Medical Image Analysis 103 (2025) 103605

Contents lists available at ScienceDirect

Medical Image Analysis

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

Review

Medical image translation with deep learning: Advances, datasets and perspectives☆

Junxin Chen [a], Zhiheng Ye [a], Renlong Zhang [b], Hao Li [c], Bo Fang [d], Li-bo Zhang [e,*], Wei Wang [f,g]

[a] School of Software, Dalian University of Technology, Dalian 116621, China
[b] Institute of Research and Clinical Innovations, Neusoft Medical Systems Co., Ltd., Beijing, China
[c] School of Computing Science, University of Glasgow, Glasgow G12 8QQ, United Kingdom
[d] School of Computer Science, The University of Sydney, Sydney, NSW 2006, Australia
[e] Department of Radiology, General Hospital of Northern Theater Command, Shenyang 110840, China
[f] Guangdong-Hong Kong-Macao Joint Laboratory for Emotion Intelligence and Pervasive Computing, Artificial Intelligence Research Institute, Shenzhen MSU-BIT University, Shenzhen 518172, China
[g] School of Medical Technology, Beijing Institute of Technology, Beijing 100081, China

ARTICLE INFO

Keywords:
Medical image translation
Deep learning
Multimodality image processing
Data augmentation

ABSTRACT

Traditional medical image generation often lacks patient-specific clinical information, limiting its clinical utility despite enhancing downstream task performance. In contrast, medical image translation precisely converts images from one modality to another, preserving both anatomical structures and cross-modal features, thus enabling efficient and accurate modality transfer and offering unique advantages for model development and clinical practice. This paper reviews the latest advancements in deep learning(DL)-based medical image translation. Initially, it elaborates on the diverse tasks and practical applications of medical image translation. Subsequently, it provides an overview of fundamental models, including convolutional neural networks (CNNs), transformers, and state space models (SSMs). Additionally, it delves into generative models such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Autoregressive Models (ARs), diffusion Models, and flow Models. Evaluation metrics for assessing translation quality are discussed, emphasizing their importance. Commonly used datasets in this field are also analyzed, highlighting their unique characteristics and applications. Looking ahead, the paper identifies future trends, challenges, and proposes research directions and solutions in medical image translation. It aims to serve as a valuable reference and inspiration for researchers, driving continued progress and innovation in this area.

Contents

☆ This work is funded by the National Natural Science Foundation of China (No. 62171114), and Xiaomi Young Talents Program.
* Corresponding author.
E-mail addresses: junxinchen@ieee.org (J. Chen), 175299084@mail.dlut.edu.cn (Z. Ye), zhangrenlong@neusoftmedical.com (R. Zhang), 2984207L@student.gla.ac.uk (H. Li), bfan0072@uni.sydney.edu.au (B. Fang), zhanglibo_academic@163.com (L.-b. Zhang), ehomewang@ieee.org (W. Wang).

https://doi.org/10.1016/j.media.2025.103605
Received 15 December 2024; Received in revised form 7 March 2025; Accepted 12 April 2025
Available online 27 April 2025

1. Introduction

Medical image translation refers to the process of transforming medical images across modalities, resolutions, and even between patients. This field has emerged as a transformative area in medical imaging, fundamentally altering how we interpret and utilize disparate medical datasets (Alotaibi, 2020; McNaughton et al., 2023a).

1.1. Medical image translation

Medical image translation is a rapidly evolving field in modern medical imaging that has garnered increasing attention due to the growing need for improved interoperability, data harmonization, and enhanced clinical decision support (Dalmaz et al., 2022; Dayarathna et al., 2023; Emami et al., 2020a). The driving force is the need to facilitate the integration, comparison and analysis of heterogeneous medical imaging data. Medical image translation can be categorized into three primary branches, including intra-modality translation, inter-modality translation, and label-based translation.

(1) *Intra-modality translation.* It refers to image conversion within the same modality, such as converting a low-dose Computed Tomography(CT) scan to a standard-dose scan.
(2) *Cross modality translation.* It involves image translation between different image modalities. Example includes converting an Magnetic Resonance Imaging(MRI) image to a CT image, or a Positron Emission Computed Tomography(PET) image to an MRI image.
(3) *Label-based translation.* This process of image conversion is guided by label information. It is able to produce additional medical images with the existing samples' labels, so as to enlarge the training datasets for downstream tasks.

In the realm of medical imaging, concepts like image generation, image reconstruction, image synthesis, and image translation might seem similar at first glance, but they have distinct characteristics.

(1) Image generation usually creates images from random elements like noise or vectors, lacking a direct link to a specific source image. It is more about generating novel visual content without relying on an existing image for reference.
(2) Image reconstruction focuses on transforming raw physical data into images, such as using X-ray attenuation coefficient measurements to reconstruct CT images. This process is based on physical data acquisition and transformation.
(3) Image synthesis broadly refers to the creation of medical images from abstract inputs, encompassing two primary paradigms: unconditional synthesis, which generates images from random vectors, and conditional synthesis, where images are produced from non-image inputs such as anatomical sketches or textual descriptions. While conditional synthesis may resemble medical image translation, the former prioritizes semantic plausibility (e.g., preserving organ presence with shape variations), whereas translation enforces strict anatomical consistency (e.g., identical lesion geometry).
(4) In contrast, medical image translation is centered around converting one medical image to another. It has a clear source-target relationship, where the output maintains significant ties to the original image, such as showing the same body parts or sharing segmentation labels. This connection is crucial for retaining medical-relevant information.

In summary, while these concepts share the common goal of creating or transforming images, medical image translation stands out for its reliance on a source medical image and the preservation of important relationships and features. Understanding these differences is essential for anyone exploring the field of medical imaging, as it helps in accurately applying the appropriate techniques for various medical applications.

1.2. Advantages and enabling technologies

Medical image translation offers a range of advantages that profoundly impact on clinical practice and medical research (Fu et al., 2020). One of the primary benefit is its ability to facilitate cross-modality integration, allowing healthcare professionals to correlate information from diverse imaging sources (Chen et al., 2022; Chourak et al., 2022). Furthermore, it facilitates data argumentation strategies, especially for limited and unbalanced datasets. By generating additional images with different features, image translation methods enhance the training of machine learning(ML) models, ultimately leading to improved performance in image analysis tasks like segmentation and classification. Additionally, medical image translation finds applications in various areas, including attenuation correction, multi-modality registration, radiation therapy planning, disease prediction, and medical education, among others. Its versatility and utility make it a powerful tool in the medical imaging domain.

In the field of medical image translation, several pivotal techniques have propelled significant advancements and applications, ARs, GANs, VAEs, flow models and diffusion models. ARs generate images pixel-by-pixel or block-by-block, ensuring intricate details and coherence, making them ideal for producing high-resolution medical image. GANs, particularly conditional GANs(CGANs), excel in generating high-quality images through adversarial training of generators and discriminators, and are highly effective in cross-modality image translation tasks. VAEs generate varied and realistic medical images by learning the latent distributions of the data, enhancing the diversity of the training dataset. flow models, utilizing reversible neural networks, efficiently sample and accurately evaluate data densities, making them well-suited for cross-modality mappings that preserve data structures and attributes. Lastly, diffusion models generate high-fidelity images by progressively denoising Gaussian noise, which is suitable for enhancing image quality and details. Collectively, these techniques constitute a formidable toolkit for medical image translation, contributing to improved diagnostic accuracy and treatment planning in clinical practice.

1.3. Comparisons and contributions

Table 1 lists some related reviews, and this paper is different from them in the following aspects.

(1) There are some studies focusing on a single image translation task, such as MRI-to-CT translation (Boulanger et al., 2021) or PET-to-CT synthesis (Spadea et al., 2021). On the other hand, this paper mines the existing intra-modality and inter-modality image translation problems comprehensively.
(2) Some studies investigated a certain enabling technology (such as GAN) for image translation or synthesis (Alotaibi, 2020; Singh and Raza, 2021; Dayarathna et al., 2023). This paper provides a more comprehensive survey the existing enabling methods for medical image translation, including VAE, diffusion modeling, etc.
(3) Existing studies do not summarize relevant datasets popular in this field, which makes data sharing and replication difficult (Wang et al., 2021b; Spadea et al., 2021; Boulanger et al., 2021; Singh and Raza, 2021; Yu et al., 2020; Dayarathna et al., 2023). To this end, we are committed to collecting publicly available datasets that can be used for medical image translation to facilitate further research.
(4) Performance assessment indicators for medical image translation are also usually overlooked in the past reviews (Dayarathna et al., 2023; McNaughton et al., 2023a; Alotaibi, 2020). On the other hand, this paper comprehensively summarizes and discusses the evaluation metrics applicable to this type of task.

Table 1
Comparison with related review articles.

Article	Focus and themes	Methods	Tasks	Quality evaluation	Datasets and challenge issues
Wang et al. (2021b)	Medical imaging synthesis	Auto-encoder, U-Net, GAN	Inter-modality, Intra-modality	×	×
Spadea et al. (2021)	Generate synthetic CT (sCT)	CNN, U-Net, GAN	PET, CBCT to CT	✓	×
Boulanger et al. (2021)	Generate sCT	CNN, GAN	MRI-CT	✓	×
Singh and Raza (2021)	Medical image generation	GAN	Inter-modality, Intra-modality	✓	×
McNaughton et al. (2023a)	Medical image translation	ML, CNN, GAN	Inter-modality	✓	✓
Alotaibi (2020)	Image-to-image translation	GAN	Inter-modality, Intra-modality	✓	✓
Yu et al. (2020)	Medical image synthesis	CNN, GAN	Inter-modality, Intra-modality	✓	×
Dayarathna et al. (2023)	Medical image synthesis	CNN, GAN	Inter-modality, Intra-modality	✓	×
Ours	Medical image translation	AR, GAN, VAE, flow, diffusion	Inter-modality, Intra-modality	✓	✓

This paper strives to provide academics with a more comprehensive and in-depth picture of the current research status and future direction in medical image translation area. Our contributions are summarized as follows.

(1) A systematic review of the technical advances for medical image translation is given, covering a variety of DL methods, such as GANs, VAE, ARs, diffusion and flow models, to provide researchers with a comprehensive technical framework.
(2) The commonly used evaluation indexes in medical image translation are discussed comprehensively, and their roles and significance are analyzed. The commonly used datasets are also summarized, and their characteristics and application scenarios are listed.
(3) The advantages and challenges of medical image translation are analyzed in depth, emphasizing the unique performance and applicable scenarios in the whole medical imaging field.

1.4. Paper organization

First, we summarize the technical advances as well as the practical applications of medical image translation in Sections 2 and 3. Next, we introduce the enabling technologies of medical image translation in Section 4, and the commonly used performance metrics are presented in Section 5. The popular open datasets are introduced in Section 6. In Section 7, the future development trends and challenges are envisioned, and some possible research directions are discussed. Finally, Section 8 concludes the whole paper.

2. Advances of medical image translation

2.1. Overview

In medical imaging, regulatory mandates require patient consent for diagnostic images intended for publication or dissemination (Cunniff et al., 2000). Despite collaborative efforts to create large open-access datasets, researchers face limited accessibility due to challenges related to consent, annotation scarcity, and image quality. To address these issues, medical image translation has emerged as a promising solution, leveraging computer algorithms to generate high-fidelity, medically usable images.

Traditional data augmentation techniques, such as scaling, rotation, and elastic deformation (Simard et al., 2003), while widely used, often fail to capture the complex variations inherent in medical images. In contrast, DL-based medical image translation has made significant strides in simulating real image features and generating realistic translated images. These techniques not only enhance image quality and resolution but also fill in missing dataset information, improving diagnostic accuracy. As shown in Fig. 1, it encompasses intra-modality, cross-modality and label-based translations. Intra-modality translation alters the style or features within the same modality, useful for image enhancement and data augmentation without changing underlying semantic information. Cross-modality translation focuses on converting images between different modalities (e.g., MRI to CT), aiming to achieve information translation across different imaging devices or modalities, thereby improving the interpretability and usability of medical images. Label-based translation, on the other hand, harnesses label information to direct the generative model, ultimately elevating the quality of the resultant images.

In summary, the challenges in acquiring, annotating, and ensuring high-quality medical images have spurred the development of medical image translation techniques. By leveraging DL, these methods generate high-quality images, enrich datasets, and improve diagnostic accuracy. While traditional augmentation techniques have limitations, DL-based image-to-image translation offers a robust, versatile solution with profound implications for both research and clinical practice.

2.2. Intra-modality translation

Image translation within medical image modalities involves generating new image samples within the same medical imaging modality. This research direction aims to enrich medical image datasets and enhance the robustness and generalization performance of DL models by simulating diversity within the same imaging modality. It encompasses the study of transforming images between two different protocols in an imaging modality, for instance, between different MRI sequences, or restoring images from a low-quality protocol to a higher-quality study. Specifically, it mainly includes MRI intra-modality image translation, CT intra-modality image translation, and PET intra-modality image translation.

2.2.1. MRI intra-modality image translation

MRI is pivotal in medical imaging due to its multimodality and multi-sequence capabilities, yet faces challenges such as long acquisition times, high costs, and resolution/contrast limitations. Therefore, Image translation techniques have emerged as valuable solutions, addressing various MRI-related issues, as illustrated in Table 2, including conversion between sequence types (Wang et al., 2023b; Liu et al., 2023; Yan et al., 2022), conversion of low-field MRI to high-field MRI (Figini et al., 2020), and recovery of undersampled acquisitions (Liu et al., 2019a). These techniques not only improve image quality, but also play an important role in a variety of application scenarios.

Firstly, translation techniques can reconstruct 2D slices into 3D images, preserving the spatial structure of the MRI image, which is

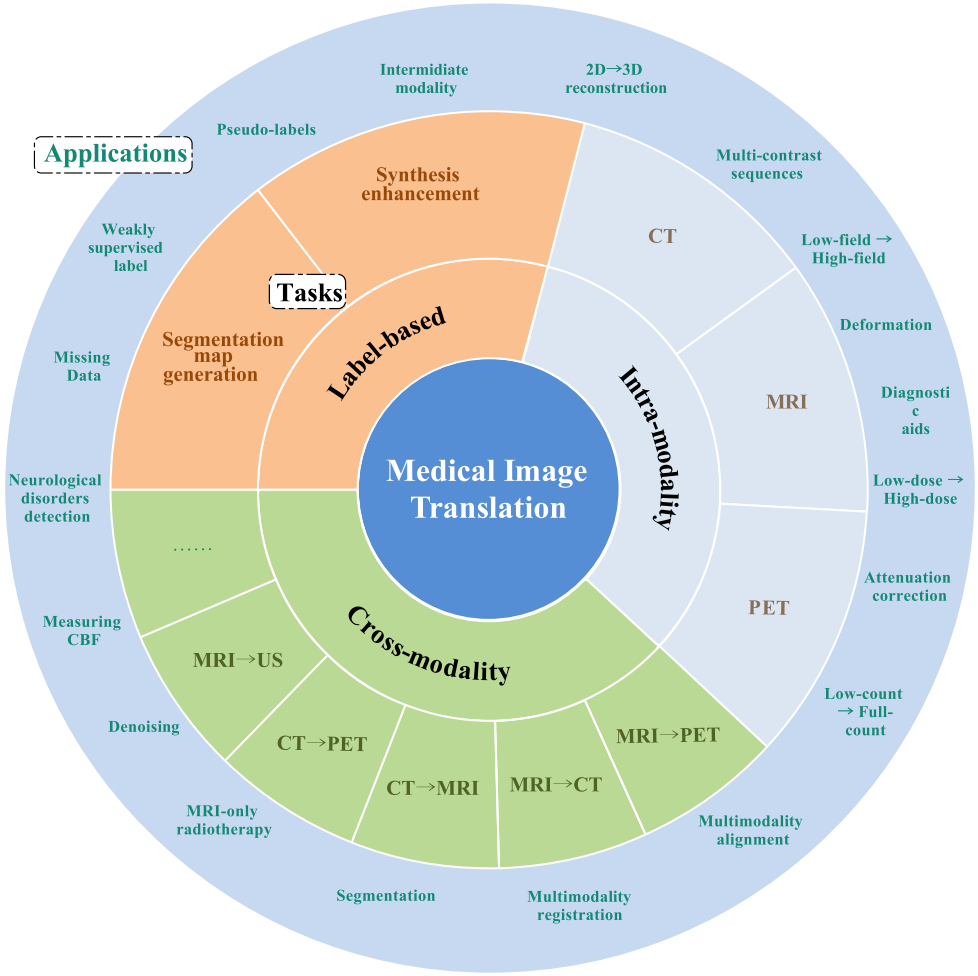

Fig. 1. Overview of tasks and applications in medical image-to-image translation.

particularly important for clinical diagnosis and surgical planning that require detailed anatomical structures. For example, Zhu et al. (2023) proposed Make-A-Volume for 3D MRI translation using 2D slices, which improves the fidelity of 3D medical image translation.

Secondly, MRI encompasses multiple contrast sequences such as T1, T2, FLAIR, PD, and DWI, and each contrast sequence possesses its unique imaging characteristics, reflecting different structures and physiological functions of the human body. Image translation techniques can convert between different contrast domains, including but not limited to T1w to T2w, T1w to FLAIR, T2w to T1w, PDw to T2w, T1 to T1c, T2w to STIR, etc., generating high-quality multi-contrast images, which can contribute to comprehensive analysis and diagnosis (Wang et al., 2023b; Pan et al., 2023a; Liu et al., 2023; Özbey et al., 2023), and fine processing techniques ensure detail preservation, enhancing diagnostic effectiveness. Moreover, the translation technique can also fuse multiple contrast MRI images to provide comprehensive lesion information, improving the diagnostic accuracy (Jiang et al., 2023; Yan et al., 2022). For example, Li et al. (2019) developed a scalable multimodality method to generate T1c using T1, T2 and Flair, which achieved remarkable results.

When dealing with low-field MRI data, low-field MRI devices, although less costly and safer, usually do not have the same image quality as high-field MRI (Figini et al., 2020). Through translation techniques, low-field MRI images can be converted to high-field MRI images, improving the resolution and contrast of the images and thus providing more accurate diagnostic information.

For the problem of deformation in data, MRI images may be affected by patient movement and other factors during acquisition, resulting in image deformation, and translation technology can be used to correct these deformations and improve the quality and reliability of images (Liu et al., 2019a).

In the alignment of multimodality MRI images, image misalignment is also a common problem, and translation technology can be used to avoid it by generating aligned images, thus ensuring the accurate superposition of multimodality images (Lin et al., 2022). Advanced DL algorithms preserve semantic and texture information, ensuring diagnostic consistency and reliability (Mao et al., 2022; Vaidya et al., 2022).

2.2.2. CT intra-modality image translation

CT is renowned for its high-resolution imaging and rapid scanning capabilities, rendering it an essential tool in clinical diagnosis. However, CT scanning confronts several formidable challenges. These include the relatively high radiation dose administered to patients, the presence of noisy metal artifacts that can obscure anatomical details, restricted image quality, and limited datasets (Lu et al., 2010; Tian et al., 2011; Gjesteby et al., 2016; Khodarahmi et al., 2019). Traditional strategies like using contrast enhancers to improve image visibility or adjusting radiation doses to manage the radiation risk come with their own set of drawbacks. For instance, contrast agents may pose risks to patients with certain pre-existing conditions, and altering radiation doses can lead to increased image noise (Yu et al., 2009). Although efforts have been made to correct metal artifacts and implement other post-processing techniques, their efficacy remains constrained, especially in complex clinical scenarios (Lu et al., 2010).

Table 2
Summary of articles on MRI intra-modality image translation.

Paper	Translation type	Methods	Target	Body region	Evaluation metrics
Che et al. (2025)	T1, T2 , Flair → T1CE	CNN, SSM	Reducing contrast agent using	Brain	SSIM, FSIM, PSNR, DSC
Yang and Wang (2025)	Labeled → Unlabeled	CycleGAN	Domain Adaptation	Brain	DSC, AASD, VS
Chen et al. (2024c)	Translation among T1, T2, Flair, PD	CNN, Transformer, SSM	2D data translation	Brain	PSNR, SSIM
Galati et al. (2024)	Multi-source MRI	Federated learning	Intermediate modality translation	Brain	DSC
Atli et al. (2024)	Translation among T1, T2, Flair, PD	CNN, SSM	2D data translation	Brain	PSNR, SSIM
Arslan et al. (2024)	Translation among T1, T2, Flair, PD	Recursive diffusion bridge	2D data translation	Brain	PSNR, SSIM
Zhao et al. (2024)	Masked → Unmasked	Transformer, DiffGAN	2D data translation	Brain	PSNR, SSIM
Chen et al. (2024b)	Translation between T1 and T2	Transformer	2D data translation	Brain	PSNR, SSIM, MAE
Dalmaz et al. (2024)	Translation among T1, T2, Flair, PD	Federated learning, GAN	Generalizable translation	Brain	PSNR, SSIM, FID
Gui et al. (2024)	T1, T2, Flair → T1Gd	Transformer, conditional AR	Obtaining contrast-enhanced MRI	Brain	SSIM, PSNR, DSC
Bevilacqua et al. (2024)	Low-resolution → High-resolution	LDM	Generating high-resolution images	Brain	SSIM, PSNR
Lu and Chen (2024)	Translation between T1 and T2	CycleGAN with MFC-CIT and MVL	Enhancing cross domain information performance	Brain	SSIM, MAE, PSNR, FID, IFC
Xing et al. (2024)	T1, T2 → T1-Gd, T2-Flair	Cross-conditioned diffusion	Modality completion	Brain	PSNR, SSIM, MAE
Friedrich et al. (2024)	Low-Resolution → High-resolution	3D diffusion with IDWT	Overcoming memory issues in generating high-resolution images	Brain	FID, MS-SSIM
Siddiquee et al. (2024)	Unlabeled healthy and diseased image → Healthy images	GAN	Unsupervised learning	Brain	AUC, AUCp, FID
Siam et al. (2024)	Translation between 3T1 MRI and 7T1 MRI	CycleGAN	Improving image resolution	Brain	SSIM, PSNR, NMSE, NMAE
Hamghalam and Simpson (2024)	Low contrast → High contrast	CGANs	Data argumentation	Brain	Acc, DSC, Sens, PPV
Pandey et al. (2024)	T1 → T2	CycleGAN	Multimodality translation	Brain	Subjective evaluation
Xu et al. (2024)	Translation between Lesion samples and disease-free samples	CycleGAN	Data argumentation	Brain	ACC, Recall, F1, AUC, Sens, DSC, mIOU
Kim and Park (2024)	T1 → T2, T2 → Flair, T1 → PD	3D LDM	3D data translation	Brain	PSNR, NMSE, SSIM
Jain et al. (2023)	T1 → T2,Flair	CGAN	3D data translation	Brain	ACC
Wang et al. (2023a)	Translation on different condition	Multiple models with spatial-intensity transform	Matching target image domain	Brain	RMSE, DSSIM, FID, PRD, Age error
Jiang et al. (2023)	Translation between T2, T1ce, T1 and Flair	2D LDM	Multimodality translation	Brain	SSIM, PSNR
Zhu et al. (2023)	T1 → T2	3D LDM	3D data translation	Brain	MAE, SSIM, PSNR
Wang et al. (2023b)	T1→FLAIR, FLAIR→T1, T1→PD, PD→T1	2D DDPM	2D data translation	Brain	MSE, SSIM, PSNR, MI, FID
Pan et al. (2023a)	T1→ T2, T2 → T1, T1 → Flair, FLAIR→T1	CG-DDPM	2D data translation	Brain	MAE, MSSIM, PSNR
Liu et al. (2023)	Missing data imputation among T1, T2, PD	MMT	2D data translation	Brain	SSIM, PSNR, LPIPS
Özbey et al. (2023)	T1 → T2, T2 → T1, T1 → PD, T1 → PD, PD → T1, PD→ T2	SynDiff	2D data translation	Brain	SSIM, PSNR
Yan et al. (2022)	T1 → T2,PD → PD-FS	Swin Transformer-based GAN	Multimodality translation	Brain	MAE, SSIM, PSNR, Experts
Dalmaz et al. (2022)	T1, T2 → PD, T1, T2 → Flair	ResViT	2D data translation	Brain	SSIM, PSNR
Vaidya et al. (2022)	T1 → T2	pTransGAN	2D data translation	Brain	MSE, SSIM, PSNR, LPIPS, UQI, VIF
Mao et al. (2022)	DWI (diffusion weighted images) → T2	AN-GAN	2D data translation	Brain	MSE, SSIM, PSNR, FSIM
Lin et al. (2022)	T1 → T2	NEDNet	2D data translation	Brain	MAE, SSIM, PSNR
Osman and Tamam (2022)	Translation among T1, T2, Flair	2D U-Net	2D data translation	Brain	MAE, MSE, SSIM, PSNR
Kamli et al. (2020)	T1 pre/post, T2, FLAIR → Tumor volume	GAN	Tumor volume prediction	Brain	Recall, precision, DSC
Bui et al. (2020)	Translation between T1 and T2	Flow	2D data translation with unpaired data	Brain	MSE, PSNR, SSIM
Armanious et al. (2020)	Motion-corrupted MRI	End-to-end CGAN with CasNet	Motion correction	Brain	SSIM, PSNR, MSE, VIF, UQI, LPIPS

Table 3
Summary of articles on CT to MRI image translation.

Paper	Translation type	Methods	Target	Body region	Evaluation metrics
Ji and Chung (2024)	Translation between MRI and CT	Diffusion	Intermediate modality translation	Brain	DSC, ASD
Chen et al. (2024b)	Translation between MRI and CT	Transformer	2D data translation	Brain	PSNR, SSIM, MAE
Chen et al. (2024a)	Translation between MRI and CT	ICycle-GAN	Data augmentation	Liver, abdomen	SSIM, PSNR, NMAE, FID
Kang et al. (2023)	Translation between MRI and CT	GAN	2D data translation	Brain	DSC
McNaughton et al. (2023b)	CT → MRI	CycleGAN, U-Net	Multimodality registration	Brain	MAE, MSE, PSNR, SSIM, DSC
Hong et al. (2022)	CT → MRI	U-GAT-IT	2D data translation	Lumbar	SSIM, PSNR, Experts
Feng et al. (2022)	CT → MRI	GAN	2D data translation	Brain	Experts
Paavilainen et al. (2021)	CT → MRI	Pix2pix	Multimodality registration	Prostate	KID, FID, DSC
Kalantar et al. (2021)	CT → MRI	CycleGAN, UNet++	2D data translation	Pelvis	PSNR, SSIM, Experts, MAE, MSE
Kieselmann et al. (2021)	CT → MRI	CycleGAN	Segmentation	Head, neck	DSC, HD, MSD
Dai et al. (2021)	CT → MRI	GAN	Segmentation	Head, neck	DSC, HD95, MSD
Li et al. (2020b)	CT → MRI	CycleGAN, U-Net	Multimodality registration	Brain	MAE, SSIM, PSNR
Li et al. (2020a)	CT → MRI	CycleGAN, Pix2Pix, U-Net	Multimodality registration	Brain	MAE, MSE, SSIM, PSNR
Jiang and Veeraraghavan (2020)	Translation between MRI and CT	VAE	2D data translation, segmentation	Brain	MAE, T-SNE cluster distances
Dong et al. (2019b)	CT → MRI	CycleGAN	Segmentation	Pelvis	DSC, HD, MSD
Rubin and Abulnaga (2019)	CT → MRI	CGAN	Segmentation	Brain	DSC, HD
Jin et al. (2019)	CT → MRI	CycleGAN	2D data translation with paired/unpaired data	Brain	MAE, PSNR, SSIM

The emergence of ML, particularly DL, has opened up promising avenues for addressing these long-standing issues. Learning-based image reconstruction represents a fascinating intersection between ML and image computing. Traditional image reconstruction techniques, firmly grounded in the principles of image computing, rely on mathematical models of image formation and physical laws. However, with the infusion of ML concepts, a new breed of learning-based methods has emerged. These innovative methods harness ML algorithms, particular neural networks, to optimize the reconstruction process. By doing so, they blend the data-driven learning power of ML with the fundamental principles of image formation and processing from image computing (Kang et al., 2017).

In the context of CT image processing, image reconstruction algorithms play a pivotal role in enhancing image quality and resolution. Traditional iterative and statistical reconstruction methods have been effective to some extent. However, they often encounter difficulties when dealing with complex anatomical structures. Examples of such challenging structures include the intricate network of blood vessels in the human vasculature, the complex bone-joint interfaces in the spine, which are characterized by highly variable densities and irregular shapes, and the elaborate soft tissue organ arrangements in the abdomen. These structures have subtle details that are difficult for traditional methods to accurately reconstruct, frequently resulting in artifacts and the loss of fine-scale features (Padole et al., 2015; Niu and Zhu, 2012).

Low-dose CT (LDCT) scanning, which aims to reduce the radiation exposure to patients, typically leads to the introduction of noise and a degradation in image quality. ML algorithms, especially those based on DL, have attracted significant attention for their potential to improve LDCT images. Techniques such as denoising and reconstruction are employed to mitigate the negative impacts of low-dose scans (Harms et al., 2016; Zhang et al., 2013; Huang et al., 2024; Yin et al., 2023). However, traditional ML methods often fall short when faced with CT images that exhibit large scale variations and intricate structures. This is primarily due to their relatively limited ability to extract complex features from the data. DL stands out in the realm of LDCT recovery, mainly because of its data-driven approach. It can automatically learn image features and model parameters. For example, Kang et al. (2017) introduced a deep CNNs framework specifically designed for LDCT reconstruction. This framework combines a deep CNN with a directional wavelet approach, which not only enhances the denoising capabilities but also shortens the reconstruction times.

Furthermore, DL-based intra-modality image translation for CT offers substantial advantages. It can handle multi-contrast domain data, generating high-quality comprehensive images by transforming and fusing different contrast domains (Zhou et al., 2025). This capability is invaluable for comprehensive diagnostic analysis. In challenging body regions, translation techniques can complement and enhance image information, leading to improved clarity, contrast, and more accurate lesion feature identification. Additionally, CT image translation can expand limited datasets. By enhancing data diversity and representativeness, it boosts the training effectiveness and generalization ability of DL models. Numerous DL methods have been proposed to tackle these challenges (Jin et al., 2017; Yi and Babyn, 2018; Lee et al., 2018). For example, Dong et al. (2019a) presented a DL reconstruction framework for incomplete-data CTs. This framework integrates DL techniques with classical Filtered Back-Projection (FBP) reconstruction algorithms, offering potential solutions for X-CTs image reconstruction from incomplete data.

2.2.3. PET intra-modality image translation

PET holds an important role in nuclear medicine imaging, with extensive applications in functional diagnosis and therapeutic monitoring. However, it also faces challenges, notably limited image quality and high radiation dose. The employment of image translation techniques within PET modalities demonstrates significant potential, mainly including transforming non-attenuation-corrected PET (NAC PET) into attenuation-corrected PET (AC PET) (Yang et al., 2019; Shiri et al., 2019) and the mapping of low-count PET to full-count PET (Sanaat et al., 2020; Gong et al., 2018; Kaplan and Zhu, 2019).

NAC PET images often exhibit artifacts and degraded quality due to the absence of attenuation correction. A common remedy is to generate translated CT from NAC PET for attenuation correction during PET reconstruction. Recently, advanced methodologies have been introduced that leverage image translation techniques, particularly DL, to directly convert NAC PET into high-quality AC PET. These methods restore the details and contrast lost due to the lack of attenuation correction, thereby enhancing image quality and diagnostic precision (Yang et al., 2019). Shiri et al. demonstrated promising results by acquiring attenuation-corrected PET images directly from non-attenuation-corrected images using a convolutional encoder–decoder network (Shiri et al., 2019).

Low-count PET, characterized by better motion control and reduced radiation dose, finds wide application in pediatric PET scanning and radiotherapy response assessment. Nevertheless, the statistics of low-count PET can lead to increased image noise, decreased contrast-to-noise ratio and significant bias in uptake measurements. Since reducing radiation dose alters underlying biological and metabolic processes, post-processing operations like denoising alone are ineffective for reconstructing standard or full-count PET from low-count PET. DL image translation techniques offer a solution by mapping low-count PET to full-count PET, generating high-quality images that not only reduce noise but also restore the accuracy of local tracer uptake values. This approach preserves the benefits of low-count PET in terms of motion control and radiation dose while enhancing image clarity and contrast, making it more suitable for pediatric PET scanning and radiotherapy response assessment.

Research on mapping low-count PET to full-count PET has focused on two approaches: image space and projection space. In the image space approach, neural networks are directly applied to PET images to learn and reconstruct image features, effectively improving the image quality, and this is the method used in the majority of studies in comparison. In the projective space approach, operating directly in the space of the original projected data and using techniques like neural networks to process and enhance the projected data can help to generate more accurate and clearer PET images. Sanaat et al. (2020) used an improved 3-dimensional U-Net model to directly process the projected space data, which improves the performance of PET image translation. Both methods have their own advantages and disadvantages. The image space approach is advantageous in its simplicity, requiring fewer computational resources, and has been extensively studied, showing consistent improvements in image quality. However, it has the disadvantage of operating in a lower semantic space, which can lead to the loss of fine details and the potential introduction of artifacts if the neural network does not generalize well to all types of images. On the other hand, the projection space approach offers higher accuracy by operating in a higher semantic space, providing more precise control and higher flexibility to handle a wide range of data and adapt to different imaging conditions. However, it is more computationally intensive, requires more sophisticated algorithms, and typically needs larger and more diverse datasets for training, which can be challenging to obtain.

2.3. Cross modality translation

2.3.1. CT to MRI

The intricacies and significance of medical image translation stem from the fundamental differences between various imaging modalities. CT and MRI represent the two main medical imaging techniques, each offering distinct benefits in terms of visualizing anatomical structures, contrast, and specific lesion depiction. The direct comparison or fusion of images from these modalities is complicated due to their differing physical properties and image acquisition methods. Firstly, CT and MRI exhibit notable disparities in radiation levels, contrast capabilities, and spatial resolution. Secondly, each modality has its own constraints in capturing pathological features and tissue structures. To fully harness the information from both modalities and enhance the overall quality and comprehensiveness of medical images, cross-modality image translation has emerged as a vital tool. As illustrated in Table 3, cross-modality translation from CT to MRI is commonly used for diagnostic aids in clinical diseases (Hong et al., 2022; Kalantar et al., 2021; Feng et al., 2022), multimodal registration (Simonovsky et al., 2016; Li et al., 2020b,a; McNaughton et al., 2023b; Paavilainen et al., 2021), and subsequent medical image segmentation tasks (Dong et al., 2019b; Kieselmann et al., 2021; Dai et al., 2021; Rubin and Abulnaga, 2019).

CT and MRI are indispensable in diagnosing numerous clinical conditions, with MRI often being preferred due to its higher resolution, which aids in differentiating lesions and organs. Nevertheless, there are instances where CT is chosen for clinical assessments, despite its limitations in providing a definitive diagnosis. In such cases, translated MRI images generated from CT images can greatly assist in the comprehensive evaluation of the disease. Hong et al. (2022) developed a GAN-based model to generate MRI images of the lumbar spine from CT images, aiding in the diagnosis and assessment of degenerative spinal disorders. Kalantar et al. (2021) devised a DL framework to generate pelvic T1-weighted MRIs from existing clinical program CT repositories to support MR-Linac therapy. Feng et al. (2022) proposed a cross-modality CT-to-MRI image generation algorithm for acute ischemic stroke by combining radiomics with GAN in order to ensure the golden time for treatment of acute ischemic stroke patients.

For medical imaging datasets, acquiring pairs of MR and CT images is not an easy task. It can take a long time to collect patients scanned by MR and CT scanners. Moreover, alignment between MR and CT images must also be accurate to generate paired MR-CT datasets. Cross-modality image alignment between MRI and CT is challenging because different imaging mechanisms result in a high degree of variability in the appearance of tissues or organs, leading to a lack of common rules for comparing such images. Additionally, CT and MRI images differ in intensities, voxel sizes, image orientations, and fields of view, making multimodality alignment more complex than unimodal alignment (Simonovsky et al., 2016). To address these challenges, researchers have employed GAN as well as CNN to generate MRI from CT, reducing the alignment errors and enriching the anatomical information to aid in subsequent diagnosis (Li et al., 2020b,a; McNaughton et al., 2023b; Paavilainen et al., 2021).

Automated segmentation algorithms are invaluable in clinical diagnostics and radiotherapy treatments, as manual segmentation is time-consuming and subject to operator's knowledge, experience and skill. While DL offers a solution to this problem, training such model usually requires large datasets, which are often scarce in medical imaging. Therefore, translating medical images to augment segmentation algorithms is highly beneficial. Dong et al. (2019b) leveraged the high-quality soft tissue information provided by translated MRI to aid in multi-organ segmentation on pelvic CT images. Kieselmann et al. (2021) utilized a large number of publicly available annotated CT images to generate translated MR images, which were then used to train a CNN to segment the parotid glands on MR images of both the head and the cephalic region. Dai et al. (2021) developed a DL-based automated OAR depiction method to outline organs in head and neck cancer patients. Rubin and Abulnaga (2019) used the generated MR data input to perform ischemic stroke lesion segmentation. This approach of synthesizing artificial examples to support limited training datasets is equally advantageous for numerous other tasks in medical imaging.

When considering CT to MRI image translation, it is essential to note a significant challenge: replicating the soft tissue contrast characteristic of MRI from CT images. CT images predominantly capture bone-soft tissue contrast, and it is uncertain whether all the information required to accurately represent soft tissue contrast in MRI is present in CT. Although numerous studies have explored this translation, some research has shown that this task is highly ill-posed, leading to potentially inaccurate results in soft tissue representation. This limitation must be taken into account when assessing the performance and clinical applicability of CT-to-MRI translation methods. It serves as an area that requires further research and innovation to improve the quality and reliability of such translations.

2.3.2. MRI to CT

Differences in physical properties and image representations between imaging modalities pose challenges for medical image processing and analysis, with CT playing a key role in radiation treatment planning with its excellent display of calcium and bone tissue, while MRI offering unique advantages in lesion diagnosis and evaluation with its high contrast of soft tissue and neural structures.

Table 4
Summary of articles on MRI to CT image translation.

Paper	Translation type	Methods	Target	Body region	Evaluation metrics
Ji and Chung (2024)	Translation between MRI and CT	diffusion	Intermediate modality translation	Brain	DSC, ASD
Atli et al. (2024)	T1, T2 → CT	CNN, SSM	2D data translation	Brain	PSNR, SSIM
Arslan et al. (2024)	T1, T2 → CT	Recursive diffusion bridge	2D data translation	Brain	PSNR, SSIM
Chen et al. (2024b)	Translation between MRI and CT	Transformer	2D data translation	Brain	PSNR, SSIM, MAE
Phan et al. (2024)	MRI → PET/CT	U-Net, transformer, CycleGAN	2D data translation	Brain	MAE, SSIM, PSNR
Das et al. (2024)	MRI → CT	U-Net and GAN	2D data translation	Brain	MI, VIFF, SCD, etc
Luo et al. (2024)	MRI → CT	Target-Guided diffusion	2D data translation	Brain	PSNR, SSIM, FID, Expert
Lu and Chen (2024)	Translation between MRI and CT	CycleGAN with MFC-CIT and MVL	Enhancing cross domain information performance	Abdomen	SSIM, MAE, PSNR, FID, IFC
Hognon et al. (2024)	MRI → CT	GAN with Contrastive learning	Reducing acquisition shift	Brain	MAE, RMSE, SSIM
Pan et al. (2023b)	MRI → CT	Transformer, DDPM	Radiation therapy	Brain	MAE, PSNR, SSIM, NCC
Kang et al. (2023)	Translation between MRI and CT	GAN	2D data translation	Brain	DSC
Karimzadeh and Ibragimov (2023)	MRI → CT	Pix2pix, SwinUNet	Reducing radiation	Brain	MAE, PSNR, SSIM
Li et al. (2023b)	MRI → CT	Frequency-Decoupled diffusion	Maintaining anatomical structure	Brain, pelvis	FID, SSIM, MSE, MAE
Arbabi et al. (2023)	MRI → CT	3D U-Net	3D data translation	Knee	Diagnostic accuracy
Zhao et al. (2023a)	MRI → CT	Self-attention ResUNet	MRI-only radiation therapy	Brain	MAE, Dosimetric
Nijskens et al. (2023)	MRI → CT	Pix2pix	MRI-only radiation therapy	Brain	MAE, GPR
Zhao et al. (2023b)	MRI → CT	Compensation-CycleGAN	MRI-only radiation therapy	Head, neck	MAE, PSNR, SSIM
Sun et al. (2023)	MRI → CT	DU-CycleGAN	MRI-only radiation therapy	Brain	MAE, PSNR, SSIM
Florkow et al. (2022)	MRI → CT	3D U-Net	3D data translation	Hip	Regional analysis
Morbée et al. (2022)	MRI → CT	U-Net	2D data translation	Hip	Regional analysis
Ahangari et al. (2022)	MRI → CT	3D U-Net	Attenuation correction	Whole body	MAE, Regional analysis, Correlation
Zimmermann et al. (2022)	MRI → CT	3D U-Net	MRI-only radiation therapy	Head, neck	MAE, SSIM, Dosimetric
Chen et al. (2022)	MRI → CT	GAN	MRI-only radiation therapy	Head, neck	MAE, Dosimetric
Bambach and Ho (2022)	MRI → CT	Light U-Net, VGG-16 U-Net	Reducing radiation	Head, neck	MAE, MSE
Lyu and Wang (2022)	MRI → CT	Diffusion, Score-Matching	2D data translation	Pelvis	SSIM
Hsu et al. (2022)	MRI → CT	Pix2pix	MRI-only radiation therapy	Prostate	MAE, PSNR, SSIM,Dosimetric
Park et al. (2022)	MRI → CT	Pix2pix	MRI-only radiation therapy	Brain	MAE, SSIM, Dosimetric
Chourak et al. (2022)	MRI → CT	BDM,GAN	MRI-only radiation therapy	Prostate	MAE, ME, MAPE, DSC
Wang et al. (2022)	MRI → CT	CycleCUT	MRI-only radiation therapy	Brain	MAE, PSNR, SSIM
Paavilainen et al. (2021)	MRI → CT	Pix2pix	2D data translation	Prostate	KID, FID, DSC
Morbée et al. (2021)	MRI → CT	3D U-Net	3D data translation	Lumbar	Regional analysis
Jans et al. (2021)	MRI → CT	3D U-Net	3D data translation	Sacroiliac ioint	Diagnostic accuracy
Willemsen et al. (2021)	MRI → CT	CNN	Reducing radiation	Lower arm	Surgical Planning Errors
Shi et al. (2021)	MRI → CT	GAN,CNN	MRI-only radiation therapy	Brain	MAE, PSNR, SSIM
Olberg et al. (2021)	MRI → CT	GAN	MRI-only radiation therapy	Abdomen	MAE, DSC
Gholamiankhah et al. (2021)	MRI → CT	ResNet, GAN	MRI-only radiation therapy	Brain	MAE, SSIM, PSNR
Kang et al. (2021)	MRI → CT	U-Net,CycleGAN	MRI-only radiation therapy	Pelvis	MAE, RMSE, PSNR, SSIM
Bourbonne et al. (2021)	MRI → CT	Pix2pix	MRI-only radiation therapy	Brain	Dosimetric
Liu et al. (2021a)	MRI → CT	GAN, CNN	MRI-only radiation therapy	Brain	Dosimetric, Registration
Yuan et al. (2021)	MRI → CT	ResU-Net	MRI-only radiation therapy	Brain	ME, MAE, MSE

(*continued on next page*)

Table 4 (*continued*).

Paper	Translation type	Methods	Target	Body region	Evaluation metrics
Brou Boni et al. (2021)	MRI → CT	AugCGAN	MRI-only radiation therapy	Pelvis	MAE, Dosimetric
Song et al. (2021)	MRI → CT	U-Net, CycleGAN	MRI-only radiation therapy	Head, neck	MAE, SSIM, PSNR
Koh et al. (2021)	MRI → CT	CGAN	Attenuation correction	Brain	MAE,DSC
Kazemifar et al. (2020)	MRI → CT	GAN	MRI-only radiation therapy	Brain	MAE
Maspero et al. (2020)	MRI → CT	CGAN	MRI-only radiation therapy	Brain	MAE, Dosimetric
McKenzie et al. (2020)	MRI → CT	CycleGAN	Multimodality registration	Head, neck	Registration
Emami et al. (2020a)	MRI → CT	Attention-GAN	MRI-only radiation therapy	Brain	MAE
Peng et al. (2020)	MRI → CT	CGAN	MRI-only radiation therapy	Head, neck	MAE, Dosimetric
Yang et al. (2020)	MRI → CT	CAE-GAN	Multimodality registration	Head, neck	MAE, PCC, SLPD
Masoudi et al. (2020)	MRI → CT	CGAN	Segmentation	Abdomen	Segmentation
Fu et al. (2020)	MRI → CT	CGAN, CycleGAN	MRI-only radiation therapy	Abdomen	MAE, Dosimetric
Cusumano et al. (2020)	MRI → CT	Pix2pix	MRI-only radiation therapy	Thorax	MAE, ME, Dosimetric
Boni et al. (2020)	MRI → CT	Pix2pixHD	MRI-only radiation therapy	Pelvis	MAE
Liu et al. (2020)	MRI → CT	U-Net	MRI-only radiation therapy	Abdomen	MAE, Dosimetric
Qian et al. (2020)	MRI → CT	RU-ACGAN	Attenuation correction	Abdomen	MAPD, RMSE, CC
Kläser et al. (2019)	MRI → CT	CNN	Attenuation correction	Brain	MAE, PET reconstruction
Olberg et al. (2019)	MRI → CT	Pix2pix, U-Net, ASPP	MRI-only radiation therapy	Thorax	RMSE, SSIM, PSNR,Dosimetric
Lei et al. (2019)	MRI → CT	CycleGAN	MRI-only radiation therapy	Brain	MAE, PSNR, NCC
Liu et al. (2019b)	MRI → CT	CycleGAN	MRI-only radiation therapy	Prostate	MAE, Dosimetric
Wang et al. (2019)	MRI → CT	U-Net	MRI-only radiation therapy	Head, neck	MAE, ME
Kläser et al. (2018)	MRI → CT	CNN	Attenuation correction	Brain	MAE, PET reconstruction

Translating MRI to CT images can provide physicians with more comprehensive and enriched information to improve the accuracy of radiation treatment planning and perform more precise lesion localization and evaluation. DL models like GANs and diffusions are widely used for image translation tasks across modalities from MRI to CT. These models excel at learning and capturing the intricate nonlinear relationships between modalities, generating realistic images that possess corresponding modality features. Furthermore, some studies have combined traditional image processing techniques to enhance the quality and alignment of translated images.

Cross-modality conversion techniques from MRI to CT have shown significant value in a number of areas, as illustrated in Table 4, including assisted diagnosis (Kazemifar et al., 2020; Zimmermann et al., 2022), MRI-only radiotherapy (Sun et al., 2023; Yang et al., 2018; Shi et al., 2021; Olberg et al., 2021; Nijskens et al., 2023; Gholamiankhah et al., 2021), attenuation correction (Ahangari et al., 2022; Qian et al., 2020; Kläser et al., 2018, 2019), and image alignment (McKenzie et al., 2020; Yang et al., 2020). Firstly, it can be used to aid in tumor contouring during radiation treatment planning. By providing more accurate information about anatomical structures, translated CT enables physicians to locate tumor margins more precisely, providing more precise targets for radiotherapy. Secondly, CT translated from MRI can be used for organ localization, especially for some complex anatomical structures like head and neck, providing clearer auxiliary information to enhance the precision of radiotherapy (Kazemifar et al., 2020; Zimmermann et al., 2022; Maspero et al., 2020; Chen et al., 2022). In addition, translated CT can be used for dose calculation, the process of calculating the radiation dose during radiation treatment planning, ensuring accurate tissue density information and consequently, the precision and safety of radiation doses (Sun et al., 2023; Yang et al., 2018; Shi et al., 2021; Olberg et al., 2021; Nijskens et al., 2023; Gholamiankhah et al., 2021), which is critical to improving treatment outcomes and reducing damage to normal tissue.

In terms of attenuation correction, translated CT is indispensable for certain medical applications, such as PET imaging. Since the X-ray uptake of tissues is density-dependent and therefore requires accurate attenuation correction of the CT image, MRI cannot provide attenuation information for tissues, translated CT technology generates virtual images with CT-like density information from MRI, enabling accurate attenuation correction for PET images (Ahangari et al., 2022; Qian et al., 2020; Kläser et al., 2018, 2019). This attenuation correction is essential for calculating radiotherapy dose distributions and supporting precise treatment planning.

Furthermore, medical image translation also plays an important role in image alignment. Direct MR-CT image alignment is a challenging task due to the differences in contrast and characteristics between two modalities. Translated CT technology overcomes this challenge by translating MRI images into virtual CT images, which share similar density information with actual CT images. This virtual CT image serves as a bridge for matching MR images with real CT images, simplifying the task of MR-CT image alignment and improving both the accuracy and efficiency of MR-CT image alignment. Studies have shown that the method of translating CT from MRI outperforms traditional direct MR-CT image alignment methods in terms of accuracy and stability (McKenzie et al., 2020; Yang et al., 2020).

2.3.3. MRI to PET

MRI to PET translation has emerged as a promising approach to overcome the clinical limitations of PET imaging. PET acquisition presents inherent challenges: high operational costs and invasive radioactive tracer injections that induce radiation exposure risks and

Table 5
Summary of articles MRI to PET image translation.

Paper	Translation type	Methods	Target	Body region	Evaluation metrics
Phan et al. (2024)	MRI → PET/CT	U-Net, transformer, CycleGAN	2D data translation	Brain	MAE, SSIM, PSNR
Das et al. (2024)	MRI → PET/CT	U-Net and GAN	2D data translation	Brain	MI, VIFF, SCD, etc
Vega et al. (2024)	MRI → PET	CGAN	Reducing radiation collection	Brain	SSIM, PSNR, Expert
Jang et al. (2023)	MRI → PET	TauPETGen(LDM)	2D data translation	Brain	MMSE
Xie et al. (2023)	MRI → PET	JDAM	2D data translation	Brain	PSNR, SSIM
Sun et al. (2022)	PET, MRI → PET	bi-c-GAN	Generating high-quality data	Head	PSNR, NMSE, SSIM, CNR
Hussein et al. (2022)	MRI → PET	3D CNN	Measuring CBF	Brain	PSNR, SSIM
Zhang et al. (2022)	MRI → PET	STFNet	Denoise	Brain	RSME, PSNR, SSIM, PCC
Rajagopal et al. (2022)	MRI → PET	3D residual U-Net	3D data translation	Whole body	AC
Sikka et al. (2021)	MRI → PET	GLA-GAN	Multimodality diagnostics for AD	Brain	MAE, SSIM, PSNR
Wang et al. (2021a)	PET, MRI → PET	CNN	Generating F-FDG PET diagnostic images	Whole body	PSNR, NMSE, SSIM
Lin et al. (2021)	MRI → PET	3D RevGAN	Multimodality diagnostics for AD	Brain	RMSE, PSNR, SSIM
Hu et al. (2021)	MRI → PET	BMGAN	2D data translation	Brain	MAE, PSNR, MSSSIM, FID
Kao et al. (2021)	MRI → PET	U-Net with CCA, ESIT	2D data translation	Brain	MAE, PSNR, SSIM
Emami et al. (2020b)	MRI → PET	FREA-Unet	2D data translation	Brain	MAE, SSIM, PSNR
Shin et al. (2020)	MRI → PET	GANBERT	2D data translation	Brain	PSNR, SSIM, RSME
Wei et al. (2019)	MRI → PET	Sketcher-Refiner GAN	Measuring tissue myelin content	Brain	MSE, PSNR
Sun et al. (2019)	MRI → PET	Flow, VAE	Reducing radiation	Brain	PSNR, SSIM
Sikka et al. (2018)	MRI → PET	3D U-Net	Multimodality diagnostics for AD	Brain	MAE, PSNR, SSIM
Wang et al. (2018)	PET, MRI → PET	3D LA-GANs	Generating high-quality data	Brain	PSNR, SSIM

potential adverse reactions. These practical constraints critically restrict PET accessibility, resulting in limited baseline datasets (typically < 200 paired cases) in most medical centers. By generating PET images from widely available MRI scans, this computational strategy aims to democratize PET-derived functional information while mitigating clinical risks. Nevertheless, the scarcity of high-quality PET-MRI paired data remains a key bottleneck, particularly for modeling rare pathologies where PET ground truth acquisition is ethically constrained.

To overcome these limitations, as illustrated in Table 5, researchers have proposed various models to translate PET images from MRI data, a promising approach given the absence of anatomical information in PET scans. Notable models for PET translation include the frequency-aware attention Unet (FREA-Unet) proposed by Emami et al. (2020b), the 3D end-to-end translation network called Bidirectional Mapping GAN (BMGAN) proposed by Hu et al. (2021), a new joint diffusive attention model with a joint probability distribution and an attention strategy called JDAM proposed by Xie et al. (2023), and 3D Residual U-Net proposed by Rajagopal et al. (2022).

There are three main types of PET imaging: AV45, AV1451, and FDG. AV45 is used to measure amyloid protein uptake in the brain, crucial for studying biomarkers linked to neurological disorders, especially Alzheimer's disease. AV1451 assesses tau protein aggregation, aiding in the understanding of neurodegenerative disease progression. FDG, as a fluorodeoxyglucose tracer, provides comprehensive information on glucose metabolism patterns, revealing neurological function and disease states. The choice of PET imaging techniques depends on the needs of the specific research or clinical needs and the focus on specific biomarkers. Researchers have also developed methods to translate these three types of PET images. Shin et al. (2020) used Transformers' bidirectional encoder representation (BERT) algorithm to generate AV45-PET, Jang et al. (2023) proposed TauPETGen, a novel LDM-based text-conditional image translation method to generate tau PET, and Wang et al. (2021a) used CNN to generate F-FDG PET.

While diagnostic-quality PET images typically require a full dose of tracer, radiation exposure poses potential health risks, especially for patients undergoing multiple scans. To address radiation concerns, some researchers have attempted to reduce the tracer dose during PET scans (e.g., by using half the full dose), although this increases noise and decreases imaging detail and quantification, compromising image quality (Alessio et al., 2012). Therefore, the translation of high-quality PET images (Sun et al., 2022; Wang et al., 2018) from low-dose images and the denoising (Zhang et al., 2022) of PET are of great clinical importance.

In addition, translated PET images aid clinical diagnosis. For instance, predicting PET-derived myelin content maps from multimodality MRI is valuable for understanding multiple sclerosis (MS) pathophysiology, monitoring disease progression, evaluating treatment efficacy, reducing treatment costs, and avoiding radiation exposure (Wei et al., 2019). In Alzheimer's disease (AD), a chronic and progressive neurological disorder, translating PET images from multimodality MRI for AD classification enhances diagnosis by leveraging complementary information (Sikka et al., 2021, 2018; Lin et al., 2021).

2.4. Label-based image translation

Constraining segmentation maps can enhance the precision and realism of translated medical images. This technique finds several applications in medical image translation (Jin et al., 2018). For instance, in CT image translation, segmentation maps guide generator networks to model organ shapes and locations more accurately. Similarly, in MRI image translation, they aid in generating images with precise tissue structures and organ boundaries. This approach is also valuable for simulating images under diverse pathological conditions, thereby providing realistic and diverse data for medical research and education.

Conditional segmentation maps can be generated from GANs or from pre-trained segmentation networks by treating generation as a two-stage process (Costa et al., 2017; Guibas et al., 2017). Mok and Chung (2019) introduced an innovative automated data enhancement method that utilizes a CGAN to generate high-resolution MRI images for brain tumor segmentation. This method operates in a coarse-to-fine manner, with a particular emphasis on ensuring clear tumor boundaries in the translated images.

Boschet et al. (2024) performed conversion between annotated and unannotated modalities, utilizing the SysDiff framework to integrate GANs and diffusion models. By leveraging GANs to model the

transitions between denoising process steps, they enhanced sampling efficiency and mode coverage. Ultimately, pseudo-labels were generated, providing useful initial masks for the annotation process of medical images.

Additionally, research has delved into harnessing weakly supervised labels, like classification tags, to direct models. In clinical practice, image-level categorical labels (indicating normal or diseased states) are more prevalent than detailed segmentation masks. Consequently, investigating methods to leverage these weak labels for medical image segmentation holds immense value. Boschet et al. (2024) introduced a GAN-based approach that initializes lesion areas using Class Activation Mapping (CAM) guided by weakly supervised labels. They subsequently refined lesion boundaries via a Complementary Branch Fusion Network (CBFNet), while preserving semantic consistency in non-lesion regions. Furthermore, the incorporation of a cycle consistency loss ensured the reversibility of the mappings, bolstering the robustness of both image generation and segmentation. Ultimately, their method achieved performance on par with fully supervised baselines, albeit typically requiring supplementary model training to effectively utilize the weak labels.

3. Applications

3.1. Aid diagnosis

Medical image translation contributes significantly to enhancing diagnostics by furnishing physicians with additional insights and support. Specifically, this technology provides powerful assistance in the diagnosis of a variety of diseases: In lung cancer diagnosis, translated CT images facilitate clearer visualization of lung cancer lesions, thereby promoting early detection and in-depth lesion analysis. In cardiovascular disease treatment, by generating MRI and CT images, doctors can achieve more precise diagnoses and treatment planning. These high-resolution cardiac images enable the identification of atherosclerosis, myocardial infarction and other lesions (Florkow et al., 2022; Jans et al., 2021; Morbée et al., 2022). For cerebral neurological diseases, enhanced MRI images improve the quality of brain imaging, aiding doctors in the precise localization and analysis of tumors, strokes, and other neurological conditions. In orthopedic diseases, translated X-ray, CT or MRI images provide clear depictions of bone structures, supporting the diagnosis of orthopedic conditions and facilitating surgical planning (Kalantar et al., 2021; Hong et al., 2022; Feng et al., 2022). In breast cancer detection, translated mammography and MRI images are utilized to aid in early detection.

These varied applications highlight the importance of medical image translation in improving physicians' comprehension of patient conditions, formulating treatment strategies, and boosting diagnostic accuracy (Morbée et al., 2021). Through generating realistic translated images, medical image translation provides new perspectives and tools to aid in diagnosis.

3.2. Missing data

Medical image translation can serve as an effective measure to supplement missing data by converting existing modality data into the missing modality data, thereby alleviating the issue of missing data, particularly the common problem of missing modalities in multimodality analysis. In MRI and CT scans, patient images may have incomplete data due to various factors (Wang et al., 2023b; Pan et al., 2023a). Medical image translation technology can reconstruct the missing portions, thereby restoring the full MRI or CT image. This ensures that physicians have access to comprehensive anatomical information, enabling accurate diagnoses. Similarly, PET images may suffer from missing data owing to scanning time limits or patient movement. With medical image translation, the missing parts of the PET image can be filled in to provide more complete metabolic information, facilitating precise disease assessment.

In radiation treatment planning, certain areas of the patient may not have complete image information due to postural changes or other factors. Medical image translation technology can generate virtual images of missing areas for better planning of radiation treatment programs. In addition, super-resolution translation technology can generate images with higher resolution through medical image translation, allowing physicians to discern anatomical structures and lesions with greater clarity (Dalmaz et al., 2022; Lin et al., 2022; Osman and Tamam, 2022).

These illustrative applications underscore the versatility and efficacy of medical image translation in managing missing data. By supplementing incomplete information, this technology is poised to enhance the integrity and utility of medical images, offering more reliable support for medical diagnosis and treatment.

3.3. Attenuation correction

Medical image translation is pivotal in attenuation correction for PET imaging, particularly when actual CT images are unavailable. Accurate attenuation correction is essential for obtaining accurate metabolic information in PET scan. Since PET and CT are usually performed jointly, virtual CT images can be generated by translation techniques for performing attenuation correction of PET images (Kläser et al., 2018, 2019). This method learns the intricate relationship between MR and CT during training, producing translated CT images that accurately depict tissue density and attenuation characteristics. These translated CTs can then be directly utilized for PET attenuation correction, thereby enhancing the quantitative accuracy of PET imaging.

Additionally, medical image translation can address motion artifacts in PET imaging. The relatively long scanning time in PET can lead to patient movement, resulting in artifacts that degrade image quality. By generating CT images and integrating them with motion correction techniques, these artifacts can be effectively mitigated, improving overall imaging quality. Taken together, the application of medical image translation for PET attenuation correction offers a robust tool to enhance the precision and accuracy of PET images, with significant potential for advancements in clinical practice (Qian et al., 2020; Ahangari et al., 2022).

3.4. Multimodality registration

Medical image translation is significant for multimodality image alignment, as it enhances alignment accuracy by generating translated images with consistent features and corresponding structures.

In brain diseases diagnosis, the fusion of MR and CT images is essential to obtain more comprehensive information. Medical image translation enables the creation of translated CT images that closely resemble actual MR images in structure and feature, facilitating improved image contrast and precise lesion localization. The fusion of PET and MR images offers a wealth of biological and anatomical information. Medical image translation aids in generating translated MR images that correspond to PET images, simplifying the alignment process and enabling accurate localization of abnormal areas, thereby boosting the accuracy of disease diagnosis. In nuclear medicine, the joint use of SPECT and CT images is widespread. Medical image translation generates translated CT images with comparable structural and anatomical details, ensuring precise alignment of SPECT and CT images. This, in turn, aids in the exact localization of radiotracer distribution (McKenzie et al., 2020; Yang et al., 2020). In addition, medical image translation plays an important role in the fusion of Ultrasound(US) and MRI. By producing translated MR images with similar structural and morphological characteristics, it facilitates accurate alignment of US and MR images, providing finer guidance for navigational procedures. To address the significant differences between various imaging modalities,

it is effective to convert these diverse modalities into a unified intermediate pseudo-modality (Ma et al., 2024). This approach substantially reduces the modality gap and simplifies the associated registration challenges. A more in-depth exploration of this concept and its connection to "intermediate" translations can be found in Section 7.3.

In conclusion, the application of medical image translation in multimodality image alignment offers a powerful tool for fusing different imaging modalities, and it is anticipated to become increasingly important in clinical multimodality image analysis.

3.5. MRI-only radiation therapy

In radiation treatment planning, CT images are conventionally required for dose calculation and treatment plan design. Notably, MRI provides richer soft-tissue contrast, making it particularly suitable for tumor treatment in the brain, pelvis, and other areas. Consequently, utilizing MRI for treatment planning has become an appealing alternative.

Medical image translation has emerged as a solution to facilitate radiation treatment planning using MRI by generating translated CT images, thereby eliminating the need for actual CT acquisitions (Liu et al., 2019b; Brou Boni et al., 2021; Song et al., 2021; Peng et al., 2020). This approach not only streamlines the process but also minimizes patient radiation exposure while preserving the benefits of MRI. For example, some studies have used DL methods such as GANs to translate MRI images into corresponding CT images for brain radiation treatment planning. By training these networks on a vast dataset of paired MRI and CT images, the generated CT images can closely mimic the structural and morphological characteristics of real CT images (Hsu et al., 2022; Brou Boni et al., 2021). This enables physicians to carry out radiation treatment planning directly on MRI images, without the necessity of additional CT scans.

The application has demonstrated promising results in radiation therapy for various regions, including the brain and pelvis. Through medical image translation, MRI-only radiation therapy emerges as a more convenient and patient-centric approach, while ensuring the precision and viability of treatment plans. The advancement of this technology offers an innovative pathway to enhance the accuracy and patient experience in radiation therapy.

3.6. Segmentation

Medical image translation has diverse applications in segmentation tasks, offering tough support for image analysis and diagnosis.

Firstly, in tumor segmentation, researchers leverage medical image translation techniques to generate MRI images containing different types of tumors by methods such as GANs. These translated images enhance the training datasets for tumor segmentation algorithms, thereby improving their generalization and robustness. For organ segmentation, medical image translation can generate translated images with distinct organ boundaries. By conditioning on segmentation maps, CT or MRI images with varying anatomical structures can be translated, providing a rich set of training samples for organ segmentation models (Fu et al., 2020; Dong et al., 2019b; Rubin and Abulnaga, 2019). In the case of lesion segmentation, medical image translation aids in modeling the wide range of lesion characteristics, including shape, size, and location. The synthesized images significantly enhance the ability of lesion segmentation algorithms to generalize across different lesion features.

In addition, medical image translation facilitates the fusion of multimodality information. By translating images from different modalities into a single modality, a more comprehensive dataset is provided for the segmentation task, ultimately improving the accuracy of the segmentation model.

Lastly, medical image translation also serves as a valuable tool for data augmentation. By generating additional images, the original training data can be expanded, enabling the model to better adapt to complex clinical scenarios and enhancing its robustness (Kieselmann et al., 2021; Dai et al., 2021).

Overall, these applications highlight the versatility and effectiveness of medical image translation in segmentation tasks. By translating diverse datasets, researchers and physicians can gain deeper insights into various clinical challenges, ultimately improving image analysis and automated diagnosis.

4. Enabling technologies

4.1. Fundamental architectural paradigms

4.1.1. CNN-based models

CNNs (LeCun et al., 1998), built upon compact filters, have long been the cornerstone of computer vision tasks. As shown in Fig. 2, their convolutional layers, sliding small filters over images, excel at detecting local features like edges and textures. This local processing prowess allows CNNs to precisely represent the fine-grained details in medical images, which is crucial for tasks such as identifying small lesions or subtle anatomical structures.

However, CNNs encounter significant limitations when dealing with long-range context. Their local receptive fields restrict their ability to capture dependencies and relationships over large distances within an image. In medical imaging, understanding the long-range context is often essential for grasping the overall structure of organs or the relationship between different anatomical regions. For example, in a full-body CT scan, traditional CNNs may struggle to analyze the connection between different organs, as they are not well-equipped to handle long-range dependencies.

In the medical image translation field, U-Net (Ronneberger et al., 2015) and ResNet (He et al., 2016) are among the most widely-used CNN-based architectures. The attention-enhanced U-Net, exemplified by the SARU (Self-attention ResUNet) framework (Zhao et al., 2023a), demonstrates superior capability in refining local feature extraction for MRI-to-CT translation. By integrating dual attention mechanisms(spatial and channel attention modules) into a hybrid U-Shaped architecture with residual connections, this model achieves precise tissue boundary reconstruction (e.g., skull margins and cutaneous interfaces) critical for BNCT dose calculation. ResNet, with its residual learning paradigm, was further optimized in Zhao et al. (2023a) via the Attentional ResBlock integrating depthwise separable convolution. This architecture reduced parameter counts by 8× compared to standard U-Net while maintaining high-resolution reconstructions, effectively mitigating vanishing gradients.

Moreover, recent research (Atli et al., 2024) reveals another significant shortcoming of CNNs in medical image translation: their bottleneck in modality conversion. Different-modality medical images, such as MRI and CT, have distinct imaging principles and feature representations. CNNs have difficulty effectively capturing and fusing these highly different features, which leads to a negative impact on the quality and accuracy of the converted images. For example, when converting MRI images to CT images, CNN models often fail to accurately restore the characteristic features of CT images. This results in the converted images having lower diagnostic accuracy, as important CT-specific features like bone density representation may be lost or misrepresented.

4.1.2. Transformer-based models

Transformers (Vaswani et al., 2017), which rely on attention-driven filters, have revolutionized the field of natural language processing and have also made significant inroads into computer vision, including medical image translation. As shown in Fig. 2, the self-attention mechanism in transformers allows the model to weigh the importance of different parts of the input sequence (or in the case of images, different patches) when making predictions. This results in enhanced sensitivity

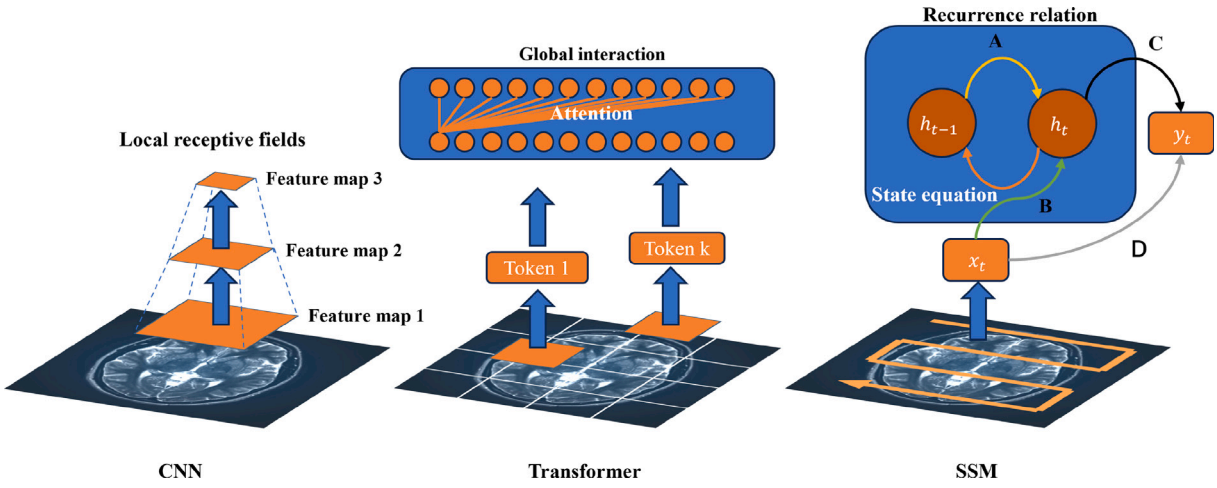

Fig. 2. The core mechanism comparison among CNN, Transformer and SSM.

to long-range context, as the model can directly capture relationships between distant elements in the image.

Nevertheless, transformers are not without drawbacks. The self-attention mechanism, while powerful, comes with a high computational cost due to its quadratic complexity with respect to the sequence length (Öztürk et al., 2024). This can make training and inference time-consuming, especially when dealing with large medical images. Additionally, in the context of small-scale datasets, transformers may overfit due to their high capacity and the complex nature of the self-attention operation. This overfitting can lead to poor generalization performance, where the model performs well on the training data but fails to generalize to new, unseen medical images.

Vision Transformer (ViT) (Dosovitskiy et al., 2020) and Swin Transformer (Liu et al., 2021b) are two prominent transformer-based architectures applied in medical image analysis. In Dalmaz et al. (2022), residual connections were incorporated into ViT architecture to enhance multimodality medical image synthesis capabilities. Through the residual vision transformers framework, ResViT effectively captured cross-modality dependencies in tasks like MRI-to-CT translation, demonstrating superior anatomical structure preservation compared to conventional CNN-based methods. Swin Transformer, with its hierarchical architecture and shifted window attention, reduces the computational complexity while still maintaining the ability to capture both local and global information. The Swin Transformer-based GAN (Yan et al., 2022) introduced a U-shaped architecture with shifted window self-attention blocks, achieving state-of-the-art(SOTA) performance in multimodality medical image translation tasks. This hybrid architecture combined the structural modeling advantages of Swin Transformer with GAN's adversarial training strategy, generating images with higher SSIM scores and finer pathological details compared to pure CNN implementations.

4.1.3. State space model

In the domain of medical image translation, State Space Models (SSMs) (Gu et al., 2021) offer a novel and powerful approach. SSMs are based on the concept of representing a system's state over time or space. As shown in Fig. 2, they use a state vector to summarize the internal state of the system at a given moment, and this state evolves according to a set of state-transition equations. In the context of medical image translation, SSMs treat medical images as a sequence of states. For example, in the case of a time-series of MRI images, each time-step's image can be considered a state of the system. The state-transition equations in SSMs are designed to capture the relationships between consecutive images in the sequence, allowing the model to understand how the image features change over time. This is different from traditional methods that may struggle to model these temporal or long-range dependencies effectively. Compared to CNNs constrained by local receptive fields and transformer sensitive to the quadratic complexity of self-attention operators, SSMs achieve a paradigm-shifting equilibrium between capturing global contextual information and preserving local feature precision through recurrent modeling with linear complexity. This characteristic endows them with unique clinical value, enabling efficient processing of long-term dynamic data (e.g., MRI temporal sequence reconstruction) while precisely maintaining anatomical details of subtle lesions in CT images, thus establishing a theoretical framework for real-time medical image analysis and diagnostic decision-making.

Recent studies substantiate the core advantages of SSMs in medical image processing. Taking LDCT denoising as an exemplar, the DenoMamba (Öztürk et al., 2024) framework demonstrates for the first time that a unified SSM architecture can simultaneously capture multi-scale spatial correlations of anatomical structures and complementary inter-channel feature relationships through innovative fusion of spatial-channel state-space modeling. This approach not only overcomes the denoising limitations of traditional models in complex tissue regions but also elevates performance to new heights while preserving pixel-level details, revealing SSMs' potent adaptability to heterogeneous medical imaging features. Similarly, in multimodality medical image translation, frameworks like I2I-Mamba (Atli et al., 2024) significantly enhance missing modality imputation quality by integrating channel mixing mechanisms with cross-modality long-range dependency modeling through SSM layers, while retaining convolutional modules' sensitivity to local features. Such studies demonstrate that SSMs can either serve as standalone architectures to surpass performance bottlenecks of traditional methods or synergize with existing technologies to create complementary advantages, thereby unlocking novel possibilities for core medical imaging tasks including reconstruction, registration, and segmentation.

Despite their immense potential, SSMs remain in their infancy for medical image translation applications. Current research predominantly focuses on backbone architecture validation, leaving significant gaps in data diversity (e.g., model generalizability across multi-center and cross-device contexts), task compatibility (e.g., 3D volumetric data processing and multi-task joint optimization), and theoretical interpretability (e.g., mapping relationships between state transition processes and anatomical features). Future advancements must prioritize three directions: First, developing adaptive mechanisms to address data heterogeneity stemming from varying dose levels and acquisition protocols; Second, integrating self-supervised learning paradigms to mitigate reliance on paired medical data annotations; Third, exploring synergistic frameworks with emerging generative technologies like diffusion models, where the fusion of SSMs' deterministic modeling and

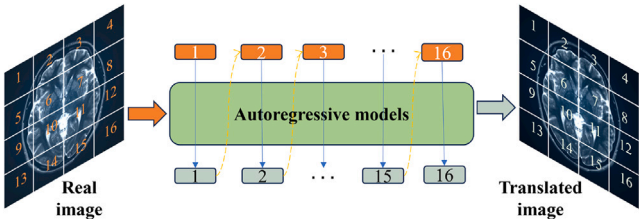

Fig. 3. The basic architecture of AR for medical image translation.

probabilistic generative processes could propel medical image translation towards ultra-high fidelity. These breakthroughs will critically influence the practical clinical utility of intelligent imaging diagnostic systems.

In the analysis of 140 papers included in our survey, CNNs remain the most prevalently adopted architecture, appearing in 92 studies as foundational components of their model frameworks. Transformers, on the other hand, have experienced a significant surge in adoption starting from 2020, with 46 papers incorporating transformer-based architectures in recent years. Notably, SSMs represent an emerging trend, with 6 studies (Che et al., 2025; Zhou et al., 2025; Chen et al., 2024c; Öztürk et al., 2024; Atli et al., 2024; Huang et al., 2024) published since 2024, demonstrating their potential as a novel alternative for sequence modeling tasks. These findings highlight the evolutionary trajectory of medical image translation research, where CNNs maintain dominance while newer architectures like transformers and SSMs are gaining traction due to their unique capabilities in capturing long-range dependencies and improving computational efficiency.

4.2. Generative modeling paradigms

4.2.1. Autoregressive model

AR model (Van Den Oord et al., 2016) is a prevalent technique for modeling data distributions and generating new samples. As shown in Fig. 3, it posits that the probability distribution of generated data can be expressed as a linear combination of past observations. In the context of a first-order AR model (AR(1)), the tth observation x_t is generated based on its predecessor x_{t-1} and an error term ε_t

$$x_t = \phi \cdot x_{t-1} + \varepsilon_t, \tag{1}$$

where ϕ denotes the autoregressive coefficient, signifying the linear relationship between current and past observations, and ε_t represents the error term adhering to a specific distribution.

AR find extensive application in generative tasks, including language modeling and image generation. For instance, in medical image generation, AR can be employed to generate high-resolution PET images from low-dose counterparts by predicting each pixel value sequentially using already generated pixels. This approach ensures image coherence and consistency, enabling the recovery of high-quality images while minimizing patient radiation exposure.

AR learn the data distribution from training samples to generate new instances that conform to this distribution. The loss function is formulated as

$$\mathcal{L}_{\mathrm{AR}} = \sum_{t=1}^{T} \log P(x_t | x_{<t}), \tag{2}$$

where $x_{<t}$ represents all past observations up to time t, and $P(x_t|x_{<t})$ is the conditional probability of x_t given these observations.

In medical image translation, AR offer the advantages of producing high-quality, coherent images through their pixel-by-pixel generation process. They are versatile and applicable to various medical imaging tasks. Moreover, the sequential generation provides better interpretability. However, AR models also suffer from high computational complexity and lengthy training times, particularly for high-resolution images. Their heavy reliance on historical data makes them prone to overfitting with limited datasets. Additionally, the inherent sequential nature of the generation process results in slower performance, rendering them less suitable for real-time or interactive clinical applications.

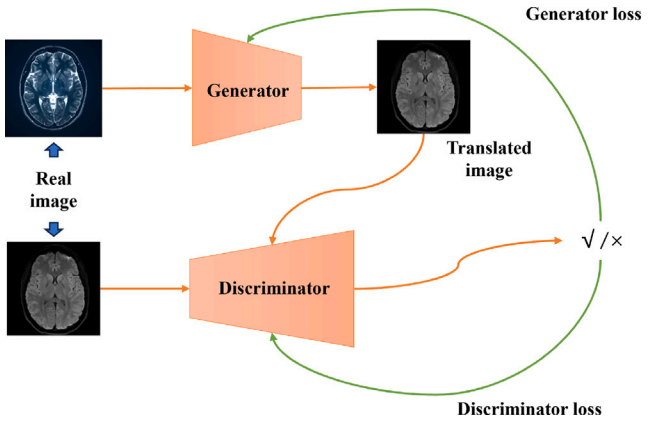

Fig. 4. The basic architecture of GAN for medical image translation.

4.2.2. Generative adversarial network

GAN (Goodfellow et al., 2014) is a DL framework that consists of two parts, Generator (G) and Discriminator (D) networks, aiming at generating realistic samples by adversarial learning. As shown in Fig. 4, the Generator network is responsible for generating "fake" samples similar to the real samples, while the Discriminator network is focused on distinguishing the real and generated "fake" samples. They compete with each other by adversarial learning to achieve the goal of training a generative model that can generate high-quality samples. The overall objective of GAN is

$$\begin{aligned}\min_G \max_D \mathcal{L}_{\mathrm{GAN}}(G, D) =& \mathbb{E}_{x\sim p_{\mathrm{data}}(x)}[\log D(x)] \\ &+ \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))],\end{aligned} \tag{3}$$

where x represents the real data sample, $p_{\mathrm{data}}(x)$ represents the real data distribution, z represents the noise, and $p_z(z)$ represents the potential spatial distribution.

Numerous GAN variants, such as CGAN (Mirza and Osindero, 2014), Pix2pix (Isola et al., 2017), DualGAN (Yi et al., 2017), StarGAN (Choi et al., 2018), CycleGAN (Zhu et al., 2017), UNIT (Liu et al., 2017), MUNIT (Huang et al., 2018), AguGAN (Almahairi et al., 2018), BigGAN (Brock et al., 2018), and GANILLA (Hicsonmez et al., 2020), have emerged. GANs have demonstrated transformative potential in medical image translation through their distinct architectural advantages. CycleGAN addresses the scarcity of paired medical data by leveraging unpaired cycle-consistent learning for cross-modality translation (e.g., CT-MRI synthesis). Pix2pix, built on conditional adversarial frameworks, enables precise pixel-level alignment in tasks requiring strict structural correspondence, such as lesion segmentation, albeit with inherent dependency on paired datasets. StarGAN extends multi-domain translation capabilities via unified domain-label encoding, efficiently synthesizing diverse pathological states (e.g., normal-to-diseased lung tissue) to augment diagnostic datasets. StyleGAN enhances cross-device generalization by disentangling anatomical content (e.g., organ morphology) from stylistic imaging attributes (e.g., scanner-specific textures).

However, persistent challenges, including training instability (mode collapse), insufficient cross-domain generalization, and incoherent 3D volume generation, necessitate integrating diffusion models for enhanced fidelity, contrastive learning for domain adaptation, and

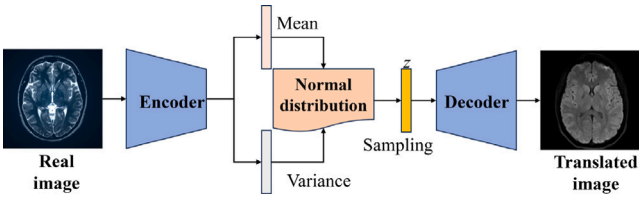

Fig. 5. The basic architecture of VAE for medical image translation.

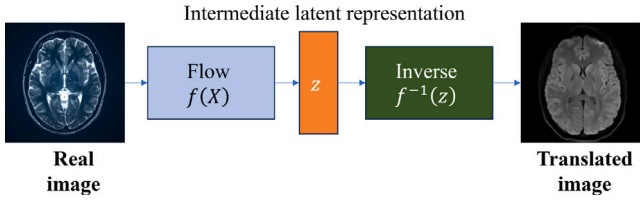

Fig. 6. The basic architecture of flow model for medical image translation.

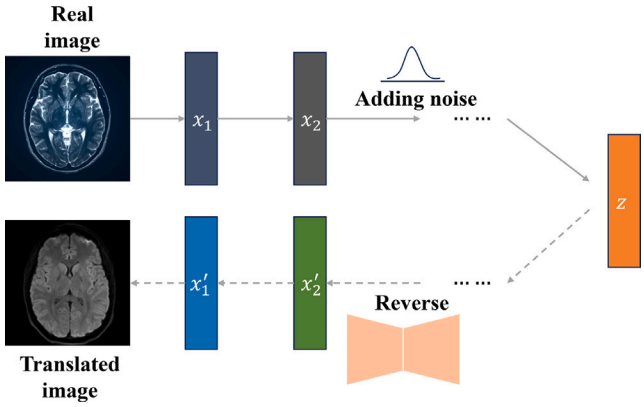

Fig. 7. The basic architecture of diffusion for medical image translation.

lightweight deployment strategies. Addressing these limitations is critical to facilitating the reliable translation of medical imaging technologies from laboratory research to clinical deployment.

4.2.3. Variational auto-encoder

VAE (Kingma and Welling, 2013) is a generative model that combines the ideas of self-encoder and probabilistic graphical models. The main goal of VAE is to learn the latent representation of the data, i.e., to learn the hidden variable representation of the data distribution.

As shown in Fig. 5, training VAE consists of two main stages, encoder training and decoder training. In encoder training, real samples are fed into the encoder network, and the mean and variance parameters of the latent distribution are estimated. Then, the hidden variables are sampled from the latent distribution and fed into the decoder network for reconstruction. In the training of the decoder, reconstructed samples are compared with the original ones, and the decoder network's parameters are adjusted by error back propagation to improve the reconstruction quality as much as possible. The VAE's overall objective is

$$\mathcal{L}_{\text{VAE}} = -\mathbb{E}_{q(z|x)}[\log p(x|z)] + \text{KL}[q(z|x)||p(z)], \tag{4}$$

where x represents the input data, $q(z|x)$ denotes the encoder, z denotes the latent variable, $p(x|z)$ denotes the decoder, and $p(z)$ denotes the prior distribution for z.

Currently, there are many variants of VAE, such as GMVAE (Dilokthanakul et al., 2016), VaDE (Jiang et al., 2016), VQ-VAE (Van Den Oord et al., 2017), S-VAE (Davidson et al., 2018), and S3VAE (Zhu et al., 2020), among others. VAEs and their advanced variants offer distinct advantages for medical image translation through architectural innovations. GMVAE leverages Gaussian mixture priors to model multimodality data distributions, enabling cross-modality translation with enhanced semantic coherence. VaDE integrates variational inference with deep clustering, disentangling anatomical content from imaging artifacts to improve cross-device generalization. VQ-VAE employs discrete latent codes via vector quantization, preserving high-frequency details (e.g., vascular structures) while mitigating blurry outputs. S-VAE constrains latent spaces to geometric manifolds (e.g., hyperspheres), optimizing 3D medical volume generation through intrinsic geometric priors. S3VAE introduces hierarchical latent variables to dynamically model temporal or pathological progression. Despite their mathematical interpretability and diversity generation via probabilistic sampling, VAEs face challenges including blurred reconstructions (due to KL-divergence constraints) and training instability, which necessitate hybrid frameworks (e.g., VAE-GAN) or diffusion-enhanced priors. Future advancements may focus on causal reasoning for intervention simulation and lightweight architectures for real-time clinical deployment.

4.2.4. Flow model

The Flow Model (Dinh et al., 2014) is a probabilistic model that simulates data distributions by learning the data generation process. In Fig. 6, flow model starts with a prior distribution (e.g., Gaussian or uniform) and transforms it into a posterior distribution through a series of invertible transformations, such as affine and orthogonal functions. These transformations form a transformation flow, consisting of multiple layers that map prior samples to output samples while preserving dimensionality and probability distribution. Specifically, the process begins with sampling initial data from the prior distribution. This data serves as a latent representation of the input image. Subsequently, these data are gradually transformed within the transformation flow. During the forward transformation, the data pass through each transformation layer in sequence, resulting in a representation in the intermediate latent space. This representation is closely linked to the distribution of the input medical image through invertible transformations. For the reverse transformation, it leverages the invertibility of the transformations. Starting from the representation in the intermediate latent space, it passes through the inverse transformation functions in reverse order. Ultimately, the target medical image is generated, achieving the conversion from a medical image of one modality or with certain features to another.

Training is done via maximum likelihood estimation, tuning model parameters to maximize the likelihood of generated data. Inference involves mapping samples from the posterior back to the prior through inverse transformations. The loss function is given by

$$\mathcal{L}_{\text{Flow}} = -\mathbb{E}_{p(x)}[\log q(z)] + \text{KL}[q(z)||p(z)], \tag{5}$$

where x is the input data, z is the latent variable, and $p(x)$ and $q(z)$ are their respective probability distributions.

Advantages of the Flow Model include its interpretability due to reversible transformation functions and high-quality, diverse generated samples. In medical image translation, it effectively maps source image features (e.g., MRI) to target image features (e.g., CT) in the hidden space.

Challenges include the large model capacity required for multiple invertible transformations, leading to longer training times, and the computational overhead of computing inverse transformations for large models.

4.2.5. Diffusion model

Diffusion models (Sohl-Dickstein et al., 2015) is powerful in medical image translation. Differing from flow models that operate in the

hidden space, they directly work on the image space, generating data by simulating the gradual evolution of data points over continuous time steps. As shown in Fig. 7, beginning with an initial prior distribution, a reversible diffusion operator is employed in a repetitive manner. This iterative process of applying the operator systematically morphs the prior distribution into a posterior distribution that closely mimics the characteristics of the actual real data.

During training, diffusion models rely on a distinct training mechanism known as score matching (Huang et al., 2021). The forward diffusion process gradually adds noise to the original image x_0, with

$$x_t = \sqrt{(\overline{\alpha}_t)}x_0 + \sqrt{(1-\overline{\alpha}_t)}\epsilon, \tag{6}$$

where $\overline{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$, β_s is the noise schedule parameter, and $\epsilon \sim \mathcal{N}(0, I)$. In the reverse process of diffusion models, the goal is to estimate x_{t-1} from x_t. According to the principles of diffusion, the conditional probability $P\left(x_{t-1} \mid x_t\right)$ is related to the prediction of the added noise. Given the predicted noise $\epsilon_\theta(x_t, t)$, we can estimate x_{t-1}. The loss function,

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

is based on score matching. Since the score function $\nabla_{x_t} \log p(x_t)$ is proportional to the negative of the noise term $\epsilon_\theta(x_t, t)$, the loss function can be interpreted as a weighted score matching objective (Song et al., 2020). In contrast to other inference models, such as GANs which use adversarial loss between a generator and a discriminator, or VAEs which optimize a variational lower bound related to reconstruction error and KL divergence, diffusion models' score-matching-based loss function directly focuses on the accuracy of noise prediction in the forward and reverse diffusion processes. This allows diffusion models to better capture the complex data distributions in medical image modalities and generate high-quality translated images.

Diffusion models can be categorized into different types. Denoising Diffusion Probabilistic Models (DDPMs) (Ho et al., 2020) learn a multi-step denoising transformation. Their forward process adds noise to the target image, and the reverse process, guided by the source image, denoises it. They can recover image structure from noisy data but may face issues as the denoising transformation may not match the required source-to-target one. Score-based Models model the data distribution's gradient (score function) and use Langevin dynamics for sampling (Song and Ermon, 2019). They are flexible in handling complex data distributions in medical image modalities and can work with different data types. Latent Diffusion Models (LDMs) (Rombach et al., 2022), such as Stable Diffusion, are a type of diffusion model. While diffusion models generally work on the image space, LDMs operate in a low-dimensional latent space. They first compress the high-dimensional images into a low-dimensional latent representation using a VAE. In this latent space, the diffusion process takes place, which reduces computational complexity. After the diffusion in the latent space, the latent representation is then decoded back into the image space. This allows for efficient high-quality translations between modalities. Recent research, such as the Self - Consistent Recursive Diffusion Bridge (SelfRDB) (Arslan et al., 2024), advances diffusion models in medical image translation. SelfRDB's novel forward process with a source-modality soft-prior better captures source characteristics, improving generalization and information transfer. Its self-consistent recursive estimation procedure enhances reverse-step sampling accuracy, overcoming issues like weak source-guidance and sub-optimal sampling in traditional models.

Diffusion models in medical image translation achieve high-quality conversions by adding and removing Gaussian noise over multiple steps. This preserves details and reduces noise. Their advantages include interpretability, the ability to capture time-dependent changes, and good performance in sample generation and interpolation. However, diffusion models have challenges. Their training and inference are computationally complex and time-consuming, and their performance depends on the diffusion operator and parameter settings.

Among the 140 articles included in our study, the GAN architecture is the most prevalently used, with 92 articles adopting it, and it is frequently applied in cross-modality translation. The diffusion model started to emerge in 2022, and up to now, 27 articles have utilized it, indicating a new trend. As an established architecture, the VAE has been used in 17 articles, mainly around 2020. The flow and AR architectures have been adopted in relatively few studies. Specifically, only 1 article was found to use the AR architecture (Gui et al., 2024), and 3 articles used the flow architecture (Sun et al., 2019; Bui et al., 2020; Le et al., 2021).

4.3. Hybrid architectures

The hybrid architecture gradually becomes a new trend in medical image translation. As illustrated in Table 6, each architecture has its own characteristics, and the core of hybrid architecture lies in integrating the complementary advantages of diverse DL paradigms, thereby overcoming the unique challenges encountered during the process of medical image translation. The translation tasks involve transforming medical images from one modality to another, which requires addressing complex challenges such as multi-scale feature modeling, modality alignment, and high-fidelity synthesis. Multi-scale feature modeling necessitates that the model be capable of capturing a wide range of image features from the microscopic to the macroscopic levels. In medical images, both the subtle tissue textures and the larger organ structures contain vital diagnostic information. Modality alignment is dedicated to resolving the disparities between different imaging modalities, ensuring that the anatomical structures and functional information in the images can accurately correspond during the conversion process. High-fidelity translation, meanwhile, aims to generate images that are highly similar to real medical images, which is of utmost importance for applications such as auxiliary diagnosis and surgical planning.

In the process of tackling these challenges, the integration of basic architectures is a critical component. Combining the local feature extraction ability of CNNs with the global context modeling ability of like Transformers and SSMs has become the cornerstone of medical image analysis. CNNs, with their convolutional and pooling layers, can effectively extract local detailed features in images, such as the edges and textures of lesions. However, CNNs have certain limitations in capturing the relationships between distant elements in the image and the overall context information. In contrast, transformers and SSMs, can model the image information globally and comprehend the long-range dependencies among different parts of the image. For example, when analyzing brain MRI images, transformer can capture the spatial relationships between different brain regions, which are essential for diagnosing brain diseases. Therefore, by combining the two, the model cannot only accurately identify local lesions but also grasp the overall condition comprehensively, significantly enhancing the accuracy and reliability of medical image analysis (Öztürk et al., 2024; Atli et al., 2024).

In addition to the integration of basic architectures, the fusion of different generative models has also brought about new breakthroughs in medical image translation, such as the integration of diffusion models and GANs. Based on the principle of step-by-step denoising, the diffusion model gradually restores clear images by performing reverse diffusion on noisy images over multiple time steps, ensuring the structural stability of the generated images. GANs, on the other hand, enhance the perceptual authenticity of the images through the adversarial training between the generator and the discriminator. Özbey et al. (2023) proposes a novel method SynDiff based on adversarial diffusion modeling for medical image translation, demonstrating superior performance compared to GAN and diffusion models in multi-contrast MRI and MRI-CT translation. Zhao et al. (2024) presents a Local Vision Transformer based adversarial diffusion model for accelerating MRI

Table 6
Generative models for medical image translation.

Model	Generator type	Generated image quality	Training stability	Training speed	Advantage	Disadvantage
AR	CNN/RNN /Transformer	High	Medium	Fast	High quality of generated images, parallelized training	Localized generation, faster training
VAE	CNN/RNN /Transformer	Medium	Medium	Fast	Generate better diversity, better mathematical interpretability	The generated images may be blurrier than GAN and the training process may be more complicated
GAN	CNN/Transformer	High	Low	Slow	Excellent performance in image-to-image tasks. High image quality and diversity	Training instability, mode crash issues
Flow	CNN/RNN /Transformer	High	High	Slow	Accurate likelihood estimation, high image quality and diversity	High computational cost and design constraints of the network structure
Diffusion	CNN/Transformer	High	High	Slow	High image quality and diversity	Slow generation process and resource intensive training process

reconstruction, which performs outstandingly on multiple datasets. These studies have further advanced the development of medical image translation technology.

Furthermore, transformer-based diffusion models, such as DiT (Peebles and Xie, 2023), fully exploit the advantages of the attention mechanisms of the transformer. When dealing with large-sized and high-resolution medical images, these mechanisms can effectively capture the long-range dependencies in the images. For example, when analyzing whole-body PET-CT images, DiT can focus on the potential connections between different organs, assisting doctors in having a more comprehensive understanding of the patient's condition. Pan et al. (2023b) proposes a 3D Shifted-window (Swin) Transformer-based Denoising Diffusion Probabilistic Model (MC-IDDPM) for generating sCT from MRI. Hybrid models like ControlNet (Zhang et al., 2023) introduce control signals. By utilizing edge maps or other feature maps, they can guide the model to perform more accurate image translation. In medical image translation, this means that it is possible to convert images from one modality to another more accurately while preserving the key anatomical structures and lesion information. Multimodality hybrid models further expand the boundaries of medical image analysis. By integrating text-based medical information or other modality information, such as medical records and diagnostic reports, these models can provide more semantic information for image analysis, thereby improving the accuracy of diagnosis (Chen et al., 2024b). For example, by combining the patient's medical history and symptom descriptions, multimodality hybrid models can interpret medical images more accurately and avoid misdiagnosis and missed diagnosis.

Looking ahead, the research directions of the hybrid architecture are likely to focus on two key areas. On the one hand, lightweight hybrid design has attracted significant attention, especially with the aid of neural architecture search technology. Neural architecture search automatically searches for the optimal network structure within a vast architectural space, which can reduce the number of model parameters and computational complexity while ensuring model performance. This is of great significance for deploying DL models on medical devices with limited resources, such as portable US diagnostic devices, enabling real-time image analysis and diagnosis. On the other hand, task-specific customization for rare disease groups will also become a research hotspot. Due to the small number of patients and complex disease characteristics of rare diseases, traditional general models often struggle to make accurate diagnoses. By constructing hybrid models tailored to rare diseases and combining the medical image data and clinical knowledge of specific diseases, the accuracy and pertinence of diagnosis can be improved. For example, for certain genetic diseases, constructing hybrid models using multimodality data (such as genetic data and medical image data) is expected to uncover the early characteristics of the diseases, enabling more precise diagnosis and treatment.

5. Performance indicators

In this section, we delve into the performance indicators in the field of medical image translation. As shown in Fig. 8, these indicators can be classified into three categories, based on the specific image properties they accentuate and the associated evaluation methodologies:

(1) Intensity-based metrics focus on the pixel-level information of the image. This category encompasses the mean square error (MSE), peak signal-to-noise ratio (PSNR), mean absolute error (MAE), structural similarity index (SSIM), feature similarity index (FSIM), learned perceptual image patch similarity (LPIPS), Inception Score (IS), and Fréchet Inception Distance (FID). These metrics mainly assess the similarity or difference between the original and translated images from aspects like pixel values, brightness, and contrast.
(2) Geometric fidelity metrics center around the shape and structural features of the image. Metrics in this category include the normalized cross-correlation (NCC), Hausdorff distance (HD), and high-visibility signal-to-noise ratio (HVSNR). These metrics mainly assess the similarity or difference between the original and translated images from aspects related to shape, structure, and how well the geometric properties are preserved.
(3) Statistical metrics, such as the Pearson correlation coefficient (PCC), Mutual Information(MI) are utilized for statistical analysis. In medical image translation, it can help analyze the relationship between different features in the original and translated images, providing insights into the overall relationship between the two images from a statistical perspective.

Each type of these metrics offers unique advantages in evaluating medical image translation. Intensity-based metrics are excellent for quickly and quantitatively assessing the basic performance of a translation method at the pixel level. They can precisely capture fine-grained details in the image, which is crucial for applications like image reconstruction and denoising. For example, in CT image denoising, intensity-based metrics can effectively measure how well the denoising algorithm preserves the original intensity values of different tissues. Geometric fidelity metrics are indispensable when it comes to evaluating the preservation of image shape and structure. In tasks like organ segmentation and multimodality image registration, maintaining the geometric integrity of organs is of utmost importance. Geometric fidelity metrics can accurately detect any structural changes or misalignments, ensuring the reliability of subsequent clinical applications. Statistical metrics, on the other hand, provide a global perspective on the relationship between the original and translated images. They can reveal hidden patterns and associations that are not easily observable through other metrics. This helps researchers understand the overall consistency and generalizability of the translation method across different datasets or

Table 7
Common metrics used to evaluate medical image translation.

Evaluation methods	Characterization	Application scenarios
SSIM	Consider the structure, brightness and contrast of the image	Image compression, image enhancement, image restoration
PSNR	Pixel-based error accumulation	Image compression, super resolution
Experts	Expert evaluation	Medical imaging, image restoration
MAE	Average absolute error	Generalized image quality evaluation
DSC	Dice similarity coefficient, a measure of shape overlap	Medical image segmentation
HD	Hausdorff distance, maximum distance between farthest points	Medical image segmentation
MSD	Average surface distance	Medical image
PCC	Pearson correlation coefficient	Statistical analysis, image registration
Dosimetric	dosimetric evaluation	radiotherapy program
IS	Inception score	Image generation
FID	Frechet inception distance	Image generation
HVSNR	Human visual system-based signal-to-noise ratio	High dynamic range images
LPIPS	Learned perceptual image patch similarity	Image compression, super resolution, image generation
MI	Mutual information	Statistical analysis, image registration

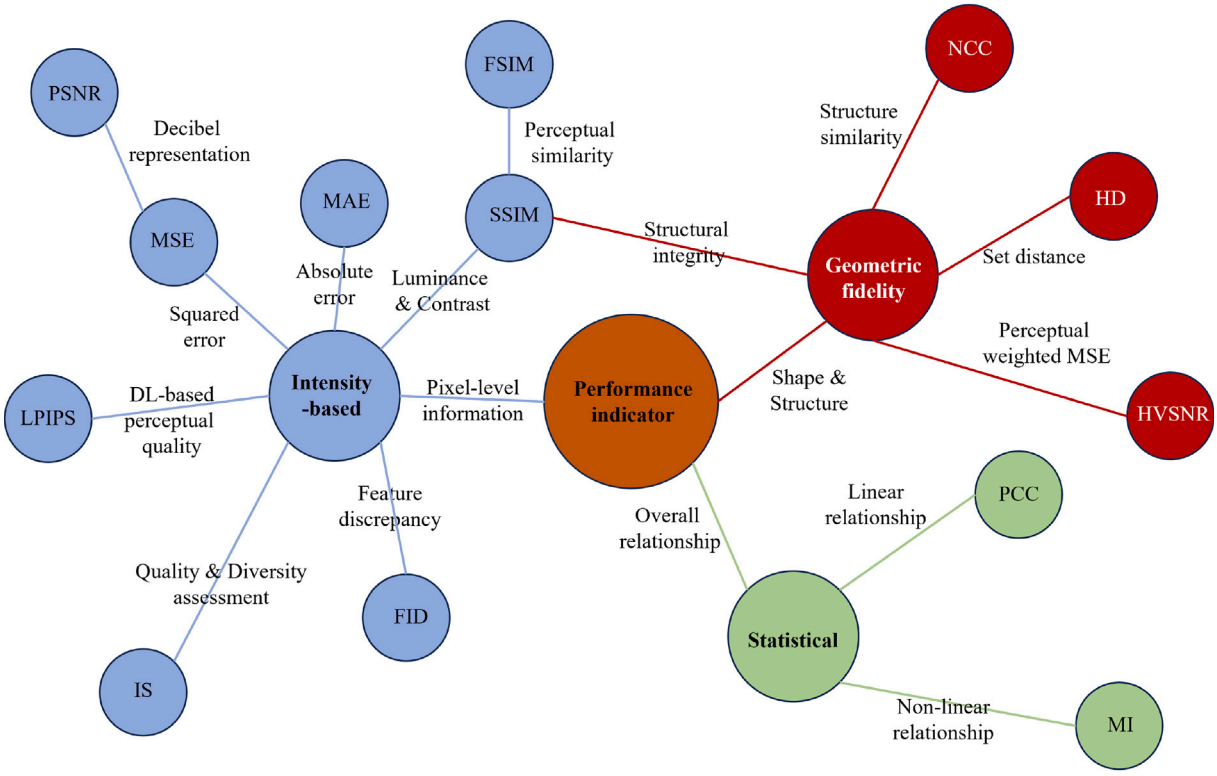

Fig. 8. Overview of metrics for medical image translation.

patient populations, facilitating more informed decisions in algorithm development and optimization.

By comprehensively considering these three types of metrics, we can conduct a more thorough and accurate evaluation of the geometric, structural, and statistical characteristics of the translated images in comparison to the originals. This multi-dimensional evaluation is essential for a comprehensive understanding of the performance of medical image translation algorithms. In Table 7, the commonly used indicators are summarized here.

5.1. Intensity-based metrics

5.1.1. MSE

MSE (Wang and Bovik, 2009) is a basic metric for evaluating the performance of statistical models and ML algorithms. It measures the average squared difference between predicted and actual values. Mathematically, it is defined as

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{8}$$

where n is the number of observations, y_i is the actual value, and $\hat{y}_i$ is the predicted value.

In image quality assessment, MSE quantifies the discrepancy between the generated and original images, aiding in algorithm comparison and quality monitoring. However, MSE's sensitivity to outliers may necessitate supplementary evaluation methods for a comprehensive assessment.

5.1.2. PSNR

PSNR (Wang et al., 2004) is a widely used metric for evaluating image and video quality, typically expressed in decibels (dB). It is computed based on the MSE and is defined as follows:

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{(2^b-1)^2}{\mathrm{MSE}}\right), \tag{9}$$

where b denotes the number of bits of the image pixel data type, for example, in an 8-bit unsigned integer (uint8) type image, $b = 8$, and (2^b-1) is 255; MSE represents the mean squared difference between the original and the processed image pixels. So PSNR can be considered as

a decibel (dB) representation of the MSE, aimed at providing a more distinguishable numerical scale for assessing image quality.

A higher PSNR value indicates a closer resemblance between the processed image and the original, thereby suggesting superior image quality. Generally, PSNR values fall within the range of 20 to 50 dB, with values exceeding 30 dB often deemed indicative of good image quality. However, PSNR is also an average reflection of image quality just like MSE and may not capture all facets of image quality. Consequently, in many instances, a comprehensive evaluation necessitates the incorporation of additional metrics alongside subjective assessments to fully gauge the overall image quality.

5.1.3. MAE

MAE (Wang et al., 2004) is a variant of MSE. It evaluates the accuracy of regression models by averaging the absolute differences between actual observations and model predictions. Its mathematical formula is given by

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{10}$$

where n denotes the number of observations, y_i represents the actual value of the ith observation, and $\hat{y}_i$ represents the model's predicted value for the ith observation. A lower MAE value indicates higher model accuracy, and MAE is robust to outliers since it considers only the absolute errors.

5.1.4. SSIM

SSIM (Wang et al., 2004) is a unique metric that evaluates both intensity-based and structural aspects of image similarity. It comprehensively assesses the similarity between a processed image and its reference, considering luminance, contrast, and structural integrity, making it more closely aligned with human perception of image quality.

SSIM is often grouped with intensity-based metrics due to its consideration of pixel-level information like luminance and contrast. However, it also significantly incorporates structural aspects through the covariance term in its formula, which is a key factor bridging the gap between intensity-based and geometric fidelity evaluations.

It is usually calculated between −1 and 1, where 1 means that the two images are exactly the same, 0 means that there is no similarity, and −1 means that they are completely different.

The mathematical formula for SSIM is

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{11}$$

where μ_x and μ_y denote the means of images x and y, respectively, reflecting the average luminance of the two images; σ_x and σ_y represent the standard deviations of images x and y, which correspond to the respective contrasts of the images; σ_{xy} signifies the covariance between images x and y, reflecting the joint variation relationship between their pixel values and indicating the structural integrity; and C_1 and C_2 are small constants introduced for stability, typically taking on very small positive values. Based on the aforementioned information, SSIM can comprehensively reflect the similarity between two images from multiple perspectives.

5.1.5. FSIM

FSIM (Zhang et al., 2011) is an improvement based on SSIM, placing particular emphasis on perceptual similarity. It evaluates image quality by extracting a comprehensive set of perceptual features, encompassing image structure, texture, contrast, brightness, and more. To achieve this, FSIM employs a multi-scale Gabor filter for feature extraction, mimicking the way the human eye perceives images across various scales. By considering feature similarities at multiple scales, FSIM accounts for the fact that image quality issues can manifest at different resolutions. Ultimately, it integrates these feature similarity measurements across scales to produce a definitive image quality score. The mathematical formula for FSIM is

$$\mathrm{FSIM}(I,K) = \frac{2\cdot\mu_I\cdot\mu_K + C_1}{\mu_I^2 + \mu_K^2 + C_1}\cdot \frac{2\cdot\sigma_{IK} + C_2}{\sigma_I^2 + \sigma_K^2 + C_2}\cdot\frac{\rho_{IK} + C_3}{\sigma_I\cdot\sigma_K + C_3}, \tag{12}$$

where I and K represent the two input images respectively, μ_I and μ_K represent the local mean of images I and K, σ_I and σ_K represent the local variance of images I and K respectively, σ_{IK} represent the local covariance of images I and K. ρ_{IK} represent the local correlation coefficients of images I and K, and C_1, C_2, and C_3 are constants used to stabilize the calculation and are used to avoid the case where the denominator is zero.

5.1.6. LPIPS

LPIPS (Zhang et al., 2018) is an intensity-based similarity metric designed to enhance image quality assessment and measure image similarity. It primarily focuses on perceptual quality and relies largely on intensity information of images.

The core idea of LPIPS is to leverage deep neural networks to mimic human visual perception, thereby more accurately gauging the similarity between images. This method employs pre-trained CNNs, such as VGG or ResNet, to extract feature representations of image blocks. These features capture various visual attributes of an image, including both low-level visual features (closely related to pixel intensities) and high-level semantic features.

By comparing the feature representations of image chunks, LPIPS computes the perceptual distance, which better reflects how human vision perceives the similarity based on the intensity-related aspects of the images. Through a careful selection of the feature extraction layer and the computation of perceptual distance, LPIPS can be customized to meet the specific requirements of different tasks, offering more accurate and reliable metrics in the field of image processing. The mathematical formulation of LPIPS is given by

$$\mathrm{LPIPS}(I,K) = \sum_l w_l\cdot\left\|\phi_l(I) - \phi_l(K)\right\|, \tag{13}$$

where I and K denote the two input images, $\phi_l(I)$ and $\phi_l(K)$ represent the feature representations of images I and K at layer l, respectively, ϕ_l signifies the feature extraction function at layer l, and w_l denotes the weight assigned to the features at layer l. Note that the Euclidean norm ($\|\cdot\|$) is used to measure the distance between the feature representations.

5.1.7. IS

The Inception Score (IS) (Salimans et al., 2016) is an intensity-based metric designed to assess the quality and diversity of images generated by GANs. It mainly relies on intensity information from the images and focuses on pixel-level details.

IS leverages the feature extraction capabilities of CNNs (Inception network) to compute the category distribution of the generated images. The core principle of IS is to quantify the quality of generated images by evaluating their diversity and fidelity based on intensity-related aspects. A higher IS score indicates that the generated images exhibit greater diversity and fidelity in terms of intensity features.

One of the advantages of IS is its ability to operate without requiring labeled real images or reference adversarial samples, thus making it widely applicable. This simplicity allows for a quick quantitative assessment of how well the image generation process preserves intensity-related features. For example, in medical image translation scenarios such as generating synthetic X-ray images from ultrasound data, IS can rapidly gauge whether the generated X-ray images have a diverse and accurate distribution of intensity values, similar to real X-ray images.

However, IS is not a universal evaluation metric. It may show insensitivity in certain specific image generation tasks. The results obtained

using IS can be influenced by the choice of the generative model and the Inception network used for feature extraction. Consequently, in practical applications, IS is often combined with other evaluation metrics to provide a more comprehensive evaluation of image quality.

The mathematical formulation of the Inception Score is given by

$$\mathrm{IS}(G) = \exp\left(\mathbb{E}_{\mathbf{x}\sim G}\left[D_{\mathrm{KL}}\left(p(y|\mathbf{x}) \,\middle\|\, p(y)\right)\right]\right), \tag{14}$$

where G denotes the image generator that produces the image $\mathbf{x}$, $p(y|\mathbf{x})$ represents the conditional probability distribution of the label y given the generated image $\mathbf{x}$, $p(y)$ denotes the marginal probability distribution of the label y, and D_{KL} denotes the Kullback–Leibler divergence, which measures the discrepancy between the two probability distributions.

5.1.8. FID

FID (Heusel et al., 2017) is an intensity-based metric for evaluating the similarity between generated images and real images. Fundamentally, it is based on the intensity information of images, evaluating the statistical similarity between feature distributions derived from image intensities, although it can indirectly reflect some geometric properties.

Its core concept involves measuring the distributional differences between generated and real images by comparing their feature distributions. A lower FID score indicates that the distribution of the generated image is more similar to that of the real image in terms of intensity-related features, thus serving as an indicator of the generative model's performance.

As an intensity-based metric, FID has the advantage of simplicity in evaluating how well the image generation process preserves intensity-related features. It does not require labeling or image pairing of real images, but only two data distributions. This makes FID an effective tool widely used in image generation tasks, such as medical image translation for generating synthetic MRI from CT, to quickly and quantitatively assess the quality and fidelity of the generated images in terms of intensity.

However, the results of FID may be affected by the choice of feature extraction network and the nature of the generation task. Different pre-trained networks may extract different intensity-related features, leading to variations in FID scores. So, care needs to be taken in choosing the appropriate settings when using it.

The mathematical formula for FID is

$$\mathrm{FID}(P, Q) = \|\mu_P - \mu_Q\|^2 + \mathrm{Tr}(C_P + C_Q - 2\sqrt{C_P \cdot C_Q}), \tag{15}$$

where P and Q denote the real image distribution and the generated image distribution; μ_P and μ_Q denote the mean vectors of P and Q, respectively; C_P and C_Q denote the covariance matrices of P and Q, respectively; Tr is a trace calculation operation. These mean vectors and covariance matrices reflect the center position and variation of the intensity-based feature distributions.

5.2. Geometric fidelity metrics

5.2.1. NCC

NCC (Wang et al., 2004) is a robust technique employed in image processing and computer vision to gauge the similarity between two signals or images, particularly in the context of template matching. By computing the normalized cross-correlation, NCC effectively mitigates the impact of signal amplitude variations, thereby constraining the similarity metric within the interval $[-1, 1]$. Specifically, -1 indicates complete negative correlation, 1 signifies completely identical, and 0 means no linear correlation. This method incorporates the mean and standard deviation of the signals, yielding a comprehensive similarity measure.

Mathematically, NCC is expressed as

$$\mathrm{NCC}(X, Y) = \frac{\sum_{i=1}^{n}(X[i] - \mu_X) \cdot (Y[i] - \mu_Y)}{\sigma_X \cdot \sigma_Y}, \tag{16}$$

where X and Y represent the two images or signals under comparison, μ_X and μ_Y denote their respective means, and σ_X and σ_Y indicate their standard deviations. This formulation ensures that NCC remains invariant to linear transformations of the signals, such as scaling and shifting, making it particularly suitable for applications involving images with varying brightness and contrast. Especially in medical imaging, due to factors such as imaging equipment and scanning angles, the brightness and contrast of the same organ in different images may vary. However, the relative geometric structure of the organ remains stable. NCC can accurately assess the similarity of organs in different images while ignoring these intensity changes. Its core lies in focusing on the shape and structural features of the images, rather than merely paying attention to the pixel intensity values.

5.2.2. HD

Hausdorff Distance (HD) (Baudrier et al., 2007) is used to quantify the similarity between two sets. It is determined by identifying the greatest distance from any point in one set to the nearest point in the other set, and then taking the maximum of these distances. This approach ensures that HD captures the worst-case scenario of matching between the two sets, yielding a more robust similarity measure.

In medical imaging, HD is particularly useful for comparing organ contours across different images, aiding in diagnosis and surgical planning. Mathematically, HD is defined as

$$H(A, B) = \max\left(\sup_{a\in A}\inf_{b\in B}\|a - b\|, \sup_{b\in B}\inf_{a\in A}\|a - b\|\right), \tag{17}$$

where sup represents the supremum (maximum value), inf represents the infimum (minimum value), a and b are points in sets A and B, respectively, and $\|a-b\|$ denotes the Euclidean distance between points a and b. This formulation provides a comprehensive measure of the dissimilarity between the two sets.

5.2.3. HVSNR

HVSNR (Wang et al., 2015) evaluates the quality of images or videos with a particular emphasis on incorporating the perceptual characteristics of the human eye. Traditional signal-to-noise ratio (SNR) methods often fall short in fully capturing the human eye's varying perception of signals across different frequencies and brightness levels. In contrast, HVSNR employs a weighted MSE approach to more accurately simulate human visual perception.

The calculation of HVSNR is grounded in the MSE between the original and the degraded signals. However, it goes beyond simple MSE by applying different weights to the MSE based on the frequency components and pixel positions, thereby better reflecting the nuances of human visual perception. By taking into account the perceptual characteristics of the human visual system, HVSNR offers a more precise evaluation of image or video quality compared to traditional SNR methods. The mathematical formula for HVSNR is

$$\mathrm{HVSNR}(I, I') = 10 \cdot \log_{10}\left(\frac{\sum w_i \cdot I_i^2}{\sum w_i \cdot N_i^2}\right), \tag{18}$$

where I and I' represent the original image and its translated version, respectively, I_i and N_i denote the intensity of the original image and the noise image at the ith pixel position, respectively, and w_i denotes the weight of the ith pixel position.

5.3. Statistical metrics

5.3.1. PCC

PCC (Pearson, 1896), is a statistical measure widely used in the medical image translation domain to evaluate the linear relationship between two variables. It offers a quantitative assessment of the degree of linear correlation, which is of great value when analyzing data related to medical images. PCC remains invariant to data scale and location, allowing for the comparison of variables with different

magnitudes or offsets without being influenced by unit changes or data value shifts.

Mathematically, PCC is defined as

$$r_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{19}$$

where $X = (x_1, x_2, \ldots, x_n)$ and $Y = (y_1, y_2, \ldots, y_n)$ are the two sets of variables, $\bar{x}$ is the mean of X, and $\bar{y}$ is the mean of Y. The value of r_{XY} ranges from -1 to 1, with 1 indicating a perfect positive linear correlation, -1 representing a perfect negative linear correlation, and 0 implying no linear correlation.

In medical image translation, PCC has several application scenarios. It can assess global intensity correlations between real and synthesized images, like in MRI-to-CT conversion. It also verifies medical signal consistency in dynamic imaging and analyzes multimodality data covariance without a gold standard. However, PCC has limitations. It only captures linear relationships, ignoring non-linear ones. As a global statistic, it overlooks spatial structure information and is sensitive to outliers, which can be problematic in images with artifacts or noise. To use PCC rationally, it is recommended to combine it with other metrics like SSIM. And PCC suits global gray-level linear consistency checks but needs to be cautious in tasks involving local structures or non-linear mappings.

5.3.2. MI

MI (Shannon, 1948) is a fundamental concept in information theory that quantifies the amount of information that one random variable contains about another random variable. In medical image analysis, MI serves as a core metric for assessing the non-linear dependence between multimodality images, outperforming PCC in capturing complex associations.

For discrete random variables X and Y with probability mass functions $P(X)$ and $P(Y)$ and joint probability mass function $P(X, Y)$, Mutual Information is defined as

$$I(X;Y) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P(x,y)\log\frac{P(x,y)}{P(x)P(y)} \tag{20}$$

For continuous random variables X and Y with probability density functions $p(x)$ and $p(y)$ and joint probability density function $p(x, y)$, it is defined as

$$I(X;Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}\,dx\,dy \tag{21}$$

The value of mutual information is always non-negative, and it is zero if and only if the two variables are statistically independent. Larger values of mutual information indicate a stronger dependence or more shared information between the variables.

In medical images, MI is widely applied. In multimodality image registration, it maximizes the MI between images to optimize alignment, like in MRI-T1/T2 registration or PET-CT fusion. For cross-modality translation, such as MRI-to-CT conversion, it checks information retention. In functional imaging, it analyzes non-linear couplings between fMRI and EEG, or complementary information in multi-parameter MRI. As a supplement to the DSC, it evaluates segmentation consistency in areas with varying intensity distributions.

However, MI has limitations. It overlooks spatial structures, as high MI can coexist with wrong spatial correspondence. Its calculation is sensitive to histogram bin size and noise, and the original MI lacks an upper bound for easy comparison. To address these, it is recommended to combine it with other metrics like HD or use its improved variants.

5.4. Summary

In this section, we delve into the performance indicators in the field of medical image translation. These indicators can be classified into three categories: intensity-based, geometric fidelity, and statistical metrics. However, it should be noted that accurately evaluating the quality of medical image translation is still a challenging task.

Traditional evaluation metrics such as PSNR, MAE, and MSE face significant limitations when dealing with medical images. In medical images, particularly CT images, minute details like tiny shadows carry great diagnostic importance. However, these minor changes result in only marginal differences in traditional metrics, rendering them ineffective in precisely measuring the quality of medical image translation. Additionally, the absence of reliable ground truth in medical image translation tasks further undermines the accuracy and reliability of evaluations based on these traditional metrics. To address these issues, we focus on two key strategies: integrating Region of Interest (ROI) analysis with evaluation metrics and incorporating these metrics into the model training process.

Regarding ROI-based metric calculation, as discussed in Nakhaie and Shokouhi (2011), we can isolate and analyze specific ROIs containing critical details. In CT images, for example, we can define ROIs around suspected lesions or areas with subtle anatomical changes. By calculating PSNR, MAE, or MSE within these ROIs, the impact of small changes becomes more pronounced. This approach allows for a more meticulous assessment of translation quality in diagnostically relevant regions, enhancing the sensitivity of the evaluation metrics.

For integrating metrics into the model training, according to Ma et al. (2021), using metrics as part of the loss function can guide the model to better preserve important details. We can define a weighted loss function that includes traditional metrics calculated within ROIs. A weighted combination of PSNR or MAE values within ROIs can be added to the overall loss function during training. This encourages the model to minimize differences in these critical regions, leading to better preservation of small details in the translated images.

In addition to these technical improvements, we also recognize the value of expert knowledge in the evaluation process. Medical images are highly specialized and complex, and medical experts possess in-depth knowledge that can identify subtle features crucial for diagnosis. We can organize medical experts to conduct pairwise comparisons between original and translated medical images to obtain expert preference information (Prashnani et al., 2018). Although there are concerns about the subjectivity of expert evaluation, it can still serve as a supplementary reference. By using the proportion of experts choosing a certain translated image as a way to initially quantify expert evaluations, and combining it with the more objective methods of ROI-based metric calculation and metric-integrated training, we can establish a more comprehensive evaluation system.

In summary, by integrating ROI-based metric calculation, metric-integrated training, and incorporating expert knowledge in a complementary manner, we aim to develop a more accurate evaluation system for medical image translation. This system can overcome the limitations of traditional evaluation metrics, better capture the quality of translated images, especially in terms of preserving diagnostically important details, and ultimately contribute to the advancement of medical image translation technology.

6. Open datasets

6.1. SynthRAD 2023

Cone-beam CT (CBCT) is crucial in image-guided therapy, yet its quality is often hampered by artifacts. To overcome this, sCT technology has emerged, leveraging the superior soft-tissue contrast of MRI for enhanced treatment planning, particularly for vulnerable patient groups. SynthRAD2023 is a dual-task challenge designed to advance sCT technology, MRI to sCT generation (facilitating MR-RT) and CBCT to sCT generation (facilitating IGART). The dataset covers imaging data of patients who received radiation therapy in the brain or pelvic region. It is predominantly adult-based and no gender restrictions. Task 1 utilizes paired CT and MRI images from treatment planning, while Task

2 requires CT and CBCT images for patient localization. The datasets for Task 1 and Task 2 did not necessarily contain the same patients due to differences in image acquisition across tasks. Overall, the dataset contains 1080 image pairs (720 training, 120 validation, 240 testing).

In Karimzadeh and Ibragimov (2023), the SynthRad 2023 dataset of 360 brain and pelvis MR-CT image pairs was used. After preprocessing and splitting, a Pix2Pix-SwinUnet-based GAN model was trained and evaluated. Metrics showed its potential in generating sCT from MR, aiming to cut radiation exposure and costs.

6.2. BraSyn

The BraSyn 2023 dataset, built upon the RSNA-ASNR-MICCAI BraTS 2021 dataset, is a retrospective compilation of multi-institutional mpMRI scans of brain tumors acquired under standard clinical settings. These scans exhibit diverse image qualities due to variations in equipment and imaging protocols, mirroring the distinct clinical practices of different institutions. Each tumor subregion has been meticulously annotated by expert neuroradiologists. BraSyn offers a comprehensive set of four MRI modalities: T1-weighted images (both contrast-enhanced and non-enhanced), T2-weighted images, and FLAIR images. In the validation and testing phases, participants face the challenge of synthesizing a randomly excluded ("dropout") modality for each subject, without access to the corresponding segmentation mask. To evaluate the realism of the synthesized images, SSIM is used to compare them against actual clinical images of both the tumor region and healthy brain tissue.

In Gui et al. (2024), the BraSyn 2023 dataset underwent preprocessing to prepare it for the evaluation of the Conditional Autoregressive Vision Model (CAVM). CAVM innovatively employs an autoregressive approach, leveraging LLaMA-style Transformer blocks for dose-variant autoregression and token decomposition. Comparing with four SOTA models, and evaluated using metrics like SSIM, PSNR, and DSC, CAVM outperformed in both tumor regions and healthy tissue, validating its effectiveness in synthesizing contrast-enhanced brain tumor MRI images.

6.3. crossMoDA

Domain adaptation (DA) has garnered significant interest in the medical imaging field due to its ability to enhance the versatility of ML algorithms across diverse clinical scenarios. The crossMoDA Challenge introduces the inaugural large-scale, multi-class dataset specifically designed for unsupervised cross-modality domain adaptation. Unlike earlier crossMoDA iterations, which featured multi-institutional data collected under controlled conditions for radiosurgery planning and were limited to a 2-class segmentation task (tumor and cochlea), the 2023 version expands the segmentation challenge by incorporating heterogeneous, multi-institutional data. This dataset encompasses London SC-GK data, Tilburg SC-GK data, and UK MC-RC data, primarily consisting of T1-weighted and T2-weighted images. The primary objective is to accomplish cross-modality segmentation from contrast-enhanced T1 (ceT1) to high-resolution T2 (hrT2) images. Participants are tasked with training their models using unpaired annotated ceT1 scans and unannotated hrT2 scans, sourced from pre- and postoperative time points, respectively.

In Yang and Wang (2025), researchers addressed unsupervised cross-modality domain adaptation in medical image segmentation using the crossMoDA dataset. They proposed a method integrating image translation and self-training. For image translation, CycleGAN was used to translate unpaired ceT1 scans to hrT2 scans. In self-training, a 3D full-resolution U-Net from nnU-Net was trained with generated pseudo hrT2 images and real labels, and then iteratively retrained with pseudo labels from real hrT2 images. This approach achieved promising results on the crossMoDA 2022 challenge validation set, demonstrating its effectiveness in such tasks.

6.4. OASIS-3

OASIS-3 is a longitudinal dataset that integrates multimodality neuroimaging, clinical assessments, cognitive evaluations, and biomarkers for the study of normal aging and Alzheimer's disease. It comprises 2,842 MR imaging sessions, encompassing a wide range of sequences including T1-weighted (T1w), T2-weighted (T2w), FLAIR, ASL, SWI, time-of-flight, resting-state BOLD, and DTI. Many of these MR sessions are supplemented with volume segmentation files produced by FreeSurfer processing. Additionally, OASIS-3 offers over 2,157 PET imaging scans utilizing various tracers such as PIB, AV45, and FDG. These PET images are accompanied by post-processing files from the Pet Unified Pipeline (PUP), providing researchers with a rich and diverse dataset to explore the intricacies of aging and Alzheimer's disease.

In Jain et al. (2023), the RadiomicsGAN framework was applied to the OASIS-3 dataset. GAN was used to de-identify, unify, balance, and augment the data. RadiomicsGAN, inspired by CGAN and Radiomics features, translated T1-weighted MRI images to T2 and FLAIR formats. It extracted various Radiomics features from T2 images and used a ResNet-inspired chain of generators in a modified CGAN. The translated images were fed into a 3D-CNN for brain degeneration prediction, showing high accuracy and the method's efficacy.

6.5. ADNI-4

ADNI-4 marks the fourth iteration of a comprehensive, longitudinal research endeavor focused on Alzheimer's disease (AD) and its associated dementias. This initiative integrates a wide range of imaging modalities, encompassing structural MRI, PET scans, and potentially other advanced imaging technologies, alongside an extensive battery of clinical assessments, including cognitive evaluations, behavioral ratings, and biomarkers. The core objective of ADNI-4 is to deepen our insight into the progression and underlying mechanisms of AD and related disorders by tracking participants' clinical trajectories, imaging alterations, and biomarker changes over time. This endeavor seeks to facilitate earlier and more accurate diagnosis, while also accelerating the discovery and development of efficacious therapeutic interventions. Distinguished by its emphasis on inclusivity and diversity, ADNI-4 actively engages underrepresented populations, employing culturally sensitive and linguistically appropriate methods to ensure broad participation. The dataset, which is continuously being collected and expanded, encompasses a rich array of clinical, imaging, biomarker, and sociocultural data, providing an invaluable resource for researchers exploring the multifaceted nature of Alzheimer's disease and its associated conditions.

In Kamli et al. (2020), researchers used the ANDI dataset. They proposed Synthetic Medical Image Generator (SMIG) and Tumor Growth Predictor (TGP). SMIG, based on GAN, uses ANDI-sourced healthy brains and TCIA-sourced tumor volumes to generate synthetic MRI images for data augmentation and anonymization, keeping tumor size intact for accurate growth prediction. TGP, inspired by convolutional autoencoders, is a full-convolutional network predicting 90-day-later tumor volumes from initial scans.

6.6. IXI

The IXI database (Information eXtraction from Images) serves as a comprehensive multimodality medical image repository tailored for structural and functional brain studies. It encompasses a range of imaging techniques, including structural MRI (sMRI), functional MRI (fMRI), and magnetic resonance spectroscopy (MRS). These images facilitate detailed brain anatomical analyses, investigations of brain functional activities, and the detection of brain tissue metabolite concentrations. To ensure data consistency and comparability, the IXI database employs standardized acquisition protocols for image collection. The database

features various image modalities, notably T1, T2, PD, MRA, and DTI, which enable researchers to conduct thorough brain analyses by leveraging the information from different imaging techniques. This multimodality approach allows for in-depth exploration of brain structure, function, and metabolism, as well as facilitating cross-modality image translation studies. Because of its modalities, it is also often used for cross-modality image translation.

In Özbey et al. (2023), researchers applied the SynDiff model to the IXI dataset for medical image translation. SynDiff consists of a diffusive module and a non-diffusive module. The diffusive module uses a fast diffusion process with a source-conditional adversarial projector. It sets large step-size diffusion formulas and a reverse diffusion sampling mechanism guided by source images. The non-diffusive module estimates source images paired with target images using two generator-discriminator pairs based on adversarial losses, guiding the diffusive module. The two modules are jointly trained via cycle-consistency loss for unsupervised learning. As a result, SynDiff outperforms other GAN and diffusion models in generating high-quality images for multi-contrast MRI translation tasks.

6.7. CHAOS

The Combined (CT-MR) Healthy Abdominal Organs Segmentation (CHAOS) dataset is a specialized medical image segmentation resource aimed at advancing the research and development of computerized algorithms for the segmentation of healthy abdominal organs. This dataset integrates CT and MR imaging modalities, offering a more holistic view of the abdominal organs. Comprising CT and MR scans from various institutions, the CHAOS dataset captures the abdominal region of healthy individuals, encompassing key organs such as the liver, spleen, pancreas, and kidneys. Given the inherent challenges posed by the abdominal area's flexible anatomical structure and the distinct image features captured by different modalities, the CHAOS dataset is frequently utilized for cross-modality translation tasks between CT and MRI, enabling researchers to leverage the strengths of both imaging techniques for improved segmentation accuracy and analysis.

In Jiang and Veeraraghavan (2020), researchers applied their proposed method to the MRI data in the CHAOS. They introduced a unified cross-modality feature disentangling approach for multi-domain image translation and multi-organ segmentation. This method utilized a VAE to disentangle image content and style. It made the style feature encoding to match a Gaussian prior, transformed the extracted style into a latent style scaling code, and modulated the generator to generate multimodality images. Meanwhile, multiple losses were calculated to constrain the multi-domain image-to-image translation process. In the segmentation part, separate multi-organ segmentation networks were trained for each target modality and optimized with cross-entropy loss. A joint distribution structure discriminator was also introduced, treating images and segmentation probability maps as a joint distribution and calculating domain mismatches through adversarial training to further constrain multi-domain translation and improve segmentation performance. This method demonstrated good results in multi-domain image translation and multi-organ segmentation tasks.

6.8. ACDC

The Automated Cardiac Diagnostic Challenge (ACDC) Database is a specialized resource dedicated to advancing research in cardiac medical imaging, particularly in the development of automated diagnostic methods. This database offers a rich collection of multimodality cardiac MRI images, encompassing the left ventricle, right ventricle, left ventricular myocardium, right ventricular myocardium, and additional heart structures, thereby facilitating comprehensive analysis of cardiac structure and function. Accompanying these images are detailed annotations that provide crucial information on cardiac anatomy and function. These annotations serve as invaluable material for supervised learning, empowering researchers to develop and refine automated diagnostic algorithms.

In Bevilacqua et al. (2024), researchers utilized the MRI images from the ACDC dataset for the medical image super-resolution task. They proposed using LDM conducting noise-adding and noise-removing operations in the latent space to reduce computational cost and conditions the model input via a cross-attention mechanism. During training, the LDM pre-trained on the OpenImages dataset was fine-tuned following the Super-Resolution via Repeated Refinement (SR3) training pipeline. The results showed that LDM achieved better similarity values in the medical image super-resolution task, generating images with better details and a higher degree of match with the original high-resolution images.

6.9. BraTS 2024

BraTS (Brain Tumor Segmentation Challenge) is an international competition dedicated to advancing brain tumor segmentation through medical image processing and artificial intelligence in neuroscience and medicine. Its primary objective is to achieve precise segmentation of brain tumors from multimodality MRI. This competition emphasizes the segmentation of intricate brain tumor substructures, which includes multiple challenges such as identifying the tumor core and peritumoral edema. BraTS leverages extensive, publicly accessible brain MRI datasets, allowing participants to train and validate their algorithms on a diverse range of cases. These datasets typically encompass various imaging modalities (e.g., T1, T2, FLAIR, etc.), providing a comprehensive view of each case. Due to the rich multimodality data available, some researchers also use it for cross-modality translation of medical images, such as T1 to T2 or T2 to T1.

In Xing et al. (2024), researchers conducted research on medical image cross-modality translation using the BraTS and proposed the Cross-conditioned Diffusion Model (CDM). The model consisted of three components: the Modality-specific Representation Model (MRM), the Modality-decoupled Diffusion Network (MDN), and the Cross-conditioned UNet (C-UNet). The MRM learned the distribution of target modalities, the MDN models this distribution and improved efficiency, and the C-UNet generated target-modality images by combining source modalities with the target distribution. On the BraTS dataset, CDM outperformed several SOTA methods in terms of image-generation quality metrics. Ablation experiments also confirmed the effectiveness of each module.

6.10. MMWHS

The MMWHS (Multimodality Whole Heart Segmentation) Challenge is an esteemed international competition aimed at advancing research in cardiac medical image segmentation. The MMWHS Challenge offers a comprehensive dataset comprising 120 multimodality cardiac images, evenly split between 60 cardiac CT/CTA and 60 cardiac MRI scans. These images, which capture the entire heart and its vital substructures, were acquired in real clinical settings and utilized for clinical diagnosis. Given the diverse sources of these images, there is variation in image quality, with some being of relatively lower quality. As a result, many researchers have leveraged this dataset for cross-modality translation between CT and MRI, aiming to enhance image quality and improve segmentation accuracy.

In Kang et al. (2023), researchers conducted research on cross-modality medical image domain adaptation using the MMWHS dataset. The proposed image translation model consisted of structure and texture encoders, a generator, discriminators, etc. It was trained by optimizing the GAN, mutual information, and texture co-occurrence losses. On the MMWHS dataset, the images generated by this model were used to train a segmentation model. Compared with multiple methods, it shows better segmentation performance, demonstrating the effectiveness of this method in cross-modality medical image domain adaptation tasks.

6.11. FDG-PET/CT

The DG-PET/CT dataset is a comprehensive medical imaging resource that includes PET and CT images. It encompasses patients with histologically confirmed malignant melanoma, lymphoma, or lung cancer, as well as negative control patients who underwent FDG-PET/CT scans at two large medical centers. The dataset comprises 1014 studies, with each case (whether for training or testing) featuring a 3D whole-body FDG-PET volume, a corresponding 3D whole-body CT volume, and a 3D binary mask of tumor foci that have been manually segmented on the FDG-PET volume.

In Phan et al. (2024), researchers focused on unpaired medical image translation using data related to FDG-PET/CT. Considering the issues of Transformer in unpaired image synthesis, they proposed the UNet Structural Transformer (UNest) architecture, consisting of Structural Transformer (ST) blocks and a convolutional decoder with skip connections. The ST block classified foreground and background using a lightweight patch classifier and masks from Segment-Anything Model (SAM), applying structural and local attention mechanisms respectively. The foreground structural attention aggregated context within the anatomy, while the background local attention enables effective feature exchange. Compared with multiple SOTA methods in FDG-PET/CT-related medical image translation tasks, UNest shows better translation accuracy, demonstrating its effectiveness.

6.12. AANLIB

ANLIB, also known as Harvard Whole Brain Atlas, is a comprehensive medical imaging resource developed collaboratively by Massachusetts General Hospital and Harvard Medical School. This database boasts an extensive collection of brain structure images, encompassing MRI, CT, PET, and SPECT modalities. It serves as a versatile tool for various types of neuroimaging research. The image data are meticulously labeled and categorized, facilitating users' ability to identify and study diverse brain structures and diseases with ease.

In Das et al. (2024), researchers conducted research on multimodality medical image fusion using data from AANLIB. They proposed an end-to-end multimodality medical image fusion method based on a content-aware GAN. The method consisted of a generator similar to the U-Net architecture and two discriminators. The generator extracts and fuses features through a series of downsampling, upsampling layers and skip-connections. The two discriminators distinguished between the generated fused image and the source images. Compared with several DL-based fusion methods, this method performed better in subjective visual effects and quantitative indicators, effectively fusing image information.

6.13. Ultra-low dose PET 2024

Ionizing radiation exposure is a major concern in PET imaging, restricting its use in multiple scenarios. The 2024 Ultra-Low Dose PET(UDPET) aims to resolve this issue. It uses advanced computational algorithms to reconstruct high-quality images from data collected during low-dose scans, with the objective of minimizing radiation exposure to levels similar to those during a transatlantic flight. The dataset encompasses whole-body 18F-FDG PET images from 1447 subjects. These images are sourced from two commercial total-body PET systems: Siemens Biograph Vision Quadra (n = 387) and United Imaging uExplorer (n = 1060). All data was acquired in list mode, enabling the simulation of different acquisition times by rebinding the data. The dataset includes simulated low-statistics data corresponding to low-dose PET with dose-reduction factors (DRF) of 4, 10, 20, 50, and 100, along with full-dose images. The low-dose PET images are created by subsampling from the full-scan, ensuring perfect alignment with the full-dose PET counterparts. Researchers are leveraging the ULD-PET scanner and applying sophisticated algorithms to enhance the quality of images obtained from these low-dose scans. This research shows great promise for advancing PET imaging technology and reducing the radiation risks associated with medical imaging examinations for patients.

In Xue et al. (2024), researchers used the UDPET to address the issue of improving ultra-low-dose PET imaging quality. They proposed the WaveNet method decomposing ultra-low-dose PET images into multiple frequency components via wavelet transform (WT) and inputting these components into a 3D-UNet-structured neural network for frequency-domain denoising. Additionally, it replaced the downsampling layer with 3D discrete wavelet transform (DWT) to reduce information loss and preserve texture details. During training, a traditional U-Net model was used as a baseline, and a customized scoring system that included global metrics like normalized root MSE (NRMSE) and PSNR, as well as local indices, was used to evaluate image quality. Experimental results showed that WaveNet outperformed U-Net at all DRF levels and demonstrated good generalizability on cross-scanner datasets, indicating its effectiveness in restoring high-quality images from ultra-low-dose PET scans.

In the studies included in this review, a diverse array of datasets serves as the bedrock for innovation and discovery in the dynamic landscape of medical image research. Among them, SynthRAD is leveraged in 8 research undertakings, BraSyn in 6 investigations, and crossMoDA in 15 studies. The IXI dataset, a staple in the field, is incorporated into as many as 31 research projects, while ADNI and BraTS are respectively utilized in 17 and 28 studies, each contributing significantly to the body of knowledge. CHAOS and ACDC are employed in 11 and 21 research works, respectively, playing crucial roles in specific research domains. OASIS-3 and MMWHS are involved in 5 and 7 studies, offering unique insights for targeted research. FDG-PET/CT is utilized in 15 research efforts, and AANLIB in 4, with the Ultra-low Dose PET 2024 dataset being applied in 3 studies. These datasets, each with distinct characteristics, collectively fuel a broad spectrum of research directions, driving the advancement of medical image analysis.

7. Challenges and discussions

7.1. Performance evaluation metrics

In Section 5, we introduced a comprehensive set of evaluation metrics, including MSE, PSNR, MAE, SSIM, FSIM, NCC, HD, LPIPS, HVSNR, IS, FID, PCC and MI. These evaluation metrics are crucial for accessing the performance and effectiveness of image translation tasks, each tailored to specific characteristics and application contexts.

Traditional image quality metrics like MSE, PSNR, and MAE are straightforward and easy to comprehend. However, they may lack sensitivity to perceptual image qualities aligned with the human visual system. Consequently, these metrics are typically best suited for tasks such as image reconstruction or denoising.

SSIM and FSIM, in contrast, incorporate structural image information, aligning more closely with human perception. They effectively evaluate image realism and similarity, making them ideal for image restoration or enhancement tasks.

NCC measures the linear correlation between images, while HD assesses geometric distance. Both are adept at reflecting overall image similarity and differences, particularly in medical image alignment and matching applications.

LPIPS, HVSNR, IS, and FID are metrics rooted in DL models. LPIPS, which considers feature space distance, aligns well with human perception but can be computationally intensive. It is particularly suitable for tasks like image style transfer. HVSNR, which accounts for image signal-to-noise ratio, measures image clarity and noise levels, making it apt for medical image reconstruction or enhancement. IS and FID, on the other hand, evaluate the distribution difference between generated and real images, providing a comprehensive assessment of generated

Table 8
Common datasets used in medical image translation.

Dataset	Modalities	Data size	Body part	Task	Address
SynthRAD 2023 (Huijben et al., 2024)	MRI, CT, CBCT	540 Paired MRI-CT, 540 CBCT-CT Sets	Brain	MRI → sCT , CBCT → sCT	https://synthrad2023.grand-challenge.org/
BraSyn (Li et al., 2023a)	FLAIR, T1, T1ce, T2	2040 MRI scans	Brain	Synthesis of missing MRI modalities	https://www.med.upenn.edu/cbica/brats/
crossMoDA (Dorent et al., 2023)	ceT1, hrT2	379 MRI scans	Cochlear implant	ceT1 → hrT2, segmentation	https://crossmoda-challenge.ml/
IXI (2015)	T1, T2, PD, MRA, DTI	600 MR images	Brain	T1, T2 → PD, T1, PD → T2, T2, PD → T1, T2 → PD, PD → T2	https://brain-development.org/ixi-dataset/
ADNI-4 (Rivera Mindt et al., 2024)	MRI, PET, CT	ADNI-1, ADNI-GO, ADNI-2, ANDI-3 and keep being expanded	Brain	Alzheimer's disease research	https://adni.loni.usc.edu/about/adni4/
BraTS 2024 (de Verdier et al., 2024)	FLAIR, T1, T1ce, T2	4500 cases	Brain	Brain glioma challenge	https://www.synapse.org/Synapse:syn53708249
CHAOS (Valindria et al., 2018)	T1-DUAL , T2-SPIR, CT	40 CTs, 80 T1-DUALs and 40 T2-SPIRs	Abdomen	Abdominal organ segmentation	https://chaos.grand-challenge.org/
ACDC (Bernard et al., 2018)	MRI	150 cases	Cardiac	Cardiac diagnosis	https://opendatalab.com/OpenDataLab/ACDC
OASIS-3 (LaMontagne et al., 2019)	MRI, PET	2842 MR	Brain	Alzheimer's disease research	https://www.oasis-brains.org/
MMWHS (Zhuang, 2018)	CT, MRI	120 multimodality cardiac images	Cardiac	Cardiac segmentation	https://zmiclab.github.io/zxh/0/mmwhs/
FDG-PET/CT (Gatidis et al., 2022)	PET, CT	1014 studies (900 patients)	whole body	Tumor and lung cancer research	https://autopet.grand-challenge.org/Dataset/
AANLIB (Summers, 2003)	MRI, CT, PET, SPECT	Normal and 5 major diseases	Brain	Brain disease research	https://www.med.harvard.edu/aanlib/
Ultra-low Dose PET 2024	PET	1447 case 18F-FDG PET subjects	Whole body	PET synthesis	https://udpet-challenge.github.io/

image quality and diversity. They are commonly used to evaluate generative models, such as GANs.

Despite their utility, challenges and discussions persist regarding medical image translation assessment metrics. Existing metrics may not fully capture the nuances of medical image characteristics. Traditional pixel-level metrics like MSE and SSIM, while useful, may not always reflect medical image specifics. Thus, there is a need for metrics tailored to medical imagery. Furthermore, the selection and standardization of metrics require deeper investigation. Different medical image translation tasks may necessitate distinct evaluation metrics, yet a unified standard and index system are currently lacking. Establishing a comprehensive, reliable evaluation system is crucial for accurate model performance assessment and comparative analysis (Hong et al., 2022). Additionally, the interpretability and comprehensibility of metrics are important considerations. Medical image translation results must be understandable to medical professionals and researchers. Therefore, metrics should intuitively reflect model performance and be easily interpretable. Lastly, the robustness and consistency of evaluation metrics across datasets and application scenarios are vital. Metrics should maintain reliability to enable accurate model performance evaluation and comparison, fostering further development and application in medical image translation.

When selecting metrics, it is essential to consider task-specific characteristics and datasets. For medical image translation, factors such as image clarity, structural similarity, and distribution differences with real images must be considered. Different metrics have distinct foci and applications, necessitating careful selection based on specific needs. Simultaneously, leveraging multiple, correlated metrics can provide a more holistic assessment of model performance.

7.2. Open datasets

Open datasets serve as pivotal resources in the realm of medical image translation, catalyzing advancements in algorithmic research and application development. Nevertheless, their utilization is fraught with several challenges and considerations.

Firstly, data quality and annotations accuracy are paramount in determining model performance and the reliability of outcomes. Medical images, owing to their intricate nature and variability, may harbor inconsistencies in quality or labeling inaccuracies within datasets. These discrepancies can precipitate instability during model training and result in erroneous outputs. Consequently, there is an imperative need for rigorous monitoring and assurance of data quality.

Furthermore, data privacy and security constitute critical concerns that demand meticulous attention. Stringent privacy safeguards must be established to ensure the confidentiality and protection of patient data.

Lastly, the finite nature of datasets poses constraints on the efficacy of model training and deployment. As illustrated in Table 8, there are relatively few available datasets. DL models necessitate extensive data for stable training. Given the high costs and potential radiation risks associated with medical image acquisition, open medical datasets are often limited, potentially insufficient for all medical image translation endeavors. In specialized domains or tasks, there may be a dearth of adequate data volume and diversity. For instance, supervised learning necessitates precisely aligned source and target domain images, which can further restrict the effectiveness of model training and application. Consequently, it is crucial to explore innovative data sources and data augmentation techniques to enhance existing datasets. Semi-synthetic

datasets offer an effective solution to these challenges, complemented by patch-based training methods (Boni et al., 2020), transfer learning (Boulanger et al., 2021), and multi-task learning approaches that have also been devised to tackle these issues (Arbabi et al., 2023).

7.3. Future applications

The potential avenues for future research in this literature can be broadly categorized into three key areas: DL networks, datasets, and validation of results.

With regard to DL-based methods for medical image translation, a significant portion of existing studies has concentrated on MRI translation. Consequently, there exists an opportunity to delve deeper into image contrast translation using diverse strategies. Additionally, while supervised-based methods have yielded impressive results in image transformation through the use of paired data, there remains ample room for architectural enhancements in unsupervised methods.

Both CNN-based and GAN-based approaches have demonstrated competitive performance, yet they are inherently limited in their ability to capture long-range dependencies between image regions due to their constrained receptive fields. Recently introduced visual transformer-based methods have addressed this limitation, surpassing many traditional DL-based image translation techniques. These transformer-based approaches have been applied to a limited extent in MRI-PET translation and MRI contrast translation studies (Boni et al., 2020; Brou Boni et al., 2021; Hu et al., 2021). Furthermore, emerging transformer-based, contrast-learning-based, and diffusion-based methods have exhibited superior performance in image translation compared to conventional network architectures. Notably, diffusion-based methods have shown particularly promising results. Currently, models with enhanced interpretability and a focus on low-level features are urgently needed in the medical field. Interpretability is crucial for promoting acceptance and trust in medical clinical applications. Recently, novel models such as Kolmogorov-Arnold Networks (KANs) have demonstrated potential in medical image generation tasks (Li et al., 2024), providing an innovative approach to improving interpretability in medicine. Low-level features, as direct manifestations of human body details, often signify significant changes in body structure and function through subtle variations. However, images generated by existing models still fall short of original images in terms of detail representation and diagnostic accuracy, as perceived by doctors (Salmanpour et al., 2024). Therefore, developing models that are more focused on extracting low-level features is particularly important for enhancing the accuracy and practicality of medical image analysis.

In terms of datasets, a handful of studies have proposed novel approaches to overcome the constraints posed by data scarcity, such as the generation of semi-synthetic data. This represents a promising research direction aimed at alleviating the degradation of model performance due to insufficient data (Bernard et al., 2018; Mao et al., 2022). Moreover, leveraging multiple image sequences from different institutions could enhance the generalizability of the translation methods.

With respect to the validation of translation results, there are certain areas of the current research that merit further investigation. Although current evaluation metrics can intuitively reflect the quality of medical images in quantitative analysis, they often fail to capture sufficient detail in qualitative assessments, particularly those conducted by experts. This limitation underscores the need to explore image quality evaluation standards that more closely align with the criteria used by medical professionals. For instance, while quantitative metrics might provide a clear numerical score, they may overlook subtle nuances and diagnostic details that are crucial to expert evaluations. Therefore, developing evaluation standards that better reflect the expert perspective is essential for ensuring that medical images meet the high standards required for accurate diagnosis and patient care.

Furthermore, there are fundamental differences in translation methods regarding whether they can handle multiple tasks and whether they can be deployed at different imaging sites without significant generalization errors. This is a problem that needs to be deeply analyzed and is closely related to the efforts to mitigate these problems in recent years. In the field of medical image translation, different methods show significant differences in multi-task. Some methods are only designed to perform a single translation task, such as converting MRI images into CT images. However, in actual clinical applications, it is often necessary to handle multiple tasks at the same time, such as performing image translation and disease diagnosis at the same time. For example, in a comprehensive clinical diagnosis scenario, it is not only necessary to convert images of different modalities but also to make accurate diagnoses and analyses based on the translated images. This multi-task ability is of great significance for improving medical efficiency and accuracy. At the same time, the deployment of medical image translation methods at different imaging sites also faces challenges. Due to the differences in equipment, imaging protocols, and patient populations at different imaging sites, the data distribution is heterogeneous, which affects the generalization performance of the model. In previous studies, although some methods have been proposed to reduce the adverse effects of data heterogeneity on model performance, such as using content-contrast disentanglement methods (such as StyleGAN, etc.) and multi-site model training techniques, these methods still have some limitations in practical applications. A recent study Dalmaz et al. (2024) proposed a personalized federated learning method for multi-contrast MRI synthesis. This method effectively addresses the implicit and explicit heterogeneity problems in multi-site datasets by using personalized blocks and partial network aggregation, and improves the generalization performance of the model. This provides a valuable reference for us to solve the problems of multi-task and cross-site deployment of medical image translation methods. In the future, we need to further explore the multi-task ability and cross-site deployment performance of medical image translation methods, explore more effective ways to reduce the adverse effects of data heterogeneity on model performance, and improve the accuracy and reliability of medical image translation to provide stronger support for clinical applications.

Recently, there has been a notable trend in medical image translation towards avoiding “direct” translation and instead focusing on “intermediate” translations. In this emerging approach, the emphasis is on creating an intermediate representation between the source and target images. The quality of the target image takes a secondary role, as the primary goal is to use this intermediate representation to boost the performance of the intended medical task, such as segmentation or diagnosis. This trend aligns well with the concept of converting diverse imaging modalities into a unified intermediate pseudo-modality (Ma et al., 2024). By doing so, the modality gap is substantially reduced, and registration challenges are simplified. For example, in multimodality image alignment, this intermediate representation can serve as a common ground, making it easier to align different modalities. A diffusion-based unsupervised domain adaptation framework exemplifies this trend (Ji and Chung, 2024). By leveraging a coupled structure-preserving diffusion model, it generates intermediate images that enable better cross-domain knowledge transfer, eventually showing excellent results in abdominal multi-organ segmentation experiments. This is crucial as it helps in adapting models across different medical image domains, improving the generalization ability of algorithms. Another relevant approach is the federated image segmentation method (Galati et al., 2024). Through federated learning, it constructs a multimodality data factory. This factory creates an intermediate latent representation for feature fusion, which is essential for more accurate segmentation. The ability to fuse features from multiple modalities at an intermediate stage allows for a more comprehensive understanding of the image data, leading to good performance in tasks such as multi-scanner cardiac MRI segmentation, multimodality skull stripping, and multi-organ vascular segmentation.

This trend of intermediate translations holds great significance. It offers a new way to address the challenges in medical image translation,

such as domain shift and limited data. By focusing on the intermediate representation, models can potentially better capture the underlying relationships between different modalities or data sources. This not only improves the performance of medical image-related tasks but also has the potential to enhance the interpretability of models, as the intermediate steps can provide more insights into the translation process. Overall, it paves the way for more effective and accurate medical image translation techniques in the future.

8. Conclusion

This review has comprehensively explored the latest developments in medical image translation techniques and related datasets. We began by presenting the practical applications and research progress of medical image translation, covering aspects like intra-modality, cross-modality, and label-based translations, as well as their real-world uses in diagnosis, radiotherapy, and data handling. Next, we provided an in-depth overview of DL-based methods for generating pseudo-CT, MR, and PET images. This included summarizing diverse network architectures, analyzing the latest generative network designs, loss functions, and discussing the datasets used as well as evaluation results. Subsequently, we detailed the key DL models in medical image translation, such as GANs, VAEs, ARs, diffusion Models, and flow Models, along with their principles, advantages, and limitations. We also conducted a thorough discussion on the evaluation metrics in medical image translation, highlighting their unique features and typical applications. Additionally, we analyzed the commonly used datasets, emphasizing their characteristics and applications. Furthermore, we delved into the challenges faced in this field and proposed potential solutions. Through a comparative analysis of different image translation approaches, we pinpointed their respective strengths and weaknesses, and put forward directions for future research. Ultimately, the goal of this review is to contribute to the continuous development of medical image translation techniques. By synthesizing the current SOTA knowledge and identifying areas for improvement, we hope to inspire more accurate and innovative methods, thus facilitating progress in this crucial area of medical imaging research.

CRediT authorship contribution statement

Junxin Chen: Writing, Supervision, Funding acquisition, Conceptualization. **Zhiheng Ye:** Writing – review & editing, Formal analysis, Data curation. **Renlong Zhang:** Writing – original draft, Resources, Investigation. **Hao Li:** Investigation. **Li-bo Zhang:** Supervision. **Wei Wang:** Supervision.

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.

Acknowledgments

This work is funded by the National Natural Science Foundation of China (No. 62171114), and Xiaomi Young Talents Program.

Data availability

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

References

Ahangari, S., Beck Olin, A., Kinggård Federspiel, M., Jakoby, B., Andersen, T.L., Hansen, A.E., Fischer, B.M., Littrup Andersen, F., 2022. A deep learning-based whole-body solution for PET/MRI attenuation correction. EJNMMI Phys. 9 (1), 1–15.

Alessio, A., Vesselle, H., Lewis, D., Matesan, M., Behnia, F., Suhy, J., de Boer, B., Maniawski, P., Minoshima, S., 2012. Feasibility of low-dose FDG for whole-body TOF PET/CT oncologic workup.

Almahairi, A., Rajeshwar, S., Sordoni, A., Bachman, P., Courville, A., 2018. Augmented CycleGAN: Learning many-to-many mappings from unpaired data. In: International Conference on Machine Learning. PMLR, pp. 195–204.

Alotaibi, A., 2020. Deep generative adversarial networks for image-to-image translation: A review. Symmetry 12 (10), 1705.

Arbabi, S., Foppen, W., Gielis, W.P., van Stralen, M., Jansen, M., Arbabi, V., de Jong, P.A., Weinans, H., Seevinck, P., 2023. MRI-based synthetic CT in the detection of knee osteoarthritis: Comparison with CT. J. Orthop. Res.®.

Armanious, K., Jiang, C., Fischer, M., Küstner, T., Hepp, T., Nikolaou, K., Gatidis, S., Yang, B., 2020. MedGAN: Medical image translation using GANs. Comput. Med. Imaging Graph. 79, 101684.

Arslan, F., Kabas, B., Dalmaz, O., Ozbey, M., Çukur, T., 2024. Self-consistent recursive diffusion bridge for medical image translation. arXiv preprint arXiv:2405.06789.

Atli, O.F., Kabas, B., Arslan, F., Demirtas, A.C., Yurt, M., Dalmaz, O., Çukur, T., 2024. I2I-Mamba: Multi-modal medical image synthesis via selective state space modeling. arXiv preprint arXiv:2405.14022.

Bambach, S., Ho, M.L., 2022. Deep learning for synthetic CT from bone MRI in the head and neck. Am. J. Neuroradiol. 43 (8), 1172–1179.

Baudrier, E., Millon, G., Nicolier, F., Seulin, R., Ruan, S., 2007. Hausdorff distance-based multiresolution maps applied to image similarity measure. Imaging Sci. J. 55 (3), 164–174.

Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al., 2018. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans. Med. Imaging 37 (11), 2514–2525.

Bevilacqua, V., Di Marino, A., Di Nardo, E., Ciaramella, A., De Falco, I., Sannino, G., 2024. Cross-domain super-resolution in medical imaging. In: 2024 IEEE Symposium on Computers and Communications. ISCC, IEEE, pp. 1–6.

Boni, K.N.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 (7), 075002.

Boschet, A., Collin, A., Katoch, N., Cohen-Adad, J., 2024. Unpaired modality translation for pseudo labeling of histology images. In: MICCAI Workshop on Deep Generative Models. Springer, pp. 54–63.

Boulanger, M., Nunes, J.C., Chourak, H., Largent, A., Tahri, S., Acosta, O., De Crevoisier, R., Lafond, C., Barateau, A., 2021. Deep learning methods to generate synthetic CT from MRI in radiotherapy: A literature review. Phys. Medica 89, 265–281.

Bourbonne, V., Jaouen, V., Hognon, C., Boussion, N., Lucia, F., Pradier, O., Bert, J., Visvikis, D., Schick, U., 2021. Dosimetric validation of a GAN-based pseudo-CT generation for MRI-only stereotactic brain radiotherapy. Cancers 13 (5), 1082.

Brock, A., Donahue, J., Simonyan, K., 2018. Large scale GAN training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096.

Brou Boni, K.N., 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 (6), 3003–3010.

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: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23. Springer, pp. 728–737.

Che, Z., Zhang, Z., Wu, Y., Wang, M., 2025. Disentangle and Then Fuse: A cross-modal network for synthesizing Gadolinium-Enhanced brain MR images. IEEE Trans. Circuits Syst. Video Technol..

Chen, Y., Lin, H., Zhang, W., Chen, W., Zhou, Z., Heidari, A.A., Chen, H., Xu, G., 2024a. ICycle-GAN: Improved cycle generative adversarial networks for liver medical image generation. Biomed. Signal Process. Control. 92, 106100.

Chen, X., Luo, S., Pun, C.M., Wang, S., 2024b. Medprompt: Cross-modal prompting for multi-task medical image translation. In: Chinese Conference on Pattern Recognition and Computer Vision. PRCV, Springer, pp. 61–75.

Chen, S., Peng, Y., Qin, A., Liu, Y., Zhao, C., Deng, X., Deraniyagala, R., Stevens, C., Ding, X., 2022. MR-based synthetic CT image for intensity-modulated proton treatment planning of nasopharyngeal carcinoma patients. Acta Oncol. 61 (11), 1417–1424.

Chen, S., Zhang, R., Liang, H., Qian, Y., Zhou, X., 2024c. Coupling of state space modules and attention mechanisms: An input-aware multi-contrast MRI synthesis method. Med. Phys..

Choi, Y., Choi, M., Kim, M., Ha, J.W., Kim, S., Choo, J., 2018. StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8789–8797.

Chourak, H., Barateau, A., Tahri, S., Cadin, C., Lafond, C., Nunes, J.C., Boue-Rafle, A., Perazzi, M., Greer, P.B., Dowling, J., et al., 2022. Quality assurance for MRI-only radiation therapy: A voxel-wise population-based methodology for image and dose assessment of synthetic CT generation methods. Front. Oncol. 12, 968689.

Costa, P., Galdran, A., Meyer, M.I., Abramoff, M.D., Niemeijer, M., Mendonça, A.M., Campilho, A., 2017. Towards adversarial retinal image synthesis. arXiv preprint arXiv:1701.08974.

Cunniff, C., Byrne, J.L., Hudgins, L.M., Moeschler, J.B., Olney, A.H., Pauli, R.M., Seaver, L.H., Stevens, C.A., Figone, C., 2000. Informed consent for medical photographs. Genet. Med. 2 (6), 353–355.

Cusumano, D., Lenkowicz, J., Votta, C., Boldrini, L., Placidi, L., Catucci, F., Dinapoli, N., Antonelli, M.V., Romano, A., De Luca, V., et al., 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.

Dai, X., Lei, Y., Wang, T., Zhou, J., Roper, J., McDonald, M., Beitler, J.J., Curran, W.J., Liu, T., Yang, X., 2021. Automated delineation of head and neck organs at risk using synthetic MRI-aided mask scoring regional convolutional neural network. Med. Phys. 48 (10), 5862–5873.

Dalmaz, O., Mirza, M.U., Elmas, G., Ozbey, M., Dar, S.U., Ceyani, E., Oguz, K.K., Avestimehr, S., Çukur, T., 2024. One model to unite them all: Personalized federated learning of multi-contrast MRI synthesis. Med. Image Anal. 94, 103121.

Dalmaz, O., Yurt, M., Çukur, T., 2022. ResViT: Residual vision transformers for multimodal medical image synthesis. IEEE Trans. Med. Imaging 41 (10), 2598–2614.

Das, M., Gupta, D., Bakde, A., 2024. An end-to-end content-aware generative adversarial network based method for multimodal medical image fusion. In: Data Analytics for Intelligent Systems: Techniques and Solutions. IOP Publishing Bristol, UK, pp. 7–1–7–10.

Davidson, T.R., Falorsi, L., De Cao, N., Kipf, T., Tomczak, J.M., 2018. Hyperspherical variational auto-encoders. arXiv preprint arXiv:1804.00891.

Dayarathna, S., Islam, K.T., Uribe, S., Yang, G., Hayat, M., Chen, Z., 2023. Deep learning based synthesis of MRI, CT and PET: Review and analysis. Med. Image Anal. 103046.

de Verdier, M.C., Saluja, R., Gagnon, L., LaBella, D., Baid, U., Tahon, N.H., Foltyn-Dumitru, M., Zhang, J., Alafif, M., Baig, S., et al., 2024. The 2024 brain tumor segmentation (BraTS) challenge: Glioma segmentation on post-treatment MRI. arXiv preprint arXiv:2405.18368.

Dilokthanakul, N., Mediano, P.A., Garnelo, M., Lee, M.C., Salimbeni, H., Arulkumaran, K., Shanahan, M., 2016. Deep unsupervised clustering with gaussian mixture variational autoencoders. arXiv preprint arXiv:1611.02648.

Dinh, L., Krueger, D., Bengio, Y., 2014. Nice: Non-linear independent components estimation. arXiv preprint arXiv:1410.8516.

Dong, J., Fu, J., He, Z., 2019a. A deep learning reconstruction framework for X-ray computed tomography with incomplete data. PLoS One 14 (11), e0224426.

Dong, X., Lei, Y., Tian, S., Wang, T., Patel, P., Curran, W.J., Jani, A.B., Liu, T., Yang, X., 2019b. Synthetic MRI-aided multi-organ segmentation on male pelvic CT using cycle consistent deep attention network. Radiother. Oncol. 141, 192–199.

Dorent, R., Kujawa, A., Ivory, M., Bakas, S., Rieke, N., Joutard, S., Glocker, B., Cardoso, J., Modat, M., Batmanghelich, K., et al., 2023. CrossMoDA 2021 challenge: Benchmark of cross-modality domain adaptation techniques for vestibular schwannoma and cochlea segmentation. Med. Image Anal. 83, 102628.

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

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, pp. 188–193.

Emami, H., Liu, Q., Dong, M., 2020b. FREA-UNet: frequency-aware U-Net for modality transfer. arXiv preprint arXiv:2012.15397.

Feng, E., Qin, P., Chai, R., Zeng, J., Wang, Q., Meng, Y., Wang, P., 2022. MRI generated from CT for acute ischemic stroke combining radiomics and generative adversarial networks. IEEE J. Biomed. Heal. Inform. 26 (12), 6047–6057.

Figini, M., Lin, H., Ogbole, G., Arco, F.D., Blumberg, S.B., Carmichael, D.W., Tanno, R., Kaden, E., Brown, B.J., Lagunju, I., et al., 2020. Image quality transfer enhances contrast and resolution of low-field brain MRI in african paediatric epilepsy patients. arXiv preprint arXiv:2003.07216.

Florkow, M.C., Willemsen, K., Zijlstra, F., Foppen, W., van der Wal, B.C., van der Voort van Zyp, J.R., Viergever, M.A., Castelein, R.M., Weinans, H., van Stralen, M., et al., 2022. MRI-based synthetic CT shows equivalence to conventional CT for the morphological assessment of the hip joint. J. Orthop. Res.® 40 (4), 954–964.

Friedrich, P., Wolleb, J., Bieder, F., Durrer, A., Cattin, P.C., 2024. WDM: 3D wavelet diffusion models for high-resolution medical image synthesis. In: MICCAI Workshop on Deep Generative Models. Springer, pp. 11–21.

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 0.35 T MR images using generative adversarial networks for MR-only liver radiotherapy. Biomed. Phys. Eng. Express 6 (1), 015033.

Galati, F., Cortese, R., Prados, F., Lorenzi, M., Zuluaga, M.A., 2024. Federated multi-centric image segmentation with uneven label distribution. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 350–360.

Gatidis, S., Hepp, T., Früh, M., La Fougère, C., 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 (1), 601.

Gholamiankhah, F., Mostafapour, S., Arabi, H., 2021. Deep learning-based synthetic CT generation from MR images: comparison of generative adversarial and residual neural networks. arXiv preprint arXiv:2103.01609.

Gjesteby, L., De Man, B., Jin, Y., Paganetti, H., Verburg, J., Giantsoudi, D., Wang, G., 2016. Metal artifact reduction in CT: where are we after four decades? IEEE Access 4, 5826–5849.

Gong, K., Guan, J., Liu, C.C., Qi, J., 2018. PET image denoising using a deep neural network through fine tuning. IEEE Trans. Radiat. Plasma Med. Sci. 3 (2), 153–161.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. Adv. Neural Inf. Process. Syst. 27.

Gu, A., Goel, K., Ré, C., 2021. Efficiently modeling long sequences with structured state spaces. arXiv preprint arXiv:2111.00396.

Gui, L., Ye, C., Yan, T., 2024. CAVM: Conditional autoregressive vision model for contrast-enhanced brain tumor MRI synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 161–170.

Guibas, J.T., Virdi, T.S., Li, P.S., 2017. Synthetic medical images from dual generative adversarial networks. arXiv preprint arXiv:1709.01872.

Hamghalam, M., Simpson, A.L., 2024. Medical image synthesis via conditional GANs: Application to segmenting brain tumours. Comput. Biol. Med. 170, 107982.

Harms, J., Wang, T., Petrongolo, M., Zhu, L., 2016. Noise suppression for energy-resolved CT using similarity-based non-local filtration. In: Medical Imaging 2016: Physics of Medical Imaging, vol. 9783, SPIE, pp. 1075–1082.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S., 2017. GANs trained by a two time-scale update rule converge to a local nash equilibrium. Adv. Neural Inf. Process. Syst. 30.

Hicsonmez, S., Samet, N., Akbas, E., Duygulu, P., 2020. GANILLA: Generative adversarial networks for image to illustration translation. Image Vis. Comput. 95, 103886.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. Adv. Neural Inf. Process. Syst. 33, 6840–6851.

Hognon, C., Conze, P.H., Bourbonne, V., Gallinato, O., Colin, T., Jaouen, V., Visvikis, D., 2024. Contrastive image adaptation for acquisition shift reduction in medical imaging. Artif. Intell. Med. 148, 102747.

Hong, K.T., Cho, Y., Kang, C.H., Ahn, K.S., Lee, H., Kim, J., Hong, S.J., Kim, B.H., Shim, E., 2022. Lumbar spine computed tomography to magnetic resonance imaging synthesis using generative adversarial network: Visual turing test. Diagnostics 12 (2), 530.

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, 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 (1), 145–157.

Huang, C.W., Lim, J.H., Courville, A.C., 2021. A variational perspective on diffusion-based generative models and score matching. Adv. Neural Inf. Process. Syst. 34, 22863–22876.

Huang, X., Liu, M.Y., Belongie, S., Kautz, J., 2018. Multimodal unsupervised image-to-image translation. In: Proceedings of the European Conference on Computer Vision. ECCV, pp. 172–189.

Huang, J., Zhong, A., Wei, Y., 2024. A new visual State Space Model for low-dose CT denoising. Med. Phys. 51 (12), 8851–8864.

Huijben, E.M., Terpstra, M.L., Pai, S., Thummerer, A., Koopmans, P., Afonso, M., van Eijnatten, M., Gurney-Champion, O., Chen, Z., Zhang, Y., et al., 2024. Generating synthetic computed tomography for radiotherapy: SynthRAD2023 challenge report. Med. Image Anal. 97, 103276.

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, pp. 4306–4312.

Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1125–1134.

Jain, M., Rai, C., Jain, J., 2023. RadiomicsGAN: Image augmentation and translation using the conditional gan framework for enhanced prediction of brain degeneration. In: 2023 10th International Conference on Computing for Sustainable Global Development. INDIACom, IEEE, pp. 507–513.

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. arXiv preprint arXiv:2306.11984.

Jans, L.B., Chen, M., Elewaut, D., Van den Bosch, F., 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 (2), 343–349.

Ji, W., Chung, A.C., 2024. Diffusion-based domain adaptation for medical image segmentation using stochastic step alignment. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 188–198.

Jiang, L., Mao, Y., Wang, X., Chen, X., Li, C., 2023. CoLa-Diff: Conditional latent diffusion model for multi-modal MRI synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 398–408.

Jiang, J., Veeraraghavan, H., 2020. Unified cross-modality feature disentangler for unsupervised multi-domain MRI abdomen organs segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23. Springer, pp. 347–358.

Jiang, Z., Zheng, Y., Tan, H., Tang, B., Zhou, H., 2016. Variational deep embedding: An unsupervised and generative approach to clustering. arXiv preprint arXiv:1611.05148.

Jin, C.B., Kim, H., Liu, M., Jung, W., Joo, S., Park, E., Ahn, Y.S., Han, I.H., Lee, J.I., Cui, X., 2019. Deep CT to MR synthesis using paired and unpaired data. Sensors 19 (10), 2361.

Jin, K.H., McCann, M.T., Froustey, E., Unser, M., 2017. Deep convolutional neural network for inverse problems in imaging. IEEE Trans. Image Process. 26 (9), 4509–4522.

Jin, D., Xu, Z., Tang, Y., Harrison, A.P., Mollura, D.J., 2018. CT-realistic lung nodule simulation from 3D conditional generative adversarial networks for robust lung segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11. Springer, pp. 732–740.

Kalantar, R., Messiou, C., Winfield, J.M., Renn, A., Latifoltojar, A., Downey, K., Sohaib, A., Lalondrelle, S., Koh, D.M., Blackledge, M.D., 2021. CT-based pelvic T1-weighted MR image synthesis using UNet, UNet++ and cycle-consistent generative adversarial network (Cycle-GAN). Front. Oncol. 11, 665807.

Kamli, A., Saouli, R., Batatia, H., Naceur, M.B., Youkana, I., 2020. Synthetic medical image generator for data augmentation and anonymisation based on generative adversarial network for glioblastoma tumors growth prediction. IET Image Process. 14 (16), 4248–4257.

Kang, S.K., An, H.J., Jin, H., Kim, J.i., 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 (3), 263–271.

Kang, M., Chikontwe, P., Won, D., Luna, M., Park, S.H., 2023. Structure-preserving image translation for cross-modality medical imaging domain adaptation. Available at SSRN 4387006.

Kang, E., Min, J., Ye, J.C., 2017. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. Med. Phys. 44 (10), e360–e375.

Kao, C.H., Chen, Y.S., Chen, L.F., Chiu, W.C., 2021. Demystifying T1-MRI to FDGˆ 18 18-PET image translation via representational similarity. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24. Springer, pp. 402–412.

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 (5), 773–778.

Karimzadeh, R., Ibragimov, B., MRI to synthesis-CT generation using Pix2Pix Framework.

Kazemifar, S., Barragán Montero, A.M., Souris, K., Rivas, S.T., Timmerman, R., Park, Y.K., Jiang, S., Geets, X., Sterpin, E., Owrangi, A., 2020. Dosimetric evaluation of synthetic CT generated with GANs for MRI-only proton therapy treatment planning of brain tumors. J. Appl. Clin. Med. Phys. 21 (5), 76–86.

Khodarahmi, I., Isaac, A., Fishman, E.K., Dalili, D., Fritz, J., 2019. Metal about the hip and artifact reduction techniques: from basic concepts to advanced imaging. In: Seminars in Musculoskeletal Radiology, vol. 23, no. 03, Thieme Medical Publishers, pp. e68–e81.

Kieselmann, J.P., Fuller, C.D., Gurney-Champion, O.J., Oelfke, U., 2021. Cross-modality deep learning: contouring of MRI data from annotated CT data only. Med. Phys. 48 (4), 1673–1684.

Kim, J., Park, H., 2024. Adaptive latent diffusion model for 3D medical image to image translation: Multi-modal magnetic resonance imaging study. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7604–7613.

Kingma, D.P., Welling, M., 2013. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

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: Third International Workshop, SASHIMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 3. Springer, pp. 61–70.

Kläser, K., Varsavsky, T., Markiewicz, P., Vercauteren, T., Atkinson, D., Thielemans, K., Hutton, B., Cardoso, M.J., Ourselin, S., 2019. Improved MR to CT synthesis for PET/MR attenuation correction using imitation learning. In: International Workshop on Simulation and Synthesis in Medical Imaging. Springer, pp. 13–21.

Koh, H., Park, T.Y., Chung, Y.A., Lee, J.H., Kim, H., 2021. Acoustic simulation for transcranial focused ultrasound using GAN-based synthetic CT. IEEE J. Biomed. Heal. Inform. 26 (1), 161–171.

LaMontagne, P.J., Benzinger, T.L., Morris, J.C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A.G., et al., 2019. OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. pp. 2012–2019, MedRxiv.

Le, N., Sorensen, J., Bui, T.D., Choudhary, A., Luu, K., Nguyen, H., 2021. Pairflow: Enhancing portable chest X-ray by flow-based deformation for covid-19 diagnosing. In: 2021 IEEE International Conference on Image Processing. ICIP, IEEE, pp. 215–219.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.

Lee, H., Lee, J., Kim, H., Cho, B., Cho, S., 2018. Deep-neural-network-based sinogram synthesis for sparse-view CT image reconstruction. IEEE Trans. Radiat. Plasma Med. Sci. 3 (2), 109–119.

Lei, Y., Harms, J., Wang, T., Liu, Y., Shu, H.K., Jani, A.B., Curran, W.J., Mao, H., Liu, T., Yang, X., 2019. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. Med. Phys. 46 (8), 3565–3581.

Li, H.B., Conte, G.M., Anwar, S.M., Kofler, F., Ezhov, I., van Leemput, K., Piraud, M., Diaz, M., Cole, B., Calabrese, E., et al., 2023a. The brain tumor segmentation (BraTS) challenge 2023: Brain MR image synthesis for tumor segmentation (BraSyn). ArXiv.

Li, W., Li, Y., Qin, W., Liang, X., Xu, J., Xiong, J., Xie, Y., 2020a. Magnetic resonance image (MRI) synthesis from brain computed tomography (CT) images based on deep learning methods for magnetic resonance (MR)-guided radiotherapy. Quant. Imaging Med. Surg. 10 (6), 1223.

Li, Y., Li, W., Xiong, J., Xia, J., Xie, Y., et al., 2020b. Comparison of supervised and unsupervised deep learning methods for medical image synthesis between computed tomography and magnetic resonance images. BioMed Res. Int. 2020.

Li, C., Liu, X., Li, W., Wang, C., Liu, H., Liu, Y., Chen, Z., Yuan, Y., 2024. U-kan makes strong backbone for medical image segmentation and generation. arXiv preprint arXiv:2406.02918.

Li, H., Paetzold, J.C., Sekuboyina, A., Kofler, F., Zhang, J., Kirschke, J.S., Wiestler, B., Menze, B., 2019. DiamondGAN: unified multi-modal generative adversarial networks for MRI sequences synthesis. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part IV 22. Springer, pp. 795–803.

Li, Y., Shao, H.C., Qian, X., Zhang, Y., 2023b. FDDM: Unsupervised medical image translation with a frequency-decoupled diffusion model. arXiv preprint arXiv:2311.12070.

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, pp. 1–5.

Lin, W., Lin, W., Chen, G., Zhang, H., Gao, Q., Huang, Y., Tong, T., Du, M., Initiative, A.D.N., 2021. Bidirectional mapping of brain MRI and PET with 3D reversible GAN for the diagnosis of Alzheimer's disease. Front. Neurosci. 15, 646013.

Liu, M.Y., Breuel, T., Kautz, J., 2017. Unsupervised image-to-image translation networks. Adv. Neural Inf. Process. Syst. 30.

Liu, X., Emami, H., Nejad-Davarani, S.P., Morris, E., Schultz, L., Dong, M., K Glide-Hurst, C., 2021a. Performance of deep learning synthetic CTs for MR-only brain radiation therapy. J. Appl. Clin. Med. Phys. 22 (1), 308–317.

Liu, F., Feng, L., Kijowski, R., 2019a. MANTIS: Model-augmented neural network with incoherent k-space sampling for efficient MR parameter mapping. Magn. Reson. Med. 82 (1), 174–188.

Liu, L., Johansson, A., Cao, Y., Dow, J., Lawrence, T.S., Balter, J.M., 2020. Abdominal synthetic CT generation from MR Dixon images using a U-net trained with 'semi-synthetic'CT data. Phys. Med. Biol. 65 (12), 125001.

Liu, Y., Lei, Y., Wang, Y., Shafai-Erfani, G., Wang, T., Tian, S., Patel, P., Jani, A.B., McDonald, M., Curran, W.J., et al., 2019b. Evaluation of a deep learning-based pelvic synthetic CT generation technique for MRI-based prostate proton treatment planning. Phys. Med. Biol. 64 (20), 205022.

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. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10012–10022.

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.

Lu, N., Chen, Y., 2024. Multi-category domain-dependent feature-based medical image translation. Vis. Comput. 40 (6), 4519–4538.

Lu, B., Lu, H., Palta, J., 2010. A comprehensive study on decreasing the kilovoltage cone-beam CT dose by reducing the projection number. J. Appl. Clin. Med. Phys. 11 (3), 231–249.

Luo, Y., Yang, Q., Liu, Z., Shi, Z., Huang, W., Zheng, G., Cheng, J., 2024. Target-guided diffusion models for unpaired cross-modality medical image translation. IEEE J. Biomed. Heal. Inform..

Lyu, Q., Wang, G., 2022. Conversion between CT and MRI images using diffusion and score-matching models. arXiv preprint arXiv:2209.12104.

Ma, X., Anantrasirichai, N., Bolomytis, S., Achim, A., 2024. PMT: Partial-modality translation based on diffusion models for prostate magnetic resonance and Ultrasound image registration. In: Annual Conference on Medical Image Understanding and Analysis. Springer, pp. 285–297.

Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., Martel, A.L., 2021. Loss odyssey in medical image segmentation. Med. Image Anal. 71, 102035.

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, 1058487.

Masoudi, S., Anwar, S.M., Harmon, S.A., Choyke, P.L., Turkbey, B., Bagci, U., 2020. Adipose tissue segmentation in unlabeled abdomen MRI using cross modality domain adaptation. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society. EMBC, IEEE, pp. 1624–1628.

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.

McKenzie, E.M., Santhanam, A., Ruan, D., O'Connor, D., Cao, M., Sheng, K., 2020. Multimodality image registration in the head-and-neck using a deep learning-derived synthetic CT as a bridge. Med. Phys. 47 (3), 1094–1104.

McNaughton, J., Fernandez, J., Holdsworth, S., Chong, B., Shim, V., Wang, A., 2023a. Machine learning for medical image translation: A systematic review. Bioengineering 10 (9), 1078.

McNaughton, J., Holdsworth, S., Chong, B., Fernandez, J., Shim, V., Wang, A., 2023b. Synthetic MRI generation from CT scans for stroke patients. BioMedInformatics 3 (3), 791–816.

Mirza, M., Osindero, S., 2014. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.

Mok, T.C., Chung, A.C., 2019. Learning data augmentation for brain tumor segmentation with coarse-to-fine generative adversarial networks. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4. Springer, pp. 70–80.

Morbée, L., Chen, M., Herregods, N., Pullens, P., Jans, L.B., 2021. MRI-based synthetic CT of the lumbar spine: Geometric measurements for surgery planning in comparison with CT. Eur. J. Radiol. 144, 109999.

Morbée, L., Chen, M., Van Den Berghe, T., Schiettecatte, E., Gosselin, R., Herregods, N., Jans, L.B., 2022. MRI-based synthetic CT of the hip: can it be an alternative to conventional CT in the evaluation of osseous morphology? Eur. Radiol. 32 (5), 3112–3120.

Nakhaie, A.A., Shokouhi, S.B., 2011. No reference medical image quality measurement based on spread spectrum and discrete wavelet transform using ROI processing. In: 2011 24th Canadian Conference on Electrical and Computer Engineering. CCECE, IEEE, pp. 000121–000125.

Nijskens, L., van den Berg, C.A., Verhoeff, J.J., Maspero, M., 2023. Exploring contrast generalisation in deep learning-based brain MRI-to-CT synthesis. Phys. Medica 112, 102642.

Niu, T., Zhu, L., 2012. Accelerated barrier optimization compressed sensing (ABOCS) reconstruction for cone-beam CT: phantom studies. Med. Phys. 39 (7Part2), 4588–4598.

Olberg, S., Chun, J., Choi, B.S., Park, I., Kim, H., Kim, T., Kim, J.S., Green, O., Park, J.C., 2021. Abdominal synthetic CT reconstruction with intensity projection prior for MRI-only adaptive radiotherapy. Phys. Med. Biol. 66 (20), 204001.

Olberg, S., Zhang, H., Kennedy, W.R., Chun, J., Rodriguez, V., Zoberi, I., Thomas, M.A., Kim, J.S., Mutic, S., Green, O.L., et al., 2019. Synthetic CT reconstruction using a deep spatial pyramid convolutional framework for MR-only breast radiotherapy. Med. Phys. 46 (9), 4135–4147.

Osman, A.F., Tamam, N.M., 2022. Deep learning-based convolutional neural network for intramodality brain MRI synthesis. J. Appl. Clin. Med. Phys. 23 (4), e13530.

Özbey, M., Dalmaz, O., Dar, S.U., Bedel, H.A., Özturk, Ş., Güngör, A., Çukur, T., 2023. Unsupervised medical image translation with adversarial diffusion models. IEEE Trans. Med. Imaging.

Öztürk, Ş., Duran, O.C., Çukur, T., 2024. DenoMamba: A fused state-space model for low-dose CT denoising. arXiv preprint arXiv:2409.13094.

Paavilainen, P., Akram, S.U., Kannala, J., 2021. Bridging the gap between paired and unpaired medical image translation. In: MICCAI Workshop on Deep Generative Models. Springer, pp. 35–44.

Padole, A., Ali Khawaja, R.D., Kalra, M.K., Singh, S., 2015. CT radiation dose and iterative reconstruction techniques. Am. J. Roentgenol. 204 (4), W384–W392.

Pan, S., Chang, C.W., Peng, J., Zhang, J., Qiu, R.L., Wang, T., Roper, J., Liu, T., Mao, H., Yang, X., 2023a. Cycle-guided denoising diffusion probability model for 3D cross-modality MRI synthesis. arXiv preprint arXiv:2305.00042.

Pan, S., Wang, T., Qiu, R.L., Axente, M., Chang, C.W., Peng, J., Patel, A.B., Shelton, J., Patel, S.A., Roper, J., et al., 2023b. 2D medical image synthesis using transformer-based denoising diffusion probabilistic model. Phys. Med. Biol. 68 (10), 105004.

Pandey, M., Kunda, N.S.S., Kc, P., Singh, T., Nair, R.R., 2024. Bridging the modality gap: Generative adversarial networks for T1-T2 MRI image translation. In: 2024 15th International Conference on Computing Communication and Networking Technologies. ICCCNT, IEEE, pp. 1–5.

Park, S.H., Choi, D.M., Jung, I.H., Chang, K.W., Kim, M.J., Jung, H.H., Chang, J.W., Kim, H., Chang, W.S., 2022. Clinical application of deep learning-based synthetic CT from real MRI to improve dose planning accuracy in Gamma Knife radiosurgery: a proof of concept study. Biomed. Eng. Lett. 12 (4), 359–367.

Pearson, K., 1896. VII. Mathematical contributions to the theory of evolution.—III. regression, heredity, and panmixia. Philos. Trans. R. Soc. Lond. Ser. A, Contain. Pap. A Math. Or Phys. Character (187), 253–318.

Peebles, W., Xie, S., 2023. Scalable diffusion models with transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4195–4205.

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.

Phan, V.M.H., Xie, Y., Zhang, B., Qi, Y., Liao, Z., Perperidis, A., Phung, S.L., Verjans, J.W., To, M.S., 2024. Structural attention: Rethinking transformer for unpaired medical image synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 690–700.

Prashnani, E., Cai, H., Mostofi, Y., Sen, P., 2018. Pieapp: Perceptual image-error assessment through pairwise preference. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1808–1817.

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 Comput. 18, 211–226.

Rajagopal, A., Natsuaki, Y., Wangerin, K., Hamdi, M., An, H., Sunderland, J.J., Laforest, R., Kinahan, P.E., Larson, P.E., Hope, T.A., 2022. Synthetic PET via domain translation of 3-D MRI. IEEE Tans. Radiat. Plasma Med. Sci. 7 (4), 333–343.

Rivera Mindt, M., Arentoft, A., Calcetas, A.T., Guzman, V.A., Amaza, H., Ajayi, A., Ashford, M.T., Ayo, O., Barnes, L.L., Camuy, A., et al., 2024. The Alzheimer's Disease Neuroimaging Initiative-4 (ADNI-4) Engagement Core: A culturally informed, community-engaged research (CI-CER) model to advance brain health equity. Alzheimer' s Dement..

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B., 2022. High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10684–10695.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, pp. 234–241.

Rubin, J., Abulnaga, S.M., 2019. CT-To-MR conditional generative adversarial networks for ischemic stroke lesion segmentation. In: 2019 IEEE International Conference on Healthcare Informatics. ICHI, IEEE, pp. 1–7.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X., 2016. Improved techniques for training GANs. Adv. Neural Inf. Process. Syst. 29.

Salmanpour, M.R., Mousavi, A., Xu, Y., Weeks, W.B., Hacihaliloglu, I., 2024. Do high-performance image-to-image translation networks enable the discovery of radiomic features? Application to MRI synthesis from ultrasound in prostate cancer. In: International Workshop on Advances in Simplifying Medical Ultrasound. Springer, pp. 24–34.

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 (9), 1388–1396.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27 (3), 379–423.

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: First Workshop, DGM4MICCAI 2021, and First Workshop, DALI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 1. Springer, pp. 3–13.

Shin, H.C., Ihsani, A., Mandava, S., Sreenivas, S.T., Forster, C., Cha, J., Initiative, A.D.N., 2020. GANBERT: Generative adversarial networks with bidirectional encoder representations from transformers for MRI to PET synthesis. arXiv preprint arXiv:2008.04393.

Shiri, I., Ghafarian, P., Geramifar, P., Leung, K.H.Y., Ghelichoghli, M., Oveisi, M., Rahmim, A., Ay, M.R., 2019. Direct attenuation correction of brain PET images using only emission data via a deep convolutional encoder-decoder (Deep-DAC). Eur. Radiol. 29, 6867–6879.

Siam, Z.S., Hasan, R.T., Chowdhury, M.H., Sumon, M.S.I., Reaz, M.B.I., Ali, S.H.B.M., Mushtak, A., Al-Hashimi, I., Zoghoul, S.B., Chowdhury, M.E., 2024. Improving MRI resolution: A cycle consistent generative adversarial network-based approach for 3T to 7T translation. IEEE Access.

Siddiquee, M.M.R., Shah, J., Wu, T., Chong, C., Schwedt, T.J., Dumkrieger, G., Nikolova, S., Li, B., 2024. Brainomaly: Unsupervised neurologic disease detection utilizing unannotated T1-weighted brain MR images. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7573–7582.

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: Third International Workshop, SASHIMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 3. Springer, pp. 80–89.

Sikka, A., Virk, J.S., Bathula, D.R., et al., 2021. MRI to PET cross-modality translation using globally and locally aware GAN (GLA-GAN) for multi-modal diagnosis of alzheimer's disease. arXiv preprint arXiv:2108.02160.

Simard, P.Y., Steinkraus, D., Platt, J.C., et al., 2003. Best practices for convolutional neural networks applied to visual document analysis. In: Icdar, vol. 3, no. 2003, Edinburgh.

Simonovsky, M., Gutiérrez-Becker, B., Mateus, D., Navab, N., Komodakis, N., 2016. A deep metric for multimodal registration. In: Medical Image Computing and Computer-Assisted Intervention-MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part III 19. Springer, pp. 10–18.

Singh, N.K., Raza, K., 2021. Medical image generation using generative adversarial networks: A review. Heal. Inform.: A Comput. Perspect. Heal. 77–96.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S., 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning. PMLR, pp. 2256–2265.

Song, Y., Ermon, S., 2019. Generative modeling by estimating gradients of the data distribution. Adv. Neural Inf. Process. Syst. 32.

Song, L., Li, Y., Dong, G., Lambo, R., Qin, W., Wang, Y., Zhang, G., Liu, J., Xie, Y., 2021. Artificial intelligence-based bone-enhanced magnetic resonance image—A computed tomography/magnetic resonance image composite image modality in nasopharyngeal carcinoma radiotherapy. Quant. Imaging Med. Surg. 11 (12), 4709.

Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B., 2020. Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456.

Spadea, M.F., Maspero, M., Zaffino, P., Seco, J., 2021. Deep learning based synthetic-CT generation in radiotherapy and PET: a review. Med. Phys. 48 (11), 6537–6566.

Summers, D., 2003. Harvard Whole Brain Atlas: www. med. harvard. edu/AANLIB/home. html. J. Neurol. Neurosurg. Psychiatry 74 (3), 288–288.

Sun, B., Jia, S., Jiang, X., Jia, F., 2023. Double U-Net CycleGAN for 3D MR to CT image synthesis. Int. J. Comput. Assist. Radiol. Surg. 18 (1), 149–156.

Sun, H., Jiang, Y., Yuan, J., Wang, H., Liang, D., Fan, W., Hu, Z., Zhang, N., 2022. High-quality PET image synthesis from ultra-low-dose PET/MRI using bi-task deep learning. Quant. Imaging Med. Surg. 12 (12), 5326.

Sun, H., Mehta, R., Zhou, H.H., Huang, Z., Johnson, S.C., Prabhakaran, V., Singh, V., 2019. Dual-glow: Conditional flow-based generative model for modality transfer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10611–10620.

Tian, Z., Jia, X., Yuan, K., Pan, T., Jiang, S.B., 2011. Low-dose CT reconstruction via edge-preserving total variation regularization. Phys. Med. Biol. 56 (18), 5949.

Vaidya, A., Stough, J.V., Patel, A.A., 2022. Perceptually improved T1-T2 MRI translations using conditional generative adversarial networks. In: Medical Imaging 2022: Image Processing, vol. 12032, SPIE, pp. 505–511.

Valindria, V.V., Pawlowski, N., Rajchl, M., Lavdas, I., Aboagye, E.O., Rockall, A.G., Rueckert, D., Glocker, B., 2018. Multi-modal learning from unpaired images: Application to multi-organ segmentation in CT and MRI. In: 2018 IEEE Winter Conference on Applications of Computer Vision. WACV, IEEE, pp. 547–556.

Van Den Oord, A., Kalchbrenner, N., Kavukcuoglu, K., 2016. Pixel recurrent neural networks. In: International Conference on Machine Learning. PMLR, pp. 1747–1756.

Van Den Oord, A., Vinyals, O., et al., 2017. Neural discrete representation learning. Adv. Neural Inf. Process. Syst. 30.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

Vega, F., Addeh, A., Ganesh, A., Smith, E.E., MacDonald, M.E., 2024. Image translation for estimating two-dimensional axial amyloid-beta PET from structural MRI. J. Magn. Reson. Imaging 59 (3), 1021–1031.

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., et al., 2021a. Artificial intelligence enables whole-body positron emission tomography scans with minimal radiation exposure. Eur. J. Nucl. Med. Mol. Imaging 48, 2771–2781.

Wang, Z., Bovik, A.C., 2009. Mean squared error: Love it or leave it? A new look at signal fidelity measures. IEEE Signal Process. Mag. 26 (1), 98–117.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612.

Wang, S., Jin, K., Lu, H., Cheng, C., Ye, J., Qian, D., 2015. Human visual system-based fundus image quality assessment of portable fundus camera photographs. IEEE Trans. Med. Imaging 35 (4), 1046–1055.

Wang, T., Lei, Y., Fu, Y., Wynne, J.F., Curran, W.J., Liu, T., Yang, X., 2021b. A review on medical imaging synthesis using deep learning and its clinical applications. J. Appl. Clin. Med. Phys. 22 (1), 11–36.

Wang, Y., Liu, C., Zhang, X., Deng, W., 2019. Synthetic CT generation based on T2 weighted MRI of nasopharyngeal carcinoma (NPC) using a deep convolutional neural network (DCNN). Front. Oncol. 9, 1333.

Wang, C.J., Rost, N.S., Golland, P., 2023a. Spatial-intensity transforms for medical image-to-image translation. IEEE Trans. Med. Imaging 42 (11), 3362–3373.

Wang, J., Yan, B., Wu, X., Jiang, X., Zuo, Y., Yang, Y., 2022. Development of an unsupervised cycle contrastive unpaired translation network for MRI-to-CT synthesis. J. Appl. Clin. Med. Phys. 23 (11), e13775.

Wang, Z., Yang, Y., Sermesant, M., Delingette, H., Wu, O., 2023b. Zero-shot-learning cross-modality data translation through mutual information guided stochastic diffusion. arXiv preprint arXiv:2301.13743.

Wang, Y., Zhou, L., Yu, B., Wang, L., Zu, C., Lalush, D.S., Lin, W., Wu, X., Zhou, J., Shen, D., 2018. 3D auto-context-based locality adaptive multi-modality GANs for PET synthesis. IEEE Trans. Med. Imaging 38 (6), 1328–1339.

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.

Willemsen, K., Ketel, M.H., Zijlstra, F., Florkow, M.C., Kuiper, R.J., van der Wal, B.C., Weinans, H., Pouran, B., Beekman, F.J., Seevinck, P.R., et al., 2021. 3D-printed saw guides for lower arm osteotomy, a comparison between a synthetic CT and CT-based workflow. 3D Print. Med. 7 (1), 1–12.

Xie, T., Cao, C., Cui, Z., Guo, Y., Wu, C., Wang, X., Li, Q., Hu, Z., Sun, T., Sang, Z., et al., 2023. Synthesizing PET images from high-field and ultra-high-field MR images using joint diffusion attention model. arXiv preprint arXiv:2305.03901.

Xing, Z., Yang, S., Chen, S., Ye, T., Yang, Y., Qin, J., Zhu, L., 2024. Cross-conditioned diffusion model for medical image to image translation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 201–211.

Xu, Z., Tang, J., Qi, C., Yao, D., Liu, C., Zhan, Y., Lukasiewicz, T., 2024. Cross-domain attention-guided generative data augmentation for medical image analysis with limited data. Comput. Biol. Med. 168, 107744.

Xue, S., Liu, F., Wang, H., Zhu, H., Sari, H., Viscione, M., Sznitman, R., Rominger, A., Guo, R., Li, B., et al., 2024. A deep learning method for the recovery of standard-dose imaging quality from ultra-low-dose PET on wavelet domain. Eur. J. Nucl. Med. Mol. Imaging 1–11.

Yan, S., Wang, C., Chen, W., Lyu, J., 2022. Swin transformer-based GAN for multi-modal medical image translation. Front. Oncol. 12, 942511.

Yang, J., Park, D., Gullberg, G.T., Seo, Y., 2019. Joint correction of attenuation and scatter in image space using deep convolutional neural networks for dedicated brain 18F-FDG PET. Phys. Med. Biol. 64 (7), 075019.

Yang, H., Qian, P., Fan, C., 2020. An indirect multimodal image registration and completion method guided by image synthesis. Comput. Math. Methods Med. 2020.

Yang, H., Sun, J., Carass, A., Zhao, C., Lee, J., Xu, Z., Prince, J., 2018. 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: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer, pp. 174–182.

Yang, T., Wang, L., 2025. Image translation-based unsupervised cross-modality domain adaptation for medical image segmentation. arXiv preprint arXiv:2502.15193.

Yi, X., Babyn, P., 2018. Sharpness-aware low-dose CT denoising using conditional generative adversarial network. J. Digit. Imaging 31, 655–669.

Yi, Z., Zhang, H., Tan, P., Gong, M., 2017. DualGAN: Unsupervised dual learning for image-to-image translation. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2849–2857.

Yin, Z., Xia, K., Wang, S., He, Z., Zhang, J., Zu, B., 2023. Unpaired low-dose CT denoising via an improved cycle-consistent adversarial network with attention ensemble. Vis. Comput. 39 (10), 4423–4444.

Yu, L., Liu, X., Leng, S., Kofler, J.M., Ramirez-Giraldo, J.C., Qu, M., Christner, J., Fletcher, J.G., McCollough, C.H., 2009. Radiation dose reduction in computed tomography: techniques and future perspective. Imaging Med. 1 (1), 65.

Yu, B., Wang, Y., Wang, L., Shen, D., Zhou, L., 2020. Medical image synthesis via deep learning. In: Deep Learning in Medical Image Analysis: Challenges and Applications. Springer, pp. 23–44.

Yuan, J., Fredman, E., Jin, J.Y., Choi, S., Mansur, D., Sloan, A., Machtay, M., Zheng, Y., 2021. Monte Carlo dose calculation using MRI based synthetic CT generated by fully convolutional neural network for Gamma Knife radiosurgery. Technol. Cancer Res. Treat. 20, 15330338211046433.

Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O., 2018. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 586–595.

Zhang, L., Rao, A., Agrawala, M., 2023. Adding conditional control to text-to-image diffusion models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3836–3847.

Zhang, R., Thibault, J.B., Bouman, C.A., Sauer, K.D., Hsieh, J., 2013. Model-based iterative reconstruction for dual-energy X-ray CT using a joint quadratic likelihood model. IEEE Trans. Med. Imaging 33 (1), 117–134.

Zhang, L., Xiao, Z., Zhou, C., Yuan, J., He, Q., Yang, Y., Liu, X., Liang, D., Zheng, H., Fan, W., et al., 2022. Spatial adaptive and transformer fusion network (STFNet) for low-count PET blind denoising with MRI. Med. Phys. 49 (1), 343–356.

Zhang, L., Zhang, L., Mou, X., Zhang, D., 2011. FSIM: A feature similarity index for image quality assessment. IEEE Trans. Image Process. 20 (8), 2378–2386.

Zhao, S., Geng, C., Guo, C., Tian, F., Tang, X., 2023a. SARU: A self-attention ResUNet to generate synthetic CT images for MR-only BNCT treatment planning. Med. Phys. 50 (1), 117–127.

Zhao, Y., Wang, H., Yu, C., Court, L.E., Wang, X., Wang, Q., Pan, T., Ding, Y., Phan, J., Yang, J., 2023b. Compensation cycle consistent generative adversarial networks (Comp-GAN) for synthetic CT generation from MR scans with truncated anatomy. Med. Phys. 50 (7), 4399–4414.

Zhao, X., Yang, T., Li, B., Yang, A., Yan, Y., Jiao, C., 2024. DiffGAN: an adversarial diffusion model with local transformer for MRI reconstruction. Magn. Reson. Imaging 109, 108–119.

Zhou, X., Wu, J., Zhao, H., Chen, L., Zhang, S., Wang, G., 2025. GLFC: Unified global-local feature and contrast learning with Mamba-Enhanced UNet for synthetic CT generation from CBCT. arXiv preprint arXiv:2501.02992.

Zhu, Y., Min, M.R., Kadav, A., Graf, H.P., 2020. S3vae: Self-supervised sequential vae for representation disentanglement and data generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6538–6547.

Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2223–2232.

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. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 592–601.

Zhuang, X., 2018. Multivariate mixture model for myocardial segmentation combining multi-source images. IEEE Trans. Pattern Anal. Mach. Intell. 41 (12), 2933–2946.

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 (2), 218–227.