et al., 2019b; Liu et al., 2020b; Oulbacha and Kadoury, 2020; Koh et al., 2022; Wang et al., 2022a).

Due to the computational cost of deep network structures with volumetric data, several studies (Li et al., 2018; Spadea et al., 2019; Hsu et al., 2022; Sun et al., 2022a) have employed 2D images with innovative approaches to preserve 3D structural information of images. Li et al. (2018) proposed a triple orthogonal 2D fully convolutional neural network (FCN) to retain structural information in synthetic CT images. The proposed model's architecture consists of a triple orthogonal network with three 2D parallel CNNs for axial, coronal, and sagittal image planes, trained in parallel for each plane to synthesize one CT image using a linear combination of generated images. The concept of the triple orthogonal model was also used by Spadea et al. (2019) and Maspero et al. (2020) with three separate U-Nets. To mitigate

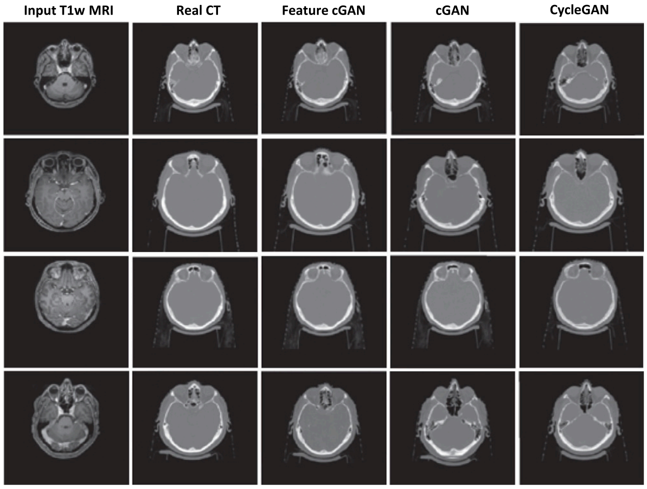

Fig. 9. The first column shows the source MR images, and the second column shows the ground truth CT images. Other columns are synthesized CT images for different deep learning networks, including feature-invariant cGAN, cGAN and CycleGAN for the head and neck imaging datasets. The proposed feature cGAN accurately matches the ground truth contrary to other deep models, which are less reliable visually.
Source: Image from Touati et al. (2021) on feature-invariant CT synthesis.

the high computational requirement in 3D networks, Oulbacha and Kadoury (2020) proposed a 3D volume-based training method comprising four neighbouring 2D slices stacked together, capturing nearby 3D features without relying on computationally heavy 3D convolutions.

- Capturing features in challenging body regions

Prediction of CT images from the head and neck MRI images is challenging due to the variability of patient anatomies and imaging protocols (Qi et al., 2020; Klages et al., 2019; Touati et al., 2021; Ang et al., 2022). It is, therefore, essential to identify high-level features such as contours and give them more weight in the models. Touati et al. (2021) proposed a novel feature invariant method to match common structural details of synthesized and real CT images of the neck and head regions for high-frequency components. This method identifies high-level features, including contours, to generate images with similar high-frequency feature distribution in selected regions. This was achieved by encoding the deep feature space from the CNN using quantized histograms of intensity distribution and edge attributes. Fig. 9 illustrates the comparison of the proposed method in HU space. Synthesize of thorax images is also challenging due to the heterogeneous nature of lungs in terms of electron density and the difficulty involved in modelling lesions (Lenkowicz et al., 2022). By employing a hybrid approach that incorporates assigning bulk density values to the Gross Tumour Volume, they have successfully demonstrated the ability of synthetic thorax CT imaging to accurately replicate the intricate details of challenging lung regions.

Fu et al. (2020) demonstrated the first deep learning-based CT synthesis for radiotherapy treatment planning in the liver. Cusumano et al. (2020) and Lapaeva et al. (2022) also proposed the utilization of synthetic CT for pelvis and abdomen regions. Compared to other body regions, head CT imaging was widely used in paediatric CT scans to identify cranial abnormalities in child patients. Paediatric patients also have thinner skull bones compared to adults and a much smaller contrast between bone and soft tissues, making the synthetic task more challenging (Boroojeni et al., 2022). However, most deep learning-based CT synthesis from MRI for adults was not developed to identify small bony structures such as fractures and sutures. Boroojeni et al. (2022) proposed a novel synthetic CT method for paediatric cranial bone imaging with a patch-based residual-UNet. They have used two patch-based networks for whole-head (NetWH) and bone or air (NetBA). The complete pseudo-CT with the higher spatial resolution was obtained by combining synthesized CTs from NetWH and NetBA.

- Dealing with multi-contrast domain data

In the context of mapping between MRI and CT images, numerous GAN-based CT synthesis approaches rely on mapping CT images with single MR sequences (Hiasa et al., 2018; Armanious et al., 2019; Dong et al., 2019; Ge et al., 2019). However, these methods do not address the variability in MR contrasts and protocols, which arise a common issue with the generalizability of the deep network. Augmented CycleGAN (AugCycleGAN) is an innovative method that addresses this limitation by improving the generalizability of CycleGANs to use multicentric data (Almahairi et al., 2018). Building upon this concept, Boni et al. (2021) proposed an augmented CycleGAN designed to translate multiple MR images to a single CT image. In the AugCycleGAN framework, the generator is augmented by a latent space (Z_{mri}) that serves as an additional input space for missing details, without learning the $CT \rightarrow MRI$ mapping using a single input. Consequently, MR-to-CT mapping has been developed as a CT synthesis method without a dedicated MR contrast.

In a recent contribution, Zhou et al. (2023) unveiled a cascade-based multi-modality synchronous generation network, designed to leverage a single MRI modality for multiple synthetic tasks. This innovative cGAN-based approach incorporates a shared encoder, paired with a multi-branch decoder in the generator, facilitating the synthesis of multi-modality MRI images from a single T1w MRI. Subsequently, a composite image – derived from both synthesized MRIs and the input MRI – is employed for the generation of corresponding CT images. Notably, the method offers enhanced information crucial for crafting realistic CT images by astutely harnessing intermediately generated MRI images.

- Dealing with limited datasets

Qian et al. (2020) introduced a novel GAN architecture by adding a classifier to the discriminator model, improving the network's stability and accuracy when training with limited samples. Abu-Srhan et al. (2021) proposed using knowledge transfer from a non-medical pre-trained model to resolve the problem with limited paired datasets. A well-performing non-medical pre-trained model was chosen as the source for the proposed network and then transferred to learn the model using both paired and unpaired data. Li et al. (2021a) and Wang et al. (2022a) also highlighted the use of transfer learning on pre-trained CycleGANs to improve the model's generalizability to different datasets and modalities. Li et al. (2020) proposed a shallower U-Net-based generator using image patches with 50% overlap between patches instead of the whole image to augment the data. They further made the U-Net model shallower to train with fewer image data.

3.1.2. Model design based on network architectures

- Spatial transformation and attention

Most recent medical image translation approaches focus on the effectiveness of translating the whole image instead of a specific region of interest (Chen et al., 2018a; Florkow et al., 2020; Kazemifar et al., 2019; Ranjan et al., 2022; Wang et al., 2022a). However, this may result in poor image quality in localized areas with deformed, blurry, or impure textures. Inspired by the human attention mechanism, Emami et al. (2020a) proposed the attention-based GAN method to compute spatial attention within the discriminator to assist the generator in focusing on regions with considerable distance between real and synthetic CT images. However, training an attention-gated discriminator is difficult due to the problem with gradient saturation. Kearney et al. (2020) proposed an alternative solution by using a variational autoencoder to impose an attention mechanism in the discriminator while allowing it to progress with the generator. Yang et al. (2020) incorporated a self-attention block in the generator for modelling non-local relationships for a broad spatial region of interest. Chen et al. (2021) proposed a novel solution with a Target Aware Generative Adversarial Network, which includes a generator with two encoder–decoder blocks. One stream translates the whole image, while the other focuses on a target region.

Dovletov et al. (2022) proposed a Grad-CAM guided attention method for image translation. Grad-CAM is a gradient-based visualization method that produces interpretable localization maps of a given image. The proposed method guides the U-Net-based generator model using the mean squared error (MSE) between CT class-specific localization maps obtained from real and synthesized images using a pre-trained classifier with CT and MR to make the translation network focus on a specific region. This approach was further extended with a double Grad-CAM-based method with a greater focus on the bone structures by using the pre-trained classifier guidance twice (Dovletov et al., 2023).

- Mapping multi-scale feature

Retaining high-frequency details in the synthesized CT images is a major challenge when synthesizing quality images. Standard loss functions including $\mathcal{L}_1$ loss perform well in the context of low-frequency image contents but not for high-frequency details (Shi et al., 2021). Armanious et al. (2020) and Ang et al. (2022) proposed a novel approach to capturing high and low-frequency components in image target modalities by uniquely combining adversarial frameworks with non-adversarial losses. The proposed generator model is penalized pixel-wise and perceptually using an adversarial discriminator while employing a pre-trained feature extractor to ensure that the translated images match the desired target image in texture, style, and content (Armanious et al., 2020). Shi et al. (2021) directly used a frequency-supervised network to retain high-frequency information during modelling. The approach consists of a base network for synthetic CT generation and a frequency-supervised synthesis network as a refinement module to improve the quality of high-frequency features in generated CTs. Cao et al. (2021) proposed a U-Net-based approach to preserve low-level features of images by enhancing the semantic information of images with the advantage of U-Net's multi-scale feature fusion. To maximize the mutual information between source and target images in CycleGANs, Park et al. (2020) proposed a Contrastive Unsupervised Translation (CUT) network by learning a cross-domain similarity function. Building on the CUT architecture, Jiangtao et al. (2021) developed a novel deep network that replaced the generator's residual blocks with dense blocks inspired by the Dense-CycleGAN network (Lei et al., 2019b).

- Modelling 3D spatial information

Lei et al. (2019b) and Liu et al. (2020b) proposed a dense CycleGAN-based method with a patch-to-patch translation to utilize 3D spatial information with less memory requirement. Additionally, deep neural networks require large training datasets and the learning of numerous parameters to obtain high-quality synthesized images. Zeng and Zheng (2019) addressed this problem by proposing a hybrid GAN network with a 3D generator and 2D discriminator networks. A 3D generator with fully-connected CNN learns spatial information and resolves discontinuity across image slices, while a 2D discriminator reduces the memory burden for training the networks.

- Transformer and Diffusion based models

Vision Transformers and Diffusion-based methods have be used in recent literature for CT synthesis due to their potential ability to generate high-fidelity images (Vaswani et al., 2017; Ho et al., 2020). Hybrid architectural structures with CNNs and Transformers were introduced to capture the multi-level information from MR images and synthesize CT with improved intensity and structural details (Li et al., 2023c). Li et al. (2023b) used an adaptive multi-sequence fusion network to model both voxel and context-wise correlation between different MRI contrasts. This includes the combination of convolution with cross-attention modules to exploit both local and contextual information across the multiple image contrasts. Li et al. (2023c) and Zhao et al. (2023) also demonstrated that hybrid synthesis approaches with CNNs and Residual Transformers were able to capture both local texture details and the global correlation between MR and CT images. Lyu and Wang (2022) proposed a denoising diffusion probabilistic model and score-matching model conditioned on MR images for generating more realistic CT images.

Li et al. (2023a) also proposed a similar MRI-conditioned Diffusion-based approach by inserting refined null-space contents (Schwab et al., 2019; Wang et al., 2022) from sparse view CT into the reverse denoising process. By manipulating the denoising steps in CT measurements, this approach has offered precise guidance in the inverse process of the Diffusion model. Moreover, Pan et al. (2023a) integrated a shifted-window transformer (Swin Transformer) (Liu et al., 2021b) based encoder decoder in the denoising process of 3D DDPM, conditioned on MRI images. This method has obtained significant performance compared to GAN-based methods capturing complex structural details of high dimensional data with the sequential process followed in the denoising steps.

- Summary of pseudo CT

Table 1 presents an overview of deep learning-based pseudo-CT methods categorized by their feature-wise attributes, summarizing the most relevant studies in each category while highlighting their novelties, network types, loss functions, input modalities, and data regions.

3.2. Synthetic MR

Multimodal MR image synthesis employing deep learning techniques has shown promising results (Dai et al., 2020; Yang et al., 2021). Based on the current state-of-the-art methods for mapping between MR modalities, MR synthesis can be classified into single-modality and multi-modality-based deep learning strategies (Chartsias et al., 2018; Dar et al., 2019; Li et al., 2019b; Dai et al., 2020; Zhan et al., 2021; Yurt et al., 2022; Zhang et al., 2022c; Kawahara et al., 2023). Most MR contrast synthesis techniques rely on GANs for image-to-image translation between MR contrasts, utilizing 2D U-Net-based generators and PatchGAN discriminators, with a few exceptions (Chartsias et al., 2018; Salem et al., 2019; Osman and Tamam, 2022; Liu et al., 2021c; Hu et al., 2022) employing U-Net and variational autoencoder (VAE) based methodologies. Moreover, there is growing interest in Transformer and Diffusion-based methods for MRI synthesis from the recent literature (Dalmaz et al., 2022; Zhang et al., 2022a; Özbey et al., 2022; Zhu et al., 2023).

3.2.1. Model design based on input domain

- Preserving 3D information

In contrast to traditional adversarial-based approaches for MRI synthesis, Uzunova et al. (2020) proposed a 3D multi-scale patch-based method for generating high-resolution MRI contrasts, leveraging a low-resolution GAN followed by a series of high-resolution GANs at each resolution level. Mao et al. (2022) introduced a novel technique to extract deep semantic information from high-frequency details, which are subsequently incorporated with feature maps within the decoder network. In addition to feature-wise enhancements in the networks, Zhao et al. (2021) proposed an MRI-Trans-GAN that employs larger 3D patches in sagittal and coronal axes and smaller patches in the vertical axis to reduce memory usage and maintain the relationship among adjacent slices.

Table 1
Overview of DL-based Pseudo CT.

Feature	Highlights	Architecture	Loss function	Input data and region	Reference
Mitigate discontinuity between image slices	Dense block generator with 3D patch based training	3D CycleGAN	pl -norm distance, Termed mean p distance + Gradient loss	T1w, T2w - MRI, Brain	Lei et al. (2019b)
	Utilize 3D volume consisting of 4 neighbouring 2D slices		AL* + CCL#	T2w - MRI, Lumbar spine	Oulbacha and Kadoury (2020)
	Adversarial 3D FCN with Auto-Context Model	3D cGAN	AL + GDL	MRI, Brain, Pelvic	Nie et al. (2018)
	3D patches based GAN network		AL, MPD, GDL.	3T, T1w - MRI, Brain	Koh et al. (2022)
	Improve continuity between slices	2.5D U-Net	Weighted MSE	3T, T1w - MRI, Abdomen	Liu et al. (2020a)
		2.5D CycleGAN	AL + CCL + Shape consistency loss	MRI, Brain	Sun et al. (2022a)
Hybrid Architecture	3D Generator and a 2D Discriminator	(2D, 3D) cGAN	AL + $\mathcal{L}_1$ loss	3T, T1w - MRI, Head	Zeng and Zheng (2019)
	Hybrid network associated with Auxiliary Classifier-GAN and cGAN, ResNet, U-net	2D cGAN	AL + $\mathcal{L}_1$ loss + Softmax loss	MRI, Abdomen	Qian et al. (2020)
Triple Orthogonal Networks	Three 2D parallel CNNs for separately for axial, coronal, and sagittal image plane	2D U-Net	Tissue focused MSE	3T, T1w - MRI, Brain	Li et al. (2019a)
			*MAE	T1w - MRI, Head	Spadea et al. (2019)
Attention Guided Learning	Generates more accurate and sharper images through the production of attention masks.	2D CycleGAN	AL + $\mathcal{L}_1$ loss + structure, gradient, content-based, KL divergence, and softmax.	3T, T2w - MRI, Brain	Abu-Srhan et al. (2021)
	Attention mechanism to help the discriminator attend to specific anatomy within an image		AL + CCL	3T, T1w - MRI, Head, Neck, and Brain	Kearney et al. (2020)
	Self-attention strategy to identify the most informative components	3D CycleGAN	AL + CCL + GDL	FDG PET/CT, whole-body	Dong et al. (2019)
	Grad-CAM guided attention based network	2D U-Net	MSE + Grad-CAM similarity Loss	T1-w MRI, Cranial	Dovletov et al. (2022)
	Computes the spatial attention in discriminator to draws attention of the generator to regions that has more difference	2D cGAN	AL + $\mathcal{L}_1$ loss	1T, T1w - MRI, Brain	Emami et al. (2020a)
Hybrid Objective Functions	Enforce visual realism, structural consistency between image domains		AL + Style loss + Perceptual loss	T2w - MRI, Head, Neck	Ang et al. (2022)
	Utilize non-AL functions analogous to the perceptual and style transfer losses	2D CycleGAN	AL + CCL + Style loss + Perceptual loss	PET/CT, Brain	Armanious et al. (2019)

(*continued on next page*)

- Dealing with multi-contrast domain data

The majority of MR syntheses are performed within two contrasts, such as generating T1-weighted (T1w) MRI from T2-weighted (T2w) MRI (Dar et al., 2019; Yu et al., 2019; Bui et al., 2020; Uzunova et al., 2020; Kong et al., 2021; Zhao et al., 2021). Nevertheless, multi-modal image synthesis has also been investigated in MRI contrast synthesis to capitalize on the anatomical features acquired from multi-modal MRI. This emphasizes relevant properties among contrasts and yields high-quality synthesis results (Chartsias et al., 2018; Li et al., 2019b; Zhou et al., 2020a; Sharma and Hamarneh, 2020). Chartsias et al. (2018) showed a modality-invariant latent representation-based learning method that mapped all input modalities into a shared latent space using an encoder, latent fusion, and a decoder-based architecture. The fusion step combines the encoder representation of each modality into a single fusion, and the fused representation integrates the unique features of each modality, resulting in a robust model capable of accommodating missing input modalities. Zhan et al. (2021) proposed a latent representation-based cGAN for multi-modality MR synthesis with a similar generator network. They employed a latent space processing network (LSPN) to combine the features from the modalities using a ResNet, which returns a latent representation of the target modality.

By replacing the LSPN in the proposed architecture, Zhan et al. (2022) introduced a gate mergence (GM) mechanism for the integration of features from the encoders for each modality, which consists of fusion methods such as Add, Conv-cat, and Cat-Conv to combine the features of modalities. Moreover, GM can enhance crucial information, such as edges or texture, and eliminate irrelevant noises from the modalities. To map multiple image sequences to one target modality, Li et al. (2019b) proposed a CycleGAN-based approach by concatenating the input modalities into a multi-channel input and the target output modality also into a multi-channel with respect to the input modalities. The mapping results in a diamond-shaped topology with two generators and discriminators called DiamondGAN. In contrast to DiamondGAN, Dai et al. (2020) proposed a unified GAN-based method for single-to-multiple contrast mapping, utilizing a single generator to

Table 1 (*continued*).

High-frequency Feature Identification	Focus on high-frequency CT image parts through a decomposition layer	3D U-Net	MAE	MRI, Head	Shi et al. (2021)
	Multiscale-invariant representation to encodes the feature space at different image resolution levels	2D cGAN	AL + $\mathcal{L}_1$ loss	3T, T1w - MRI, Head, Neck	Touati et al. (2021)
	Achieve significantly better local translation for the target area	2D CycleGAN	AL + CCL + Crossing loss + Shape consistency loss	T1w, T2w - MRI, Abdomen	Chen et al. (2021)
Multi Sequence Inputs	CT synthesis using multicentric data with the possibility of using several MRI sequences		AL + CCL + Marginal matching loss	3T, 1.5T, T2w - MRI, Bladder, Rectum, Femoral Head	Boni et al. (2021)
	Adaptive multi-sequence fusion network on cross-modality attention	2D CNN	AL + $\mathcal{L}_1$ loss	T1w, T2w - MRI, Brain	Li et al. (2023b)
	Using multi-channel input with an independent feature extraction networks	2D cGAN	AL + $\mathcal{L}_1$ loss	1.5T, Recontrast T1w , postcontrast T1w with fat-saturation, and T2w MRI, Head	Tie et al. (2020)
	Using multiMR sequences on sCT		AL + $\mathcal{L}_1$ loss	T1w, T2w, T1C, T1DixonC-water, Head, Neck	Qi et al. (2020)
		3D U-Net	MAE	T1w, T2w - MRI, Head	Zimmermann et al. (2022)
Preserve Structure-consistency	A modality independent neighbourhood descriptor and position based selection strategy for selecting training images	2D CycleGAN	AL + CCL + Structure-consistency loss	1.5T - MRI, Brain	Yang et al. (2018b)
Transfer Learning based synthesis	Overcome limited dataset		AL + CCL	1.5T, T1w, T2w - MRI, Pelvic	Wang et al. (2021)
	An adapted model, generalizable to small and different MR protocols	2.5D CycleGAN	AL + CCL	1.5T, T1-FLAIR, T1-POST, T2w - MRI, Brain	Li et al. (2021a)
Transformer based synthesis	Hybrid Convolution and Transformer based multi-scale image synthesis	2D CNN + Transformer	AL + $\mathcal{L}_1$ loss + MS-SSIM	3T, T1w - MRI, Head, Neck	Li et al. (2023c)
	Residual Transformer Conditional GAN for multi-level feature extraction		AL + Feature reconstruction loss + MSE	T2w - MRI, Pelvic	Zhao et al. (2023)
Diffusion based synthesis	Conditional Diffusion and score based networks for CT synthesis	2D DDPM	MSE	T2w - MRI, Pelvic	Lyu and Wang (2022), Li et al. (2023a)

*AL : Adversarial Loss #CCL : Cycle Consistency Loss.

translate the source image to the target image and reconstruct the original source image from the generated image and its modality label. A contrast modulator modifies the encoder and decoder parameters to adapt them to different contrasts and is tuned using Filter Scaling and Conditional Instance Normalization.

- Dealing with low-field MRI data

DL-based conversion of low-field (LF) to high-field (HF) MRI offers a more efficient solution for synthesizing high-resolution MR images with enhanced anatomical details. Nevertheless, few studies have proposed deep learning-based solutions for high-field MR synthesis employing various supervised learning approaches. These proposed solutions use deep CNN-based networks to learn the detailed mapping between low-field and high-field images (Lin et al., 2019). Lin et al. (2019), Figini et al. (2020), Lin et al. (2023) presented an Image Quality Transfer(IQT) framework for enhancing LF MRI (0.36T) by improving the contrast and spatial resolution of the images. These approaches utilize a stochastic LF image simulator as a forward model to capture their uncertainty and the variations with respect to its high-field images and utilize anisotropic U-Net for the IQT inverse problem. A recent study (Bagheri and Uludag, 2023) in LF MRI explored the use of multi-contrast LF MRI images to synthesize HF MRI images in one or multiple contrasts. The study proposed a U-Net-based method, which was successfully applied to generate 3T MR images from 0.5T MRI images.

- Dealing with deformations in data

Bui et al. (2020) extended the flow-based method by employing temporal information between consecutive image slices to learn the displacement between slices using additional registration networks for each domain. However, paired image contrast with supervised learning approaches for MRI synthesis is heavily afflicted by misalignments in the paired data, which may cause displacements within the synthesized images (Li et al., 2019b; Kong et al., 2021). Addressing this primary concern, Kong et al. (2021) treated misalignments as noisy labels and converted the problem into a method with noisy labels called RegGAN. This approach offers a standard image translation and registration solution, where a generator is trained alongside a registration network to find an optimal solution for both tasks (Bui et al., 2020; Kong et al., 2021). In contrast to Bui et al. (2020), a single registration network (CNN) was trained to mitigate misaligned noise in the images by predicting the deformable vector field. However, a fundamental issue with most studies is their explicit consideration of the correlation between MR contrasts. To address this, Lin et al. (2022) represented MRI modalities as non-linear embeddings with respect to their atlas and learned the deformation feature across modalities. Both modality-specific atlases and multi-modal deformation are then utilized for image synthesis.

3.2.2. Model design based on network architectures

- Mapping multi-contrast features

While multi-modality MRI synthesis offers the benefit of learning from shared features among multiple image contrasts, notably when the features are weakly represented in individual source modalities, Yurt et al. (2021) identified a significant concern regarding the complete disregard for one-to-one translation between image pairs. This issue predominantly emerges when critical features are unique to a specific input contrast and when unique details accurately predict the target image, leading to suboptimal performance in multi-modality image translation. Yurt et al. (2021) developed a multi-stream GAN that utilizes shared and complementary image features from different modalities to capitalize on the advantages of both single and multi-modality image synthesis. This approach combines shared feature maps from many-to-many translation and complementary features from one-to-one translation using a fusion network, resulting in higher quality and sharper synthesized images. Furthermore, to ensure one-to-one mapping for unique translation between unpaired images with GANs, Grover et al. (2019) introduced a novel flow-based generative model that guarantees exact cycle consistency and learns a shared latent space between domains while integrating adversarial learning with the same maximum likelihood. Shen et al. (2021) proposed a many-to-many GAN framework by learning semantic information across the modalities in shared representation and domain-specific features to handle the problem with random missing data.

- Transformer-based models

Compared to traditional adversarial-based generative methods, Vision Transformers have recently garnered increased attention due to their promising performance and ability to model contextual data representation in medical imaging tasks, such as image translation, registration, and segmentation. Dalmaz et al. (2022) introduced the first adversarial model with a Transformer-based generator for medical image synthesis, referred to as Residual Vision Transformers (ResViT). They proposed a novel aggregated residual transformer (ART) incorporating a transformer and CNN module with skip connections to extract contextual and local features from input features. The generator in the proposed architecture follows an encoder-bottleneck-decoder path-based module, where the information bottleneck consists of a stack of ART blocks. The primary function of these ART blocks is to combine low-level input features with their contextual, local, and local-contextual features.

Yan et al. (2022) and Pan et al. (2022) introduced a recent Transformer-based study using Swin Transformers (ST) to address border artefacts caused by smaller size patches in Vision Transformers. They presented a GAN-based model for multi-modal image translation with an ST generator (MMTrans), and Yan et al. (2022) extended it with an ST registration network. The ST Generator (STG) consists of residual ST blocks (RSTB), each containing ST layers (STL) for local and cross-window interaction-based learning. A U-shaped registration network with STLs was employed to learn the deformable vector field. With the proposed architecture, the generator could generate images with the same style and content details as the target image. Error map results indicate that MMTrans outperforms other competitive methods, including RegGAN, proposed by Kong et al. (2021). Li et al. (2022) proposed a Transformer-based network with edge-aware pre-training for MR synthesis to preserve both intensity and edge information at the same time. Liu et al. (2023) proposed a novel Multi-contrast Multi-scale Transformer based approach to develop a sequence-to-sequence MRI prediction method by taking variable length combination of MRI contrasts and synthesizing missing contrasts as output.

- Diffusion-based models

Score-based generative models have demonstrated a remarkable ability to efficiently sample the target distribution through stochastic diffusion techniques. Meng et al. (2022) proposed the first score-based generative model for cross-modality MRI synthesis by introducing a classifier-free conditional Diffusion-based approach. The proposed score-based reverse generation of any MRI contrast uses the guidance from the other remaining contrasts in the denoising process by using a single network for learning cross-modal score functions. Yoon et al. (2022) proposed the first sequential-aware Diffusion-based approach for MRI synthesis by exploring the temporal dependency of sequential data as conditioning prior. This has enabled the generation of longitudinal images and imputation of missing data in multi-frame cardiac and longitudinal brain MRIs.

In contrast to the conventional conditioning methods in Diffusion models, Pan et al. (2023b) proposed a novel cycle-guided DDPM by using two identical denoising networks for stable preservation of anatomical details in data for lesion identification. While many Diffusion-based approaches for MRI synthesis utilize 2D networks to manage computational complexity, this might lead to a potential volumetric inconsistency between image slices. 3D DDPM may overcome this issue. However, it comes with a high computation and memory cost. Aiming to overcome computational overhead, Zhu et al. (2023) employed Latent Diffusion Models (LDM) to benefit from its low-dimensional data representation to obtain 3D image synthesis and improve the volumetric consistency between image slices. They have introduced a novel 3D volumetric data synthesis using 2D backbones to overcome the computation overhead by utilizing a series of volumetric layers in the 2D network. Jiang et al. (2023) has also utilized LDM for many-to-one translation between MRI images to overcome the excessive memory consumption with DDPMs for multi-modal MRI synthesis. This was the first study on Diffusion-based many-to-one MRI synthesis with an adapting condition weights approach to balance the multi-conditioning guidance.

Conditional Diffusion models require well-aligned pair data for training, which is a challenging task to obtain. To address this limitation, conditional denoising using unsupervised learning methods is explored. Özbey et al. (2022) proposed an adversarial Diffusion model for high-fidelity image synthesis (SynDiff), representing the first unsupervised-based MRI synthesis approach using a Diffusion model. SynDiff utilizes a cycle-consistent framework with non-diffusive and diffusive modules. This method has achieved superior results in synthesized image quality and high fidelity compared to other GAN-based approaches. Fig. 10 provides a qualitative comparison of the results obtained with the proposed SynDiff approach. Furthermore, Wang et al. (2023) proposed a zero-shot learning-based unsupervised method using stochastic Diffusion models. They utilized statistical feature-wise homogeneity for conditioning the reverse diffusion process instead of using conditioning-based guidance from a data domain. This has provided an efficient method for bridging the source and target image modalities by capturing their local statistical attributes.

- Summary of synthetic MR

Table 2 overviews deep learning-based synthetic MR approaches, categorized by their primary feature attributes. This summary outlines the most relevant studies in each category, emphasizing the main novelties and details regarding network type, loss function, and MRI translations.

3.3. Synthetic PET

Deep CNNs and GANs have been extensively employed for PET synthesis from MRI due to their outstanding feature extraction capabilities and exceptional performance in image synthesis (Sikka et al., 2018; Hu et al., 2019; Shin et al., 2020b; Hussein et al., 2022; Rajagopal et al.,

Table 2
Overview of DL-based MRI synthesis.

Feature	Highlights	Architecture	Loss function	MRI translation	Reference
Single input to output MRI synthesis	Multi-contrast MRI synthesis which preserves intermediate-to-high frequency details	2D cGAN	AL* + $\mathcal{L}_1$ loss + Perceptual loss	T1 → T2, T2 → T1	Dar et al. (2019)
		2D U-Net	MSE	Translation between T1, T2, Flair	Osman and Tamam (2022)
Multi-modal image synthesis	Unified and scalable network for multiple-to-one mapping among	2D CycleGAN	AL + CCL#	T1+T2+Flair→ T1c , T1+T2+Flair→ DIR	Li et al. (2019b)
	Provides single-input multi-output MRI image synthesis using an unified GAN	2D cGAN	AL + Classification loss + CCL + consistency loss	T2 → T1, T2 → T1c, T2 → Flair, T1 → T1c, T1 → T2, T1 → Flair	Dai et al. (2020)
	Multiple-input single-output MRI synthesis method with latent space processing network.		AL + $\mathcal{L}_1$ loss + GDL	T1 → Flair, T1 + T2 → Flair, T1 + T1c + T2 → Flair	Zhan et al. (2021)
	Mixture of multiple one-to-one streams and a joint many-to-one stream.		AL + $\mathcal{L}_1$ loss	T2, FLAIR→T1, T1, FLAIR→T2, T1, T2→FLAIR	Yurt et al. (2021)
	An unified hyper-GAN to facilitate the translation between different contrast pairs		AL+ CCL + Common space loss + Classification loss	Different translation between T1, T1Gd, T2 and Flair	Yang et al. (2021)
	Multi-contrast, multi-scale missing data imputation		AL + $\mathcal{L}_1$ loss + Reconstruction loss	Missing data imputation between T1, T2, PD	Liu et al. (2023)
	Multi-input multi-output MRI synthesis	2D U-Net	MAE, L1-norm	T1 → T2, T1 → Flair	Chartsias et al. (2018)
Avoid misalignments in image pairs	Utilize the temporal information between consecutive slices	2D cGAN	AL + Temporal consistency loss + Registration loss + Total variation loss	T1 → T2	Bui et al. (2020)
	Image translation with registration to address the problem with well aligned data pairs.		AL	T1 → T2	Kong et al. (2021)
	Learn MRI modalities as a non-linear embedding with respect to their own atlas		AL + $\mathcal{L}_1$ loss + Consistency loss + Perceptual Loss	T1 → T2	Lin et al. (2022)
Memory efficient synthesis	High resolution 3D image generation in a memory-efficient way multi-scale GAN-based approach	3D cGAN	AL + $\mathcal{L}_1$ loss	T1 → T2, T2 → T1	Uzunova et al. (2020)
	GAN with less memory than conventional big-patch 3D methods	3D CycleGAN	AL + CCL + Perceptual loss	T1 → T2, T2 → T1	Zhao et al. (2021)
Preserve semantic, texture information	Use an adaptive normalization after adding high-level semantic information	2D CycleGAN	AL + CCL	DWI (diffusion-weighted images) → T2	Mao et al. (2022)
	Minimizes individual pixel losses and non-AL to transfer texture between domains	2D cGAN	AL + $\mathcal{L}_1$ loss + Style loss + Content loss	T1 → T2	Vaidya et al. (2022)
Transformer based generators	Residual transformer blocks to preserve and context	2D ViT	AL + $\mathcal{L}_1$ loss	T1, T2 → PD, T1, T2 → Flair	Dalmaz et al. (2022)
	Swin Transformer-based GAN for multi-modal image translation	2D Swin Transformer cGAN	AL + $\mathcal{L}_1$ loss + Reconstruction loss	T1 → T2, PD → PD-FS	Yan et al. (2022)
Diffusion based synthesis	Adversarial DDPM	Diffusion-based 2D CycleGAN	AL + CCL	T1 → T2, T2 → T1, T1 → PD, T1 → PD, PD → T1, PD → T2	Özbey et al. (2022)
	Cycle-guided DDPM	3D DDPM	MAE	T1 → T2, T2 → T1, T1 → Flair, FLAIR→T1	Pan et al. (2023b)
	Mutual Information Guided Stochastic Diffusion	2D DDPM	MAE	T1→FLAIR, FLAIR→T1, T1→PD, PD→T1	Wang et al. (2023)
	3D data synthesis by leveraging 2D backbones	3D LDM	MAE	T1 → T2	Zhu et al. (2023)
	Conditional Multi-modality MRI synthesis	2D LDM	MAE	Translation between T2, T1ce, T1 and Flair	Jiang et al. (2023)
High field MRI synthesis	Anisotropic U-Net	2D U-Net	MSE	LF T1 (0.36T) → HF T1	Lin et al. (2019)
				T1 images (0.36T) → 1.5T	Figini et al. (2020)

*AL : Adversarial Loss   # CCL : Cycle Consistency Loss

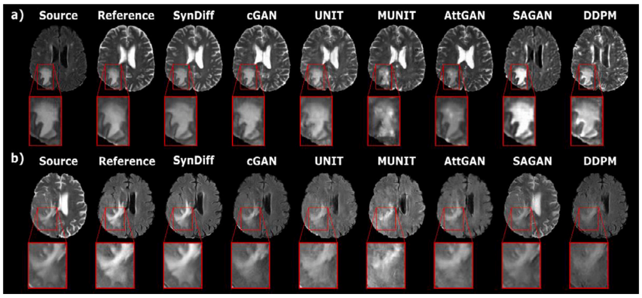

Fig. 10. Results are shown for (a) $FLAIR \rightarrow T2$ (b) $T2 \rightarrow FLAIR$ tasks using BraTs dataset. SynDiff shows a lower artefact level and detailed structure compared to other methods.
Source: Image from Özbey et al. (2022) on qualitative results for translation between MRI contrasts.

2023; Ouyang et al., 2023). Furthermore, Transformer and Diffusion based approaches are demonstrated with high-quality PET synthesis (Shin et al., 2020a; Zhang et al., 2021; Xie et al., 2023; Jang et al., 2023). In the synthesis of FD PET images from LD images, a common emphasis in deep learning-based approaches has been to either employ multi-channel-based input networks or treat each input modality as a new task for processing MRI or CT-accompanied PET images, thereby combining both functional and morphological details (Wang et al., 2019; Chen et al., 2019). The majority of studies have focused on GAN-based approaches (Xue et al., 2021; Zhao et al., 2020b; Lei et al., 2019a; Kaplan and Zhu, 2019; Lei et al., 2020) for estimating FD PET images, while a few studies (Chen et al., 2019; Häggström et al., 2019; Lu et al., 2019; Sanaat et al., 2020; Luo et al., 2021; Dutta et al., 2022; Zeng et al., 2022; Hu and Liu, 2022) have investigated U-Net and Transformer-based networks.

3.3.1. Model design based on input domain

- Preserving 3D information

Sikka et al. (2018) introduced the first global and non-linear-based method for synthesizing PET from MRI using a 3D convolutional U-Net model, capturing the global correlation between MRI and PET images. Hu et al. (2021) and Lin et al. (2021) addressed the accuracy of 3D-based generator networks over synthesizing 2D image slices. They proposed a novel 3D bidirectional mapping GAN network (BMGAN) that learns forward mapping and backward mapping to return PET images to the latent space. This approach promotes consistency between the latent space and PET images, embedding detailed semantics into the high-dimensional latent space and synthesizing perceptually similar PET images.

- Dealing with multi-frequency data and projection space data

To address the complexity of PET images, which exhibit high variance in frequency levels, these DL-based solutions have proven advantageous in synthesizing accurate PET images with diverse frequency levels (Shin et al., 2020a). The texture-wise differences between MRI and PET significantly impact the synthesized images. MRI exhibits more texture-wise details, and PET displays more complex features that vary across different frequencies. Emami et al. (2020b) suggested separately addressing the different frequency levels of the images to preserve more realistic features in the synthesized images. Their proposed frequency-aware U-Net (FREA-UNet) independently synthesized low- and high-level features by weighing the frequencies' importance levels and assigning higher weights to the most relevant regions. Fig. 11 illustrates the comparison of synthesis results from FREA-UNet with other state-of-the-art deep generative networks.

In contrast to other methods that utilize image space data to generate full-dose PET scans, Sanaat et al. (2020) proposed a 3D U-Net-based PET synthesis approach for both the image and projection spaces. The primary objective of this study was to analyse the quality of synthesized PET scans in both low-dose PET images and sinograms. Xue et al. (2021) proposed a GAN-based projection space method and employed a prior domain transfer operation to convert sinograms to images before inputting them into the network. Häggström et al. (2019) also synthesized projection space data with a CNN-based encoder–decoder network for image synthesis. Chen et al. (2019) and Lu et al. (2019) proposed similar 2D and 3D U-Net-based approaches in the image space.

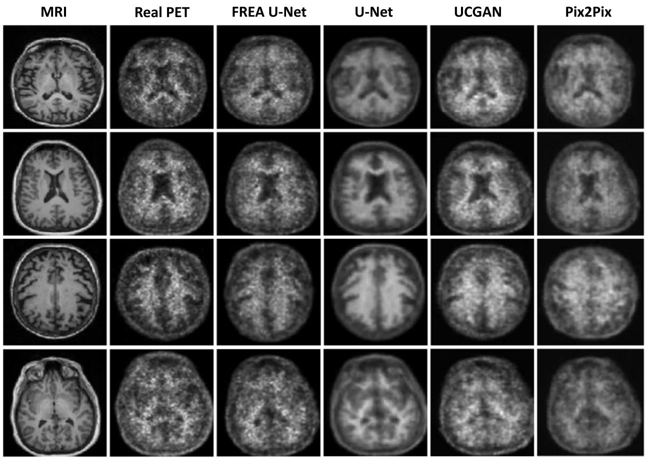

Fig. 11. As FREA-UNet synthesizes low and high-frequency scales separately, the synthesized images have high spatial resolution and sharp boundary regions compared to the generated image from other networks.
Source: Image from Emami et al. (2020b) for qualitative comparison of synthesis PET results from FREA-UNet and other deep networks, UNet, cGAN with UNet architecture (UCGAN) and pix2pix.

- Dealing with multi-modal input

Considering the benefits of multi-modality-based PET synthesis, Wang et al. (2019) proposed a full-dose FDG-PET with low-dose FDG-PET accompanied by T1w MRI and diffusion tensor image. Since different image modalities vary in image locations and contribute differently, a locality adaptive-based fusion method was introduced as an alternative to applying an identical filter in each imaging modality and combining the generated feature maps. A separate module was added to fuse the multi-modality images and further preserve context information by integrating high-level image appearance with low-level details. In contrast to the conventional CycleGAN-based approaches for image synthesis, Zhao et al. (2020b) proposed a supervised CycleGAN network for FD PET synthesis using an additional supervised loss.

3.3.2. Model design based on network architectures

- Mapping multi-scale features

To support locally and globally aware PET synthesis, Sikka et al. (2021) developed a novel GAN-based method called GLA-GAN, which consists of two sub-networks: global and local modules. The global sub-network is a generator that maps the entire MR image to a latent space feature. At the same time, the local module comprises K generator networks, each operating on K disjoint image patches covering the entire MR image. To refine the anatomical details of the synthesized images, Wei et al. (2019) introduced a novel Sketcher-Refiner-based network for learning the myelin content of brain PET images from multimodal MRIs. This approach employed a sketcher to generate initial anatomical details, followed by a refiner to further enhance the generated images with the myelin content of the brain.

- Task-driven learning models

In the reviewed literature on FD PET synthesis, the majority of methods focused on single-task learning (SDL) as their deep learning approaches. In contrast, Sun et al. (2022b) introduced a novel bi-task-based method utilizing LD PET/MRI, considering MRI as an additional task rather than another channel within a single task. In the approach proposed by Sun et al. (2022b), two decoders and two encoders were employed for each LD PET and MRI input, sharing a latent space that allowed the outputs of each latent space to serve as inputs to the decoders. Given that MRI assists FD PET synthesis, a task conditioned on MRI is the secondary task, while the LD-conditioned task is the primary task. The $\mathcal{L}_1$ loss and GAN adversarial loss of the secondary task was regarded as bias loss in the training process. The weight ratio between primary and secondary task losses was set from 0.1:1 to 10:1. Zhou et al. (2022) also proposed a task-driven approach using a segmentation-guided style-based GAN method, employing segmentation as a secondary task. The style-based generator model comprises a mapping network, a noise module, adaptive instance normalization, and a synthesis network. This configuration controls hierarchical features to generate more realistic image textures, and the segmentation-guided strategy enhances the accuracy of image translation in regions of interest. Similarly, Fei et al. (2022) proposed a bidirectional contrastive GAN-based method with mild cognitive impairment classification of synthesized PET images as the secondary task.

- Transformer and Diffusion based models

Recently, Transformer and Diffusion based methods have received considerable interest in the context of deep learning-based PET synthesis. Zhang et al. (2021) proposed a low-dose PET/MRI denoising method using spatial adaptive and Transformer-based fusion networks for PET synthesis. They have incorporated a global attention mechanism in Transformers to establish appropriate pixel-wise relationships between MRI and PET images. Shin et al. (2020a) proposed a novel approach utilizing a BERT-based network to handle the wide range of floating intensity values (i.e. radiotracer uptake values) in PET images. In the proposed method, the images are summarized and quantized into sequences that closely resemble text data for training with BERT. MRI and PET image sequences are concatenated and separated with [SPE] and [CLS] tokens, akin to BERT-based text data processing. The BERT model's objective is to predict the real PET using the provided masked MRI and masked synthesized PET sequences.

With the advances in Diffusion based networks, Xie et al. (2023) proposed a high-field MRI-conditioned Diffusion-based network. Further, Jang et al. (2023) extended MRI-conditioned PET synthesis by guiding the reverse diffusion process with textual descriptions. The combination of MRI prior and text description, such as gender, scan time, cognitive test scores, and amyloid status, was utilized to guide the denoising process with a Latent Diffusion Network.

Transformer-based approaches have used in recent studies on FD PET synthesis. To capture local spatial image details, Luo et al. (2021) introduced a novel Transformer-based approach with a CNN-based encoder and decoder, thereby enhancing the extraction of long-range semantic information between the encoder and decoder. This has efficiently used global and local semantic details using low-level spatial structures captured from CNNs. However, Zeng et al. (2022) highlighted that utilizing Transformers as a bottleneck module in the network reduced their direct interaction with CNNs, leading to possible semantic ambiguity. This arises from Transformers' limited ability to capture low-level details effectively. With the motivation of this limitation, Zeng et al. (2022) have combined CNNs and Transformers for FD PET synthesis inspired by convolutional Vision Transformers. Further, Hu et al. (2022) proposed a residual Swin Transformer-based regularizer to synthesize FD PET from LD sinograms.

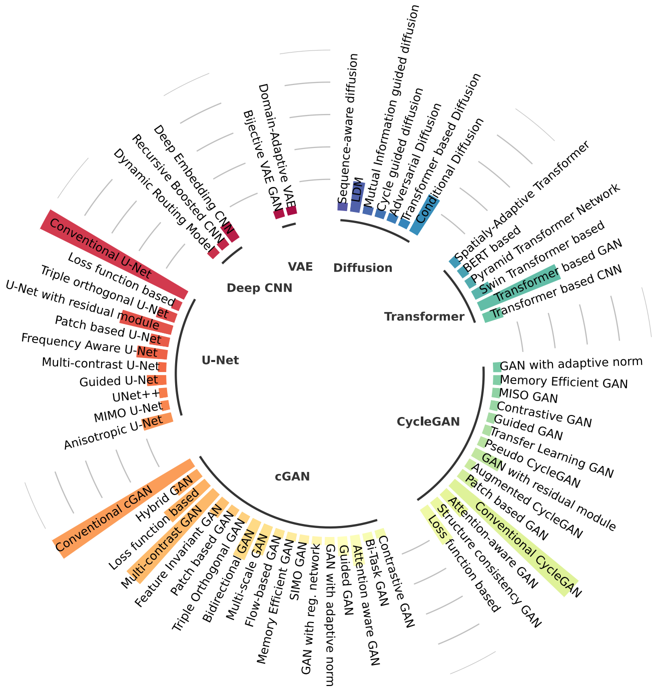

Fig. 12. An overview of deep neural network architectures for medical image synthesis in the literature.

- Summary of synthetic PET

Table 3 presents an overview of deep-learning-based synthetic PET approaches grouped by their feature-wise innovations. This summary highlights the most relevant studies in each category, emphasizing the main contributions with details of network type, loss function, input data, and regions.

4. Analysis of networks

4.1. Network models

Deep-learning-based architectures predominantly employed in medical image synthesis include CNN networks, U-Nets, GAN-based networks, Transformer-based networks and, more recently, Diffusion-based architectures. CNNs gained popularity due to their exceptional performance in computer vision tasks, particularly in medical imaging. These techniques (Chartsias et al., 2018; Fu et al., 2019; Lin et al., 2019; Florkow et al., 2020; Sanaat et al., 2020; Hussein et al., 2022) have demonstrated their ability to synthesize images in a target domain from a distinct source domain and generate high-resolution images from low-resolution counterparts. The literature reveals that cGANs and CycleGANs have been most extensively utilized for synthetic MR and pseudo-CT generation. In contrast, cGANs and U-Nets have been commonly employed for synthetic PET synthesis. Moreover, there has been a notable increase in interest towards transformer and Diffusion-based architectures for medical image synthesis.

The literature's deep neural networks for medical image synthesis can be clustered based on their architecture types to support different input and feature-level enhancements. As shown in Fig. 12 multi-contrast networks include Single Input Multiple Output (SIMO), Multiple Input Single Output (MISO), and Multiple Input Multiple Output (MIMO) networks for image synthesis. Further, Multi-scale architectures specify the feature level differences in the modalities at different scales, including their variations of contribution for synthesizing the target modality. Frequency-invariant networks match common regions

Table 3
Overview of DL-based synthetic PET.

Feature	Highlights	Architecture	Loss function	Input data and region	Reference
Capture the global and local features	Global and non-linear cross-modal approach	3D U-Net	Binary cross entropy loss	MRI, Brain	Sikka et al. (2018)
	Globally and locally aware GAN	3D cGAN	AL* + $\mathcal{L}_1$ loss + SSIM loss		Sikka et al. (2021)
	Sketcher-Refiner generative network		AL + $\mathcal{L}_1$ loss		Wei et al. (2019)
Frequency-aware synthesis	Optimizing low/high frequency scales of the image with separately	2D U-Net	L1 norm		Emami et al. (2020b)
Bi-directional mapping	Embed the semantic information of PET images into the high-dimensional latent space.	3D cGAN	AL + $\mathcal{L}_1$ loss + Perceptual loss		Hu et al. (2021)
	3D reversible GAN		AL		Lin et al. (2021)
Handle a wide range of floating intensity	Adopt BERT with NSP and MLM objectives, to reproduce highly biased intensity values.	2D cGAN	AL + $\mathcal{L}_1$ loss + MLM loss		Shin et al. (2020a)
Transformer based synthesis	Spatial adaptive and transformer fusion network	2D ViT	$\mathcal{L}_1$ loss	MRI, Brain	Zhang et al. (2021)
Diffusion based synthesis	Joint diffusion with high field MRI	2D DDPM	MAE	MRI, Brain	Xie et al. (2023)
	Textual description and MRI conditioned diffusion	2D LDM	MAE	MRI, Brain	Jang et al. (2023)
LD to FD mapping	Locally adaptive multimodality GAN	3D cGAN	AL + L1 norm loss	PET/MRI, Brain	Wang et al. (2019)
	3D cGANs based refinement module		AL + L1 norm loss	PET, Brain	Wang et al. (2018)
	Applying a low-pass filter to the low-dose image to remove noise, sectioning the body and train a model for each region.	2D cGAN	AL + MSE	PET/CT, Whole-body	Kaplan and Zhu (2019)
	Residual CNN with middle fusion of input modalities.	2D U-Net	Weighted attention loss	PET/MRI, Whole-body	Wang et al. (2021b)
	Using projection space data		MSE	PET/CT, Whole-body	Häggström et al. (2019)
		3D U-Net	MSE	PET/CT, Brain	Sanaat et al. (2020)
		2D CycleGAN	AL + CCL# + $\mathcal{L}_1$ loss	PET/CT, Whole-body	Xue et al. (2021)
	Task-driven learning	2D cGAN	AL + L1 norm loss + structural loss + bias loss	PET/MRI, Head	Sun et al. (2022b)
		3D cGAN	AL + Content loss + Segmentation loss	PET/CT, Whole-body	Zhou et al. (2022)
	Transformer based synthesis	3D ViT-GAN	AL + $\mathcal{L}_1$ loss	PET/MRI, Brain	Luo et al. (2021), Zeng et al. (2022)
		2D Swin Transformer	MSE	PET, Brain	Hu and Liu (2022)

*AL : Adversarial Loss # CCL : Cycle Consistency Loss

in the source and target images in high-frequency components to improve the robustness of the synthesized images, and frequency-aware networks separately optimize different frequency levels in the input modalities. Deep neural architectures with registration(reg.) networks consist of an additional registration network that handles misaligned data distribution adaptively. Guided networks guide the image translation process using an auxiliary classifier, and the generative models with residual modules take advantage of residual network structure to capture high-level context from the image contrasts. Further, to preserve more 3D structural information with 2D inputs, orthogonal networks independently utilize three generative networks for orthogonal planes.

4.1.1. CNNs

CNN-based networks gained immense popularity due to their capability to learn the underlying relationships within different input domains since their introduction. To apply conventional CNNs to learn the end-to-end mapping between different input domains, Xiang et al. (2018) proposed a novel embedding strategy by decomposing the CNN model into transform and reconstruction stages. The transform stage is responsible for forwarding the feature maps derived from the input modality and synthesizing the target modality in the reconstruction stage. The embedding block estimates the target modality given the latest feature maps by processing through convolution and concatenation layers. A series of embedding blocks helps to fill the gap between source and target modalities by refining features in each block. For utilizing 3D-based CNNs with an efficient parameter count, Kläser et al. (2018) proposed a learning method of 3D representation of input data and mapping to the target domain through a series of 1D convolutions without using non-linear activation. The proposed architecture consists of two networks for synthesizing the initial target modality and improving the synthesis through a second network by adding residuals. The outputs of the second networks are fed back to the network to update the weights, and all the weights are shared within the network to make it less computational and to adapt to limited datasets. A similar residual-based concept was also utilized to address the gradient degradation problems with deep CNN networks Li et al. (2019a).

4.1.2. VAEs

VAEs have been considered a generative modelling framework that learns the continuous distribution of latent variables in contrast to traditional autoencoders (Kingma and Welling, 2013). However, one of the common limitations of conventional VAE-based generative models is they tend to produce blurry results due to the limited expressiveness of the framework, injected noise or imperfect loss criteria such as L1 or L2. Therefore, the synaesthetic quality of VAEs still lags behind GANs, and very few studies have employed VAEs for medical image synthesis from the recent literature. In order to improve the perceptual quality of the generated image (texture-wise with VAEs), Liu et al. (2021c) adopted an additional adversarial training to the network. They have introduced unpaired cycle reconstruction into the VAE by creating VAE-GAN, which helped to mitigate the contrast and texture-wise differences between real and synthesized images. Further, Hu et al. (2022) introduced unsupervised VAE-based domain adaptive methods when the target image domain is not accessible. The proposed VAE-based network by Hu et al. (2022) was first trained on the source domain, and the KL divergence between a target distribution and the prior distribution of synthesized images was computed in the adaptation stage.

4.1.3. U-Nets

In the literature, CNN-based generator architectures encompass U-shaped encoder–decoder networks featuring various enhancements to the network structure (Kläser et al., 2018; Bahrami et al., 2019; Li et al., 2019a; Bahrami et al., 2019; Jans et al., 2021). During network development, several factors are considered, including the characteristics and availability of training data, learning of both global and local features between image domains, and retention of valuable information throughout the training process to produce high-quality images.

Numerous studies have employed conventional U-Net architecture for image synthesis, with a few introducing novel architectural features to enhance model performance and generate more realistic images (Bahrami et al., 2019; Spadea et al., 2019; Florkow et al., 2020). In the expanding path of the U-Net architecture, unpooling layers have frequently been replaced by deconvolution layers, as the latter can produce dense feature maps compared to unpooling layers (Fu et al., 2019). This reduces computational memory requirements, as unpooling layers demand substantial memory to track max-pooling indices. Drawing inspiration from ResNet’s architecture, skip connections in U-Net architectures were substituted with residual shortcuts to decrease memory consumption further (Fu et al., 2019; Zhao et al., 2022). In this network architecture, the number of feature maps is doubled in U-Net with skip connections. It concatenates the encoder feature maps with upsampled features, while residual shortcuts add the two feature maps without altering the number of feature maps (Fu et al., 2019). By integrating U-Net and encoder–decoder structures, Bahrami et al. (2020) proposed a novel efficient CNN by replacing convolutional layers with building structures. The building structure consists of convolutional layers followed by batch normalization and Scaled Exponential Linear Unit (SeLU) activation. Contrasting other CNN networks that utilize ReLU activation, Bahrami et al. (2020) emphasized the importance of substituting ReLU with SeLU in specific scenarios. Although ReLU is beneficial in deep neural networks for addressing vanishing gradient problems, it risks getting stuck on the negative side due to mapping the negative inputs to zero, known as the dead state or dying ReLU effect. Since SeLU provides different output than zero for negative inputs, it promotes self-normalizing during training, which helps overcome dead states. Moreover, Fu et al. (2019) and Hu et al. (2019) replaced batch normalization with instance normalization, as small batch sizes might cause less accurate estimation, reducing batch normalization’s effectiveness. With the advancement of multi-task deep learning-based algorithms, Hussein et al. (2022) proposed a multi-task CNN architecture with distinct sub-networks for multi-contrast MR images, which are trained concurrently. Emami et al. (2020b) introduced a frequency-aware U-Net for separately generating high and low-frequency image scales to enhance the sharpness of synthesized images. In the proposed architecture, two layers were individually assigned for low and high-frequency paths, separated using a Gaussian filter. Moreover, Zhao et al. (2022) introduced a channel and spatial-wise attention module with U-Net for improving the edge predictions of the synthesized images. In contrast to the conventional U-Net architecture, Sreeja and Mubarak (2022) proposed a U-Net++ (Zhou et al., 2020) network that consists of convolutional layers and dense skip connections. This enables model pruning by deep supervision, bridging the semantic gap between feature maps from the encoder and decoder.

4.1.4. Generative adversarial networks

GANs have achieved state-of-the-art results in medical image synthesis by accurately learning the non-linear associations between image modalities (Ranjan et al., 2022). Most conditional GAN (cGAN)-based image synthesis approaches have employed U-Net or ResNet-based generator architectures in conjunction with PatchGAN-based discriminators (Ge et al., 2019; Lei et al., 2019b; Reimold et al., 2019; Wu et al., 2019; Boni et al., 2020; Zhao et al., 2021; Vaidya et al., 2022). Each study introduced distinct generator and discriminator architectural enhancements to synthesize more realistic images. These architectural features address synthesizing high-level image features, processing multi-dimensional data using memory-efficient methods, managing various mappings between source and target domains, and implementing novel architectural modifications to preserve contextual image features.

CNN-based encoder–decoder networks have been predominantly employed for generators, accompanied by pooling and ReLU activation layers (Qi et al., 2020). However, traditional CNNs may not ensure adequate receptive fields since pooling layers typically reduce the spatial resolution of feature maps (Nie et al., 2018). As an alternative, dilated convolutions have been extensively adopted to obtain sufficient receptive fields. LeakyReLU activation has also been frequently utilized to replace ReLU, addressing the issue of networks getting stuck in local minima when zero channels appear in the latent space early in the training process (Chartsias et al., 2018; Armanious et al., 2020; Oulbacha and Kadoury, 2020; Qian et al., 2020; Qi et al., 2020). In contrast to conventional generator networks, Armanious et al. (2020) demonstrated a novel generator architecture called CasNet, inspired by ResNets. CasNet comprises a stack of U-Net blocks with skip connections to transfer low-level information between encoder and decoder paths, enhancing the GAN network’s generalizability by passing input through several concatenated U-blocks. Although ResNet and CasNet share similar architectural structures, CasNet consists of deeper blocks with 16 convolutional layers. Furthermore, skip connections in the encoder–decoder paths alleviate the vanishing gradient problem caused by the network’s deeper structure. Qian et al. (2020) proposed a hybrid generative network incorporating both residual and shortcut connections and added a classification component to the discriminative cGAN to further stabilize training (Cao et al., 2021).

cGAN-based medical image synthesis, the majority of studies (Wei et al., 2019; Yang et al., 2018b; Kaplan and Zhu, 2019; Lei et al., 2019b; Kazemifar et al., 2019; Armanious et al., 2020; Sikka et al., 2021; Kong et al., 2021; Liu et al., 2021a; Lenkowicz et al., 2022) have employed 2D GAN architectures using two-dimensional image slices from paired data. The most prevalent representation of these proposed architectures aims to identify high-frequency patterns between real and synthesized image regions, focusing on high-level features such as contours and assigning them greater weight. Analogous to capturing high-level and low-level features, exploiting both local and global contexts of images is also crucial in medical image synthesis. Multi-module generator-based architectures were proposed to facilitate locally and globally aware image translation using GAN-based frameworks (Boni et al., 2020; Sikka et al., 2021), incorporating two separate sub-networks for local and global contexts. To identify organ outlines when mapping PET to CT, Dong et al. (2019) integrated attention gates into the U-Net generator network to eliminate uncertainty in noisy details present in skip connections. The integration of attention gates proved beneficial in highlighting the most critical features from the network's encoder path. Additionally, structural consistency GANs (Yang et al., 2020) employing spectral normalized (SN) convolutional layers provided a robust solution to stabilize both generator and discriminator models by controlling the Lipschitz constant of the entire network. This approach enhanced network stability by avoiding unusual gradients. Similarly, Zhao et al. (2020a) utilized attention aware generator to integrate local and global features via dual attention modules with minutious and global attention modules (MAM and GAM). MAM identifies interdependent feature maps for specific anatomy by improving tumour specificity and GAM encodes contextual details to local features by making the network context-aware. A dual generator architecture with transformer and CNN-based generators linked in series was also employed to resolve the limitation with CNN to extract more contextual details (Li et al., 2022a). Yu et al. (2019) introduced an edge-aware GAN that captures the structural details of images through adversarial learning of edge similarity between real and synthesized images. This approach has demonstrated promising outcomes in synthetic MRI, as it effectively learns the local intensity variations along image boundaries between different tissue contrasts, resulting in sharpening synthetic images. Furthermore, reversible generators ensured perceptually similar and more realistic synthesized images by preserving essential structural features from image modalities such as MR (Lin et al., 2021).

For fusion-based multi-modality GAN architectures, U-Net-based generator models have been employed alongside shared modality-invariant latent methods (Chartsias et al., 2018; Zhou et al., 2020a; Boni et al., 2021). The model's architecture comprises an encoder, a latent fusion, and a decoder. The fusion step merges the encoder representation of each modality into a single fusion, subsequently integrating the unique features of each. Given the variability in image locations across different modalities, Wang et al. (2019) proposed a locality adaptive-based fusion method for each imaging modality to combine the generated feature maps. Wang et al. (2021b) applied middle fusion for input integration, while others combined modalities at the input level. This approach offers the advantage of preventing the loss of information on each modality's characteristics in early fusion by integrating them after the fourth residual block in the CNN network. A similar approach was proposed by Zhou et al. (2020a) using a layer-wise fusion method to combine the multi-level and multi-modal representation of inputs. Contrastingly, Park et al. (2020) and Jiangtao et al. (2021) mapped similar feature patches in the learned space to other relative features in the dataset using a single-side translation with one generator and discriminator pair, thereby reducing training overhead. Li et al. (2023b) proposed a multi-sequence fusion network by introducing element and patch-wise fusion methods on cross-modality attention using Transformers for MRI synthesis. In the proposed fusion module, the independently encoded MRI contrast features were fused in parallel using the patch and voxel-wise cross-correlation and then combined the fused feature spaces. Similarly, using a multi-headed self-attentive method, Zhao et al. (2023) employed a Residual Transformer to perform residual concatenation of global and local features. Moreover, Unified GANs have been employed for single-input to multi-output image synthesis by utilizing a single generator that accepts both the input image and the target modality's label (Li et al., 2019b; Dai et al., 2020; Yang et al., 2021). To support GAN-based networks for multi-input to single-input images, Zhan et al. (2021) proposed multiple encoders in the generator for each modality, and a latent space processing network was employed to combine different features from the modalities and return a latent representation of the target modality without concatenation. Conversely, Tie et al. (2020) proposed a multi-path GAN featuring a contracting path in the U-Net-based generator, divided into multiple paths where each channel has its feature extraction to independently extract features from input modalities, avoiding the loss of unique features.

CycleGANs have been extensively employed in GAN-based networks utilizing unpaired datasets, with training reliant on the cycle-consistency loss function (Takamiya et al., 2023; Estakhraji et al., 2023). A novel Pseudo-Cycle Consistent Module has been introduced to the CycleGAN architecture by incorporating a target generator for each cycle, using Pseudo-Cycle Consistent Loss as the consistency constraint for the target generator model (Liu et al., 2021a). A domain controller module was proposed to prevent images in the expected domain from being inaccurately translated to another target domain, thereby ensuring domain correctness and stabilizing the generator. Zhang et al. (2022c) proposed a novel switchable GAN-based multi-contrast MRI synthesis using only one generator network, effectively reducing the computational overhead of CycleGANs with two generators. This design employed a single switchable generator model based on adaptive instance normalization (AdaIN) for image style transfer. The generator comprises an autoencoder for content synthesis and AdaIN for style adjustment between image contrasts, enabling the generator to synthesize images with varying styles.

4.1.5. *Transformer based networks*

Despite the success of CNN-based generative networks for cross-modality medical image synthesis, the use of a limited receptive field makes these networks more focused on local features rather than extracting global features with long-range dependencies. However, these global features are crucial for medical imaging as some organs can be broadly distributed within large areas, which need to consider dependencies among distinct voxels. Therefore, Transformer based medical image synthesis has gained much attention recently due to its advantage of extracting global correlation, and a combination of CNN and Transformer is popular to model both local and global features.

Zhao et al. (2023) utilized a Transformer-based deeper feature extractor as the bottleneck layer of encoder–decoder based generator network. To effectively utilize local and global features, they have constructed a residual transformer block consisting of transformer layers and one convolution layer. To mitigate the insufficient local feature extraction with the transformer module, a residual concatenation was introduced prior to computing self-attention, and this has further improved the stability of the network. The conditional image synthesis with ResViT (Dalmaz et al., 2022) has also been utilized to model the generator's bottleneck with a stack of aggregated residual Transformers (ART) with residual convolution and transformer layers. A weight-sharing method was adopted across the transformer bottleneck further to reduce the memory demand of multiple ART blocks. A similar hybrid architecture was introduced by Li et al. (2023c) as a multi-scale network to obtain local and global features with a Transformer-based bottleneck layer. These deeper Transformer-based feature bottleneck layers further reduced the enormous computational cost of ViTs. Further, employing a CNN-based encoder module extracts the low-level spatial information and provides a compact set of features for the transformer layer. In contrast to the conventional transformer layers, Luo et al. (2021) applied an encoder decoder-based transformer module in the bottleneck layer with parallel decoding. In addition, they have utilized an additional multi-head encoder–decoder attention in the decoder part, which differs from MHSA by modelling the relation between encoder output and previous MHSA, where MHSA compute attention with its own inputs. This has improved the deeper module of the network to model more long-range dependencies. In contrast to the most common way of using a transformer as the bottleneck layer, Zeng et al. (2022) proposed a novel method of combining CNN with a transformer in an effective manner by replacing linear embedding and linear projection layers in the transformer with convolutional embedding and projection.

Transformer-based generative architectures have also shown highly beneficial in exploring voxel and contextual cross-modality correlations. By focusing on these contextual correlations, Li et al. (2023b) designed a multi-sequence fusion network with cross-modality attention using a CNN-Transformer hybrid module. Similarly, Zhang et al. (2021) incorporated a transformer fusion network for fusing PET and MRI in the encoding path to give more attention to the regions of interest. In contrast to the conventional Transformer-based approaches for medical image synthesis, Shin et al. (2020a) introduced self-attention through a bidirectional transformer in the bottleneck layer.

Conventional Transformer models compute self-attention using fixed-scale image tokens. By contrast, Swin-Transformers incorporate shifted window-based attention, enabling the construction of a hierarchical representation of inputs. This deep hierarchical representation of Swin-Transformers was utilized by Yan et al. (2022) to build the generative network, which consists of multiple residual Swin transformer blocks with convolutional layers in between for feature enhancements. A multi-contrast, multi-scale transformer was utilized by Liu et al. (2023) for missing data imputation with multi-contrast MRI images by further enhancing the attention mechanism within local cross-contrast windows to model both inter and intra-contrast dependencies. Recently, Li et al. (2022) proposed a multi-scale Transformer-based network for edge-aware cross-modality synthesis, which is designed in two stages: edge-aware pretraining and multi-scale fine-tuning. They have employed ViT for modelling encoders of edge-preserving masked autoencoders and Swin Transformer-based networks in the fine-tuning stage for integrating multi-scale features from pre-trained encoders for image synthesis.

4.1.6. Denoising diffusion models

Diffusion-based networks have recently emerged as highly promising generative models. When applying these networks to cross-modality medical image synthesis, architectural innovations focus primarily on conditioning the reverse denoising process to synthesize target image modalities from their source modalities. Two prevalent conditioning methods in Diffusion models are classifier-guided and classifier-free guidance. Classifier-free guidance involves conditioning the Diffusion model in a supervised manner (Ho and Salimans, 2022). On the other hand, classifier-guided guidance incorporates conditioning information into the sampling process using an optimization step based on a classifier network (Dhariwal and Nichol, 2021). However, using a classifier for medical image synthesis poses challenges as it heavily relies on the classifier's performance, potentially affecting the overall denoising process. As a result, many conditioning-based diffusion methods for medical image synthesis have been utilized for classifier-free guidance, utilizing supervised conditioning.

The most common method of conditioning the diffusion process involves using the source image modality as conditioning prior to the reverse diffusion process. For example, Lyu and Wang (2022) proposed an MRI-conditioned CT synthesis using a score-based diffusion process with coregistered image pairs. Similarly, Xie et al. (2023), Lyu and Wang (2022), Pan et al. (2023a) and Özbey et al. (2022), Meng et al. (2022) incorporated a conditional denoising process for synthesizing PET from MRI, CT from MRI, and MRI contrast synthesis using source conditional guidance in cross-modality translation. Li et al. (2023a) introduced a guiding denoising process that leveraged MRI and sampled prior CT information for CT synthesis, aiming to capture structural and anatomical details in low-dose generated CT images. Expanding on the conditioning approach, Meng et al. (2022) supported multi-modality conditioning by enabling a unified conditional synthesis process. This allowed a single network to refine synthesized images for any missing modalities by conditioning the synthesis on all available modalities through noise distribution. Similarly, Jiang et al. (2023) proposed a multi-conditioning approach named CoLa-Diff, which involved structural guidance of brain images using anatomical masks for MRI synthesis. They have additionally implemented an auto-weight adaptation method, enabling automatic balancing of multiple conditioning inputs to leverage relevant information from each conditioning prior effectively.

In contrast to directly relying on guidance from the source modality image, Wang et al. (2022) introduced a mutual information-guided diffusion process for learning the translation from an unseen source modality to the target modality using an unsupervised zero-shot learning approach. This approach has proven beneficial in cross-modality translation, as it assumes that statistical features between the two domains are identical. Additionally, to address the issue of sequential dependency in longitudinal data, Yoon et al. (2022) incorporated a sequence-aware Diffusion model. This model enables the learning of temporal dependencies within a sequence of data, even in the presence of missing data points. Jang et al. (2023) incorporated a text conditioning process for PET synthesis and guided the reverse process with multi-conditioning with both text and source MRI images. Text conditioning provided information such as gender, cognitive test score, scant time, and amyloid status. Moreover, Pan et al. (2023b) proposed a novel cycle-guided DDPM by employing two Diffusion models that condition each other by exchanging random noise in the reverse diffusion processes. In addition, the utilization of LDMs has gained popularity due to their ability to process high-dimensional data efficiently by employing low-dimensional latent space embeddings, thereby enhancing the overall efficiency of the Diffusion models (Jang et al., 2023; Jiang et al., 2023; Zhu et al., 2023).

The most commonly utilized denoising network architecture in conventional DDPMs is U-Nets, which is trained through a guided denoising process in a supervised manner. To extend the conventional approach, Özbey et al. (2022) proposed the first adversarial Diffusion model for medical image synthesis, trained in an unsupervised manner. Moreover, unlike regular DDPMs, they adopted a fast diffusion process using a larger reverse diffusion step size. To achieve unsupervised MRI translation, they incorporated a diffusive module with source conditional sampling and a non-diffusive model for estimating the source image, which is paired with its corresponding target image. For 3D image synthesis and learning cross-modality slice-wise mapping with LDMs, Zhu et al. (2023) introduced a network consisting of a series of 2D volumetric layers in a 2D slice mapping autoencoder network. This network was further fine-tuned with 3D data, enabling the extension of the 2D Diffusion model to its volumetric counterpart with reduced computational overhead. Additionally, Pan et al. (2023a) proposed a Transformer-based DDPM using a Swin ViT as the denoising network. This architecture involved Swin attention blocks in an encoder–decoder setup, with convolutional layers deployed in high-resolution levels to capture local information. The Swin attention layers provided a refined extraction of global features, resulting in high-quality image synthesis.

4.1.7. Model training

The training processes of deep learning-based networks for medical image synthesis can be classified into supervised, unsupervised, and semi-supervised approaches. In supervised training, models are trained with paired images, whereas source domain images are pixel-wise paired with corresponding target domain images. Most of the studies in the literature focus on supervised learning for medical image synthesis (Jung et al., 2018; Fu et al., 2019; Armanious et al., 2020; Sikka et al., 2021; Sun et al., 2022b). Predominantly, U-Net and cGAN-based networks are trained in a supervised manner with paired images. CycleGANs are the most commonly used GAN-based architecture for unsupervised learning. Despite utilizing an unsupervised approach with unpaired data, the cycle-consistency loss in the network ensures that synthesized images are indistinguishable from real images. In contrast to supervised and unsupervised learning methods, Yurt et al. (2022) proposed a semi-supervised approach for MR contrast synthesis to avoid excessive reliance on fully-sampled ground truth images by training the GAN network using undersampled source and target images.

4.2. Loss functions

4.2.1. Extended $\mathcal{L}_1$ and $\mathcal{L}_2$ losses

The selection and optimization of loss functions significantly influence a model's training process's stability and overall performance (Abu-Srhan et al., 2021). A majority of the CNN-based networks for medical image synthesis, including those by Chartsias et al. (2018), Chen et al. (2018a), Sikka et al. (2018), Fu et al. (2019), Li et al. (2019a), Bahrami et al. (2019), Sanaat et al. (2020), Zimmermann et al. (2022) have employed MAE or MSE. On the other hand, GAN-based approaches such as those by Jung et al. (2018), Wang et al. (2018, 2019), Hu et al. (2019), Kaplan and Zhu (2019), Shin et al. (2020b) have utilized the conventional adversarial loss function (Eq. (1)). CycleGAN-based approaches, including those by Pan et al. (2018), Armanious et al. (2019), Bui et al. (2020), Kong et al. (2021), Zhao et al. (2021), Xue et al. (2021), have employed the adversarial loss function in conjunction with the cycle consistency loss (Eq. (3)).

Image synthesis relying exclusively on the conventional loss function may not yield consistent results. To ensure high-fidelity image synthesis between source and target domains, $\mathcal{L}_1$ loss has been employed alongside the generator loss function to enhance the sharpness and resolution of the synthesized images (Yang et al., 2018a; Dar et al., 2019; Uzunova et al., 2020; Emami et al., 2020a; Zhan et al., 2022; Yurt et al., 2021). This also ensures the synthesized image maintains the same global structure as the real image variant (Emami et al., 2018; Zeng and Zheng, 2019; Armanious et al., 2020; Cusumano et al., 2020; Touati et al., 2021). Moreover, the concept of conventional $\mathcal{L}_1$ loss has been extended in different ways, such as preserving modality invariant characteristics of the synthesized images, explicitly refining the features in distinct frequency levels and ensuring the temporal consistency between images. Emami et al. (2020b) introduced a frequency-based $\mathcal{L}_1$ loss for their proposed FREA-UNet, with separate losses for high and low-frequency scales where G_h extracts high frequency and G_l extracts a low frequency and subsequently combining these losses to obtain the final objective function (Eq. (8)),

$$\mathcal{L}_{FREA-Unet} = \mathcal{L}_1(G_h) + \mathcal{L}_1(G_l) \tag{8}$$

Shi et al. (2021) also specified an $\mathcal{L}_1$ loss for high-frequency synthesized images by minimizing the difference between synthesized high-frequency image components and their ground truths obtained through Gaussian filtering. A novel loss function based on $\mathcal{L}_1$ was introduced for undersampled data in the semi-supervised GAN-based method proposed by Yurt et al. (2022), which comprises three components: multi-coil image loss, k-space loss, and adversarial loss. The multi-coil image loss calculates the $\mathcal{L}_1$ distance between the synthesized and reference images, both subjected to the same undersampling mask. Bui et al. (2020) introduced using $\mathcal{L}_1$ loss to ensure the temporal consistency between synthesized images by measuring the warping of generated and reference image slices in their proposed architecture to explore the displacement between image slices. Chen et al. (2021) employed $\mathcal{L}_1$ loss to regularize G by enforcing to focus on a target region in their proposed target-aware network, known as crossing loss. In addition, a weighted MSE loss function has been employed by Liu et al. (2020a) to overcome the bias in the optimization of the network due to the unbalanced sizes of bulk tissue volumes in the abdomen.

4.2.2. Mean P distance and gradient consistency losses

Although pixel reconstruction loss enhances the quality of synthesized images to a certain degree, typical $\mathcal{L}_1$ or $\mathcal{L}_2$ norms or mean squared distances for generator loss functions often result in images containing blurry regions (Nie et al., 2018; Lei et al., 2019b; Liu et al., 2020b). To address this issue and obtain sharper, more realistic images, Mean P Distance Loss (MPD) (p=1.5) (Eq. (9)) and gradient losses (Eq. (10)) have been introduced (Hiasa et al., 2018; Nie et al., 2018; Zhan et al., 2022; Koh et al., 2022), given as

$$\mathcal{L}_{mpd} = \|G(x) - y\|_P \tag{9}$$

$$\begin{aligned}\mathcal{L}_{gdl} = {} & \left\| |\nabla G(x)_x| - |\nabla y_x| \right\|_2 + \left\| |\nabla G(x)_y| - |\nabla y_y| \right\|_2 \\ & + \left\| |\nabla G(x)_z| - |\nabla y_z| \right\|_2\end{aligned} \tag{10}$$

These losses are commonly employed as non-adversarial losses for CycleGANs utilizing unpaired data (Armanious et al., 2019; Li et al., 2020; Kang et al., 2021). Hiasa et al. (2018) proposed a gradient consistency loss to enhance the CycleGAN approach by improving the accuracy at image boundaries, while Lei et al. (2019b) and Liu et al. (2020b) introduced MPD loss to mitigate blurry regions in synthesized images.

4.2.3. Perceptual loss

To achieve optimal anatomical detail in synthesized images, various non-adversarial losses have been combined with adversarial losses (Armanious et al., 2020). Among the most commonly employed non-adversarial losses are style-content loss and perceptual loss, which capture high and low-frequency components of synthesized images. Perceptual loss addresses inconsistencies in high-frequency features with fine details while ensuring global consistency (Dar et al., 2019; Vaidya et al., 2022). This approach further mitigates the issue of blurry results produced by pixel reconstruction loss, as it fails to capture the perceptual quality from a human perspective (Armanious et al., 2019, 2020). Kang et al. (2021) also applied the concept of perceptual loss to minimize high-level feature differences and avoid style and content discrepancies between synthetic and input images. The perceptual loss is computed from the MAE between extracted feature representations Armanious et al. (2020), using a discriminator as a feature extractor to obtain hidden layer features (Eq. (11)):

$$P_i(G(x,z), y) = \frac{1}{F_i} \| D_i(G(x,z), x) - D_i(x, y) \|_1 \tag{11}$$

where D_i denotes the feature representation of ith hidden layer of the D and F_i denotes the dimensions of the feature space. Then the perceptual loss can be defined as (Eq. (12)),

$$\mathcal{L}_{pt} = \sum_{i=1}^{L} P_i(G(x,z), y) \tag{12}$$

Additionally, Zhou et al. (2021) employed a Tencent network to extract 2D features, achieving perceptual similarity between synthesized and real images.

As an alternative approach, Li et al. (2020), Hu et al. (2021), and Ang et al. (2022) employed a pre-trained VGG model to calculate perceptual loss by extracting features from real and synthetic images without utilizing the discriminator as a feature extractor. Style loss has been incorporated to achieve optimal detail in synthesized images, enabling the transfer of real image style to the synthesized image to match their texture and details (Armanious et al., 2020).

4.2.4. Structure similarity loss

In addition to adversarial and reconstruction losses, Sikka et al. (2021), Li et al. (2023c) proposed the utilization of the Structural Similarity Index Measure (SSIM) as a cost function to preserve structural changes in images. However, SSIM is only sensitive to the scale at the local structure. To solve this limitation, the multi-Scale Structural Similarity Index Measure (MS-SSIM) was computed at various scales. To compute MS-SSIM, downsampled images (by a factor of 2) were iteratively employed and subsequently combined accordingly (Sikka et al., 2021) as follows (Eq. (13)).

$$\begin{aligned}&\text{MS-SSIM}(G(x), y) = \\ &w(k) \cdot [l(G(x), y)] \cdot \prod_{j=1}^{s} [c_j(G(x), y)]^{\beta^j} \cdot [s_j(G(x), y)]^{\gamma^j}\end{aligned} \tag{13}$$

The terms $l(G(x), y)$, $c_j(G(x), y)$, and $s_j(G(x), y)$ represent the luminance, contrast, and structure image components, respectively, while w(s) provides the weight factor for scale k. The authors employed three resolution levels, incorporating three distinct weights for each level to ensure structural consistency across various image resolutions. The combination of $\mathcal{L}_1$ loss with non-adversarial losses yielded superior results. The amalgamation of structural consistency loss and $\mathcal{L}_1$ loss effectively managed high-frequency regions preserved by the structural consistency loss and low-frequency regions with $\mathcal{L}_1$ loss (Abu-Srhan et al., 2021).

In unsupervised learning with CycleGANs, preserving the structural details of synthesized images is challenging due to the lack of paired ground truth images for comparison. To maintain the structural features of MR images in synthesized CT images using CycleGANs, Yang et al. (2018b) introduced a structure-consistency loss that enforces voxel-wise similarity between the extracted features of synthetic and input images. This approach employed a modality-independent neighbourhood descriptor (MIND) for feature extraction across both domains. Ge et al. (2019) suggested a Mutual Information (MI) loss as a cross-modality similarity metric in CT synthesis from MR images, aiming to enhance the structural consistency between synthesized and real images using (Eq. (14))

$$\mathcal{L}_{MI} = \sum_{G(x)} \sum_{x} p(G(x), x) \log \left[\frac{p(G(x), x)}{p(G(x))p(x)} \right] \tag{14}$$

where $p(x)$ and $p(G(x))$ denotes the distribution of x and $G(x)$, $p(x, G(x))$ gives the joint distribution of x and $G(x)$. Additionally, Jiangtao et al. (2021) utilized the SSIM loss function to generate more realistic synthesized images in CT synthesis.

4.2.5. Identical loss

Besides the conventional losses employed in CycleGANs, Zhao et al. (2020b) incorporated an identity loss into the overall loss function to ensure that the generators remain unaltered when presented with a source image. In this context, A and B represent the source and target domains, while G_A and G_B denote the two generators (Eq. (15)):

$$\begin{aligned}\mathcal{L}_{id}(G_A, G_B) = \mathbb{E}_{x\sim A}[\|G_A(x) - x\|_1] \\ +\mathbb{E}_{y\sim B}[\|G_B(y) - y\|_1]\end{aligned} \tag{15}$$

4.2.6. Attention-weighted loss

In contrast to conventional non-adversarial losses employed for image synthesis in specific regions, the attention-weighted loss has been utilized to emphasize specific regions in images. Li et al. (2019a) proposed a tissue-focused weighted mean squared error (MSE)-based loss function for CT image synthesis from MR brain images. The conventional MSE function was applied with brain and tissue masks to compute the MSE between images. The brain mask eliminates the image background, while tissue masks encompass white matter, grey matter, and cerebrospinal fluid (CSF) while excluding skin and the brain skull. Moreover, attention-weighted loss has been employed for whole-body images to identify particular areas, particularly in diagnostic processes (Wang et al., 2021b; Sikka et al., 2021).

5. Overview of public datasets

Studies on medical image synthesis have employed various image modalities and datasets to generate synthesized images. For CT synthesis from MR images, different MR contrasts have been utilized in experiments with paired and unpaired CT images. Most datasets were private and not publicly available except for a few datasets (Oulbacha and Kadoury, 2020; Chen et al., 2021; Ranjan et al., 2022). A summary of publicly available datasets utilized in medical image synthesis is presented in Table 4. T1-weighted and T2-weighted images have been widely used across numerous studies due to their extensive availability and limited training requirements for a single sequence (Hiasa et al., 2018; Lei et al., 2019b; Spadea et al., 2019; Boni et al., 2020, 2021; Zimmermann et al., 2022). Some studies have employed multiple MR sequences (T1, T2, T1C, and T1DixonC-water) by developing 2 to 4-channel model inputs, where each channel incorporates a specific image sequence (Qi et al., 2020). In supervised learning approaches for pseudo-CT, source image modality and CT images were aligned using various registration methods, such as rigid or deformable registration (Li et al., 2020; Klages et al., 2019; Jiangtao et al., 2021). Nonetheless, misalignment errors during registration significantly impacted training and evaluation processes, particularly at bone and air interfaces (Xiang et al., 2018; Arabi and Zaidi, 2021).

Standard pre-processing techniques in the reviewed studies for pseudo-CT include resampling, bias correction, normalization, and geometric distortion correction (Lei et al., 2019b; Fu et al., 2020; Klages et al., 2019; Peng et al., 2020; Koh et al., 2022; Takamiya et al., 2023). Various data regions, such as the brain, neck, pelvis, lung, and abdomen, have been considered for synthesizing CT images, but most studies have utilized brain datasets (Yang et al., 2018b; Lei et al., 2019b; Armanious et al., 2020; Abu-Srhan et al., 2021; Baydoun et al., 2021; Koh et al., 2022). Compared to other regions, abdominal soft tissues exhibit significant deformation between MR and CT images due to respiration, motion, and different organ-filling statuses (Qian et al., 2020), making it challenging to obtain voxel-wise alignment between CT and MR images for the same subject. Although deformable registration methods might enhance soft tissue alignment, they have limited accuracy due to residual mismatches between scans, which can reduce the quality of synthesized CT images (Liu et al., 2020a). Liu et al. (2020a) introduced a novel data scheme called semi-synthetic CT and MRI pairs to address these limitations. Semi-synthetic CT images were generated with soft tissue contrasts from voxel-wise soft tissue classification on MR images. Bone contrasts were determined from CT images by rigidly aligning them with MR images. This approach resolved soft tissue mismatching in image pairs and provided well-aligned images for network training. Numerous studies on CT synthesis have focused on high-field MRI ($> 1T$), while Cusumano et al. (2020), Fu et al. (2020), and Lenkowicz et al. (2022) proposed CT synthesis methodologies independent of the magnetic field strength of MR images, employing 0.35T MRI. In the context of CT synthesis for PET attenuation correction, PET/CT images have been predominantly utilized to acquire paired attenuation correction PET and CT images (Dong et al., 2019; Hwang et al., 2019; Reimold et al., 2019). Florkow et al. (2020) investigated the impact of gradient echo-based MRI contrasts as 3D patches in CT synthesis. During image acquisition, one echo was obtained in the almost in-phase (aIP) and another in the almost opposed-phase (aOP). Images were subsequently reconstructed to yield in-phase, opposed-phase, fat-only, and water-only images.

The majority of PET image synthesis studies (Jung et al., 2018; Sikka et al., 2018; Hu et al., 2019; Emami et al., 2020b; Hu et al., 2021; Sikka et al., 2021; Zhang et al., 2022b) relied on the Alzheimer Disease Neuroimaging Initiative (ADNI) database, which provides publicly accessible F18-AV-45 (florbetapir) and F18-FDG (fluorodeoxyglucose) PET images, as well as co-registered T1-weighted MRI data. A few investigations (Rajagopal et al., 2023) utilized private datasets. This trend can be attributed to most PET synthesis methods in the literature aiming to address missing data issues in Alzheimer's disease classification. The ADNI database offers the advantage of pre-processed datasets, facilitating the implementation of numerous supervised-based PET synthesis methods. Several studies (Jung et al., 2018; Hu et al., 2019; Lin et al., 2021) performed manual pre-processing on raw data from the same database, employing co-registration and normalization techniques, segmentation, and resampling. For FD PET synthesis, the most commonly used datasets in the literature were accompanied by either MRI or CT (Wang et al., 2019; Kaplan and Zhu, 2019; Chen et al., 2019; Sanaat et al., 2020; Wang et al., 2021b; Sun et al., 2022b), with input modalities based on multi-modality PET/MRI or PET/CT images. The incorporation of MRI or CT enhanced anatomical details within the synthesized images, which might be overlooked if only LD images were used. The primary focus region was the human brain, while other studies (Lei et al., 2019a; Häggström et al., 2019;

Table 4
A summary of publicly available 3D datasets utilized in medical image synthesis.

Dataset	Reference	Modality	Region	Highlights	Synthetic		
					CT	MR	PET
BraTS	Bakas et al. (2018)	MRI	Brain	T1, post-contrast T1-weighted (T1Gd), T2-weighted and T2-FLAIR	×	✓	×
IXI	Dalmaz et al. (2022)			T1, T2 and PD-weighted, MRA images and Diffusion-weighted image			
ADNI	Jack et al. (2008)	MRI, fMRI, PET, diffusion MRI, resting-fMRI, task-fMRI, and MEG/EEG		Multisite longitudinal study with biomarkers for use in Alzheimer's disease			✓
HCP	Essen et al. (2013)	Diffusion MRI, Resting-fMRI, Task-fMRI, and MEG/EEG		Characterize human brain connectivity and function in healthy adults			×
MRBrainS	Mendrik et al. (2015)	MRI		Provided a framework to evaluate automatic and semiautomatic algorithms for segmentation purposes in brain MRI of elderly subjects			
ISLES	Maier et al. (2017)			Ischemic Stroke Lesion Segmentation Dataset			
fastMRI	Zbontar et al. (2019)	MR	Brain, Knee	Contains both raw MR measurements and clinical MR images			
Iseg	Wang et al. (2015)	MRI	Brain	Multi-contrast MR images of infants and corresponding labels of Grey Matter (gm) and White Matter (wm).			
MICCAI-WMH dataset	Kuijf et al. (2019)			MRI for segmentation of wm hyperintensities			
MIDAS	Bullitt et al. (2005)			MRI images from healthy subjects			
ABCD	Hagler et al. (2019)			Adolescent Brain Cognitive Development			
ACDC	Bernard et al. (2018)		Cardiac	Cardiac MRI data			
DHCP	Makropoulos et al. (2018)		Brain	Infant brain MRI scans			
BCP	Howell et al. (2019)			MRI for behavioural assessments in healthy children			
LEMON	Babayan et al. (2019)	MRI, EEG	Brain, Body	Related to physiological characteristics of the brain and body			
Gold Atlas project (GAp)	Nyholm et al. (2018)	MRI, CT	Pelvis	Deformably registered CT and MR images over the pelvic region	✓	×	
spineweb	Vrtovec et al. (2016)		Spine	Provide Vertebral features			
RIRE	West et al. (1997)	MRI, CT, PET	Cranial	Designed to compare CT-MR and PET-MR registration techniques			
CHAOS	Kavur et al. (2021)	MRI, CT	Abdomen	Segmentation of abdominal organs from CT and MRI data			
COPD	Ehrhardt et al. (2016)	CT	Thorax	Thorax CT image database			
Ultra-low Dose PET	Shi et al. (2022)	PET	Whole-body	High quality and, low-dose PET imaging	×		✓
FDG-PET/CT	Gatidis et al. (2022)			Oncologic FDG-PET/CT			

Xue et al., 2021) examined whole-body scans. These investigations included image and projection space input modalities compared to image translation between other modalities. Most input data were in image space, and a few studies employed projection-based sinogram data for FD image synthesis. Most studies utilized private datasets for model training instead of publicly available datasets.

In MR image contrast synthesis, most studies have employed publicly accessible MRI datasets to demonstrate their proposed methodologies. Commonly synthesized image contrasts include T1-weighted and T2-weighted images, with other contrasts such as FLAIR, T1c, and PD also being synthesized (Dar et al., 2019; Bui et al., 2020; Uzunova et al., 2020; Kong et al., 2021; Zhao et al., 2021). Most supervised-based MRI synthesis approaches benefit from the availability of public datasets co-registered to the same anatomical templates, compared to other synthesis methods. For high-field MR image synthesis, low-field images are simulated from high-field images due to the challenge of obtaining both low and high-field images for the same subject. A low-field simulation is characterized by using high-field images as inputs and considering the mean signal-to-noise ratio (SNR) for grey and white matters (GM and WM) in the output image (Figini et al., 2020). The proposed image acquisition procedure for brain images encompasses segmenting WM, GM, CSF, and downsampling with a Gaussian filter along the direction of slices. The target SNR of low-field images is estimated from example low-field images, and the appropriate noise variance is added to obtain the mean SNR. Finally, simulated images are validated using paired 0.36T and 1.5T images of the same subject (Figini et al., 2020).

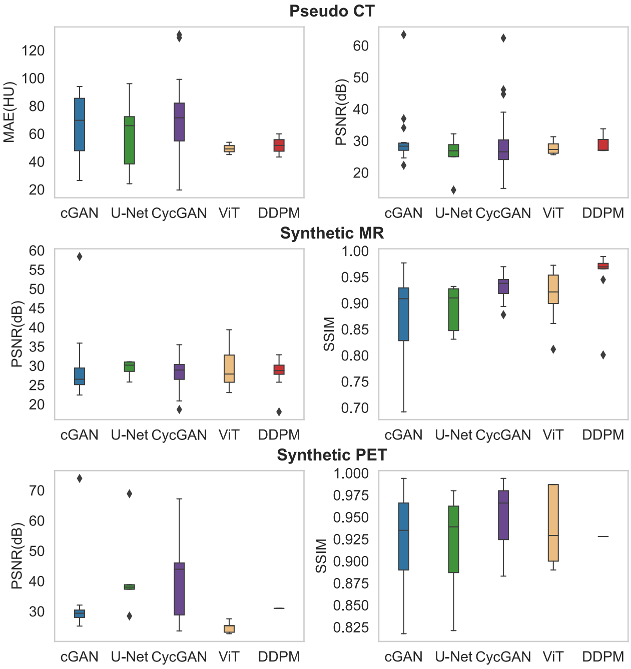

Fig. 13. Performance comparison of different image synthesis methods.

6. Performance evaluation

6.1. Evaluation based on image related metrics

The quantitative analysis of medical image synthesis most frequently employs evaluation metrics such as MAE, Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM). MAE computes the average pixel-wise difference between the synthesized and source images, PSNR assesses whether the synthesized image is a uniform projection of the real image, and SSIM concentrates on the visible structure of the images (Dai et al., 2020; Emami et al., 2020b; Sanaat et al., 2020). In addition to these metrics, MSE, Root Mean Square Error (RMSE), Frechet Inception Distance (FID), Normalized Root Mean Square Error (NRMSE), Normalized Mean Square Error (NMSE), and Pearson Correlation Coefficient (PCC), as well as Normalized Cross Correlation (NCC), are extensively utilized evaluation metrics across various studies. The number of studies employing different evaluation metrics in the literature indicates that MAE, PSNR, and SSIM are the most prevalent metrics across all image synthesis categories. MAE and PSNR are predominantly used for pseudo-CT synthesis, whereas PSNR and SSIM are commonly employed to evaluate synthetic MR and PET images. Fig. 13 summarizes evaluation results for distinct network types in each image synthesis category, including the obtained values for PSNR, SSIM and MAE.

In pseudo-CT synthesis, the mean absolute error for most studies ranges between 50 and 90 in Hounsfield Units (HU), while PSNR varies from 20 to 30 dB. However, comparing the performance of these methods based on their evaluation results is challenging, as they employ different datasets, target various body regions, and utilize distinct deep-learning techniques. Moreover, most studies use private datasets for CT synthesis, rendering their results incomparable due to the varying types of datasets used for training network models. To provide a fair representation of evaluation results, Fig. 14 displays quantitative values for CT synthesis, categorized by network type and the data region. As illustrated in the first two rows of Fig. 14, Transformers and DDPMs demonstrate superior performance in the brain region and cGAN performed better in the pelvic, achieving lower MAE and higher PSNR ranges.

Moreover, most studies have reported results within a standard range of values while comparing the mean absolute errors between real and synthetic CTs. However, a few studies (Fu et al., 2019; Hemsley et al., 2020; Zimmermann et al., 2022) have exhibited significant deviations from the norm, particularly between soft tissues and bones. For instance, Fu et al. (2019) reported larger absolute HU values for bone or air than for soft tissues. This is primarily due to the limited visibility of air and bones in MR images, which complicates the prediction process. Additionally, registration errors between CT and MRI can substantially impact intensity mapping, potentially causing shifts in air and bone tissue boundaries. Nevertheless, Fu et al. (2019) suggested that assigning a higher weight loss to bone structures during training could enhance bone accuracy. Compared to CT synthesis from MRI, utilizing PET for synthesizing CT for PET attenuation correction has presented challenges due to the images' reduced anatomical detail and spatial resolution. However, the proposed methods have achieved competitive results for CT synthesis, with less than 2% error in lesions for PET quality (Liu et al., 2018; Hwang et al., 2019; Armanious et al., 2019).

Fig. 14 (third row) presents the evaluation results for MRI synthesis and PET synthesis categorized by network type and body regions. As the summary of evaluation results indicates, Transformer and Diffusion-based approaches have demonstrated significantly better results in brain region for MRI contrast synthesis, achieving competitive outcomes regarding PSNR and SSIM. Furthermore, as illustrated in Fig. 14, the evaluation results for PET synthesis reveal that GAN-based approaches have produced better results when comparing the quality of synthesized full-dose images and PET from MRI. The choice of deep learning network has substantially impacted the quality of synthesized images for accurately learning the complex mapping between different domains. However, assessing the performance of networks trained on various datasets remains challenging. Several studies (Lei et al., 2019b; Emami et al., 2020a; Hu et al., 2019; Fu et al., 2020; Shin et al., 2020a; Lin et al., 2021; Sikka et al., 2021; Sun et al., 2022b; Xue et al., 2021; Zhao et al., 2020b) have compared their approaches with other competitive networks by using the same datasets and reproducing results to identify the limitations and accuracy of the proposed network architecture. For instance, in CT synthesis, the performance of cGANs was compared to that of CycleGANs (*CycGANs*) to investigate the accuracy of sCT for MR-only liver radiotherapy, nasopharyngeal carcinoma (NPC) IMRT planning, and MR-only treatment planning in head and neck (HN) cancer patients (Fu et al., 2020; Peng et al., 2020; Klages et al., 2019). In these studies, cGANs outperformed CycleGANs by achieving lower MAE values for synthetic CT. Nonetheless, Xue et al. (2021) demonstrated that CycleGAN outperforms other approaches in FD PET synthesis by cyclically incorporating an inverse transform. A recent study on pseudo-CT employing a Transformer-based generator model by Dalmaz et al. (2022) assessed the performance of two model variants: task-specific and unified models. The task-specific model executed a single synthesis task, with separate models constructed for each MRI contrast synthesis task. In contrast, the unified model for multi-contrast MRI achieved enhanced synthesis performance, particularly in pathological regions such as tumours and lesions, compared to task-specific models. Cao et al. (2021) and Fu et al. (2019) compared their deep learning-based approaches with conventional image synthesis methods, such as atlas-based and voxel-based. As traditional methods heavily relied on manually fine-tuned parameters and registration performance, they resulted in higher MAE than deep learning-based approaches.

Hu et al. (2021) quantitatively evaluated their proposed model, examining the effects of various hyperparameters, including depth, patch size, and the combination of different loss functions (adversarial loss, KL, perceptual loss, and $\mathcal{L}_1$). They compared the results of their 3D-based network to a 2D variant, further highlighting the superior outcomes of the 3D-based method. Sikka et al. (2021) and Zotova et al. (2021) investigated the impact of variations in the objective function

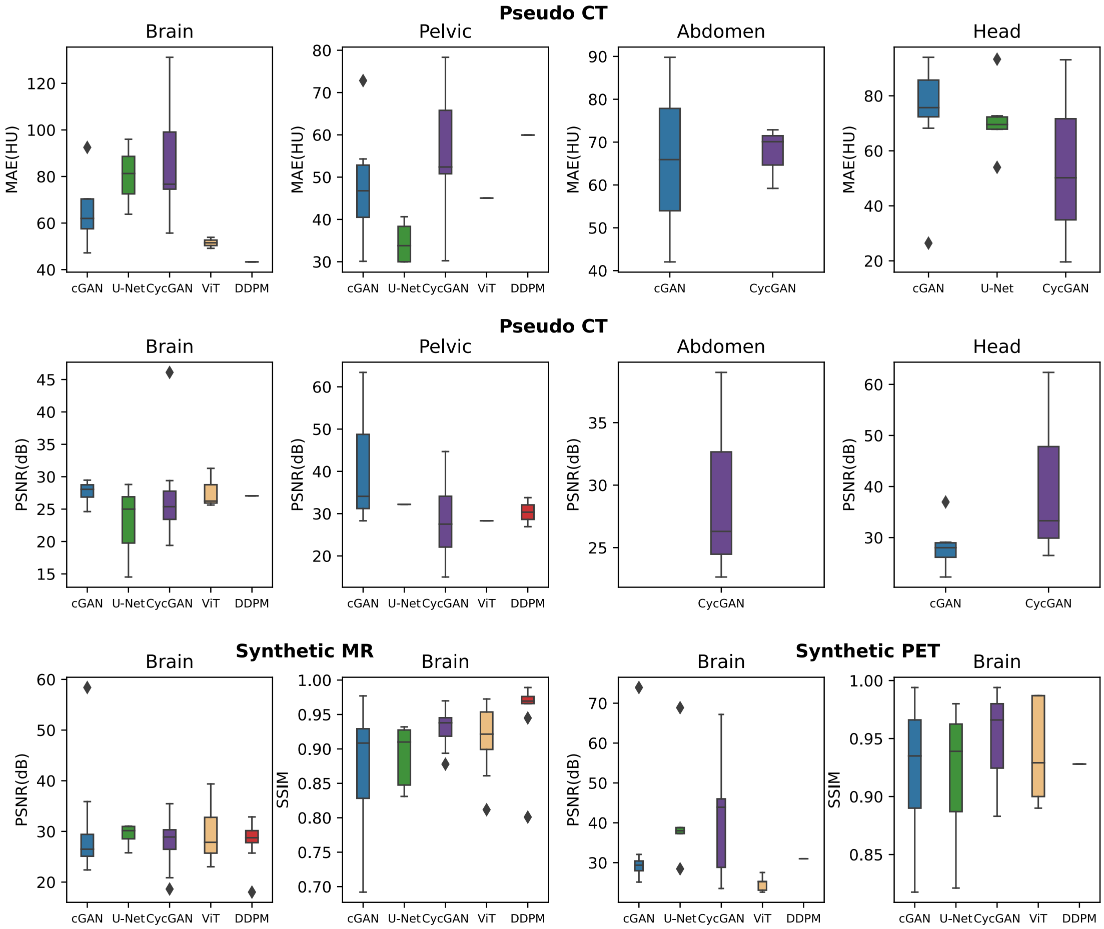

Fig. 14. Performance comparison of different image synthesis methods based on body regions.

on overall network performance. Sikka et al. (2021) demonstrated that combining GAN's adversarial loss with $\mathcal{L}_1$ loss ensures voxel fidelity in intensity values. Multi-scale-SSIM loss provides structural consistency for synthesized images at different scales. Additionally, region-wise loss based on anatomical delineation further guarantees local contextual integrity, refining region sensitivity of images for disease classification. Chartsias et al. (2018) and Yang et al. (2018a) analysed performance against different datasets, while Yang et al. (2021) evaluated the effectiveness of various training loss functions to identify the best-performing combination of adversarial loss and other structural and style losses. They found that combining adversarial losses with common space losses, such as identity loss, reconstruction loss, and contrast-classification loss, alongside cycle-consistency loss for CycleGANs yielded better results. Wang et al. (2021b) also assessed the impact of their proposed attention-weighted loss, where each significant region commonly associated with tumour occurrence was assigned a higher weight in the whole body. Their findings indicated that the weighted loss mechanism enhanced the overall performance of synthesized PET images.

Many studies have primarily focused on network design and loss function selection; however, Hu et al. (2019) evaluated the impact of batch normalization on overall performance, as inaccurate batch size estimations could hinder performance. They validated their proposed network using four normalization methods: Batch Norm, Layer Norm, Instance Norm, and Group Norm. Their findings suggest that instance normalization performs better in cross-modality translation, while most studies rely on default batch normalization. Additionally, Chen et al. (2019) and Wang et al. (2021b) assessed the impact of combining MRI with low-dose PET on image synthesis quality, demonstrating that MRI provides crucial anatomical details for synthesizing detailed FD images. As many PET synthesis approaches focus on generating missing data for Alzheimer's classification, several studies Sikka et al. (2018, 2021) evaluated the potential diagnostic utility of generated images in addition to quantitative results. While many studies assess proposed methods in image space, Sanaat et al. (2020) evaluated FD PET synthesis performance using image and projection spaces. They examined the predicted results of FD PET images and FD PET sinograms synthesized from their corresponding LD images and sinograms. The findings indicated that FD sinograms captured more patterns and anatomical details than FD images. Furthermore, Florkow et al. (2020) investigated the influence of gradient echo-MRI contrasts on CT synthesis using three main configurations. Their proposed method found that almost in-phase inputs outperformed almost opposed-phase-based input configurations, primarily due to the lower correlation between CT and MR in the opposed-phase. The opposed phase exhibited destructive interference between fat and water in MRI, which was not present in the in-phase configuration and proved more favourable for CT image synthesis.

Despite the high accuracy and computational efficiency of deep learning methods in clinical applications, incorrect predictions may occur, which pose risks for treatment planning. A possible solution involves incorporating uncertainty prediction for estimating data-dependent (DD) and model-dependent (MD) prediction errors. Hemsley et al. (2020) conducted the first study to consider an uncertainty-based evaluation method for CT synthesis. Jung et al. (2018) assessed the impact of single-axis and multiple-axis slices (augmented) on the quality of synthesized images, revealing a minor difference between single-axis and augmented methods in terms of image quality. Zotova et al. (2021) proposed a cycleGAN-based brain FDG-PET synthesis with two dataset configurations: semi-3D and 3D patch models, which were dependent on input data size. In the semi-3D approach, the model received three adjacent transverse slices corresponding to each channel as inputs, while the 3D-patch-based model utilized 3D mini patches from the original image. Although both approaches demonstrated significant results, the semi-3D method achieved slightly improved outcomes compared to 3D patches.

The majority of deep learning approaches were applied to abnormal cases, aligning with the primary objective of cross-modality medical image synthesis to reduce the necessity for acquiring multi-modal images and subjecting patients to additional radiation exposure during clinical processes. Few studies have explored their proposed methods on normal subjects (Chen et al., 2021). Consequently, the comprehensive evaluation of performance provides valuable perspectives on the proposed deep learning techniques across different pathological conditions in abnormal cases. However, few studies have taken the initiative to evaluate their methods on normal and abnormal cases. Dar et al. (2019) demonstrated their proposed MRI synthesis approach from healthy subjects and glioma patients, outperforming all other competing methods in both scenarios. Moreover, they showcased the generalizability of their model for both pathological and healthy cases. Additionally, Dalmaz et al. (2022) exhibited the effectiveness of their Transformer-based method in capturing the underlying contextual details pertaining to both healthy and pathological tissues.

When comparing the performance of networks trained on paired and unpaired data, it was observed that CycleGANs were more commonly utilized for unsupervised training, while other networks were predominantly trained in a supervised manner. Analysing the results obtained from CycleGANs and other networks, it becomes evident that supervised learning-based networks achieved better results in most cases. Moreover, Estakhraji et al. (2023) conducted a comprehensive validation of this performance gain in supervised training over unsupervised training for medical image synthesis. Their study delved into the performance of unsupervised training on larger unpaired datasets and the extent of performance improvement achievable through fine-tuning with supervised training. They validated these findings using a CycleGAN for synthetic CT from MRI data. This study revealed that even the addition of just one paired image for training could substantially enhance the quality of synthetic images, and the best results were obtained when the number of paired data was increased. Consequently, it was evident that supervised training outperforms unsupervised training in many cases.

Although the results of various studies are not directly comparable due to differences in datasets, image contrasts, and training methods, recent research on the transformer and Diffusion-based image synthesis (Dalmaz et al., 2022; Yan et al., 2022; Li et al., 2023c,b; Zhao et al., 2023; Özbey et al., 2022; Lyu and Wang, 2022) have demonstrated significant improvements compared to traditional GAN-based approaches, with enhanced contextual sensitivity. Several studies (Zhang et al., 2022a; Liu et al., 2023; Yan et al., 2022; Li et al., 2022a; Zhao et al., 2023; Luo et al., 2021; Hu and Liu, 2022) compared Transformer-based synthesis and other state-of-the-art CNN-based approaches, consistently demonstrating superior results over CNN based networks. Zhang et al. (2022a) showcased the effectiveness of their proposed Transformer-based PTNet3D in synthesizing longitudinal infant scans with varying tissue contrasts at different ages. In contrast, other CNN-based methods only achieved satisfactory results for infants under six months. Liu et al. (2023) evaluated how Transformer-based MRI synthesis outperformed other CNN-based models in missing data imputation, while Li et al. (2022a) identified the effectiveness of augmenting GANs using Transformers. Additionally, Diffusion-based methods for cross-modality synthesis showed promising outcomes in achieving high-fidelity image translation. Unlike GAN-based approaches that indirectly utilize the distribution of the target modality without evaluating the likelihood, DDPMs emerged as a better alternative, enhancing the fidelity of synthesized data (Jiang et al., 2023; Özbey et al., 2022; Lyu and Wang, 2022; Li et al., 2023a).

6.2. Evaluation based on downstream tasks

While direct evaluations traditionally focus on the quality of synthesized images, some studies have extended their evaluations to downstream tasks, such as segmentation or classification, to ascertain clinical validity. We overview evaluations concerning synthetic CT, MRI, and PET to underscore their clinical applicability.

Pseudo-CT has been predominantly employed in the context of MRI-only treatment planning. Few studies have evaluated the synthesized CT results to determine their feasibility for treatment planning purposes by focusing primarily on dose distribution estimations and segmentation accuracy rather than relying solely on direct quantitative assessments (Touati et al., 2021; Cusumano et al., 2020; Liu et al., 2020a; Boni et al., 2020). For instance, Hiasa et al. (2018) evaluated synthetic CT images derived from MRI scans of osteonecrosis patients with the segmentation accuracies for gluteus medius and minimus muscles and femur. They observed a statistically significant improved Dice similarity coefficient (DSC) for the pelvis region, yielding a value of 0.80. Similarly, Bahrami et al. (2020) achieved a DSC of 0.98, and Fu et al. (2019) obtained a DSC of 0.81 for soft tissue contrasts in cancer patients. In the context of liver segmentation as a downstream task, Chen et al. (2021) obtained a DSC of 0.97. Additionally, Boroojeni et al. (2022) reported higher accuracy in using synthesized CT images for radiation-free MR cranial bone imaging, with a DSC of 0.90. For prostate cancer patients, Kang et al. (2021) analysed the dose-volume histogram (DVH) of synthesized CT images and compared them to real CT images for planning target volume (PTV) and organs-at-risk (OAR) delineation. Their proposed approach yielded DVH parameters that were very similar between the real and synthesized CT images. Similarly, Boni et al. (2021) reported similar DVH parameter differences between synthesized and real CT images for posterior cancer patients, with a maximum relative dose difference of 1.2% in the high-dose region. Chen et al. (2018a) reported a DVH discrepancy of less than 0.87%, while Hsu et al. (2022) demonstrated a discrepancy of less than 1.0%, indicating the possibility of utilizing synthesized CT images for prostate IMRT planning.

In the context of MRI synthesis, only a limited number of studies have focused on evaluating the application of synthesized images, such as tumour segmentation. Hu et al. (2022) reported dice score values of 0.87 on a tumour dataset using synthesized cross-contrast MRI images. Further, Zhang et al. (2022a) assessed the validity of synthesized MRI contrasts in infant whole-brain segmentation. They evaluated the segmentation accuracy using DSC, Dice Average Surface (DAS), and 95% Hausdorff Distance (HD95) for synthesized MRI scans, including corrupted images. They achieved accuracy scores of 0.87, 0.21, and 0.88 for DSC, DAS, and HD95, respectively, supporting the realistic synthesis of results. Additionally, using an imputed tumour dataset, Liu et al. (2023) demonstrated the diagnostic equivalence of synthesized MRI images. Their synthesized images achieved higher DSC values greater than 0.90 for all cross-modality synthesis scenarios. In examining the impact of synthesized contrast-enhanced MRI (CEMRI) on tumour detection, Zhao et al. (2020a) exhibited a high and consistent accuracy of 89.40%, which closely approached the performance achieved with real CEMRI for identifying tumours. Salem et al. (2019) further demonstrated the validity of synthesized MRI in detecting multiple sclerosis lesions. Moreover, Lin et al. (2023) conducted volume estimation on IQT-enhanced MRI images derived from LF MRI and employed the Relative Volume Error (RVE) score to assess segmentation accuracy in seven subcortical regions by obtaining higher RVE scores.

PET image synthesis has primarily evaluated the generated images through classification tasks, as PET imaging plays a crucial role in diagnosing various conditions. For instance, Pan et al. (2018) assessed the effectiveness of synthesized PET images for Alzheimer's disease (AD) and mild cognitive impairment (MCI) classification. Their study demonstrated that the use of synthesized PET images improved brain disease classification, achieving higher sensitivity scores. Similarly, Shin et al. (2020b), Lin et al. (2021), Sikka et al. (2021), Xie et al. (2023) demonstrated the clinical utility of synthesized PET results for AD classification, showing promising results for MR-based AD diagnosis. Wang et al. (2018) also examined the capability of estimated FD PET images to maintain reliable Standardized Uptake Values (SUV) in the hippocampal regions of MCI patients. Similarly, Chen et al. (2019), Wu et al. (2019), Lu et al. (2019), Sanaat et al. (2020), Zhao et al. (2020b), Zhou et al. (2022) evaluated the SUV bias and variance between real and synthesized FD PET images derived from LD PET. This evaluation aimed to demonstrate the quality of the synthesized images, particularly concerning the accuracy of SUV measurements in comparison to real PET images. Furthermore, Wei et al. (2019) evaluated synthesized PET images for healthy and Multiple Sclerosis (MS) patients. They measured the accuracy of synthesizing myelin content in different regions of interest in MS patients. The results showed an MAE of 0.02 for MS lesions and minimal errors in other regions, indicating high accuracy of the synthesized results. Furthermore, to assess the clinical validity of synthesized PET images, Hussein et al. (2022) conducted classification analyses for Moyamoya disease, intracranial steno-occlusive disease, and stroke. Their study achieved identification accuracy higher than 90% for each condition, confirming the usability of synthetic PET images in these clinical scenarios.

7. Discussion

Medical image synthesis is a promising area of research for medical imaging, and it can potentially offer significant benefits across various medical applications, such as reducing radiation for PET and improving diagnostic confidence in MRI and CT. This review has examined numerous medical image synthesis studies from recent literature, exploring different perspectives. Each synthesis method has the potential to facilitate a range of clinical applications, for example, the generation of CT images for radiotherapy treatment planning and PET attenuation correction, PET synthesis for diagnosing various brain disorders, MRI contrast synthesis for tissue differentiation and enhanced diagnostic accuracy, and FD-PET synthesis for improved lesion detection and evaluation in oncology treatments.

Synthetic CT imaging has been widely used in oncology, specifically in the context of MRI-only radiotherapy treatment planning. It overcomes the limitation of MRI in representing cortical bone signals and enables accurate positional verification of reference images for image-guided radiotherapy. Further, its application has also expanded to include the preoperative planning for MR-only cochlear implant procedures (Fan et al., 2023). By synthesizing CT images of the temporal bone, which are typically obtained through CT scans, this technique offers an alternative when safety concerns prevent the use of CT imaging. Moreover, synthesized CT images have proven clinically valid in visualizing osseous structures, such as the sacroiliac joints, hips, and spine (Jans et al., 2021; Morbé et al., 2023). Synthetic CT imaging has also demonstrated its usefulness in detecting cranial abnormalities in pediatric patients with conditions like craniosynostosis or head trauma. Synthetic MR imaging is most commonly employed in clinical settings to obtain supplementary tissue contrast information for accurate diagnosis. These synthetic contrasts have been utilized to synthesize multiple sclerosis lesions on MR images, facilitating the generation of annotated data for research and clinical applications. Furthermore, synthetic contrast-enhanced MRI (CEMRI) has emerged as a safe and cost-effective alternative to conventional CEMRI (Zhao et al., 2020a). This provides precise tumour identification of hemangioma and hepatocellular carcinoma cases, thereby reducing the risks associated with contrast agents, particularly in patients with compromised kidney function. In the diagnosis of Alzheimer's disease, synthesized PET images are progressively replacing the joint analysis of MRI and PET. Moreover, synthesized PET-derived myelin content maps from MRI have facilitated the quantification of tissue myelin content in multiple sclerosis (Wei et al., 2019).

Moreover, a significant benefit of employing synthesized medical imaging is its ability to mitigate challenges associated with multi-modal registration for Computer-Assisted Intervention (CAI) applications. This is primarily because multi-modal registration involves the intricate task of aligning images from diverse modalities, which makes it difficult because of disparities in underlying characteristics, spatial information, and contrasts. By utilizing deep learning approaches to synthesize medical images across multiple modalities, this challenge can be effectively addressed, bridging the resolution, contrast and geometrical differences across modalities (Yang et al., 2018b; Kong et al., 2021). As synthesized modalities are modelled explicitly using their corresponding source modalities, more accurate and precise registration of images is obtained (Fu et al., 2020; Li et al., 2019a). These generated images can facilitate improved interventional applications, and overall medical imaging workflows (Liu et al., 2023). However, despite the wide-ranging applications and advantages of medical image synthesis in clinical practice, there are still common challenges in medical image synthesis that necessitate further research across various domains.

7.1. Challenges with large domain differences

Translating within modalities, for instance from T1w to T2w MRI or LD to FD PET, is generally more tractable than cross-modality mapping because of the inherent discrepancies in signal and spatial content between distinct imaging modalities. Specifically, CT synthesis from MRI data poses significant challenges; MRI does not directly yield the electron density maps intrinsic to CT synthesis. This is further complicated in regions where MRI depicts bone and air cavities with similar intensities, given their contrasting attenuation properties. The inherent limitation of conventional MR sequences in distinguishing between bone and air exacerbates the problem. Regions like the thorax also introduce complexities due to their marked heterogeneity and the intricacies in modelling lesions accurately.

The process of creating CT images from PET data is fraught with potential errors. The synthesized CT image needs to depict detailed bone structures to aid in PET attenuation correction. Moreover, it should represent soft-tissue structures accurately, ensuring effective MR motion correction and supporting subsequent tasks like segmentation and organ volume estimation. MRI, with its superior tissue contrast, especially in brain regions, showcases a wide-ranging contrast palette. This makes synthesizing one MRI contrast from another uniquely challenging, especially when compared to simpler MR-to-CT or CT-to-PET synthesis tasks, where CT or PET typically presents limited soft tissue contrast. The challenge is further amplified in multi-contrast MRI synthesis, given the imperative to map data into a unified shared representation.

Creating PET images from MRI data is also intricate, primarily because it entails predicting functional images from their structural counterparts. While MRI provides rich texture information, PET images exhibit complex textures across various frequency scales. This texture differential poses substantial challenges, often influencing the accuracy and realism of the synthesized PET images.

7.2. Challenges with architectural designs

Although deep learning-based medical image synthesis demonstrates improved learning of complex non-linear relationships between image modalities, specific challenges persist in various network architecture designs. This section discusses the identified architectural challenges in medical image synthesis and the most frequently employed solutions in the existing literature.

Conventional generative architectures: One potential problem with CNN-based image translation architectures is that they might fail to capture neighbouring details due to the fixed-size receptive fields, especially with small input sizes. The image quality of VAEs still lags behind GANs due to the model's constrained expressiveness or imperfect loss criteria. Although GANs are powerful generative models, their ability to represent intricate data and indirectly characterize data distribution without evaluating likelihood is restricted. This limitation could lead to premature convergence or model collapse. Moreover, Vision Transformers might not capture multi-scale features when using fixed-size image patches. While diffusion-based models exhibit impressive outcomes, they often alter the original data distribution of input images due to introduced random noise, neglecting the structural consistency inherent in the input data. Furthermore, the generalizability of existing architectures in the medical domain remains unexplored due to the primary concentration on training models for specific datasets from individual centres, without considering the potential benefits of leveraging shared knowledge across multiple centres.

U-Net-based architectures, featuring skip connections between the encoder and decoder, have effectively addressed challenges in capturing local details within CNN-based methods (Chen et al., 2018a; Fu et al., 2019; Scholey et al., 2022). Imposing adversarial based learning with VAEs enforces the perceptual quality of the results (Liu et al., 2021c). Multi-scale designs and pyramid networks capture features at varying resolutions, and Iterative approaches refine the outputs in stages (Li et al., 2023c, 2022; Zhang et al., 2022a; Kearney et al., 2020; Zhao et al., 2022; Kläser et al., 2018; Sikka et al., 2021). Diffusion models synthesize more complex and diverse image modalities with less risk of model collapse while different conditional and guidance strategies preserve the anatomical structure of the input domain Jiang et al. (2023). Multi-centre data challenges are addressed by developing generalist models with shared architectures, addressing centre-specific issues and introducing dynamic routing (Yang et al., 2023).

Handling diverse input dimensions: Most of the proposed network architectures use 2D independent images for translation, considering 1-channel 2D images extracted from 3D images. This approach may introduce potential discontinuity between slices and loss of voxel-wise correlation. Although 3D-based networks exhibit more accurate image synthesis, they also necessitate learning significantly more training parameters. In patch-based learning approaches, determining patch size and the overlapping ratio between 2D or 3D patches is critical and challenging. It may disrupt the feature-wise relationships between organs and tissues, ultimately reducing the quality of synthesized results. Additionally, patch size significantly impacts network memory consumption, as larger patches entail higher computational costs.

For addressing the discontinuity in 2D slices, 3D patch-based methods maintain cross-slice context (Lei et al., 2019b; Florkow et al., 2020). Attention-based methods reduce discontinuity effects by emphasizing relevant slices (Kearney et al., 2020), while the fusion of triple orthogonal images or integrating volumetric layers within 2D models fine-tunes synthesis (Li et al., 2018; Zhu et al., 2023). In patch-based architectures, optimal patch size selection, adaptive patch sizing, and overlapping patches capture both local details and broader contexts (Liu et al., 2020b; Li et al., 2020; Zhao et al., 2021; Boroojeni et al., 2022). Multi-scale patches also effectively handle spatial variations (Zhao et al., 2021)

Multi-modality based architectures: Multi-modality-based image inputs are advantageous in specific applications, such as synthesizing MRI image contrasts and generating FD PET from PET/CT and PET/MRI images. This approach may overlook the potential of deep networks to extract more complex features from different image contrasts, where each contrast contributes unique feature-wise details for creating more advanced and precise images. Even when using multi-channel-based approaches to leverage multi-modality data inputs, applying a unified kernel to convolve all modalities is not always accurate, as feature-wise information in each modality may differ at various locations. Thus, addressing the contribution of image modalities and integrating each modality's features using optimized fusion methods presents a challenge.

To efficiently fuse features from multi-modality inputs, various fusion techniques have been explored, including locality-adaptive fusion, Transformer-based fusion, early fusion, layer-wise fusion (Wang et al., 2019; Zhou et al., 2020a; Zhan et al., 2022; Li et al., 2023b). Attention mechanisms and shared latent representations further improve feature integration across modalities and enhance their contribution to the final output (Chartsias et al., 2018). Shared latent representation of multi-modalities facilitates the precise feature integration (Chartsias et al., 2018). Domain adaptation techniques ensure more precise integration of features across modalities.

Preserving local and global context: In cross-modality translation, most methods focus on capturing the local textures of images, potentially neglecting global structure and failing to synthesize finer structures. Generative architectures constructed purely from CNNs could face challenges in effectively generalizing across diverse subjects and achieving optimal feature extraction of long-range relationships. Even when attention mechanisms in spatial or channel-wise attention are incorporated to guide networks' focus to specific regions, the resulting attention maps might scatter across different areas. This outcome restricts the ability of local CNN-based features to express the broader contextual information convincingly. In the same manner, while Transformer-based networks excel at grasping the global context, their capability to extract local features is comparatively restricted in comparison to CNNs.

Encoder–decoder architectures with skip connections facilitate the network in capturing both local and global details. The adoption of multi-scale processing allows the network to grasp features at varying resolutions, and Attention-based mechanisms are instrumental in selectively focusing on more relevant features. Hybrid architectures combining both CNN and transformer networks capture local and global features simultaneously (Dalmaz et al., 2022; Zhao et al., 2023). The utilization of separate modules that focus on local and global features separately provides capturing more precise features (Sikka et al., 2021). The combination of pixel-wise losses with perceptual losses contributes to preserving both local and global structures (Dar et al., 2019; Vaidya et al., 2022).

Handling misalignments in data: The network architectures for image-to-image translation in a supervised manner necessitate well-aligned paired images to generate more precise data, as misalignment between image pairs can adversely impact the overall image synthesis process, causing unreasonable displacements. Paired images are required for specific image synthesis tasks, such as low-field to high-field MRI translation. While CycleGANs mitigate the need for well-aligned images, ensuring the structural consistency of synthesized images in this approach remains challenging. Therefore, it is challenging to satisfy the requirement of well-aligned paired data as errors might propagate through the network and result in unreasonable displacements in the synthesized images.

Integration of registration network along with the generator network effectively mitigates misalignment noise between image pairs (Kong et al., 2021). This perspective treats misaligned data as labels containing noise, framing the issue as training a model in the presence of noisy labels. Predicting deformation vector fields also aligns the paired images accurately (Bui et al., 2020). The use of modality-specific atlases and multi-modal deformation techniques enhance the image synthesis process (Lin et al., 2022).

7.3. Challenges with limited datasets

Contrary to traditional image translation methods, deep learning-based image synthesis necessitates a substantial volume of training data to accurately determine the underlying mapping between image domains. This approach helps to prevent model underfitting due to insufficient data and the production of imprecise synthesized images. Medical image translation-based studies have employed datasets with fewer samples than other computer vision image processing tasks.

Acquiring a large set of data samples in medical imaging is challenging due to high image acquisition costs, exposure to harmful radiation, and a limited number of patients. Furthermore, a portion of acquired images is lost during the cleaning process to eliminate outliers, further reducing dataset size. Supervised-based learning methods require paired images with aligned source and target images; however, obtaining precise registration between image modalities in different domains is complex and can potentially diminish the quality of synthesized images. Owing to the limited number of datasets, many studies have not tested models with unseen data, rendering the proposed methods insufficient for clinical use. Moreover, training datasets limited to single-scanner images acquired at one institution may impact the model's generalizability across different hardware settings. As medical image translation methods are involved in clinical usage for various disease diagnosis processes, it is essential to validate the quality of synthesized images for deployment in clinical applications, necessitating additional data.

Adopting a patch-based training method and various data augmentation techniques (Oulbacha and Kadoury, 2020; Li et al., 2020; Uzunova et al., 2020) effectively handles the data scarcity. Semi-synthesized datasets (Liu et al., 2020a) also overcome the data limitations. Shallow network architectures possess the capability to learn effectively from small datasets (Li et al., 2020; Baydoun et al., 2021). Transfer learning helps to mitigate data limitations by initially training the network on a large dataset from a related domain and then fine-tuning it on a limited medical dataset (Wang et al., 2021; Li et al., 2021a; Abu-Srhan et al., 2021). Moreover, Semi-supervised learning-based approaches offer a viable solution for training with limited datasets by incorporating both labelled and unlabelled data (Yurt et al., 2022). Multi-task learning is also beneficial in dealing with data limitations, as it enhances the model's ability to generalize by sharing information between tasks (Hussein et al., 2022). Domain-adaptive methods adapt the model to handle other unseen data when the target image modality is not accessible (Hu et al., 2022). Furthermore, zero-shot learning emerges as a promising solution for learning the translation of unseen source data to the target domain, providing valuable insights even when training data is scarce (Wang et al., 2023).

7.4. Challenges with computational cost

Most proposed networks employ deep architectural designs, necessitating the learning of high-dimensional data, which in turn requires computationally powerful resources with GPU support for efficient model training. Limited resources can result in longer training times. One primary reason for training models with 2D image slices is resource constraints, which can lead to the discontinuity between image slices and the loss of significant voxel-wise correlations. However, even 3D-based models trained with low computational resources might risk losing larger-scale image features. Furthermore, while novel network architectures such as Transformers and Diffusion-based models have shown promising results in medical image synthesis, they do incur a higher computational burden compared to other state-of-the-art CNN-based models.

Patch-based learning (Zhao et al., 2021) and aggregating neighbouring 2D slices to capture 3D features (Oulbacha and Kadoury, 2020) provides a robust solution for computation burden. Weight sharing mechanisms (Kläser et al., 2018; Dalmaz et al., 2022), utilizes computationally efficient architectural components (Oulbacha and Kadoury, 2020; Zhao et al., 2022; Fu et al., 2019; Zhang et al., 2022c; Luo et al., 2021), and extending 2D models to their volumetric counterparts (Zhu et al., 2023) is also provides a less computational solution. Hybrid 2D and 3D networks and architectural enhancements (Zeng and Zheng, 2019) reduce computational and memory requirements. Moreover, learning from low-dimensional latent space embeddings rather than directly in the image space addresses computational challenges (Jang et al., 2023; Jiang et al., 2023; Zhu et al., 2023). Transformers, while computationally intensive, can be employed as bottleneck layers in combination with CNN networks (Dalmaz et al., 2022). Hierarchical Transformer-based architectures, such as Swin Transformers, exhibit lower computational intensity compared to ViTs (Yan et al., 2022). In Diffusion-based approaches, the adoption of LDMs has overcome memory and computational challenges (Zhu et al., 2023).

7.5. Future directions

Potential future research directions in the examined literature can be classified into various aspects, including deep learning networks, datasets, and result validation. Based on the analysis of deep learning-based approaches for medical image translation, most studies concentrate on CT synthesis. Consequently, there is an opportunity for further research on image contrast synthesis using diverse strategies. Additionally, many studies have focused on supervised-based methods to achieve more accurate image synthesis using paired data. While supervised techniques have demonstrated significant results in the image translation process, unsupervised-based methods still possess room for improvement in architectural enhancements. Conventional unsupervised networks, such as CycleGANs, heavily rely on the network's ability to learn specific anatomies. Integrating attention-based methods with traditional GAN-based approaches could enhance networks' performance, leading to higher-quality results.

Although both CNN-based and GAN-based approaches have achieved competitive results in medical image synthesis, these networks exhibit limitations in capturing long-range relationships among image regions due to restricted receptive fields. Recently introduced Vision-Transformer-based approaches have addressed this issue, outperforming many conventional deep learning-based methods for image synthesis. These Transformer-based methods have been applied in limited MRI-to-PET translation and MRI contrast synthesis studies. Furthermore, novel Transformer-based, Contrastive-Learning-based, and Diffusion-based approaches have demonstrated significant results for image synthesis compared to conventional network architectures. Among recent generative architectures, Diffusion-based approaches have demonstrated more promising results in medical image synthesis. In addition, Implicit Neural Representation-based approaches (Chen et al., 2023) have recently demonstrated improved performance in medical image synthesis. Consequently, a gap exists in the literature concerning these innovative network architectures for medical image synthesis tasks. It would be beneficial to refine model architectures to leverage multi-contrast images for synthesizing a given modality, capturing more valuable information from multiple modalities and developing models that can generalize to other similar multi-contrast and multi-institutional settings.

Regarding datasets, a limited number of studies have proposed innovative approaches to address the constraints posed by the scarcity of data, such as generating semi-synthetic data (Liu et al., 2020a). This represents another potential research direction aimed at mitigating model performance degradation due to insufficient data. Furthermore, employing multiple image sequences from different institutions could enhance the generalizability of the synthesis method. Concerning validating synthesized results, certain areas in current studies warrant increased attention in future research. It would be more appropriate to consistently validate results using unseen data and ensure the proposed method's generalizability in various settings, thereby establishing its credibility in clinical applications.

There is a need for clinical and multicentre validation of synthetic images. Synthetic outputs have been employed for assessing their feasibility in different clinical applications such as radiotherapy treatment planning and diagnosis of neurological diseases (Li et al., 2019a; Lin et al., 2019). CT synthesis is widely examined in MRgRT to obtain electron density details for dose calculation. This can be further sorted by synthesizing CT images in low-field MR-guided radiotherapy and treatment planning for diseases such as nasopharyngeal carcinoma, intracranial tumour, prostate cancer, cervical cancer, children with pelvic sarcomas and paediatric cranial bone imaging. Pseudo-CT in MRgRT has also been utilized for specific regions such as lung, pelvic, head and neck, brain, liver, and abdomen cases. Synthetic PET images from MRI are widely used for classifying neurodegenerative diseases such as Alzheimer's, and high-dose synthetic PET images for cancer diagnosis. Synthetic MR images are mainly useful in combining different tissue contrasts for the diagnostic process and in cases like multiple sclerosis, where longitudinal comparisons of MR images and consistent modalities at different time points are necessary (Li et al., 2019b). Nonetheless, despite each study presenting a unique approach with novel features resulting in more successful deep-learning-based medical image synthesis, there remain challenges and open questions to be addressed for future clinical evaluation, especially with multicentre studies.

8. Conclusion

This review comprehensively examines the recent literature on medical image synthesis. We have assessed various medical image synthesis techniques based on deep learning approaches for pseudo-CT, MR, and PET synthesis. The analysis commences with an overview of the most commonly employed deep learning-based networks in medical image synthesis and synthetic image contrasts. The review's core analysis emphasizes diverse deep learning approaches based on input domains and network architectures. We further analyse the network models including the most recent generative architectures from the literature and their loss functions, dataset overviews, and evaluation results. Furthermore, we have identified the challenges and frequently employed solutions from the studied literature and potential future research directions. Additionally, we have provided a summary of studies on different image synthesis methods by elucidating their unique features. Ultimately, we hope this review aids in developing more accurate and innovative medical image synthesis techniques through the extensive analysis provided within the literature.

Declaration of competing interest

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

Data availability

No data was used for the research described in the article.

Acknowledgements

ZC is funded by the Australian Research Council Linkage Program (LP170100494), Australian Research Council Discovery Program (DP210101863), and Australian Research Council Industry Fellowship Program (IM230100002). GY is supported in part by the ERC IMI (101005122), the H2020 (952172), the Medical Research Council (MRC) (MC/PC/21013), the Royal Society, United Kingdom (IEC\NSFC\211235), and the UKRI Future Leaders Fellowship (MR/V023799/1).

References

Abu-Srhan, A., Almallahi, I., Abushariah, M.A., Mahafza, W., Al-Kadi, O.S., 2021. Paired-unpaired unsupervised attention guided GAN with transfer learning for bidirectional brain MR-CT synthesis. Comput. Biol. Med. 136, 104763. http://dx.doi.org/10.1016/j.compbiomed.2021.104763.

Almahairi, A., Rajeswar, S., Sordoni, A., Bachman, P., Courville, A., 2018. Augmented cyclegan: Learning many-to-many mappings from unpaired data. URL: https://arxiv.org/abs/1802.10151.

Ang, S.P., Phung, S.L., Field, M., Schira, M.M., 2022. An improved deep learning framework for MR-to-CT image synthesis with a new hybrid objective function. In: 2022 IEEE 19th International Symposium on Biomedical Imaging. ISBI, IEEE, http://dx.doi.org/10.1109/isbi52829.2022.9761546.

Arabi, H., Zaidi, H., 2021. Identification of noisy labels in deep learning-based synthetic CT generation from MR images. In: 2021 IEEE Nuclear Science Symposium and Medical Imaging Conference. NSS/MIC, IEEE, http://dx.doi.org/10.1109/nss/mic44867.2021.9875547.

Armanious, K., Jiang, C., Abdulatif, S., Kustner, T., Gatidis, S., Yang, B., 2019. Unsupervised medical image translation using cycle-MedGAN. In: 2019 27th European Signal Processing Conference. EUSIPCO, IEEE.

Armanious, K., Jiang, C., Fischer, M., Küstner, T., Nikolaou, K., Gatidis, S., Yang, B., 2020. MedGAN: Medical image translation using GANs. Comput. Med. Imaging Graph. 79, 101684. http://dx.doi.org/10.1016/j.compmedimag.2019.101684.

Babayan, A., Erbey, M., Kumral, D., Reinelt, J.D., Reiter, A.M.F., R., J., et al., 2019. A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults. Sci. Data 6, http://dx.doi.org/10.1038/sdata.2018.308.

Bagheri, F., Uludag, K., 2023. Mr image prediction at high field strength from mr images taken at low field strength using multi-to-one translation. In: CMBES Proceedings, Vol. 45.

Bahrami, A., Karimian, A., Fatemizadeh, E., Arabi, H., Zaidi, H., 2019. A novel convolutional neural network with high convergence rate: Application to CT synthesis from MR images. In: 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference. NSS/MIC, IEEE, http://dx.doi.org/10.1109/nss/mic42101.2019.9059908.

Bahrami, A., Karimian, A., Fatemizadeh, E., Arabi, H., Zaidi, H., 2020. A new deep convolutional neural network design with efficient learning capability: Application to CT image synthesis from MRI. Med. Phys. 47, 5158–5171. http://dx.doi.org/10.1002/mp.14418.

Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R., Berger, C., Ha, S., Rozycki, M., et al., 2018. Identifying the best machine learning algorithms for brain tumor segmentation. Prog. Assess. Overall Surviv. Predict. BRATS Chall. 10, URL: https://arxiv.org/abs/1811.02629.

Baran, J., Chen, Z., Sforazzini, F., Ferris, N., Jamadar, S., Schmitt, B., Faul, D., Shah, N.J., Cholewa, M., Egan, G.F., 2018. Accurate hybrid template–based and MR-based attenuation correction using UTE images for simultaneous PET/MR brain imaging applications. BMC Med. Imaging 18, http://dx.doi.org/10.1186/s12880-018-0283-3.

Baydoun, A., Xu, K., Heo, J.U., Yang, H., Zhou, F., Bethell, L.A., Fredman, E.T., Ellis, R.J., Podder, T.K., Traughber, M.S., Paspulati, R.M., Qian, P., Traughber, B.J., Muzic, R.F., 2021. Synthetic CT generation of the pelvis in patients with cervical cancer: A single input approach using generative adversarial network. IEEE Access 9, 17208–17221. http://dx.doi.org/10.1109/access.2021.3049781.

Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., C., I., et al., 2018. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? IEEE Trans. Med. Imaging 37, 2514–2525. http://dx.doi.org/10.1109/tmi.2018.2837502.

Boni, K.N.D.B., Klein, J., Gulyban, A., Reynaert, N., Pasquier, D., 2021. Improving generalization in MR-to-CT synthesis in radiotherapy by using an augmented cycle generative adversarial network with unpaired data. Med. Phys. 48, 3003–3010. http://dx.doi.org/10.1002/mp.14866.

Boni, K.N.D.B., Klein, J., Vanquin, L., Wagner, A., Lacornerie, T., Pasquier, D., Reynaert, N., 2020. MR to CT synthesis with multicenter data in the pelvic area using a conditional generative adversarial network. Phys. Med. Biol. 65, 075002. http://dx.doi.org/10.1088/1361-6560/ab7633.

Boroojeni, P.E., Chen, Y., Commean, P.K., Eldeniz, C., Skolnick, G.B., Merrill, C., Patel, K.B., An, H., 2022. Deep-learning synthesized pseudo-CT for MR high-resolution pediatric cranial bone imaging (MR-HiPCB). Magn. Reson. Med. 88, 2285–2297. http://dx.doi.org/10.1002/mrm.29356.

Bui, T.D., Nguyen, M., Le, N., Luu, K., 2020. Flow-based deformation guidance for unpaired multi-contrast MRI image-to-image translation. In: Medical Image Computing and Computer Assisted Intervention –MICCAI 2020. Springer International Publishing, pp. 728–737. http://dx.doi.org/10.1007/978-3-030-59713-9_70.

Bullitt, E., Zeng, D., Gerig, G., Aylward, S., Joshi, S., Smith, J.K., Lin, W., Ewend, M.G., 2005. Vessel tortuosity and brain tumor malignancy. Academic Radiol. 12, 1232–1240. http://dx.doi.org/10.1016/j.acra.2005.05.027.

Cao, G., Liu, S., Mao, H., Zhang, S., 2021. Improved CyeleGAN for MR to CT synthesis. In: 2021 6th International Conference on Intelligent Informatics and Biomedical Sciences. ICIIBMS, IEEE, http://dx.doi.org/10.1109/iciibms52876.2021.9651571.

Chartsias, A., Joyce, T., Giuffrida, M.V., Tsaftaris, S.A., 2018. Multimodal MR synthesis via modality-invariant latent representation. IEEE Trans. Med. Imaging 37, 803–814. http://dx.doi.org/10.1109/tmi.2017.2764326.

Chen, K.T., Gong, E., de Carvalho Macruz, F.B., Xu, J., Boumis, A., Khalighi, M., Poston, K.L., Sha, S.J., Greicius, M.D., Mormino, E., Pauly, J.M., Srinivas, S., Zaharchuk, G., 2019. Ultra–low-dose ^{18} F-florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs. Radiology 290, 649–656. http://dx.doi.org/10.1148/radiol.2018180940.

Chen, Z., Jamadar, S.D., Li, S., Sforazzini, F., Baran, J., Ferris, N., Shah, N.J., Egan, G.F., 2018b. From simultaneous to synergistic MR-PET brain imaging: A review of hybrid MR-PET imaging methodologies. Hum. Brain Mapp. 39, 5126–5144. http://dx.doi.org/10.1002/hbm.24314.

Chen, Z., Pawar, K., Ekanayake, M., Pain, C., Zhong, S., Egan, G.F., 2022. Deep learning for image enhancement and correction in magnetic resonance imaging—state-of-the-art and challenges. J. Digit. Imaging 36, 204–230. http://dx.doi.org/10.1007/s10278-022-00721-9.

Chen, S., Qin, A., Zhou, D., Yan, D., 2018a. Technical note: U-net-generated synthetic CT images for magnetic resonance imaging-only prostate intensity-modulated radiation therapy treatment planning. Med. Phys. 45, 5659–5665. http://dx.doi.org/10.1002/mp.13247.

Chen, Y., Staring, M., Wolterink, J.M., Tao, Q., 2023. Local implicit neural representations for multi-sequence mri translation. URL: https://arxiv.org/abs/2302.01031.

Chen, J., Wei, J., Li, R., 2021. Targan: Target-aware generative adversarial networks for multi-modality medical image translation. URL: https://arxiv.org/abs/2105.08993.

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

Dai, X., Lei, Y., Fu, Y., Curran, W.J., Liu, T., Mao, H., Yang, X., 2020. Multimodal MRI synthesis using unified generative adversarial networks. Med. Phys. 47, 6343–6354. http://dx.doi.org/10.1002/mp.14539.

Dalmaz, O., Yurt, M., Cukur, T., 2022. ResViT: Residual vision transformers for multimodal medical image synthesis. IEEE Trans. Med. Imaging 41, 2598–2614. http://dx.doi.org/10.1109/tmi.2022.3167808.

Dar, S.U., Yurt, M., Karacan, L., Erdem, A., Erdem, E., Cukur, T., 2019. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. IEEE Trans. Med. Imaging 38, 2375–2388. http://dx.doi.org/10.1109/tmi.2019.2901750.

Dhariwal, P., Nichol, A., 2021. Diffusion models beat gans on image synthesis. URL: https://arxiv.org/abs/2105.05233.

Dong, X., Wang, T., Lei, Y., Higgins, K., Liu, T., Curran, W.J., Mao, H., Nye, J.A., Yang, X., 2019. Synthetic CT generation from non-attenuation corrected PET images for whole-body PET imaging. Phys. Med. Biol. 64, 215016. http://dx.doi.org/10.1088/1361-6560/ab4eb7.

Dovletov, G., Lörcks, S., Pauli, J., Gratz, M., Quick, H.H., 2023. Double grad-CAM guidance for improved MRI-based pseudo-CT synthesis. In: Informatik Aktuell. Springer Fachmedien Wiesbaden, pp. 45–50. http://dx.doi.org/10.1007/978-3-658-41657-7_13.

Dovletov, G., Pham, D.D., Lorcks, S., Pauli, J., Gratz, M., Quick, H.H., 2022. Grad-CAM guided u-net for MRI-based pseudo-CT synthesis. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society. EMBC, IEEE, http://dx.doi.org/10.1109/embc48229.2022.9871994.

Dutta, K., Liu, Z., Laforest, R., Jha, A., Shoghi, K.I., 2022. Deep learning framework to synthesize high-count preclinical PET images from low-count preclinical PET images. In: Zhao, W., Yu, L. (Eds.), Medical Imaging 2022: Physics of Medical Imaging. SPIE, http://dx.doi.org/10.1117/12.2612729.

Ehrhardt, J., Jacob, F., Handels, H., Frydrychowicz, A., 2016. Comparison of post-hoc normalization approaches for CT-based lung emphysema index quantification. In: Informatik Aktuell. Springer Berlin Heidelberg, pp. 44–49. http://dx.doi.org/10.1007/978-3-662-49465-3_10.

Emami, H., Dong, M., Glide-Hurst, C.K., 2020a. Attention-guided generative adversarial network to address atypical anatomy in synthetic CT generation. In: 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science. IRI, IEEE, http://dx.doi.org/10.1109/iri49571.2020.00034.

Emami, H., Dong, M., Nejad-Davarani, S.P., Glide-Hurst, C.K., 2018. Generating synthetic CTs from magnetic resonance images using generative adversarial networks. Med. Phys. 45, 3627–3636. http://dx.doi.org/10.1002/mp.13047.

Emami, H., Liu, Q., Dong, M., 2020b. Frea-unet: Frequency-aware u-net for modality transfer. URL: https://arxiv.org/abs/2012.15397.

Essen, D.C.V., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., 2013. The WU-minn human connectome project: An overview. NeuroImage 80, 62–79. http://dx.doi.org/10.1016/j.neuroimage.2013.05.041.

Estakhraji, S.I.Z., Pirasteh, A., Bradshaw, T., McMillan, A., 2023. On the effect of training database size for MR-based synthetic CT generation in the head. Comput. Med. Imaging Graph. 107, 102227. http://dx.doi.org/10.1016/j.compmedimag.2023.102227.

Fan, Y., Khan, M.M.R., Liu, H., Noble, J.H., Labadie, R.F., Dawant, B.M., 2023. Temporal bone CT synthesis for MR-only cochlear implant preoperative planning. In: Linte, C.A., Siewerdsen, J.H. (Eds.), Medical Imaging 2023: Image-Guided Procedures, Robotic Interventions, and Modeling. SPIE, http://dx.doi.org/10.1117/12.2647443.

Fard, A.S., Reutens, D.C., Vegh, V., 2022. From CNNs to GANs for cross-modality medical image estimation. Comput. Biol. Med. 146, 105556. http://dx.doi.org/10.1016/j.compbiomed.2022.105556.

Fei, Y., Zu, C., Jiao, Z., Wu, X., Zhou, J., Shen, D., Wang, Y., 2022. Classification-aided high-quality PET image synthesis via bidirectional contrastive GAN with shared information maximization. In: Lecture Notes in Computer Science. Springer Nature Switzerland, pp. 527–537. http://dx.doi.org/10.1007/978-3-031-16446-0_50.

Figini, M., Lin, H., Ogbole, G., Arco, F.D., Blumberg, S.B., Carmichael, D.W., Tanno, R., Kaden, E., Brown, B.J., Lagunju, I., Cross, H.J., Fernandez-Reyes, D., Alexander, D.C., 2020. Image quality transfer enhances contrast and resolution of low-field brain mri in african paediatric epilepsy patients. URL: https://arxiv.org/abs/2003.07216.

Florkow, M.C., Zijlstra, F., Willemsen, K., Maspero, M., van den Berg, C.A., Kerkmeijer, L.G., Castelein, R.M., Weinans, H., Viergever, M.A., van Stralen, M., et al., 2020. Deep learning–based mr-to-ct synthesis: the influence of varying gradient echo–based mr images as input channels. Magn. Reson. Med. 83, 1429–1441. http://dx.doi.org/10.1002/mrm.28008.

Fu, J., Singhrao, K., Cao, M., Yu, V., Santhanam, A.P., Yang, Y., Guo, M., Raldow, A.C., Ruan, D., Lewis, J.H., 2020. Generation of abdominal synthetic CTs from 0.35t MR images using generative adversarial networks for MR-only liver radiotherapy. Biomed. Phys. Eng. Express 6, 015033. http://dx.doi.org/10.1088/2057-1976/ab6e1f.

Fu, J., Yang, Y., Singhrao, K., Ruan, D., Chu, F.I., Low, D.A., Lewis, J.H., 2019. Deep learning approaches using 2d and 3d convolutional neural networks for generating male pelvic synthetic computed tomography from magnetic resonance imaging. Med. Phys. 46, 3788–3798. http://dx.doi.org/10.1002/mp.13672.

Gatidis, S., Hepp, T., Früh, M., Fougèe, C.L., Nikolaou, K., Pfannenberg, C., Schölkopf, B., Küstner, T., Cyran, C., Rubin, D., 2022. A whole-body FDG-PET/CT dataset with manually annotated tumor lesions. Sci. Data 9, http://dx.doi.org/10.1038/s41597-022-01718-3.

Ge, Y., Wei, D., Xue, Z., Wang, Q., Zhou, X., Zhan, Y., Liao, S., 2019. Unpaired mr to CT synthesis with explicit structural constrained adversarial learning. In: 2019 IEEE 16th International Symposium on Biomedical Imaging. ISBI 2019, IEEE, http://dx.doi.org/10.1109/isbi.2019.8759529.

Grover, A., Chute, C., Shu, R., Cao, Z., Ermon, S., 2019. Alignflow: Cycle consistent learning from multiple domains via normalizing flows. URL: https://arxiv.org/abs/1905.12892.

Häggström, I., Schmidtlein, C.R., Campanella, G., Fuchs, T.J., 2019. DeepPET: A deep encoder–decoder network for directly solving the PET image reconstruction inverse problem. Med. Image Anal. 54, 253–262. http://dx.doi.org/10.1016/j.media.2019.03.013.

Hagler, D.J., Hatton, S., Cornejo, M.D., Makowski, C., Fair, D.A., D., A.S., et al., 2019. Image processing and analysis methods for the adolescent brain cognitive development study. NeuroImage 202, 116091. http://dx.doi.org/10.1016/j.neuroimage.2019.116091.

Hemsley, M., Chugh, B., Ruschin, M., Lee, Y., Tseng, C.L., Stanisz, G., Lau, A., 2020. Deep generative model for synthetic-CT generation with uncertainty predictions. In: Medical Image Computing and Computer Assisted Intervention –MICCAI 2020. Springer International Publishing, pp. 834–844. http://dx.doi.org/10.1007/978-3-030-59710-8_81.

Hiasa, Y., Otake, Y., Takao, M., Matsuoka, T., Takashima, K., Carass, A., Prince, J.L., Sugano, N., Sato, Y., 2018. Cross-modality image synthesis from unpaired data using CycleGAN. In: Simulation and Synthesis in Medical Imaging. Springer International Publishing, pp. 31–41. http://dx.doi.org/10.1007/978-3-030-00536-8_4.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. URL: https://arxiv.org/abs/2006.11239.

Ho, J., Salimans, T., 2022. Classifier-free diffusion guidance. URL: https://arxiv.org/abs/2207.12598.

Howell, B.R., Styner, M.A., Gao, W., Yap, P.T., Wang, L., Baluyot, K., Yacoub, E., Chen, G., Potts, T., Salzwedel, A., Li, G., Gilmore, J.H., Piven, J., Smith, J.K., Shen, D., Ugurbil, K., Zhu, H., Lin, W., Elison, J.T., 2019. The UNC/UMN baby connectome project (BCP): An overview of the study design and protocol development. NeuroImage 185, 891–905. http://dx.doi.org/10.1016/j.neuroimage.2018.03.049.

Hsu, S.H., Han, Z., Leeman, J.E., Hu, Y.H., Mak, R.H., Sudhyadhom, A., 2022. Synthetic CT generation for MRI-guided adaptive radiotherapy in prostate cancer. Front. Oncol. 12, http://dx.doi.org/10.3389/fonc.2022.969463.

Hu, S., Lei, B., Wang, S., Wang, Y., Feng, Z., Shen, Y., 2021. Bidirectional mapping generative adversarial networks for brain mr to pet synthesis. IEEE Trans. Med. Imaging 41, 145–157.

Hu, Q., Li, H., Zhang, J., 2022. Domain-adaptive 3d medical image synthesis: An efficient unsupervised approach. URL: https://arxiv.org/abs/2207.00844.

Hu, R., Liu, H., 2022. Transem:residual swin-transformer based regularized pet image reconstruction. URL: https://arxiv.org/abs/2205.04204.

Hu, S., Yuan, J., Wang, S., 2019. Cross-modality synthesis from MRI to PET using adversarial u-net with different normalization. In: 2019 International Conference on Medical Imaging Physics and Engineering. ICMIPE, IEEE, http://dx.doi.org/10.1109/icmipe47306.2019.9098219.

Hussein, R., Zhao, M.Y., Shin, D., Guo, J., Chen, K.T., Armindo, R.D., Davidzon, G., Moseley, M., Zaharchuk, G., 2022. Multi-task deep learning for cerebrovascular disease classification and MRI-to-PET translation. In: 2022 26th International Conference on Pattern Recognition. ICPR, IEEE, http://dx.doi.org/10.1109/icpr56361.2022.9956549.

Hwang, D., Kang, S.K., Kim, K.Y., Seo, S., Paeng, J.C., Lee, D.S., Lee, J.S., 2019. Generation of PET Attenuation Map for Whole-Body Time-of-Flight ^{18} F-FDG PET/MRI Using a Deep Neural Network Trained with Simultaneously Reconstructed Activity and Attenuation Maps. J. Nucl. Med. 60, 1183–1189. http://dx.doi.org/10.2967/jnumed.118.219493.

Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., H., D., et al., 2008. The alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn. Reson. Imaging 27, 685–691. http://dx.doi.org/10.1002/jmri.21049.

Jang, S.I., Lois, C., Thibault, E., Becker, J.A., Dong, Y., Normandin, M.D., Price, J.C., Johnson, K.A., Fakhri, G.E., Gong, K., 2023. Taupetgen: Text-conditional tau pet image synthesis based on latent diffusion models. URL: https://arxiv.org/abs/2306.11984.

Jans, L.B.O., Chen, M., Elewaut, D., den Bosch, F.V., Carron, P., Jacques, P., Wittoek, R., Jaremko, J.L., Herregods, N., 2021. MRI-based synthetic CT in the detection of structural lesions in patients with suspected sacroiliitis: Comparison with MRI. Radiology 298, 343–349. http://dx.doi.org/10.1148/radiol.2020201537.

Jiang, L., Mao, Y., Chen, X., Wang, X., Li, C., 2023. Cola-diff: Conditional latent diffusion model for multi-modal mri synthesis. URL: https://arxiv.org/abs/2303.14081.

Jiangtao, W., Xinhong, W., Xiao, J., Bing, Y., Lei, Z., Yidong, Y., 2021. MRI to CT synthesis using contrastive learning. In: 2021 IEEE International Conference on Medical Imaging Physics and Engineering. ICMIPE, IEEE, http://dx.doi.org/10.1109/icmipe53131.2021.9698888.

Jung, M.M., van den Berg, B., Postma, E., Huijbers, W., 2018. Inferring pet from mri with pix2pix. In: Benelux Conference on Artificial Intelligence. URL: https://research.tue.nl/en/publications/inferring-pet-from-mri-with-pix2pix.

Kang, S.K., An, H.J., Jin, H., Kim, J.in., Chie, E.K., Park, J.M., Lee, J.S., 2021. Synthetic CT generation from weakly paired MR images using cycle-consistent GAN for MR-guided radiotherapy. Biomed. Eng. Lett. 11, 263–271. http://dx.doi.org/10.1007/s13534-021-00195-8.

Kaplan, S., Zhu, Y.M., 2019. Full-dose pet image estimation from low-dose pet image using deep learning: a pilot study. J. Digit. Imaging 32, 773–778.

Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., G., V., et al., 2021. CHAOS challenge - combined (CT-MR) healthy abdominal organ segmentation. Med. Image Anal. 69, 101950. http://dx.doi.org/10.1016/j.media.2020.101950.

Kawahara, D., Yoshimura, H., Matsuura, T., Saito, A., Nagata, Y., 2023. MRI image synthesis for fluid-attenuated inversion recovery and diffusion-weighted images with deep learning. Phys. Eng. Sci. Med. 46, 313–323. http://dx.doi.org/10.1007/s13246-023-01220-z.

Kazemifar, S., McGuire, S., Timmerman, R., Wardak, Z., Nguyen, D., Park, Y., Jiang, S., Owrangi, A., 2019. MRI-only brain radiotherapy: Assessing the dosimetric accuracy of synthetic CT images generated using a deep learning approach. Radiother. Oncol. 136, 56–63. http://dx.doi.org/10.1016/j.radonc.2019.03.026.

Kazerouni, A., Aghdam, E.K., Heidari, M., Azad, R., Fayyaz, M., Hacihaliloglu, I., Merhof, D., 2022. Diffusion models for medical image analysis: A comprehensive survey. URL: https://arxiv.org/abs/2211.07804.

Kearney, V., Ziemer, B.P., Perry, A., Wang, T., Chan, J.W., Ma, L., Morin, O., Yom, S.S., Solberg, T.D., 2020. Attention-aware discrimination for MR-to-CT image translation using cycle-consistent generative adversarial networks. Radiology: Artif. Intell. 2, e190027. http://dx.doi.org/10.1148/ryai.2020190027.

Khader, F., Müller-Franzes, G., Arasteh, S.T., Han, T., Haarburger, C., Schulze-Hagen, M., Schad, P., Engelhardt, S., Baeßler, B., Foersch, S., Stegmaier, J., Kuhl, C., Nebelung, S., Kather, J.N., Truhn, D., 2023. Denoising diffusion probabilistic models for 3d medical image generation. Sci. Rep. 13, http://dx.doi.org/10.1038/s41598-023-34341-2.

Kingma, D.P., Welling, M., 2013. Auto-encoding variational bayes. URL: https://arxiv.org/abs/1312.6114.

Klages, P., Benslimane, I., Riyahi, S., Jiang, J., Hunt, M., Deasy, J.O., Veeraraghavan, H., Tyagi, N., 2019. Patch-based generative adversarial neural network models for head and neck MR-only planning. Med. Phys. 47, 626–642. http://dx.doi.org/10.1002/mp.13927.

Kläser, K., Markiewicz, P., Ranzini, M., Li, W., Modat, M., Hutton, B.F., Atkinson, D., Thielemans, K., Cardoso, M.J., Ourselin, S., 2018. Deep boosted regression for MR to CT synthesis. In: Simulation and Synthesis in Medical Imaging. Springer International Publishing, pp. 61–70. http://dx.doi.org/10.1007/978-3-030-00536-8_7.

Koh, H., Park, T.Y., Chung, Y.A., Lee, J.H., Kim, H., 2022. Acoustic simulation for transcranial focused ultrasound using GAN-based synthetic CT. IEEE J. Biomed. Health Inf. 26, 161–171. http://dx.doi.org/10.1109/jbhi.2021.3103387.

Kong, L., Lian, C., Huang, D., Li, Z., Hu, Y., Zhou, Q., 2021. Breaking the dilemma of medical image-to-image translation. In: Neural Information Processing Systems.

Kuijf, H.J., Casamitjana, A., Collins, D.L., Dadar, M., Georgiou, A., G., M., et al., 2019. Standardized assessment of automatic segmentation of white matter hyperintensities and results of the WMH segmentation challenge. IEEE Trans. Med. Imaging 38, 2556–2568. http://dx.doi.org/10.1109/tmi.2019.2905770.

Lapaeva, M., Saint-Esteven, A.L.G., Wallimann, P., Günther, M., Konukoglu, E., Andratschke, N., Guckenberger, M., Tanadini-Lang, S., Bello, R.D., 2022. Synthetic computed tomography for low-field magnetic resonance-guided radiotherapy in the abdomen. Phys. Imaging Radiat. Oncol. 24, 173–179. http://dx.doi.org/10.1016/j.phro.2022.11.011.

Lei, Y., Dong, X., Wang, T., Higgins, K., Liu, T., Curran, W.J., Mao, H., Nye, J.A., Yang, X., 2019a. Whole-body PET estimation from low count statistics using cycle-consistent generative adversarial networks. Phys. Med. Biol. 64, 215017. http://dx.doi.org/10.1088/1361-6560/ab4891.

Lei, Y., Dong, X., Wang, T., Higgins, K., Liu, T., Curran, W.J., Mao, H., Nye, J.A., Yang, X., 2020. Estimating standard-dose PET from low-dose PET with deep learning. In: Landman, B.A., Išum, I. (Eds.), Medical Imaging 2020: Image Processing. SPIE, http://dx.doi.org/10.1117/12.2548461.

Lei, Y., Harms, J., Wang, T., Liu, Y., Shu, H.K., Jani, A.B., Curran, W.J., Mao, H., Liu, T., Yang, X., 2019b. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. Med. Phys. 46, 3565–3581. http://dx.doi.org/10.1002/mp.13617.

Lenkowicz, J., Votta, C., Nardini, M., Quaranta, F., Catucci, F., Boldrini, L., Vagni, M., Menna, S., Placidi, L., Romano, A., Chiloiro, G., Gambacorta, M.A., Mattiucci, G.C., Indovina, L., Valentini, V., Cusumano, D., 2022. A deep learning approach to generate synthetic CT in low field MR-guided radiotherapy for lung cases. Radiother. Oncol. 176, 31–38. http://dx.doi.org/10.1016/j.radonc.2022.08.028.

Li, G., Bai, L., Zhu, C., Wu, E., Ma, R., 2018. A novel method of synthetic CT generation from MR images based on convolutional neural networks. In: 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics. CISP-BMEI, IEEE, http://dx.doi.org/10.1109/cisp-bmei.2018.8633142.

Li, X., Jiang, Y., Rodriguez-Andina, J.J., Luo, H., Yin, S., Kaynak, O., 2021b. When medical images meet generative adversarial network: recent development and research opportunities. Discover Artif. Intell. 1, http://dx.doi.org/10.1007/s44163-021-00006-0.

Li, W., Kazemifar, S., Bai, T., Nguyen, D., Weng, Y., Li, Y., Xia, J., Xiong, J., Xie, Y., Owrangi, A., Jiang, S., 2021a. Synthesizing CT images from MR images with deep learning: model generalization for different datasets through transfer learning. Biomed. Phys. Eng. Express 7, 025020. http://dx.doi.org/10.1088/2057-1976/abe3a7.

Li, Y., Li, W., He, P., Xiong, J., Xia, J., Xie, Y., 2019a. CT synthesis from MRI images based on deep learning methods for MRI-only radiotherapy. In: 2019 International Conference on Medical Imaging Physics and Engineering. ICMIPE, IEEE, http://dx.doi.org/10.1109/icmipe47306.2019.9098190.

Li, H., Paetzold, J.C., Sekuboyina, A., Kofler, F., Zhang, J., Kirschke, J.S., Wiestler, B., Menze, B., 2019b. DiamondGAN: Unified multi-modal generative adversarial networks for MRI sequences synthesis. In: Lecture Notes in Computer Science. Springer International Publishing, pp. 795–803. http://dx.doi.org/10.1007/978-3-030-32251-9_87.

Li, J., Qu, Z., Yang, Y., Zhang, F., Li, M., Hu, S., 2022a. TCGAN: a transformer-enhanced GAN for PET synthetic CT. Biomed. Opt. Express 13 (6003), http://dx.doi.org/10.1364/boe.467683.

Li, X., Shang, K., Wang, G., Butala, M.D., 2023a. Ddmm-synth: A denoising diffusion model for cross-modal medical image synthesis with sparse-view measurement embedding. URL: https://arxiv.org/abs/2303.15770.

Li, Y., Wei, J., Qi, Z., Sun, Y., Lu, Y., 2020. Synthesize CT from paired MRI of the same patient with patch-based generative adversarial network. In: Hahn, H.K., Mazurowski, M.A. (Eds.), Medical Imaging 2020: Computer-Aided Diagnosis. SPIE, http://dx.doi.org/10.1117/12.2551285.

Li, Y., Xu, S., Chen, H., Sun, Y., Bian, J., Guo, S., Lu, Y., Qi, Z., 2023b. CT synthesis from multi-sequence MRI using adaptive fusion network. Comput. Biol. Med. 157, 106738. http://dx.doi.org/10.1016/j.compbiomed.2023.106738.

Li, Y., Xu, S., Lu, Y., Qi, Z., 2023c. CT synthesis from MRI with an improved multi-scale learning network. Front. Phys. 11, http://dx.doi.org/10.3389/fphy.2023.1088899.

Li, Y., Zhou, T., He, K., Zhou, Y., Shen, D., 2022. Multi-scale transformer network with edge-aware pre-training for cross-modality mr image synthesis. URL: https://arxiv.org/abs/2212.01108.

Lin, H., Figini, M., D'Arco, F., Ogbole, G., Tanno, R., Blumberg, S.B., Ronan, L., Brown, B.J., Carmichael, D.W., Lagunju, I., Cross, J.H., Fernandez-Reyes, D., Alexander, D.C., 2023. Low-field magnetic resonance image enhancement via stochastic image quality transfer. Med. Image Anal. 87, 102807. http://dx.doi.org/10.1016/j.media.2023.102807.

Lin, H., Figini, M., Tanno, R., Blumberg, S.B., Kaden, E., O., G., et al., 2019. Deep learning for low-field to high-field MR: Image quality transfer with probabilistic decimation simulator. In: Machine Learning for Medical Image Reconstruction. Springer International Publishing, pp. 58–70. http://dx.doi.org/10.1007/978-3-030-33843-5_6.

Lin, Y., Han, H., Zhou, S.K., 2022. Deep non-linear embedding deformation network for cross-modal brain MRI synthesis. In: 2022 IEEE 19th International Symposium on Biomedical Imaging. ISBI, IEEE, http://dx.doi.org/10.1109/isbi52829.2022.9761711.

Lin, W., Lin, W., Chen, G., Zhang, H., Gao, Q., Huang, Y., Tong, T., D., M., 2021. Bidirectional mapping of brain MRI and PET with 3d reversible GAN for the diagnosis of alzheimer's disease. Front. Neurosci. 15, http://dx.doi.org/10.3389/fnins.2021.646013.

Liu, Y., Chen, A., Shi, H., Huang, S., Zheng, W., Liu, Z., Zhang, Q., Yang, X., 2021a. CT synthesis from MRI using multi-cycle GAN for head-and-neck radiation therapy. Comput. Med. Imaging Graph. 91, 101953. http://dx.doi.org/10.1016/j.compmedimag.2021.101953.

Liu, F., Jang, H., Kijowski, R., Zhao, G., Bradshaw, T., McMillan, A.B., 2018. A deep learning approach for 18f-FDG PET attenuation correction. EJNMMI Phys. 5, http://dx.doi.org/10.1186/s40658-018-0225-8.

Liu, L., Johansson, A., Cao, Y., Dow, J., Lawrence, T.S., Balter, J.M., 2020a. Abdominal synthetic CT generation from MR dixon images using a u-net trained with 'semi-synthetic' CT data. Phys. Med. Biol. 65, 125001. http://dx.doi.org/10.1088/1361-6560/ab8cd2.

Liu, Y., Lei, Y., Wang, T., Zhou, J., Lin, L., Liu, T., Patel, P., Curran, W.J., Ren, L., Yang, X., 2020b. Liver synthetic CT generation based on a dense-CycleGAN for MRI-only treatment planning. In: Landman, B.A., Išgum, I. (Eds.), Medical Imaging 2020: Image Processing. SPIE, http://dx.doi.org/10.1117/12.2549265.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021b. Swin transformer: Hierarchical vision transformer using shifted windows. URL: https://arxiv.org/abs/2103.14030.

Liu, J., Pasumarthi, S., Duffy, B., Gong, E., Datta, K., Zaharchuk, G., 2023. One model to synthesize them all: Multi-contrast multi-scale transformer for missing data imputation. IEEE Trans. Med. Imaging 1. http://dx.doi.org/10.1109/tmi.2023.3261707.

Liu, X., Xing, F., Prince, J.L., Carass, A., Stone, M., Fakhri, G.E., Woo, J., 2021c. Dual-cycle constrained bijective vae-gan for tagged-to-cine magnetic resonance image synthesis. In: 2021 IEEE 18th International Symposium on Biomedical Imaging. ISBI, IEEE, http://dx.doi.org/10.1109/isbi48211.2021.9433852.

Lu, W., Onofrey, J.A., Lu, Y., Shi, L., Ma, T., Liu, Y., Liu, C., 2019. An investigation of quantitative accuracy for deep learning based denoising in oncological PET. Phys. Med. Biol. 64, 165019. http://dx.doi.org/10.1088/1361-6560/ab3242.

Luo, Y., Wang, Y., Zu, C., Zhan, B., Wu, X., Zhou, J., Shen, D., Zhou, L., 2021. 3D transformer-GAN for high-quality PET reconstruction. In: Medical Image Computing and Computer Assisted Intervention –MICCAI 2021. Springer International Publishing, pp. 276–285. http://dx.doi.org/10.1007/978-3-030-87231-1_27.

Lyu, Q., Wang, G., 2022. Conversion between ct and mri images using diffusion and score-matching models. URL: https://arxiv.org/abs/2209.12104.

Maier, O., Menze, B.H., von der Gablentz, J., Häni, L., Heinrich, M.P., L., M., et al., 2017. ISLES 2015 - a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Med. Image Anal. 35, 250–269. http://dx.doi.org/10.1016/j.media.2016.07.009.

Makropoulos, A., Robinson, E.C., Schuh, A., Wright, R., Fitzgibbon, S., B., J., et al., 2018. The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. NeuroImage 173, 88–112. http://dx.doi.org/10.1016/j.neuroimage.2018.01.054.

Mao, Y., Chen, C., Wang, Z., Cheng, D., You, P., Huang, X., Zhang, B., Zhao, F., 2022. Generative adversarial networks with adaptive normalization for synthesizing t2-weighted magnetic resonance images from diffusion-weighted images. Front. Neurosci. 16, http://dx.doi.org/10.3389/fnins.2022.1058487.

Maspero, M., Bentvelzen, L.G., Savenije, M.H., Guerreiro, F., Seravalli, E., Janssens, G.O., van den Berg, C.A., Philippens, M.E., 2020. Deep learning-based synthetic CT generation for paediatric brain MR-only photon and proton radiotherapy. Radiother. Oncol. 153, 197–204. http://dx.doi.org/10.1016/j.radonc.2020.09.029.

Mendes, J., Pereira, T., Silva, F., Frade, J., Morgado, J., Freitas, C., Negrã, E., de Lima, B.F., da Silva, M.C., Madureira, A.J., Ramos, I., Costa, J.L., Hespanhol, V., Cunha, A., Oliveira, H.P., 2023. Lung CT image synthesis using GANs. Expert Syst. Appl. 215, 119350. http://dx.doi.org/10.1016/j.eswa.2022.119350.

Mendrik, A.M., Vincken, K.L., Kuijf, H.J., Breeuwer, M., Bouvy, W.H., de Bresser, J., et al., 2015. MRBrainS challenge: Online evaluation framework for brain image segmentation in 3t MRI scans. Comput. Intell. Neurosci. 2015, 1–16. http://dx.doi.org/10.1155/2015/813696.

Meng, X., Gu, Y., Pan, Y., Wang, N., Xue, P., Lu, M., He, X., Zhan, Y., Shen, D., 2022. A novel unified conditional score-based generative framework for multi-modal medical image completion. URL: https://arxiv.org/abs/2207.03430.

Morbé, L., Vereecke, E., Laloo, F., Chen, M., Herregods, N., Jans, L.B., 2023. Common incidental findings on sacroiliac joint MRI: Added value of MRI-based synthetic CT. Eur. J. Radiol. 158, 110651. http://dx.doi.org/10.1016/j.ejrad.2022.110651.

Nie, D., Cao, X., Gao, Y., Wang, L., Shen, D., 2016. Estimating CT image from MRI data using 3d fully convolutional networks. In: Deep Learning and Data Labeling for Medical Applications. Springer International Publishing, pp. 170–178. http://dx.doi.org/10.1007/978-3-319-46976-8_18.

Nie, D., Trullo, R., Lian, J., Petitjean, C., Ruan, S., Wang, Q., Shen, D., 2017. Medical image synthesis with context-aware generative adversarial networks. In: Medical Image Computing and Computer Assisted Intervention - MICCAI 2017. Springer International Publishing, pp. 417–425. http://dx.doi.org/10.1007/978-3-319-66179-7_48.

Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., Wang, Q., Shen, D., 2018. Medical image synthesis with deep convolutional adversarial networks. IEEE Trans. Biomed. Eng. 65, 2720–2730. http://dx.doi.org/10.1109/tbme.2018.2814538.

Nyholm, T., Svensson, S., Andersson, S., Jonsson, J., Sohlin, M., Gustafsson, C., Kjellé, E., Söderström, K., Albertsson, P., Blomqvist, L., Zackrisson, B., Olsson, L.E., Gunnlaugsson, A., 2018. MR and CT data with multiobserver delineations of organs in the pelvic area-part of the gold atlas project. Med. Phys. 45, 1295–1300. http://dx.doi.org/10.1002/mp.12748.

Osman, A.F.I., Tamam, N.M., 2022. Deep learning-based convolutional neural network for intramodality brain MRI synthesis. J. Appl. Clin. Med. Phys. 23, http://dx.doi.org/10.1002/acm2.13530.

Oulbacha, R., Kadoury, S., 2020. MRI to CT synthesis of the lumbar spine from a pseudo-3d cycle GAN. In: 2020 IEEE 17th International Symposium on Biomedical Imaging. ISBI, IEEE, http://dx.doi.org/10.1109/isbi45749.2020.9098421.

Ouyang, J., Chen, K.T., Armindo, R.D., Davidzon, G.A., Hawk, K.E., Moradi, F., Rosenberg, J., Lan, E., Zhang, H., Zaharchuk, G., 2023. Predicting fdg-pet images from multi-contrast mri using deep learning in patients with brain neoplasms. http://dx.doi.org/10.1002/jmri.28837.

Özbey, M., Dalmaz, O., Dar, S.U., Bedel, H.A., Özturk, c., Güngör, A., Çukur, T., 2022. Unsupervised medical image translation with adversarial diffusion models. URL: https://arxiv.org/abs/2207.08208.

Pain, C.D., Egan, G.F., Chen, Z., 2022. Deep learning-based image reconstruction and post-processing methods in positron emission tomography for low-dose imaging and resolution enhancement. Eur. J. Nucl. Med. Mol. Imaging 49, 3098–3118. http://dx.doi.org/10.1007/s00259-022-05746-4.

Pan, S., Abouei, E., Wynne, J., Wang, T., Qiu, R.L.J., Li, Y., Chang, C.W., Peng, J., Roper, J., Patel, P., Yu, D.S., Mao, H., Yang, X., 2023a. Synthetic ct generation from mri using 3d transformer-based denoising diffusion model. URL: https://arxiv.org/abs/2305.19467.

Pan, S., Chang, C.W., Peng, J., Zhang, J., Qiu, R.L.J., Wang, T., Roper, J., Liu, T., Mao, H., Yang, X., 2023b. Cycle-guided denoising diffusion probability model for 3d cross-modality mri synthesis. URL: https://arxiv.org/abs/2305.00042.

Pan, K., Cheng, P., Huang, Z., Lin, L., Tang, X., 2022. Transformer-based t2-weighted MRI synthesis from t1-weighted images. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society. EMBC, IEEE, http://dx.doi.org/10.1109/embc48229.2022.9871183.

Pan, Y., Liu, M., Lian, C., Zhou, T., Xia, Y., Shen, D., 2018. Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis. In: Medical Image Computing and Computer Assisted Intervention –MICCAI 2018. Springer International Publishing, pp. 455–463. http://dx.doi.org/10.1007/978-3-030-00931-1_52.

Park, T., Efros, A.A., Zhang, R., Zhu, J.Y., 2020. Contrastive learning for unpaired image-to-image translation. URL: https://arxiv.org/abs/2007.15651.

Peng, Y., Chen, S., Qin, A., Chen, M., Gao, X., Liu, Y., Miao, J., Gu, H., Zhao, C., Deng, X., et al., 2020. Magnetic resonance-based synthetic computed tomography images generated using generative adversarial networks for nasopharyngeal carcinoma radiotherapy treatment planning. Radiother. Oncol. 150, 217–224.

Pozaruk, A., Pawar, K., Li, S., Carey, A., Cheng, J., Sudarshan, V.P., Cholewa, M., Grummet, J., Chen, Z., Egan, G., 2020. Augmented deep learning model for improved quantitative accuracy of MR-based PET attenuation correction in PSMA PET-MRI prostate imaging. Eur. J. Nucl. Med. Mol. Imaging 48, 9–20. http://dx.doi.org/10.1007/s00259-020-04816-9.

Qi, M., Li, Y., Wu, A., Jia, Q., Li, B., Sun, W., Dai, Z., Lu, X., Zhou, L., Deng, X., Song, T., 2020. Multi-sequence MR image-based synthetic CT generation using a generative adversarial network for head and neck MRI-only radiotherapy. Med. Phys. 47, 1880–1894. http://dx.doi.org/10.1002/mp.14075.

Qian, P., Xu, K., Wang, T., Zheng, Q., Yang, H., Baydoun, A., Zhu, J., Traughber, B., Muzic, R.F., 2020. Estimating CT from MR abdominal images using novel generative adversarial networks. J. Grid Computing 18, 211–226. http://dx.doi.org/10.1007/s10723-020-09513-3.

Radford, A., Metz, L., Chintala, S., 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. URL: https://arxiv.org/abs/1511.06434.

Rajagopal, A., Natsuaki, Y., Wangerin, K., Hamdi, M., An, H., Sunderland, J.J., Laforest, R., Kinahan, P.E., Larson, P.E., Hope, T.A., 2023. Synthetic PET via domain translation of 3d MRI. IEEE Trans. Radiat. Plasma Med. Sci. 1. http://dx.doi.org/10.1109/trpms.2022.3223275.

Ranjan, A., Lalwani, D., Misra, R., 2022. Gan for synthesizing ct from t2-weighted mri data towards mr-guided radiation treatment. Magn. Reson. Mater. Phys. Biol. Med. 35, 449–457. http://dx.doi.org/10.1007/s10334-021-00974-5.

Reimold, M., Nikolaou, K., Christian La Fougère, M., Gatidis, S., 2019. 18 Independent brain f-fdg pet attenuation correction using a deep learning approach with generative adversarial networks. Hellenic J. Nucl. Med. 22, 179–186.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2021. High-resolution image synthesis with latent diffusion models. URL: https://arxiv.org/abs/2112.10752.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Lecture Notes in Computer Science. Springer International Publishing, pp. 234–241. http://dx.doi.org/10.1007/978-3-319-24574-4_28.

Salem, M., Valverde, S., Cabezas, M., Pareto, D., Oliver, A., Salvi, J., Rovira, A., Llado, X., 2019. Multiple sclerosis lesion synthesis in MRI using an encoder–decoder u-NET. IEEE Access 7, 25171–25184. http://dx.doi.org/10.1109/access.2019.2900198.

Sanaat, A., Arabi, H., Mainta, I., Garibotto, V., Zaidi, H., 2020. Projection space implementation of deep learning–guided low-dose brain PET imaging improves performance over implementation in image space. J. Nucl. Med. 61, 1388–1396. http://dx.doi.org/10.2967/jnumed.119.239327.

Scholey, J.E., Rajagopal, A., Vasquez, E.G., Sudhyadhom, A., Larson, P.E.Z., 2022. Generation of synthetic megavoltage CT for MRI-only radiotherapy treatment planning using a 3d deep convolutional neural network. Med. Phys. 49, 6622–6634. http://dx.doi.org/10.1002/mp.15876.

Schwab, J., Antholzer, S., Haltmeier, M., 2019. Deep null space learning for inverse problems: convergence analysis and rates. Inverse Problems 35, 025008. http://dx.doi.org/10.1088/1361-6420/aaf14a.

Shamshad, F., Khan, S., Zamir, S.W., Khan, M.H., Hayat, M., Khan, F.S., Fu, H., 2022. Transformers in medical imaging: A survey. URL: https://arxiv.org/abs/2201.09873.

Sharma, A., Hamarneh, G., 2020. Missing MRI pulse sequence synthesis using multi-modal generative adversarial network. IEEE Trans. Med. Imaging 39, 1170–1183. http://dx.doi.org/10.1109/tmi.2019.2945521.

Shen, L., Zhu, W., Wang, X., Xing, L., Pauly, J.M., Turkbey, B., Harmon, S.A., Sanford, T.H., Mehralivand, S., Choyke, P.L., Wood, B.J., Xu, D., 2021. Multi-domain image completion for random missing input data. IEEE Trans. Med. Imaging 40, 1113–1122. http://dx.doi.org/10.1109/tmi.2020.3046444.

Shi, K., Guo, R., Xue, S., Wang, H., Rominger, A., Li, B., 2022. Ultra-low dose PET imaging challenge - Grand challenge. URL: https://ultra-low-dose-pet.grand-challenge.org/.

Shi, Z., Mettes, P., Zheng, G., Snoek, C., 2021. Frequency-supervised MR-to-CT image synthesis. In: Deep Generative Models, and Data Augmentation, Labelling, and Imperfections. Springer International Publishing, pp. 3–13. http://dx.doi.org/10.1007/978-3-030-88210-5_1.

Shin, H.C., Ihsani, A., Mandava, S., Sreenivas, S.T., Forster, C., Cha, J., Initiative, A.D.N., 2020a. Ganbert: Generative adversarial networks with bidirectional encoder representations from transformers for mri to pet synthesis. URL: https://arxiv.org/abs/2008.04393.

Shin, H.C., Ihsani, A., Xu, Z., Mandava, S., Sreenivas, S.T., Forster, C., Cha, J., 2020b. GANDALF: Generative adversarial networks with discriminator-adaptive loss fine-tuning for Alzheimer's disease diagnosis from MRI. In: Medical Image Computing and Computer Assisted Intervention –MICCAI 2020. Springer International Publishing, pp. 688–697. http://dx.doi.org/10.1007/978-3-030-59713-9_66.

Sikka, A., Peri, S.V., Bathula, D.R., 2018. MRI to FDG-PET: Cross-modal synthesis using 3d u-net for multi-modal Alzheimer's classification. In: Simulation and Synthesis in Medical Imaging. Springer International Publishing, pp. 80–89. http://dx.doi.org/10.1007/978-3-030-00536-8_9.

Sikka, A., Skand, Virk, J.S., Bathula, D.R., 2021. Mri to pet cross-modality translation using globally and locally aware gan (gla-gan) for multi-modal diagnosis of Alzheimer's disease. URL: https://arxiv.org/abs/2108.02160.

Skandarani, Y., Jodoin, P.M., Lalande, A., 2023. GANs for medical image synthesis: An empirical study. J. Imaging 9 (69), http://dx.doi.org/10.3390/jimaging9030069.

Spadea, M.F., Pileggi, G., Zaffino, P., Salome, P., Catana, C., Izquierdo-Garcia, D., Amato, F., Seco, J., 2019. Deep convolution neural network (DCNN) multiplane approach to synthetic CT generation from MR images—application in brain proton therapy. Int. J. Radiat. Oncol. Biol. Phys. 105, 495–503. http://dx.doi.org/10.1016/j.ijrobp.2019.06.2535.

Sreeja, S., Mubarak, D.M.N., 2022. Pseudo-CT generation from MRI images for bone lesion detection using deep learning approach. In: Pervasive Computing and Social Networking. Springer Nature Singapore, pp. 621–632. http://dx.doi.org/10.1007/978-981-19-2840-6_47.

Sudarshan, V.P., Li, S., Jamadar, S.D., Egan, G.F., Awate, S.P., Chen, Z., 2021. Incorporation of anatomical MRI knowledge for enhanced mapping of brain metabolism using functional PET. NeuroImage 233, 117928. http://dx.doi.org/10.1016/j.neuroimage.2021.117928.

Sun, B., Jia, S., Jiang, X., Jia, F., 2022a. Double u-net CycleGAN for 3d MR to CT image synthesis. Int. J. Comput. Assist. Radiol. Surg. 18, 149–156. http://dx.doi.org/10.1007/s11548-022-02732-x.

Sun, H., Jiang, Y., Yuan, J., Wang, H., Liang, D., Fan, W., Hu, Z., Zhang, N., 2022b. High-quality PET image synthesis from ultra-low-dose PET/MRI using bi-task deep learning. Quant. Imaging Med. Surg. 12, 5326–5342. http://dx.doi.org/10.21037/qims-22-116.

Takamiya, K., Iwamoto, Y., Nonaka, M., Chen, Y.W., 2023. CT brain image synthesization from MRI brain images using CycleGAN. In: 2023 IEEE International Conference on Consumer Electronics. ICCE, IEEE, http://dx.doi.org/10.1109/icce56470.2023.10043572.

Tie, X., Lam, S.K., Zhang, Y., Lee, K.H., Au, K.H., Cai, J., 2020. Pseudo-CT generation from multi-parametric MRI using a novel multi-channel multi-path conditional generative adversarial network for nasopharyngeal carcinoma patients. Med. Phys. 47, 1750–1762. http://dx.doi.org/10.1002/mp.14062.

Touati, R., Le, W.T., Kadoury, S., 2021. A feature invariant generative adversarial network for head and neck MRI/CT image synthesis. Phys. Med. Biol. 66, 095001. http://dx.doi.org/10.1088/1361-6560/abf1bb.

Uzunova, H., Ehrhardt, J., Handels, H., 2020. Memory-efficient GAN-based domain translation of high resolution 3d medical images. Comput. Med. Imaging Graph. 86, 101801. http://dx.doi.org/10.1016/j.compmedimag.2020.101801.

Vaidya, A., Stough, J.V., Patel, A., 2022. Perceptually improved t1-t2 MRI translations using conditional generative adversarial networks. In: Išgum, I., Colliot, O. (Eds.), Medical Imaging 2022: Image Processing. SPIE, http://dx.doi.org/10.1117/12.2608428.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. URL: https://arxiv.org/abs/1706.03762.

Vrtovec, T., Yao, J., Glocker, B., Klinder, T., Frangi, A., Zheng, G., Li, S. (Eds.), 2016. Computational methods and clinical applications for spine imaging. In: Lecture Notes in Computer Science, first ed. Springer International Publishing, Basel, Switzerland.

Wang, Y.R., Baratto, L., Hawk, K.E., Theruvath, A.J., Pribnow, A., Thakor, A.S., Gatidis, S., Lu, R., Gummidipundi, S.E., Garcia-Diaz, J., Rubin, D., Daldrup-Link, H.E., 2021b. Artificial intelligence enables whole-body positron emission tomography scans with minimal radiation exposure. Eur. J. Nucl. Med. Mol. Imaging 48, 2771–2781. http://dx.doi.org/10.1007/s00259-021-05197-3.

Wang, L., Gao, Y., Shi, F., Li, G., Gilmore, J.H., Lin, W., Shen, D., 2015. LINKS: Learning-based multi-source IntegratioN frameworK for segmentation of infant brain images. NeuroImage 108, 160–172. http://dx.doi.org/10.1016/j.neuroimage.2014.12.042.

Wang, T., Lei, Y., Fu, Y., Wynne, J.F., Curran, W.J., Liu, T., Yang, X., 2020. A review on medical imaging synthesis using deep learning and its clinical applications. J. Appl. Clin. Med. Phys. 22, 11–36. http://dx.doi.org/10.1002/acm2.13121.

Wang, C., Uh, J., He, X., Hua, C.H., Sahaja, A., 2021. Transfer learning-based synthetic CT generation for MR-only proton therapy planning in children with pelvic sarcomas. In: Bosmans, H., Zhao, W., Yu, L. (Eds.), Medical Imaging 2021: Physics of Medical Imaging. SPIE, http://dx.doi.org/10.1117/12.2579767.

Wang, C.C., Wu, P.H., Lin, G., Huang, Y.L., Lin, Y.C., Chang, Y.P.E., Weng, J.C., 2022a. Magnetic resonance-based synthetic computed tomography using generative adversarial networks for intracranial tumor radiotherapy treatment planning. J. Pers. Med. 12 (361), http://dx.doi.org/10.3390/jpm12030361.

Wang, Z., Yang, Y., Sermesant, M., Delingette, H., Wu, O., 2023. Zero-shot-learning cross-modality data translation through mutual information guided stochastic diffusion. URL: https://arxiv.org/abs/2301.13743.

Wang, Y., Yu, B., Wang, L., Zu, C., Lalush, D.S., Lin, W., Wu, X., Zhou, J., Shen, D., Zhou, L., 2018. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. NeuroImage 174, 550–562. http://dx.doi.org/10.1016/j.neuroimage.2018.03.045.

Wang, Y., Yu, J., Zhang, J., 2022. Zero-shot image restoration using denoising diffusion null-space model. URL: https://arxiv.org/abs/2212.00490.

Wang, Y., Zhou, L., Yu, B., Wang, L., Zu, C., Lalush, D.S., Lin, W., Wu, X., Zhou, J., Shen, D., 2019. 3D auto-context-based locality adaptive multi-modality GANs for PET synthesis. IEEE Trans. Med. Imaging 38, 1328–1339. http://dx.doi.org/10.1109/tmi.2018.2884053.

Wei, W., Poirion, E., Bodini, B., Durrleman, S., Ayache, N., Stankoff, B., Colliot, O., 2019. Predicting PET-derived demyelination from multimodal MRI using sketcher-refiner adversarial training for multiple sclerosis. Med. Image Anal. 58, 101546. http://dx.doi.org/10.1016/j.media.2019.101546.

West, J., Fitzpatrick, J.M., Wang, M.Y., Dawant, B.M., Maurer, C.R., Kessler, R.M., M., R.J., et al., 1997. Comparison and evaluation of retrospective intermodality brain image registration techniques. J. Comput. Assist. Tomogr. 21, 554–568. http://dx.doi.org/10.1097/00004728-199707000-00007.

Wu, H., Jiang, X., Jia, F., 2019. UC-GAN for MR to CT image synthesis. In: Artificial Intelligence in Radiation Therapy. Springer International Publishing, pp. 146–153. http://dx.doi.org/10.1007/978-3-030-32486-5_18.

Xiang, L., Qiao, Y., Nie, D., An, L., Lin, W., Wang, Q., Shen, D., 2017. Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. Neurocomputing 267, 406–416. http://dx.doi.org/10.1016/j.neucom.2017.06.048.

Xiang, L., Wang, Q., Nie, D., Zhang, L., Jin, X., Qiao, Y., Shen, D., 2018. Deep embedding convolutional neural network for synthesizing CT image from t1-weighted MR image. Med. Image Anal. 47, 31–44. http://dx.doi.org/10.1016/j.media.2018.03.011.

Xie, T., Cao, C., Cui, Z., Guo, Y., Wu, C., Wang, X., Li, Q., Hu, Z., Sun, T., Sang, Z., Zhou, Y., Zhu, Y., Liang, D., Jin, Q., Chen, G., Wang, H., 2023. Synthesizing pet images from high-field and ultra-high-field mr images using joint diffusion attention model. URL: https://arxiv.org/abs/2305.03901.

Xue, H., Zhang, Q., Zou, S., Zhang, W., Zhou, C., Tie, C., Wan, Q., Teng, Y., Li, Y., Liang, D., Liu, X., Yang, Y., Zheng, H., Zhu, X., Hu, Z., 2021. LCPR-net: low-count PET image reconstruction using the domain transform and cycle-consistent generative adversarial networks. Quant. Imaging Med. Surg. 11, 749–762. http://dx.doi.org/10.21037/qims-20-66.

Yan, S., Wang, C., Chen, W., Lyu, J., 2022. Swin transformer-based GAN for multi-modal medical image translation. Front. Oncol. 12, http://dx.doi.org/10.3389/fonc.2022.942511.

Yang, Q., Li, N., Zhao, Z., Fan, X., Chang, E.I.C., Xu, Y., 2018a. Mri cross-modality neuroimage-to-neuroimage translation. URL: https://arxiv.org/abs/1801.06940.

Yang, H., Sun, J., Carass, A., Zhao, C., Lee, J., Prince, J.L., Xu, Z., 2020. Unsupervised MR-to-CT synthesis using structure-constrained CycleGAN. IEEE Trans. Med. Imaging 39, 4249–4261. http://dx.doi.org/10.1109/tmi.2020.3015379.

Yang, H., Sun, J., Carass, A., Zhao, C., Lee, J., Xu, Z., Prince, J., 2018b. Unpaired brain MR-to-CT synthesis using a structure-constrained CycleGAN. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer International Publishing, pp. 174–182. http://dx.doi.org/10.1007/978-3-030-00889-5_20.

Yang, H., Sun, J., Yang, L., Xu, Z., 2021. A unified hyper-GAN model for unpaired multi-contrast MR image translation. In: Medical Image Computing and Computer Assisted Intervention –MICCAI 2021. Springer International Publishing, pp. 127–137. http://dx.doi.org/10.1007/978-3-030-87199-4_12.

Yang, Z., Zhou, Y., Zhang, H., Wei, B., Fan, Y., Xu, Y., 2023. Drmc: A generalist model with dynamic routing for multi-center pet image synthesis. URL: https://arxiv.org/abs/2307.05249.

Yi, X., Walia, E., Babyn, P., 2019. Generative adversarial network in medical imaging: A review. Med. Image Anal. 58, 101552. http://dx.doi.org/10.1016/j.media.2019.101552.

Yoon, J.S., Zhang, C., Suk, H.I., Guo, J., Li, X., 2022. Sadm: Sequence-aware diffusion model for longitudinal medical image generation. URL: https://arxiv.org/abs/2212.08228.

Yu, B., Wang, Y., Wang, L., Shen, D., Zhou, L., 2020. Medical image synthesis via deep learning. In: Advances in Experimental Medicine and Biology. Springer International Publishing, pp. 23–44. http://dx.doi.org/10.1007/978-3-030-33128-3_2.

Yu, B., Zhou, L., Wang, L., Shi, Y., Fripp, J., Bourgeat, P., 2019. Ea-GANs: Edge-aware generative adversarial networks for cross-modality MR image synthesis. IEEE Trans. Med. Imaging 38, 1750–1762. http://dx.doi.org/10.1109/tmi.2019.2895894.

Yurt, M., Dalmaz, O., Dar, S., Ozbey, M., Tinaz, B., Oguz, K., Cukur, T., 2022. Semi-supervised learning of MRI synthesis without fully-sampled ground truths. IEEE Trans. Med. Imaging 41, 3895–3906. http://dx.doi.org/10.1109/tmi.2022.3199155.

Yurt, M., Dar, S.U., Erdem, A., Erdem, E., Oguz, K.K., Çukur, T., 2021. mustGAN: multi-stream generative adversarial networks for MR image synthesis. Med. Image Anal. 70, 101944. http://dx.doi.org/10.1016/j.media.2020.101944.

Zbontar, J., Knoll, F., Sriram, A., Murrell, T., Huang, Z., Muckley, M.J., Defazio, A., Stern, R., Johnson, P., Bruno, M., Parente, M., Geras, K.J., Katsnelson, J., Chandarana, H., Zhang, Z., Drozdzal, M., Romero, A., Rabbat, M., Vincent, P., Yakubova, N., Pinkerton, J., Wang, D., Owens, E., Zitnick, C.L., Recht, M.P., Sodickson, D.K., Lui, Y.W., 2019. fastMRI: An open dataset and benchmarks for accelerated MRI. arXiv:1811.08839 [physics, stat].

Zeng, G., Zheng, G., 2019. Hybrid generative adversarial networks for deep MR to CT synthesis using unpaired data. In: Lecture Notes in Computer Science. Springer International Publishing, pp. 759–767. http://dx.doi.org/10.1007/978-3-030-32251-9_83.

Zeng, P., Zhou, L., Zu, C., Zeng, X., Jiao, Z., Wu, X., Zhou, J., Shen, D., Wang, Y., 2022. 3D CVT-GAN: A 3d convolutional vision transformer-GAN for PET reconstruction. In: Lecture Notes in Computer Science. Springer Nature Switzerland, pp. 516–526. http://dx.doi.org/10.1007/978-3-031-16446-0_49.

Zhan, B., Li, D., Wang, Y., Ma, Z., Wu, X., Zhou, J., Zhou, L., 2021. LR-cGAN: Latent representation based conditional generative adversarial network for multi-modality MRI synthesis. Biomed. Signal Process. Control 66, 102457. http://dx.doi.org/10.1016/j.bspc.2021.102457.

Zhan, B., Li, D., Wu, X., Zhou, J., Wang, Y., 2022. Multi-modal MRI image synthesis via GAN with multi-scale gate mergence. IEEE J. Biomed. Health Inf. 26, 17–26. http://dx.doi.org/10.1109/jbhi.2021.3088866.

Zhang, X., He, X., Guo, J., Ettehadi, N., Aw, N., Semanek, D., Posner, J., Laine, A., Wang, Y., 2022a. PTNet3d: A 3d high-resolution longitudinal infant brain MRI synthesizer based on transformers. IEEE Trans. Med. Imaging 41, 2925–2940. http://dx.doi.org/10.1109/tmi.2022.3174827.

Zhang, J., He, X., Qing, L., Gao, F., Wang, B., 2022b. BPGAN: Brain PET synthesis from MRI using generative adversarial network for multi-modal alzheimer's disease diagnosis. Comput. Methods Programs Biomed. 217, 106676. http://dx.doi.org/10.1016/j.cmpb.2022.106676.

Zhang, H., Li, H., Dillman, J.R., Parikh, N.A., He, L., 2022c. Multi-contrast MRI image synthesis using switchable cycle-consistent generative adversarial networks. Diagnostics 12 (816), http://dx.doi.org/10.3390/diagnostics12040816.

Zhang, L., Xiao, Z., Zhou, C., Yuan, J., He, Q., Yang, Y., Liu, X., Liang, D., Zheng, H., Fan, W., Zhang, X., Hu, Z., 2021. Spatial adaptive and transformer fusion network (STFNet) for low-count PET blind denoising with MRI. Med. Phys. 49, 343–356. http://dx.doi.org/10.1002/mp.15368.

Zhao, B., Cheng, T., Zhang, X., Wang, J., Zhu, H., Zhao, R., Li, D., Zhang, Z., Yu, G., 2023. CT synthesis from MR in the pelvic area using residual transformer conditional GAN. Comput. Med. Imaging Graph. 103, 102150. http://dx.doi.org/10.1016/j.compmedimag.2022.102150.

Zhao, S., Geng, C., Guo, C., Tian, F., Tang, X., 2022. SARU: A self-attention ResUNet to generate synthetic CT images for MR-only BNCT treatment planning. Med. Phys. 50, 117–127. http://dx.doi.org/10.1002/mp.15986.

Zhao, J., Li, D., Kassam, Z., Howey, J., Chong, J., Chen, B., Li, S., 2020a. Tripartite-GAN: Synthesizing liver contrast-enhanced MRI to improve tumor detection. Med. Image Anal. 63, 101667. http://dx.doi.org/10.1016/j.media.2020.101667.

Zhao, P., Pan, H., Xia, S., 2021. MRI-trans-GAN: 3d MRI cross-modality translation. In: 2021 40th Chinese Control Conference. CCC, IEEE, http://dx.doi.org/10.23919/ccc52363.2021.9550256.

Zhao, K., Zhou, L., Gao, S., Wang, X., Wang, Y., Zhao, X., Wang, H., Liu, K., Zhu, Y., Ye, H., 2020b. Study of low-dose PET image recovery using supervised learning with CycleGAN. PLoS One 15, e0238455. http://dx.doi.org/10.1371/journal.pone.0238455.

Zhou, X., Cai, W., Cai, J., Xiao, F., Qi, M., Liu, J., Zhou, L., Li, Y., Song, T., 2023. Multimodality MRI synchronous construction based deep learning framework for MRI-guided radiotherapy synthetic CT generation. Comput. Biol. Med. 162, 107054. http://dx.doi.org/10.1016/j.compbiomed.2023.107054.

Zhou, T., Fu, H., Chen, G., Shen, J., Shao, L., 2020a. Hi-net: Hybrid-fusion network for multi-modal MR image synthesis. IEEE Trans. Med. Imaging 39, 2772–2781. http://dx.doi.org/10.1109/tmi.2020.2975344.

Zhou, Q., Liu, Y., Hu, H., Guan, Q., Guo, Y., Zhang, F., 2021. Unsupervised multimodal MR images synthesizer using knowledge from higher dimension. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine. BIBM, IEEE, http://dx.doi.org/10.1109/bibm52615.2021.9669327.

Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2020. UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans. Med. Imaging 39, 1856–1867. http://dx.doi.org/10.1109/tmi.2019.2959609.

Zhou, Y., Yang, Z., Zhang, H., Chang, E.I.C., Fan, Y., Xu, Y., 2022. 3D segmentation guided style-based generative adversarial networks for PET synthesis. IEEE Trans. Med. Imaging 41, 2092–2104. http://dx.doi.org/10.1109/tmi.2022.3156614.

Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. URL: https://arxiv.org/abs/1703.10593.

Zhu, L., Xue, Z., Jin, Z., Liu, X., He, J., Liu, Z., Yu, L., 2023. Make-a-volume: Leveraging latent diffusion models for cross-modality 3d brain mri synthesis. URL: https://arxiv.org/abs/2307.10094.

Zimmermann, L., Knäusl, B., Stock, M., Lütgendorf-Caucig, C., Georg, D., Kuess, P., 2022. An MRI sequence independent convolutional neural network for synthetic head CT generation in proton therapy. Z. Med. Phys. 32, 218–227. http://dx.doi.org/10.1016/j.zemedi.2021.10.003.

Zotova, D., Jung, J., Lartizien, C., 2021. GAN-based synthetic FDG PET images from t1 brain MRI can serve to improve performance of deep unsupervised anomaly detection models. In: Simulation and Synthesis in Medical Imaging. Springer International Publishing, pp. 142–152. http://dx.doi.org/10.1007/978-3-030-87592-3_14.