



Review Article

A Deep Learning-Based MRI-to-CT Image Translation Using Generative Models

S. Nikitha¹ , K. Lohith² and M. J. Keerthana³ ^{1,2&3}Department of Computer Science and Engineering Jyothy Institute of Technology, Bengaluru, Karnataka, India

Article Information

Abstract

Article History

Received: 26 May 2026

Revised: 20 July 2026

Accepted: 30 August 2026

Published online: 10 September 2026

Keywords

MRI-to-CT Translation

Deep Learning

Generative Adversarial Networks

Diffusion Models

Medical Image Synthesis

Image-to-image Translation

Convolutional Neural Networks

Cross-modality Learning

Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) are both widely used medical imaging, they provide complementary diagnostic clues. Still getting both of them at the same time, ends up taking more time, more money, and for CT also involves radiation exposure, which is not small. Lately, deep learning has really pushed things forward, with cross-modality synthesis where a CT image can be created straight from an MRI scan. In this paper, there is a survey of deep learning based MRI-to-CT translation strategies, with attention on “generate it” style methods such as Generative Adversarial Networks (GANs) and diffusion models. The discussion goes into how models trade accuracy versus realism, plus what kinds of optimization tricks are usually used and what limitations show up in real practice. At the end, it also shares thoughts on upcoming research directions, like how to make the approach more stable and easier to deploy.

Correspondence*

nikitha.s@jyothyit.ac.in

ORCID

S. Nikitha

<https://orcid.org/0009-0005-5798-4834>

K. Lohith

<https://orcid.org/0009-0009-2114-0209>

M.J. Keerthana

<https://orcid.org/0009-0005-0083-865X>

© 2026 Centre for Research and Innovation (CRI). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

I. INTRODUCTION

Medical imaging plays a vital role in contemporary healthcare systems, it basically helps with diagnosis, treatment planning, and disease monitoring [1], [2]. Between the various imaging modalities, Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) are used more often, mostly because they complement each other, kind of like in tandem, not in a straight way. MRI gives strong soft-tissue contrast, so it is suitable for detecting tumours, organs and other pliable structures. Meanwhile, CT is necessary for looking at bone structures and it also provides quantitative electron density information, which is required for radiotherapy planning [1]. Even if these modalities matter, trying to get both MRI and CT scans for

just one patient brings quite a few practical hiccups. The whole process tends to bump up the clinical workload, it also drags on the scanning time, and in the end it can push healthcare costs up. On top of that, CT imaging means patients are exposed to ionising radiation, and this can create longer-term health concerns, especially when the scans are repeated [3]. Also, when we end up acquiring multiple modalities, there can be a sort of spatial misalignment because patient positioning and scanning conditions never really match perfectly, and that ends up harming the subsequent clinical tasks. [1].

The whole point here is to generate synthetic CT scans right from MRI input by using computational tricks, so there is less need for extra CT sittings. It cuts down radiation exposure

too, which helps in the end, and yeah it also makes the medical workflow simpler, while improving overall patient comfort [1], [4]. Early ways for doing MRI to CT translation leaned on atlas matching, plus those older style machine learning approaches. Basically, they tried to translate MRI into CT by using handcrafted cues and some statistical modeling [1]. But, from what people saw, these methods were kinda stuck when it came to representing intricate non linear connections, and they were also frequently unable to carry over well between datasets. The fast advancement of deep learning has, kind of, noticeably reshaped this area [3], [5]. In particular, Deep Neural Networks, especially Convolutional Neural Networks (CNNs) have shown very solid prowess in learning multistage feature depictions straight from raw data [3], [5]–[7]. And beyond that, Generative models, like Generative Adversarial Networks (GANs) plus diffusion models, have appeared as quite capable instruments for image-to-image translation jobs [3], [5], [8], [9].

GANs use an adversarial setup, with a generator and a discriminator. In general, the generator tries to synthesize realistic CT images starting from MRI inputs, and the discriminator simply checks how authentic they look. This kind of tug-of-war or competitive learning helps the whole system make high quality, visually believable images [3]. Conditional GANs, like Pix2Pix, tend to push results further along by using paired input output examples while training which makes them very effective for supervised MRI to CT translation problems [4]. More recently though, diffusion models have started getting a lot more attention as a different generative direction [5], [10], [11]. Unlike GANs, diffusion models end up making images through a kind of progressive denoising, where you begin with random noise and then slowly, step by step, refine it into something structured, like in [5]. People often mention that these models train in a more stable way and they can deliver high-fidelity images with better structural consistency than you might expect. Still, they can be more computationally heavy, so yes it's not exactly cheap. But in medical image generation, diffusion models have actually turned out pretty promising results. Also, newer work keeps looking at hybrid schemes, you know, mixing GANs with diffusion models, so the overall approach can sort of trade off computational burden for improved image quality, as discussed in [5], [12].

Furthermore, methods like attention mechanisms and structure-aware learning have been brought in, to keep anatomical consistency around and, at the same time, raise clinical dependability. Even so, problems still show up in MRI-to-CT translation. Deep learning models tend to lean on big labelled datasets, however in medical imaging those resources are usually limited [13]. Also, the computational load and memory footprint can be so heavy that they make real-time use awkward, especially in clinical workflows, or on edge devices. Another point, generalization across different datasets still isn't guaranteed, and clinical accuracy has to stay intact, not drift. So, in this setting, this survey kind of serves as a broad map of deep learning-based MRI-

to-CT image translation approaches [14], [15]. It looks over generative models, including GANs and diffusion models, and it examines their architectures, plus how they actually perform. It also touches optimization tactics and possible future directions. By gathering recent progress into one place, and highlighting which obstacles matter most, the survey aims to offer useful perspectives for both researchers and practitioners in the area of medical image generation.

II. BACKGROUND

A. Image-to-Image Translation

Image to image translation is basically the job of learning some sort of transformation mapping between two different visual realms, so that an image coming from one side becomes a matching image on the other side. In medical imaging, this often means generating CT images from MRI inputs, while still keeping the anatomical layout and the clinical meaning intact [1], [4], [16], [17].

It is not a trivial task because MRI and CT come from different imaging mechanics, their intensity patterns don't really match, and the way structures get represented can look quite distinct. Classic pipelines would lean on crafted feature extraction and atlas-based alignment. In the atlas-based style, already aligned MRI-CT pairs are taken and used to approximate the modality transformation [1].

Even if these strategies can work well in tidy or controlled settings, they tend to demand very accurate registration and can end up weak when you move across varied datasets. On top of that, changes in patient anatomy, plus differences in how the scans were acquired, tend to lower consistency even more [1]. Handcrafted feature-based approaches tried to model relationships with predefined cues like intensity gradients, texture descriptors, and some structural hints. But in practice, they had a tough time with complicated nonlinear interactions and they also needed a lot of domains know how [3], [18]. Because of that, results were often not great, and it was hard to scale them up, you know. Then deep learning came in, and especially Convolutional Neural Networks, or CNNs, it really pushed things forward. In general, CNN encoder decoder setups can learn layered feature representations directly from the data, so you no longer have to do that heavy manual feature engineering [4], [7].

Even though these methods improved the usual metrics, they still sometimes output blur, mainly because many setups lean on pixel wise loss functions. To deal with that, generative models were proposed, aiming to create sharper and more natural looking images. On top of that, conditional learning frameworks helped too, since they use paired datasets and learn a direct mapping from one domain to the other [4]. When paired samples are missing, unpaired learning strategies have been explored as well, for instance cycle consistent architectures. Still, high quality translation is not fully solved, because modality differences can be tricky and labelled data tends to be limited [14], [19]. Image to image

translation keeps being a live research topic, in particular for medical imaging, where correctness and trustworthiness matter a lot.

B. Generative Models

Generative models have become fundamental to modern image-to-image translation, enabling the synthesis of highquality and realistic images [3], [18], [20]. These models learn the underlying data distribution and generate new samples that resemble real data.

In MRI-to-CT translation, generative models aim to capture complex relationships between imaging modalities while maintaining structural consistency.

1. Generative Adversarial Networks (GANs): GANs are basically built from two neural networks, the generator and the discriminator [3], [8], [9]. The generator makes synthetic CT scans given MRI inputs, while the discriminator judges if those images look real or not. In practice both networks get trained together, inside an adversarial setup. It's like they kind of chase each other, continuously. This tricky competitive learning routine helps GANs output images that look very convincing, with lots of subtle, finegrained textures [4]. Conditional GANs or cGANs, for example Pix2Pix, push that idea further by forcing the generation to depend on some conditioning images. So instead of going fully blind, the model learns a direct correspondence from MRI to CT, and it does that using paired data [4]. GAN based methods are also popular, mainly because they can provide high quality results and still keep inference pretty quick [3]. Still, there are downsides. Training can be unstable, and that often ends up with mode collapse, meaning the generator starts repeating itself and only gives a narrow set of output variations [14]. On top of that, GANs usually demand careful hyperparameter adjustments plus well-chosen loss functions, otherwise performance can drop a lot [5].

2. Diffusion Models: Diffusion models have recently popped up as a strong alternative to GANs for image making, and yeah it's sort of become a big thing. They work by slowly adding noise to the data first, and then learning how to undo that corruption so the original image can come back [5], [10], [11], [21]. During training, the system figures out a denoising function, basically an iterative mechanism that turns random noise into images little by little, with successive refinements. One notable upside is that diffusion models tend to stay training-stable, compared to GANs [14], [22].

They're also less likely to suffer from problems like mode collapse, and they can yield diverse, high-quality results [13]. In addition, diffusion models do well at catching subtle textures and keeping spatial structure consistent, which matters a lot in medical imaging. Still though, diffusion models are computationally demanding. They need multiple repeated steps during inference [14], so without optimization they aren't ideal for real time scenarios. A lot of recent work tries to make sampling faster, using accelerated sampling

methods, and also mixing in hybrid architectures. In summary, both GANs and diffusion models end up playing crucial roles in MRI-to-CT translation. GANs typically give faster inference and better computational efficiency, while diffusion models deliver stronger visual fidelity along with more dependable training behavior. Which one to choose usually comes down to the application needs, like accuracy, speed, and what compute budget you actually have [14].

III. LITERATURE REVIEW

Deep learning-based methods have really moved forward MRI-to-CT image translation, and generative models keep showing up as kind of the main idea [3]. In that group, Generative Adversarial Networks GANs have been used a lot, mainly because they can make images that look high quality and realistic [14], [23], [24].

One of the earlier and still very influential papers here is Pix2Pix [4]. It brought in conditional GANs cGANs, for supervised image to image translation. With paired datasets, Pix2Pix learns a mapping from MRI images to CT images directly, and it tends to keep the spatial layout intact, so the output remains detailed [4]. After Pix2Pix had its success, people started to extend it to handle some weaknesses, like structural consistency and also generalization. A bunch of work tried better loss functions, plus architectural changes to raise overall image quality. For example, perceptual loss and feature matching can help keep the higher-level semantic info, and they can also reduce the blurry look. There are also multi scale discriminator setups, which try to look at both the big picture and the smaller local patterns, so the synthesis improves [4], [24], [25].

More structure aware designs have later boosted MRI-to-CT translation by explicitly guarding anatomical details [1]. As one example, SA-GAN [26] adds structural information into the training process. That way, key anatomical features, like organ boundaries, are more likely to be kept in the generated CT. This matters a lot in clinical use, because even tiny inaccuracies in structure can influence diagnosis and treatment planning, and so on. Another big research direction touches unpaired image translation methods. In real world situations, getting perfectly aligned MRI-CT pairs is usually hard, like really hard. Because of that, models such as CycleGAN try to cope by learning mappings without needing paired data [16], [27], [28]. They typically introduce cycle consistency loss, so that an image translated into the other domain can be mapped back and still resemble the starting one, which in turn makes training feel more stable and coherent, even if it is not perfect. Still, despite the flexibility, unpaired strategies might offer less exact anatomical correspondence compared with supervised approaches [14].

More recently, diffusion based models have been presented as a very hopeful alternative to GANs for image synthesis [5]. Instead of jumping directly from one image to another, these methods gradually reshape random noise into structured outputs, via a chain of iterative denoising steps. In

medical imaging, diffusion models have shown strong results when it comes to generating high quality images, and they tend to preserve structures with better reliability [13]. One major upside for diffusion models is training stability. GANs often run into things like mode collapse, and they can require careful regulation between generator and discriminator, so the overall process becomes a bit tricky. Diffusion models instead lean on a probabilistic formulation, which usually makes optimization calmer, and yields more steady outcomes across datasets [14]. They also tend to reflect the underlying data distribution more faithfully, which leads to more diverse outputs [5]. That said, diffusion models are computationally demanding, since the sampling process needs many steps in sequence. Because of this, they are often less suitable for real time use, unless you add acceleration tricks, such as faster sampling strategies or model compression [14]. Comparative investigations usually point at clear trade-offs between GANs and diffusion models. GANs provide swifter inference and reduced computational overhead, so they fit better for time sensitive scenarios, while diffusion models lean toward higher quality and stronger structural consistency, though at a higher compute cost, and this difference matters especially when deployment constraints are strict.

In the literature survey on MRI-to-CT image translation, most of the time people try to look at the existing approaches, sort of breaking things down a bit in terms of data preparation, model designs, training strategies and optimization methods. A bunch of studies point out that those parts, all together, can help with getting better accuracy, more steady results, and also higher efficiency when producing the CT images [13]. The main aim behind these works is to build models that can really map MRI images to their corresponding CT outputs while keeping the anatomical consistency, and also staying clinically meaningful [1]. Moreover, the more recent trend talks a lot about advanced generative models, like GANs, as well as diffusion-based methods, and how they can produce sharper, more realistic image synthesis results.

A. Data Preprocessing

Data preprocessing is a step that matters a lot, because the actual input quality tends to decide how well the model behaves [13]. At first, MRI and CT volumes are aligned via image registration, so there is a clear spatial match between modalities [1]. This alignment can be done with rigid registration, or deformable registration depending on what the dataset looks like, and what the use case requires.

After alignment, intensity normalization gets applied to bring pixel values into a common scale across scans. This helps lower the variability that can come from imaging hardware or acquisition protocols [1]. Noise cleanup is also done, for example Gaussian filtering along with other denoising algorithms, to make the image content look crisper and more usable [29]. Sometimes segmentation is introduced to carve out areas of interest, organs or tissue regions for instance, so the model can concentrate on the most informative structures

[13]. On top of that, data augmentation is often used with rotation, flipping, scaling, and intensity shifting, in order to enlarge the training pool and improve the ability to generalize [14].

B. Model Architectures

Deep learning models, especially generative ones, they end up being the core for MRI-to-CT translation systems [4], [22], [30]. In practice these models acquire rather involved nonlinear connections between the MRI side and the CT side.

1. GAN-based Models: Generative Adversarial Networks (GANs) are often used for image-to-image translation tasks [4]. In particular, conditional GANs (cGANs), like Pix2Pix, learn direct mappings between MRI and CT data, but they do this with paired datasets [4]. In these setups the generator, basically turns MRI inputs into synthetic CT images, and at the same time the discriminator judges how believable the result looks. Through this kind of adversarial game, the model tends to output sharper and more realistic images [3]. Some variants go further; for example, CycleGAN supports translation when there are no pairs by enforcing cycle consistency loss. Meanwhile structure aware GANs help keep anatomical, or more “body structure” related, features during the conversion [26]. Even so, GANs can be a bit touchy. They usually need careful tuning, and they may run into training instability which then lowers the overall performance [3].

2. Diffusion Models: Diffusion models are kind of a recent advancement for image creation. Basically, they work by gradually adding noise to the data, then the model tries to undo that mess, step by step, through iterative denoising [5]. When training, the model learns to predict and remove that noise little by little, so you end up with pretty high-quality images. Also diffusion models are often described as stable, and they can capture fine grained features plus structural consistency, which matters a lot in medical imaging [5]. Still, because it’s all iterative, diffusion models end up being computationally expensive, and they usually need more computation time than GANs [5].

3. Hybrid Architectures: Recent research looks into hybrid approaches that mix GANs with diffusion models, so they can borrow the strong points from both, but in a sort of unified way [14]. The goal is to keep computation cost reasonable, while still producing image synthesis that looks high-quality and crisp. Also, attention mechanisms together with multi scale feature extraction are usually woven in, to push the results further [26].

C. Training Strategy

Training a model is basically about tweaking the network parameters so the loss function, which basically quantifies the gap between generated and real CT images, gets smaller [3], [25]. For GAN-based approaches, the adversarial loss is mixed together with reconstruction objectives like L1 or

L2, so the outputs look realistic but also remain faithful, not just plausible. Sometimes a perceptual loss is added as well, to catch more abstract, high-level semantic cues [4].

With diffusion models, the learning process is oriented toward reducing the mismatch between the noise that the model predicts and the noise that is actually present. Since the whole setup is probabilistic, it tends to show stable convergence, and the generalization usually improves too. Beyond that, people commonly bring in methods like transfer learning, later-stage fine-tuning, knowledge distillation, plus hardware-aware optimization. The point is to squeeze out more performance and then adjust the model so it can run well during deployment on particular devices [14].

D. Optimization Techniques

For real-world deployment, model optimization is important, especially when we're talking about resource-constrained devices [5]. Deep neural networks usually come with big computational needs and also memory, that's kind of the whole issue [13]. Pruning helps by removing "redundant" connections, so the model becomes smaller, plus inference becomes quicker [13], [31]. Then quantization is used, it changes parameter precision down a notch, which reduces memory footprint, and it can make computation faster [14]. There are still other strategies like knowledge distillation, where a reduced size model is trained so it imitates a bigger, stronger model. Also, hardware aware optimization exists, it tunes the architecture for efficient execution on GPUs or on embedded systems, and yes this matters a lot [14].

In order to judge MRI-to-CT translation models, researchers tend to use a bunch of quantitative metrics, to see how closely the produced image, really resembles the reference CT image, given as the ground truth [13]. Usually these measurements cover separate angles, like pixel level correctness, structural consistency, and also something closer to perceptual resemblance. In other words, they are not all saying the exact same thing, but they are all related to matching the final look as much as possible.

E. Peak Signal-to-Noise Ratio (PSNR)

Peak Signal-to-Noise Ratio, usually called PSNR, is a metric people often use for judging the quality of image reconstruction. Basically, it looks at how the highest possible pixel value compares to the noise that sneaks in during the image generation process, so it gives a ratio. It is described in [13], [26].

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right) \quad (1)$$

MAE (Mean Absolute Error): Kind of measures the average of the absolute gap between predicted and the ground truth images, not exactly poetic but effective.

FID (Fre'chet Inception Distance): Evaluates how close the distributions are for generated versus real images, this is pretty common in generative models, yes. Perceptual Metrics: capture likeness using deep feature representations rather than plain pixel level differences [13] and sometimes people say it's more "human like."

where MAX represents the maximum possible pixel value of the image, and MSE is the Mean Squared Error between the generated and ground truth images. Higher PSNR numbers usually mean a better reconstruction, and less distortion too. But PSNR mainly focuses on pixel level disagreements, so it might not completely show how people actually perceive the image quality [13], [26].

F. Structural Similarity Index (SSIM)

The Structural Similarity Index, SSIM, kind of evaluates how similar two images look to a person, it does this by looking at luminance, contrast and structural data.

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

where μ_x and μ_y are the mean intensities, σ^2 and σ^2 are the variances, and σ_{xy} is the covariance between the two images. C_1 and C_2 are constants to stabilize the division. SSIM is pretty important for medical imaging, because it keeps the structural information while also reflecting perceptual quality in a better way [13], [26].

G. Root Mean Square Error (RMSE)

Root Mean Square Error, RMSE, captures sort of the average scale of pixel-by-pixel discrepancies between the generated imagery and the ground truth images, [13]. both quantitative metrics and visual quality assessment [13], [14].

GAN-based models produce realistic images and preserve spatial structure effectively [4]. Methods like Pix2Pix achieve good performance using paired data and also offer fast inference, making them suitable for real-time applications [3], [14]. However, GANs may suffer from training instability and issues like mode collapse, which can reduce output diversity [3].

Diffusion models, on the other hand, generate higher quality images with better structural consistency and detail preservation [5], [13]. They often achieve improved SSIM and RMSE scores compared to GANs [13]. However, their iterative denoising process increases computational cost and slows down inference [14].

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (I_i - \hat{I}_i)^2} \quad (3)$$

1. *Computational Cost*: where I_i (sort of the ground truth pixel value) and $\hat{I}_i$ (the predicted pixel value) are defined, and N is the overall count of pixels. Lower RMSE means better model performance. Though, kind of like PSNR RMSE is mostly locked on pixel level accuracy, and it does not really reflect perceptual quality, not even close.

2. *Additional Metrics*: Along with PSNR, SSIM, and RMSE, there are also other metrics, used to cover in a more comprehensive way the evaluation. In practice people might look at things like perceptual similarity, structure consistency, or error distributions, so they get an overall, more coherent picture. The high computational requirements of diffusion models limit their use in real-time and resource-constrained environments [13].

H. Quantitative Comparison

The comparative analysis of different model types based on common evaluation metrics [14]. GAN (Pix2Pix) models provide high PSNR and moderate SSIM, along with fast inference, making them suitable for real-time applications. Diffusion models achieve very high PSNR and high SSIM, offering superior image quality and better structural consistency. However, they require higher computational cost and slower inference [14]. Hybrid models combine the strengths of both approaches, providing high PSNR and SSIM with moderate inference time. Overall, GAN-based models balance accuracy and speed, diffusion models excel in quality, and hybrid models offer a trade-off between performance and efficiency [14].

1. *Qualitative Analysis*: GAN-generated images are sharp and visually realistic but may miss fine anatomical details [13], [14]. Diffusion-based outputs show better continuity and capture subtle structures more accurately, which is important for clinical applications [13].

I. Discussion

The choice of model depends on application requirements. GANs are preferred for faster processing, while diffusion models are suitable for high-quality image generation [13], [14]. Hybrid approaches that combine both methods show promise for balancing performance and efficiency.

IV. CHALLENGES, LIMITATIONS AND FUTURE DIRECTIONS

Even with major progress in MRI to CT image translation, there are still a few obstacles that keep these models from being really practical in day-to-day clinical environments [13]. Most deep learning approaches, especially generative ones like GANs and diffusion models, tend to demand a lot of computation, both during training, and also when running inference [14]. For diffusion models specifically, the process usually needs repeated denoising rounds, which pushes inference time up quite a bit, and also adds extra computational burden [5], [13]. So, in many cases they are

less fitting for real time clinical workflows, and they are harder to run on resource constrained hardware [14].

A. Limited Availability of Labelled Data

The performance of supervised learning models' kind of depends on having a lot of large, high-quality paired MRI-CT datasets. But collecting those is difficult, because of high costs, privacy issues, and the fact that getting perfectly aligned images, is not easy at all. Because of this bottleneck model generalization gets limited, and it can easily drift toward overfitting.

B. Training Instability in GANs

GAN-based models often end up training in a kind of unstable way, which then causes problems like mode collapse, where the generator shows only a narrow set of variations in its outputs [3]. Getting the process steady really needs careful tuning of hyperparameters, along with choices of loss functions, and overall, that makes GANs much harder to optimize, or at least it feels that way in practice.

C. Inference Speed in Diffusion Models

While diffusion models give really nice outputs, the inference process gets computationally heavy because there are multiple sampling steps [14]. So processing takes longer, in comparison with GAN based models, which ends up limiting where they can actually be used in time-sensitive scenarios [13].

D. Generalization and Robustness

When models are trained on a particular dataset they might not generalize properly to other datasets, especially if the imaging conditions shift, or the noise behaves a bit different, or the patient populations aren't the same. Making the system robust across a variety of clinical scenarios is still a big obstacle [14], and well yeah it keeps being hard.

E. Clinical Reliability

Even with quantitative metrics like PSNR and SSIM that seem to show strong performance, they do not always correspond to what is truly clinically relevant. In fact, even minor structural inaccuracies in generated CT images can have a rather significant effect on medical use cases for example radiotherapy planning [14]. Future research on MRI-to-CT translation is expected to keep chipping away at the current limitations, while also pushing for better model efficiency and accuracy, as reported in [13], [14].

F. Lightweight and Efficient Models

Building lighter-weight models, that need fewer computational resources is really important for getting them out there on edge devices and for real time applications [14]. In practice, approaches like pruning the model, quantizing

the parameters, and using knowledge distillation can be examined more deeply, so we can get more efficient implementations [13].

G. Hybrid Architectures

Mixing the strong points from GANs with diffusion models feels like a rather promising path for research [14], you know, not just theoretically. Hybrid setups can use the quick inference ability of GANs, and also the top-quality generation property of diffusion models. In that way they can somehow strike a trade-off between speed and output performance [13], which is kind of the whole idea.

H. Unsupervised and Self-Supervised Learning

To overcome the limitation of labelled data, future work can lean into unsupervised and self-supervised learning schemes [14]. In other words, these strategies are able to pull out meaningful representations from unlabelled data, and that tends to cut down on the dependence on paired datasets [13].

I. Domain Adaptation and Generalization

Improving model generalization between multiple datasets and different imaging conditions is like a second big research focus. With domain adaptation methods, the model can more or less keep the same performance across varied clinical environments [14].

J. Integration with Clinical Workflows

Future efforts should really home in on integrating MRI-to-CT translation systems into actual clinical workflows, not just keeping them stuck in experimental settings, [13]. That also means pushing for better model interpretability, making sure results are dependable, and validating the performance by means of clinical studies [14].

K. Acceleration of Diffusion Models

Optimizing diffusion models so they can run faster at inference time is, honestly, a big deal. There are ways like accelerated sampling approaches, plus efficient architecture choices, that can make these models feel more suited for real time use, or things that need low latency, if you will.

V. CONCLUSION

This paper kind of presents a comprehensive survey of deep learning-based MRI-to-CT translation methods [3]. Generative models, especially Generative Adversarial Networks (GANs) and diffusion models, have really moved the field forward, by enabling accurate and realistic picture synthesis [5]. For GAN based approaches, the main appeal is fast inference and computational efficiency, so they fit better with real time style deployments. On the other hand, diffusion models tend to deliver superior image quality, plus better structural consistency, which matters a lot for medical imaging jobs [13], [14]. Still, even with all these gains, there

are hurdles like high computational cost, limited labelled data, and weak generalization across different datasets, those remain significant obstacles [13]. Fixing these issues is basically essential for real world clinical deployment [14]. Looking ahead, future work will likely prioritize efficient and robust models, while also looking at hybrid architectures, and trying to improve how these systems integrate into clinical workflows. With continued progress, MRI to CT translation could reduce reliance on multiple imaging modalities and, in turn, help improve patient care in medical imaging scenarios [13], [14].

Declaration of Conflicting Interests

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Use of Artificial Intelligence (AI)-Assisted Technology for Manuscript Preparation

The author(s) utilised generative AI technologies solely to improve language clarity and refine grammar. No images were generated, manipulated, or altered using AI-assisted technologies.

REFERENCES

- [1] Q. Lyu and G. Wang, "Conversion between ct and mri images using diffusion and score-matching models," *IEEE Transactions*, 2020.
- [2] S. Nikitha, H. P. Sudhanva, T. A. Nayana, L. Varshini, S. V. Pandurangi, and P. Soukar, "An ai-based system for pneumonia detection in chest x-rays using deep learning models," *Asian Journal of Engineering and Applied Technology*, vol. 14, no. 1, pp. 7–13, 2025.
- [3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.
- [4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [5] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.
- [6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *ICLR*, 2015.
- [7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [8] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for gans," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [9] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional gans," in *ICLR*, 2016.
- [10] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," in *NeurIPS*, 2021.
- [11] A. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *ICML*, 2021.
- [12] H. Yang, J. Sun, H. Carass, Z. Xu, C. Zhao, and J. L. Prince, "Unpaired brain mr-to-ct synthesis using a structure-constrained cyclegan," *IEEE Transactions on Medical Imaging*, 2018.
- [13] R. Graf, J. Schmitt, and S. S. et al., "Denoising diffusion-based mri to ct image translation enables automated spinal segmentation," *European Radiology Experimental*, 2023.
- [14] A. Moschetto, L. Puglisi, A. Sargood, P. Dell'Acqua, F. Guarnera, S. Battiato, and D. Ravi, "Benchmarking gans, diffusion models, and flow matching for t1w-to-t2w mri translation," *arXiv preprint*, 2025.
- [16] Y. Jin and W. L. et al., "Deep learning based ct image generation from mri using cnn," *IEEE Access*, 2019.
- [17] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE*

- International Conference on Computer Vision (ICCV)*, 2017.
- [18] M. Nie, L. Trullo, J. Lian, C. Petitjean, and S. Ruan, "Medical image synthesis with context-aware generative adversarial networks," *Medical Image Analysis*, 2017.
 - [19] D. Kingma and M. Welling, "Auto-encoding variational bayes," in *ICLR*, 2014.
 - [20] S. Ulyanov, D. Vedaldi, and V. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *arXiv preprint*, 2016.
 - [21] J. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," in *NeurIPS*, 2019.
 - [22] Y. Song and S. Ermon, "Score-based generative modeling through stochastic differential equations," in *ICLR*, 2021.
 - [23] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015.
 - [24] K. Armanious, C. Jiang, and M. F. et al., "Medgan: Medical image translation using gans," *Computerized Medical Imaging and Graphics*, 2020.
 - [25] Z. Chatsias, T. Joyce, M. Giuffrida, and S. A. Tsaftaris, "Multimodal mr synthesis via modality-invariant latent representation," *IEEE Transactions on Medical Imaging*, 2018.
 - [26] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions*, 2017.
 - [27] H. Emami, M. Dong, S. P. Nejad-Davarani, and C. K. Glide-Hurst, "Sagan: Structure-aware gan for organ-preserving synthetic ct generation," *arXiv preprint*, 2021.
 - [28] Y. Wolterink, T. Leiner, M. Viergever, and I. Isgum, "Deep mr to ct synthesis using unpaired data," *Medical Image Analysis*, 2017.
 - [29] L. Hiasa, Y. Otake, and N. T. et al., "Cross-modality image synthesis from unpaired data using cyclegan," *Medical Imaging Conference*, 2018.
 - [30] Z. Huang and N. Wang, "Data augmentation techniques in medical imaging," *IEEE Access*, 2020.
 - [31] F. Milletari, N. Navab, and S. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3DV*, 2016.
 - [32] S. Chen, K. Ma, and Y. Zheng, "Med3d: Transfer learning for 3d medical image analysis," *arXiv preprint*, 2019.