Medical Image Translation from PET to CT using Deep Learning Techniques

by

Md. Emon (200021332)

Shaheed Rahman (200021336)

Sheikh Shahriar Kabir Munna (200021337)

Al Mahee Khan(200021342)

In Consideration of Partial Fulfillment for the Requirements of the Degree of

BACHELOR OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

Department of Electrical and Electronic Engineering

Islamic University of Technology

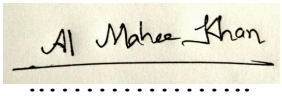
Gazipur, Bangladesh

October 2025

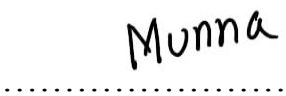
Declaration

This is to certify that this thesis titled 'Medical Image Translation of PET to CT using Deep Learning Techniques' has been carried out under the supervision of Mr. Nadim Ahmed, Assistant Professor, Department of Electrical and Electronic Engineering, Islamic University of Technology, and has not been submitted elsewhere for any academic qualification or degree.

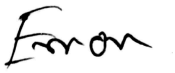
Authors



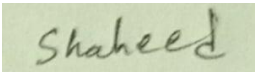
Al Mahee Khan



Sheikh Shahriar Kabir Munna



Md. Emon



Shaheed Rahman

Medical Image translation from PET to CT using Deep Learning Techniques

Approved By

.....

Mr. Nadim Ahmed

Assistant Professor

Department of Electrical and Electronic Engineering

Islamic University of Technology

Gazipur, Bangladesh

Acknowledgement

At first, we express our gratitude to Almighty Allah for His endless blessings, guidance, and mercy throughout this journey. Without His grace and kindness, it would not have been possible for us to complete this study. His strength and wisdom have been our constant source of inspiration.

We would like to extend our heartfelt thanks to our respected supervisor,Mr. Nadim Ahmed, Assistant Professor, Department of Electrical and Electronic Engineering, Islamic University of Technology (IUT). He was always there when we needed any guidance and his constant encouragement was truly instrumental in the successful completion of our task. We are especially grateful for his patience, insightful feedback, and motivation, which truly helped us to stay focused throughout the time and to take our work to the next level at every time. It has been an honor to work under his supervision and to learn from his knowledge and experience.

We would also like to thank the Department of Electrical and Electronic Engineering**,** IUT, for providing us with the facilities we needed along with the technical support and a healthy academic environment which helped us to carry out our work. This played a crucial role in completing our task.

Abstract:

Medical imaging plays a crucial role in the early detection, diagnosis, and treatment planning for various diseases, particularly cancer. Positron Emission Tomography (PET) and Computed Tomography (CT) are two common imaging techniques for cancer detection and tumor characterization. PET gives information about the metabolic activity and functional data, while CT gives detailed anatomical structure. The combination of both modalities improves the accuracy of diagnoses, but it also has some downsides, such as more radiation exposure and higher costs. To address this limitation, we propose an image translation framework for synthesizing CT images directly from PET scans using advanced deep learning techniques. This work examines three deep learning models specifically: a Conditional Generative Adversarial Network (cGAN), a Conditional Diffusion Model, and a Hybrid Diffusion-GAN architecture. The models were trained and evaluated using the QIN-Breast dataset, which contains paired PET and CT images. Quantitative evaluation metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) were used to evaluate the performance of the models. The findings indicated that the GAN-based model provided the best reconstruction quality, achieving a PSNR of 32.55 dB and an SSIM of 0.8799. The Diffusion model, while producing smooth and anatomically consistent images, resulted in a PSNR of 15.46 dB and an SSIM of 0.776. The Hybrid model effectively combined the strengths of both approaches, producing a PSNR of 21.48 dB and an SSIM of 0.833. This study shows the capacity of deep learning to improve medical imaging workflows by generating synthetic CT generation from PET data, thus reducing patient radiation exposure and associated healthcare costs. Future research will examine the extension of these models to 3D volumetric reconstructions and their integration into clinical workflows.

Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Background

Medical imaging is an essential component of modern healthcare that provides non-invasive visualization of internal organs, tissues, and biological processes . It is widely used for diagnosis of different diseases and their conditions early on, plan treatment, and keep an eye on their condition. There are many types of imaging, but Computed Tomography (CT) and Positron Emission Tomography (PET) together are especially important in tumor staging because they work well together [1]. CT scans give doctors detailed information about the patient's internal structure with anatomical information like the density of tissues, the shape of organs, and the shape of bones . For planning surgery, targeting radiotherapy, and finding the right place in the body, this structural clarity is very important and CT image provides us that . PET, on the other hand, looks at the metabolic and functional activity of tissues at the molecular level . This information is very important for finding cancer and characterizing a tumor whether a surgery is necessary, then tracking its progress, and judging how well a treatment is working .

Modern clinical practice often uses combined PET/CT scanners to make diagnoses as quickly and accurately as possible . These scanners combine functional and anatomical imaging into one workflow. The PET part finds areas with higher metabolic activity, which is usually used to characterize a tumor benign or malignant, and the CT part gives anatomical reference and helps with attenuation correction. This integration makes diagnosis and treatment planning a lot more accurate. But the combined approach does come with some problems. CT scans use more ionizing radiation, which can raise the risk of complications from radiation exposure over time [2]. And for patients this diagnosis needs to be done multiple times to track their improvements. PET/CT procedures are also expensive and need advanced infrastructure, which makes them harder to get to, especially in healthcare settings with limited resources.

Because of these problems, more and more people are interested in computational cross-modality image synthesis, which is when one type of imaging is made from another. One way to do this is to generate an equivalent CT image directly from PET scans. This could lower radiation exposure, lower the cost of imaging, and make it easier to get the information you need while keeping important anatomical information for the surgery. To achieve this Convolutional Neural Network (CNN) was used to generate an image of one modality from another modality but in recent studies it has been shown that Generative Adversarial Network performs better than traditional Convolutional Neural Network (CNN)[3]. These technologies have made it possible to create high-quality images in both natural and medical imaging. GANs are great at making realistic images because they use a Generator–Discriminator framework where the generator tries to generate an image close to the original one and the discriminator judges how close the two images are. And both the Generator and Discriminator are trained together by back propagation and updating the model [4]. Another approach is Conditional Generative Adversarial Network where the model could be used to learn multi-model [5]. Also, another crucial advantage of Conditional Generative Adversarial Network while doing image translation is not only learns the mapping from input image to output image, but also learns a loss function to train this mapping [6], [7]. Integrating Generative Adversarial Network technique in Medical images by merging the adversarial framework with non adversarial losses opens a new door for image translation technique and the special attribute of it was utilizing the style transfer losses [8]. Another modern technique of image synthesis is Diffusion Probabilistic model specially with denoising encoder which uses the technique of predicting the noise added to the image at each one step rather than predicting the whole original image [9]. Also with fewer modifications the performance of Denoising Diffusion Probabilistic Model can be improved more [10]. Also integrating transformers in Denoising Diffusion Models promises to show improved results in the case of medical image translation by preserving spatial relationships between anatomical structures of the body [11].

1.2 Motivation

The motivation for this research arises from several clinical, economic, and technological considerations.

Firstly, radiation exposure is a significant issue in medical imaging. PET scans give important information about metabolism, but they are usually done with CT scans, which add more ionizing radiation. People diagnosed with cancer need to go through this diagnosis multiple times. Now the more they go through this diagnosis the more radiation exposure they get. That may cause more health related problems in their later life. So if we can somehow generate the CT image from the PET image then we can reduce the necessity for repeated CT scans. Thus we can reduce the radiation exposure without reducing the clinical accuracy.

Secondly, these diagnosis modalities are less accessible and they are costly. It costs a lot of money to run and maintain PET/CT scanners because they need special infrastructure and trained staff. Many hospitals, especially those in low and middle income areas can not afford to do both the imaging often. So it becomes harder for patients to get important diagnostic tools. So equivalent CT image generation will cost less and make it easier for people to get accurate diagnosis, all while keeping clinically relevant image quality.

Thirdly, the deep learning technique and generative modeling is getting improved day by day. This makes it possible to translate medical images in new ways. GANs have done very well at translating both paired and unpaired images. It makes high-quality outputs in many different fields. On the other hand Diffusion models can create fine structural details with its iterative denoising process. So it preserves the anatomical accuracy as well as the structural data that are critical for clinical interpretation.

But there are some problems as well. Medical imaging datasets are often small and private. Also they are varied, and have different resolutions. As a result they need to be preprocessed so that the model gets less noise and gives accurate results. So normalization and resizing is necessary. Additionally, these generative models need to find balance between realism, structural accuracy and pixel-level fidelity. Also computational resources can become an issue if the dataset is huge and more training time.

1.3 Problem Statement

PET/CT imaging gives a lot of useful diagnostic information. But it has some downside as well such as high radiation exposure, high costs, and limited availability. Conventional computational methods could not produce CT images with adequate structural fidelity. As a

result the clinical accuracy was less. So, the problem this research is trying to solve can be summed up like this:

How can advanced deep generative models be utilized to generate structurally precise and clinically significant CT images from PET scans, thereby diminishing dependence on dual-modality imaging while maintaining diagnostic integrity?

To solve this problem, we need to create generative architectures that can keep fine anatomical details, work with small and diverse datasets, and use clinically relevant metrics to judge synthetic images. If this problem is solved correctly, it could lower healthcare costs, lower radiation exposure for patients, and make it easier for people to get advanced diagnostic imaging.

CHAPTER 2
LITERATURE REVIEW & METHODOLOGY

2.1 Literature Review

Integration of PET image and CT image is very important for medical disease recognition as it combines functional and anatomical imaging, and with more information diagnostic accuracy improves. In the case of tumor detection, staging ,therapy planning, and treatment response assessment the combined information of PET/CT is very crucial. There are some challenges as well like radiation exposure and cost to perform both medical tests [12].
In the case of breast cancer, the metabolic information from PET and anatomical detail from CT increases the accuracy in detecting primary tumors, assessing axillary lymph node involvement and identifying distance metastases. Though there are some limitations of PET and CT images in detecting small or low grade lesions due to limited spatial resolution and variable FDG uptake, this shows immense strength in staging, restaging and monitoring treatment responses [13].

Translating an image from one modality (CT) to another modality is very complex as the relationship between these is not linear, and in this task rather than conventional image synthesis methods Deep learning approaches perform better [14]. Traditional approaches for example atlas-based or segmentation-based pseudo-CT generation usually struggle to capture the full fidelity of CT image from PET image. Deep learning methods have emerged as a powerful solution for this *cross-modality* translation, ensuring data-driven nonlinear mappings to produce high-quality synthetic images [14].

Initially deep learning approaches for PET image-to-CT image were mostly based on convolutional neural networks (CNNs) trained to regress CT intensities from PET inputs. But the problem was naive CNNs optimized with pixel-wise losses (like mean squared error) tended to produce blurred results, because minimizing simple Euclidean distance leads to averaging of plausible outputs. This limitation or this major disadvantage leads for the exploration of advanced generative models, especially Generative Adversarial Networks (GANs), The speciality of Generative Adversarial Network is it can learn more perceptually

realistic mappings because of the inclusion of adversarial loss term. Here simultaneously two models learn through backpropagation . One is Generator and another is Discriminator. The core idea of Generator is to generate the image which will be very close to the original image and the discriminator is supposed to detect whether the generated image looks real or fake. The Generator model learns how to fool the discriminator by generating an image very close to the original image. And the discriminator also learns to detect the real and fake image properly. The challenging part here is to make these two models balanced. If one of them becomes very strong compared to the other, the whole model will collapse, because the generator learns from the feedback of the discriminator. The performance of the generator depends on the robustness of the discriminator. Also if the Discriminator becomes too strong, the generator gets no crucial information which part of the image is to improve to make it look like the real image. Tuning the parameters to make two models robust is the key here [4].

U-Net , specifically designed for biomedical imaging is basically a fully convolutional neural network which features a symmetric encoder-decoder structure with skip connections that directly transfer spatial information for precise localization and context preservation, and this skip connection is the key aspect of U-Net Architecture. This architecture shows remarkable performance in cell segmentation and other biomedical imaging tasks. The flexibility of this architecture across different modalities have made this one of the most popular models in computer vision for healthcare [15].

One of the most influential contributions to the field of image-to-image translation is the inclusion of Conditional Adversarial Networks. It presented the Pix2Pix framework, which remains to serve as essential for paired image translation tasks, including medical imaging applications like PET-to-CT synthesis.

The impetus for Pix2Pix arose from the recognition that numerous vision and graphics challenges—such as semantic segmentation, image colorization, sketch-to-photo transformation, and aerial-to-map conversion can be formulated as pixel-to-pixel mappings. In the past, each of these tasks needed custom algorithms and loss functions. Pix2Pix wanted to make a general-purpose framework with conditional Generative Adversarial Networks (cGANs).

In the Pix2Pix method, the generator learns how to turn an input image into an output image. cGANs are different from regular GANs because they use an input image to control the output. This makes them very good for cross-modality translation. The generator architecture was built as a U-Net. The U-Net uses encoder-decoder layers with skip connections. The skip connections let low-level spatial information go around the bottleneck. By this it makes sure that structural details from the input are kept in the output. This is important for medical imaging tasks where structural fidelity is very important.

Pix2Pix used a PatchGAN as its discriminator. It only looks at small parts of the image to see if they are real or fake, not the whole thing. This method helps the generator to make realistic local textures and small details.This makes the images less blurry than they usually are when using traditional pixel-wise losses. The U-Net generator and PatchGAN discriminator worked together to create a new type of architecture that balanced global coherence with local detail preservation.

It added a very important thing in the design of Pix2Pix's loss function. The authors added an L1 loss between the generated and ground-truth images to the adversarial loss, which makes sure the generated image looks real. The L1 term helped with pixel-level accuracy, and the adversarial term helped with sharpness and naturalness. Later more advanced structural losses were added in different work, like MS-SSIM or perceptual losses, especially for medical imaging tasks.

Isola et al. showed that Pix2Pix can be used in many different fields. It can be used in semantic segmentation of photos, aerial maps to satellite images, grayscale to colorization, and edge-to-photo synthesis. Quantitative evaluation integrated perceptual user studies with objective metrics, including pixel accuracy and segmentation-based measures. The findings indicated that Pix2Pix could produce outputs that were both realistic and structurally consistent with the input. It surpassed the conventional regression-based techniques that minimized Euclidean distance and yielded excessively smooth or indistinct results.

The Pix2Pix paper had a significant effect on research in medical imaging. Because it had a paired image translation framework. It could be used directly for tasks like translating MRI to CT, synthesizing PET to MRI, and denoising CT. Pix2Pix is a great tool for translating

PET to CT as it can perfectly align PET and CT slices into paired datasets. The U-Net architecture makes sure that anatomical structures from PET inputs are kept in the output.Then the PatchGAN discriminator makes the synthesized CT images look realistic.

The Pix2Pix tool has some downsides too. As it relied on paired datasets, it couldn't be used in situations where perfectly aligned modalities weren't available. But paired data is very rare in medical imaging. Adversarial training also made things sharper, but it sometimes added artifacts that could be a problem in clinical settings. So for solving these problems for unpaired translation a follow-up work like CycleGAN was added for more structural or perceptual constraints.

In short, the Pix2Pix framework was the first tool of modern image-to-image translation. The combination of cGANs, a U-Net generator, a PatchGAN discriminator, and hybrid L1 + adversarial losses made it a strong, generalizable solution. It has been widely used in medical imaging research since then. Pix2Pix serves as both a baseline and a benchmark for PET-to-CT synthesis. This work shows that deep generative models can be used for cross-modality medical image translation. But it also shows that more work needs to be done to make them more useful in the clinic for getting clinical accuracy [6], [7].

Cycle-Consistent Adversarial Networks (CycleGAN), another landmark in image-to-image translation. It addressed a key limitation of earlier models that is the need for paired training data. Paired data is very rare in the case of medical images because of their scanning protocols and patient mobility. CycleGAN solved this problem by allowing translation between two domains with only unpaired datasets.

The new idea added in CycleGAN was the cycle consistency loss. It makes sure that if you translate an image from domain X to Y and then back again, you should get the same image back. This stops the networks from learning random mappings. It also makes sure that important structural content stays the same. The framework has two generators and two discriminators. Each discriminator tells the difference between real images and synthetic outputs in its own area. To make it look more realistic, the model uses PatchGAN discriminators that look at small parts of the image instead of the whole thing. This makes the details sharper and the textures look more natural.

CycleGAN did very well on unpaired datasets, like changing seasons, changing objects, and changing artistic styles. It performed better than other unpaired translation techniques as it could keep the structure of the content in it.

In medical imaging techniques where paired dataset is very hard to find, CycleGAN is widely used there. Its unpaired framework makes it especially good for synthesis across different modalities. But it has some problems like structural distortions and sometimes it removes clinically significant details. It makes it less reliable for clinical use.

In short, CycleGAN is a significant advancement for unpaired datasets. As paired dataset are hard to get, it opened some new doors to image techniques for unpaired dataset. This works well with paired frameworks like Pix2Pix as well[16].

Denoising Diffusion Probabilistic Models (DDPMs), which made diffusion processes a strong new type of generative model. DDPMs basically follow the concept of non-equilibrium thermodynamics. They constantly add gaussian noise with the image and then learn to reverse this process using neural networks. Diffusion models use a Markov chain of denoising steps that are set up through variational inference. This makes training stable and based on likelihood which was a problem in the adversarial models like GANs.

DDPM basically uses denoising score matching and Langevin dynamics.The authors demonstrated that by using noise prediction in the reverse diffusion step the training gets easier. So This means that the training can be done by finding the weighted mean squared error between the true noise and the predicted noise. This makes the diffusion model optimized. Also, the iterative denoising process can be thought of as progressive image refinement. Noisy rough structure is the first output of this model then it starts to get finer later.

DDPMs got the best results on benchmark datasets like CIFAR-10 and LSUN. With this model the samples look realistic and the Fréchet Inception Distance (FID) scores were as good as GAN-based methods. The generation process avoided mode collapse, which is a common problem with GANs, and the outputs were very different from each other. However, with this method the problem is sampling is still expensive for computing power as more iterations are needed.

The benefits of DDPMs in medical imaging is that their refinement works well to keep the structural details that are very important for medical imaging. The model's probabilistic nature gives many possible reconstructions combinations, which are a common scenario in medical imaging. However, the high computational cost and long sampling times make it less useful in clinical settings, especially when compared to faster GAN-based methods.

In summary, the DDPM framework showed that diffusion models are a strong alternative to GANs by combining theoretical and empirical strength. There are still problems with efficiency, but they are an important area of research for medical image translation because they can make high-fidelity, structurally consistent images [9].

Another key improvement to the original Denoising Diffusion Probabilistic Model is done by introducing modified noise schedule, improved sampling procedure and learned variance to upgrade the sample quality and efficiency. These together are used to enable faster convergence and higher fidelity image generation. It has been shown that adjusting the training objective and parameterization reduces the gap between training and sampling objectives by interpreting diffusion models through a variational perspective [10].

There is also a fundamental trade-off in generative modeling among sample quality, mode coverage, and training stability, known as the *generative learning trilemma*. Named as Denoising Diffusion GANs (DDGANs), a hybrid model that combines the strong likelihood-based foundation of diffusion models along with the fast sampling and adversarial training benefits of Generative adversarial network. By introducing a loss function of GAN into the diffusion framework and training the generator to produce denoised samples at multiple noise levels, DDGANs achieve two important things. One is high-fidelity and the other one is diverse image generation with significantly fewer sampling steps. Experiment shows that DDGANs perform better than the traditional diffusion models in speed while maintaining or improving image quality and diversity, and thus how it effectively bridges the gap between diffusion models and GANs [17].

Another approach with the diffusion model is introducing Palette, a unified diffusion-based framework for various types of image-to-image translation tasks such as colorization, inpainting, uncropping, and style transfer. Unlike task-specific traditional

Denoising Diffusion Probabilistic Model, Palette uses a conditional denoising diffusion process that learns to iteratively refine a noisy image conditioned on the input image, enabling high-quality and consistent output generation. The model architecture usually consists of a U-Net architecture with attention mechanisms and it achieves strong generalization across tasks without requiring task-specific tuning. Experiment shows that Palette usually produces more photorealistic, diverse, and semantically consistent results than previous GAN- and regression-based methods,and thus shows the versatility and robustness of diffusion models for conditional image generation.

As the Diffusion process is computationally very expensive, to accelerate the image translation procedure keeping the quality almost same Fast-DDPM: Fast Denoising Diffusion Probabilistic Models for Medical Image-to-Image Generation was introduced which uses some technique like step distillation, adaptive noise scheduling and lightweight network architecture to deduct the number the of diffusion steps keeping the image fidelity same. Using the Fast-DDPM technique it has been more practically applicable [18].

Score-based generative modeling is a big step forward in probabilistic machine learning because it brings together different ways of combining complex data distributions.A stochastic differential equation (SDE) framework that generalizes earlier diffusion models and score-matching methods. Traditional methods like Denoising Diffusion Probabilistic Models (DDPM) and Score Matching with Langevin Dynamics (SMLD) change the distributions of data by adding noise slowly and then taking it away. The SDE perspective generalizes these discrete methods into a continuous-time framework, characterizing data corruption and denoising as forward and reverse stochastic processes regulated by time-varying drift and diffusion coefficients. The main aspect is the creation of a reverse-time SDE that lets data be made by getting rid of noise, based on the score function, which is the gradient of the log-probability density. To get sample generation, numerical SDE solvers are used to train a neural network to estimate this score across different levels of noise. This also presents a predictor-corrector (PC) sampling scheme that combines deterministic solvers with stochastic corrections to make both the stability and the quality of the samples better. They also connect the framework to neural ordinary differential equations (ODEs), which make it possible to calculate the exact likelihood and sample efficiently and deterministically through the

probability flow ODE. The empirical results show that state-of-the-art image generation is possible on datasets like CIFAR-10. For unconditional models, the Fréchet Inception Distance (FID = 2.20) and Inception Score (IS = 9.89) are the best ever. The framework is flexible because it can do more than just image synthesis. It can also do conditional tasks like inpainting and colorization. The SDE-based formulation brings together diffusion and score-matching paradigms, which makes it more coherent from a theoretical standpoint, easier to control, and better in practice for generative modeling [19].

For calculating the loss between images lots of different types of loss functions are used, as the performance of model architecture highly depends on how well we can find the difference between the original image and the generated image. While one of the most common types of loss function Mean Squared Error (MSE) fails to align with human visual perception, an alternative multi level structural similarity is introduced that weights image features according to perceptual importance across scales. Rather than just focusing on the pixel level difference this approach focuses on structural and contrast information and that leads to sharper and more perceptually faithful image synthesis. As it preserves edges, textures and structural integrity it is considered an effective approach in the image generation [20].

One of the very fundamental loss functions used in Deep learning techniques is Structural Similarity Index Measure (SSIM), this one is a groundbreaking method for assessing image quality that moves beyond traditional approaches, It is mainly based on the premise that the human visual system is highly sensitive to structural information rather than absolute pixel differences, The core functions of SSIM losses are three. Luminance, contrast and structural similarity. By these three components it compares local patterns of pixel intensities. SSIM correlates much more closely with human visual perception than other existing metrics, and it provides a more reliable and interpretable measure for image compression, transmission and restoration of quality evaluation [21].

An extended version of Structural Similarity Index Measure (SSIM) is the Multiscale Structural Similarity Index Measure (MS-SSIM). This version evaluates the image quality across multiple spatial resolutions, and it analyzes structural similarity at progressively reduced image scales, incorporating all 3 components - luminance, contrast and structure at

each level to capture both global and local image distortions. Viewing distance and image texture complexity has been possible for this multiscale approach [22].

If we are about to compare two popular evaluating functions SSIM and PSNR, PSNR mostly measures pixel wise intensity differences and that's why it often fails to align with visual perception. While SSIM evaluates image quality based on structural information. Though SSIM is a more reliable and perceptually meaningful measure of image quality than PSNR , using these two while evaluating the performance of the model might be helpful to have more information about the model performance [23].

Another loss function that replaces the traditional pixel wise losses is VGG Perceptual loss. This loss measures difference in image content and style based on deep feature activations, enabling the model to produce visually appealing. Unlike pixel based metrics these perceptual losses capture semantic and structural similarity in a more efficient way which makes a major advancement in image synthesis, restoration and generative modelling [24].

2.2 Dataset Description

These experiments used a public dataset from The Cancer Imaging Archive on Breast Cancer Patients named QIN-Breast consisting of numbers of PET images with their corresponding CT images [25].

Table 2.1: Property of QIN_Breast Dataset

Property	Description
Types	PET(Functional) + CT (Anatomical) + MR
Number of Images	102451
Subject	68
Studies	216
Format	DICOM(.dcm) files

Image Type	16-bit Grayscale
Purpose	Image Translation (From PET to CT)

2.3 Methodology Overview:

The workflow we have followed for this task

Dataset Preprocessing -> Model Training (GAN, Diffusion, Hybrid) -> Image Generation -> Evaluation

Three types of model approach were taken based on the current state of art and evaluated to compare which one is performing better in the given conditions .

2.3.1 Generative Adversarial Network (GAN) Model:

Generative Adversarial Network has been proven architecture in existing deep learning techniques for image translation where there are two separate methods driven by Generator and Discriminator works simultaneously to generate an image [4].

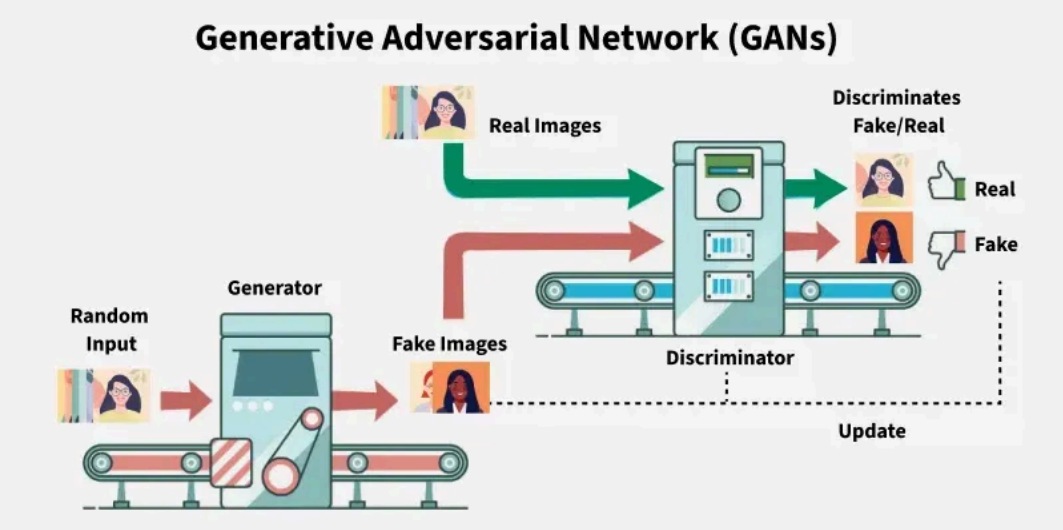


Figure 2.1 : Generative Adversarial Network [Source : [26]]

2.3.1.1. Generator:

This one is used to generate the image and gradually tries to update the quality of the image with the feedback from the discriminator. One of the special types of Generator is

U-Net Generator specifically for medical purposes, the key aspect of this type of generator is encoder and decoder with skip connection to retain spatial information [4], [5], [15].

2.3.1.2. Discriminator:

This one is used to detect whether the generated image from the generator is real or fake, but the key point here is, how accurately the discriminator can detect the fakeness of the generator helps to improve the performance of the generator. One of the improved version of discriminator is PatchGAN Discriminator, the speciality is instead of detecting the whole image as fake or real, it divides the generated image from generator into some patches, for example (70*70) or (30*30), and then it tries to predict whether that patch looks real or fake, and sends the feedback, in this way, Generator gets more information about which part of the image to improve [4], [6], [7].

2.3.1.3. Loss Functions:

Loss functions are very critical here, as deep learning techniques are used here, and there will be backpropagation to reduce the loss by updating the weights. In the case of medical images, it is not about the raw pixel difference, rather focusing on the difference of medical information and try to reconstruct that is more important, various types of loss functions are there, Adversarial loss, L1 loss, VGG Perceptual loss, Multi scale SSIM loss, and each of the losses has some special ability to extract some crucial information, and combining them to have a weighted loss function , where the weights of are taken by tuning can be a big step [4], [22], [24].

2.3.2 Diffusion Model :

A very promising model in the deep learning field where an image is turned into a noise first by adding Gaussian noise from Gaussian Distribution and then from that noise reconstructing the image [9]. It has two separate processes.

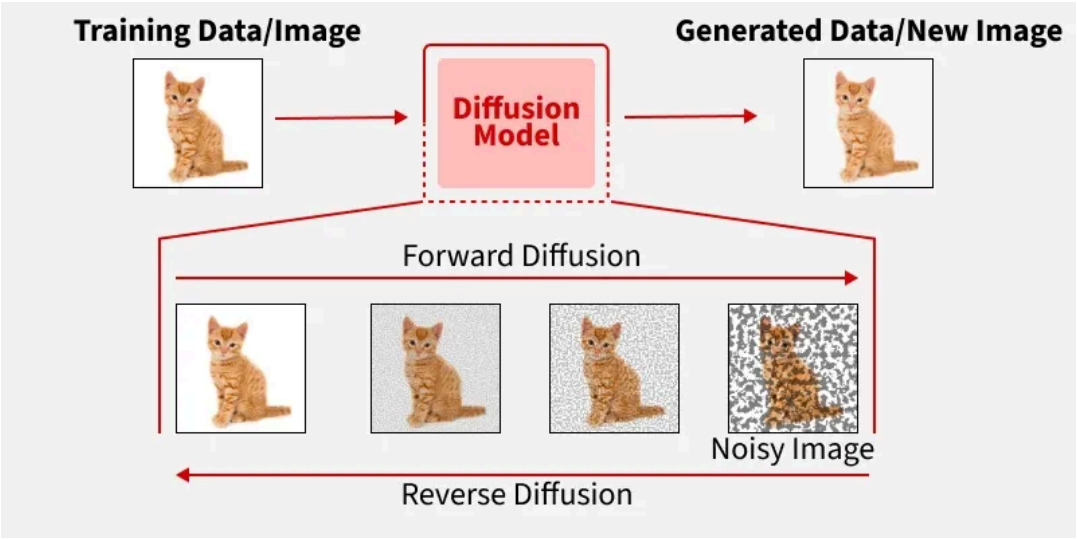


Figure 2.2 : Diffusion Model [source : [27]]

2.3.2.1. Forward Process:

In that process, the raw pixel values of the image are added with random values from the gaussian distribution as noise and in this process the image gets completely distorted and the distorted images are used for the next process [9].

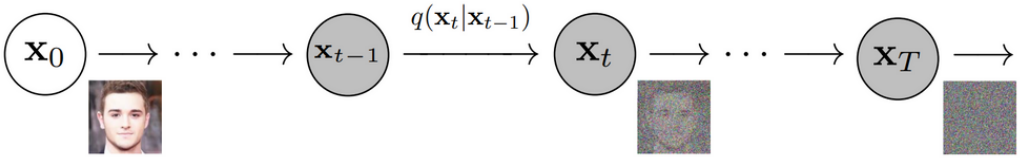


Figure 2.3 : Forward Process of Diffusion [source: [27]]

2.3.2.2. Reverse Sampling Process:

Now from that noise taken from the forward process goes through a process of noise reduction and reconstructing the original image, and that is how the model learns how to generate an image from a noise or random values [9].

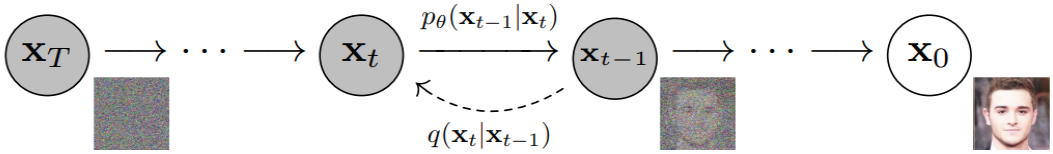


Figure 2.4 : Reverse Process of Diffusion [source : [27]]

2.3.3 Hybrid Diffusion-GAN Architecture:

To combine all the advantages of GAN and Diffusion to complement the lacking of each model, we try to make a hybrid model.

1. Conditional U-Net : It acts as the Generator
2. PatchGan Discriminator : It works as the discriminator to give feedback for better guidance about which part to improve in generating images .
3. Loss Functions : Combined Weighted Loss functions is used which might be tuned to get the best result

2.3.4 Training Configuration:

Kaggle Notebook is used for the running these models where the GPU support of Kaggle notebook (GPU T4*2, GPU P100, TPU v5e-8) [28].

Computational Resources are very important in running deep neural networks. In the process of training the model, all the weights of all the neurons get updated by backpropagation to reduce the loss functions and the number of neurons are huge as an image holds lots of pixel values. And the robustness of the model depends on how much data is feeded through the model.

2.3.5 Evaluation Metric:

To evaluate the performance of the model, alongside with visualization two are the most important metric are :

2.3.5.1. Peak Signal to Noise Ratio (PSNR):

This reflects the pixel wise accuracy.

$$MSE = \frac{\sum_{m,n}\left[I_1(m,n) - I_2(m,n)\right]^2}{M \cdot N} \tag{2.1}$$

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{\text{MSE}}\right) \tag{2.2}$$

Where ,

$I_1(m, n)$: Pixel intensity at position (m, n) in the first image.

$I_2(m, n)$: Pixel intensity at the same position (m, n) in the second image (e.g., the compressed or distorted image).

m, n: The row (m) and column (n) indices specifying a pixel's location in the image.

M: The total number of rows (height) of the image.

N: The total number of columns (width) of the image.

MSE: Mean Squared Error. The average of the squares of the differences between corresponding pixels in two images. A lower MSE means less error.

MAX_I: The maximum possible pixel value in the image. For an 8-bit image, this is 255 .

These represent the ideal formula of PSNR [23].

2.3.5.2. Structural Similarity Index Measurement (SSIM):

This shows the structural and perceptual similarity.

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{2.3}$$

SSIM(x, y): The Structural SIMilarity index between two local image windows, x and y .

x, y: Two image patches (small windows) of the same size, taken from the same location in the two images being compared.

μ_x: The mean (average) pixel value of image patch x. This represents the luminance (brightness) of the patch.

μ_y: The mean (average) pixel value of image patch y.

σ_x: The standard deviation of the pixel values in patch x. This represents the contrast of the patch.

σ_y: The standard deviation of the pixel values in patch y.

σ_{xy}: The covariance between patches x and y. This measures how much the pixel values in the two patches change together and represents the structural correlation.

C1, C2: Small constants added to stabilize the division and prevent it from being undefined when the denominators are very close to zero.

These represent the procedure of obtaining SSIM [21], [23].

CHAPTER 3

IMPLEMENTATION AND EXPERIMENTS

The chosen QIN breast Dataset to build has been gone through 3 types of implementation by 3 different types of deep learning approaches.

1. Conditional GAN
2. Conditional Diffusion
3. 2.5D Hybrid Diffusion + Weak Edge Gan

3.1 Dataset Preprocessing

1. **Z-matching per patient:** To make the model robust and learn from the best pattern between PET and CT images, a selective approach was taken for the final dataset making which will go through the model. Though the QIN-Breast consist of paired dataset of PET and CT, not all the images are well aligned, For Generative Adversarial Network and even for diffusion and hybrid model, dataset should be well aligned meaning, the pet and ct image should be corresponding to each other, here z is termed as the thickness of the PET and CT slice, A threshold value for each type of model has been decided for the preparation of dataset, Only the slice thickness below that threshold value are taken as final dataset for the model training.

2. **CSV Files:** Diffusion Model often prefers CSV files consisting of the data that matched the condition of z value requirements. Hybrid model also prefers CSV file to work with

3.2 Image Transforming,Resizing and Augmentation:

3.2.1 Transforming:

1. CT Images : For Diffusion and Hybrid Model, a new type of Unit is used named as HU unit which is universal for CT images [29]. For GAN Model, HU was not needed, slice wise min-max [0,1] was sufficient.

2. PET Images: SUV is used to calculate PET images for maintaining the global standard [30]. Though only for diffusion it was needed, in GAN architecture it wasn't mandatory.

3.2.2 Resizing:

About resizing the image, the image was 512*512. But due to lack of computational resources, it was reduced to 256*256.

3.2.3 Augmentation

Data Augmentation was not followed here, as the dataset was already enough for the given computational resources. In a theoretical sense, data augmentation is a very good approach to feed more data in the training process of a model. For Alignment CT and PET images should be well aligned in axes also.

3.2.4 Conditioning Variants

Sobel Edging is used in diffusion and in hybrid 2.5D slices was taken as PET input to give more details to the CT image about the surrounding of that particular PET image.

3.3 Model Architecture

3.3.1 Conditional GAN

1. **Generator** : U-Net was used as Generator, Eight down block and Eight Up block was used , through convolutional neural network the pixel values are flattened into one set of input for the neural network, for the activation function LeakyRelu was used in down blocks and Relu was used in up blocks, and in final output section Tanh is used, skip connections was there to retain spatial details [4], [7], [15].

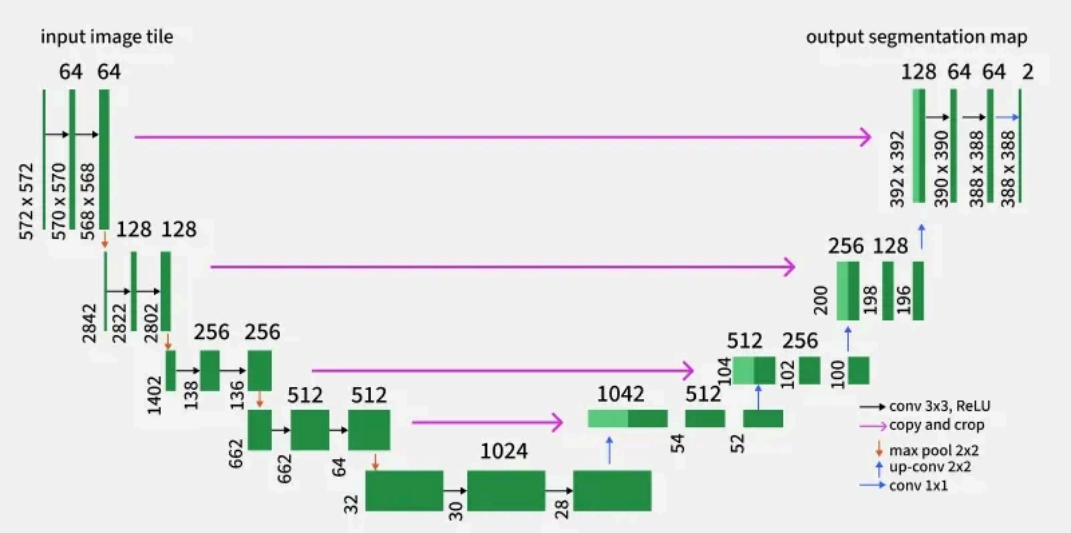


Figure 3.1 : Sample U-Net Architecture [Source : [31]]

2. **Discriminator** : For the Discriminator Part, PatchGAN Discriminator was used to give more precise feedback, Conv Stacks (64->128->256->512 stride pattern (2,2,2,1). Then Conv ->1 logit map, no sigmoid function was used here [4], [7].

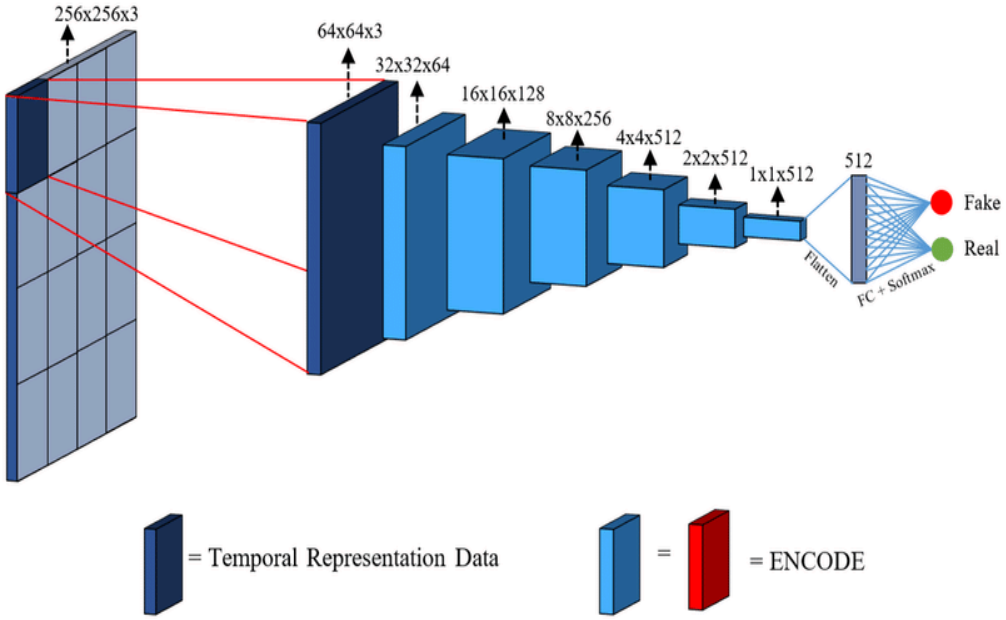


Figure 3.2 : PatchGan Discriminator [source : [32]]

3.3.2 Loss Functions:

1. **Discriminator Loss** : For discriminator only Binary Cross entropy loss was used, Binary Cross Entropy is calculated as the difference between the actual class/label and the predicted probabilities, the lower the binary cross entropy loss is the better the model is performing [4].

$$L_D = -(1/2)\,[\log(D(x,y)) + \log(1 - D(x, G(x)))] \tag{3.1}$$

2. **Generator Loss**: For the generator a combined weighted loss function is used consisting of Adversarial loss, L1 loss, MS SSIM loss and VGG Perceptual Loss.
   - **Adversarial Loss:** It defines the generator based on how well it is performing on fooling the discriminator from distinguishing real and fake images [4]. The formula of adversarial loss is as follows

$$L_G = -\frac{1}{N}\sum_{i=1}^{N} \log(D(x_i, G(x_i))) \tag{3.2}$$

   - **L1 Loss:** Calculated the difference between raw pixel values of generated image and target image [7]. The formula for L1 loss is as follows

$$L_{L1} = \frac{1}{N}\sum_{i=1}^{N} \left\| y_i - G(x_i) \right\|_1 \tag{3.3}$$

   - **SSIM:** To retain the structural similarity, to emphasize the structural difference between two images, this loss is calculated [21]. The formula for SSIM loss is as follows:

$$L_{SSIM} = 1 - \frac{(2\mu_{G(x)}\mu_y + C_1)(2\sigma_{G(x)y} + C_2)}{(\mu^2_{G(x)} + \mu^2_y + C_1)(\sigma^2_{G(x)} + \sigma^2_y + C_2)} \tag{3.4}$$

   - **VGG Perceptual Loss:** Trained on Image Net to extract features from images [24]. The formula of VGG Perceptual Loss is as follows

$$L_{perceptual} = \frac{1}{N}\sum_{i=1}^{N} \left\| \phi_j(G(x_i)) - \phi_j(y_i) \right\|_2^2 \tag{3.5}$$

The final loss function was used as the weighted sum of these 4 loss functions, and the weight of each loss function is a tunable parameter.

Total_Generator_Loss = Adversarial_Loss + 100*L1_loss + 10*SSIM_Loss + 1.0*VGG_Perceptual_Loss

3.3.3 Optimization

For optimization the parameter was taken through experiments by tuning which combination works better. The parameter works good for us

Generator - Adam (G, 1e-4, Beta = (0.5,0.999)),

Discriminator - Adam (D, 4e-4, Beta = (0.5,0.99)),

StepLR(step=30,gamma=0.1) for both Generator and Discriminator

AMP with separate GradScaler for G and D

3.4 Conditional Diffusion

3.4.1 U-Net as Generator and for conditioning:

- In channels : 3 [x(t) + PET + Sobel (PET)]
- Blocks : ResBlk (GroupNorm(8), SiLU as activation function, with explicit skip channel sizes Down and Up blocks, and to predict the final noise which is supposed to be added with image comes from Conv2d(base-1).

3.4.2 Schedule and Objectives

Cosine Beta schedule with T =400,For Faster convergence and sharper supervision under short epoch constraints auxiliary x0 - L1 is used.

Forward Diffusion Process :

$$x_t = \sqrt{(\alpha^-_t)}\, x_0 + \sqrt{(1-\alpha^-_t)}\, \varepsilon \qquad (3.6)$$

Here,

x_0 = tanh(CT)

And at each timestep t, the noised image is x_t

x_0 is the normalized ground truth CT slice in tanh space

$\bar{\alpha}$ controls how much noise is added in each step.

3.4.3 Training Objective (Loss Functions)

While reconstructing the image,the model basically tries to predict the added noise used to distort the image, and to predict that it followed the ε-MSE term which is combined with L1 and reconstruction loss.

$$L = ||\varepsilon\hat{} - \varepsilon||_2^2 + \lambda_L 1 \cdot || \tag{3.7}$$

Where,

λ_L1 = 15

$$x\hat{}_0 = (x_t - \sqrt{(1 - \alpha^-_t)} \cdot \varepsilon\hat{}) / \sqrt{(\alpha^-_t)}$$

ε-MSE: It actually measures how accurately the network predicts the true Gaussian noise ε that was added to x_0. It is minimized means the training is stable and denoising fidelity is improving.

3.4.4 Training parameters, AMP, clipping, EMA

For these parameter the values were taken as follows

Adam (lr=1e-4,Beta = (0.9,0.999)),

AMP on,

grad clip = 1,

optional step cam per epoch (1600)

EMA (0.999) maintained every step,

validation runs on EMA weights.

3.4.5 Inference

DDIM = 150 steps.

ETA = 0

PET and edges are precomputed and concatenated with input x(t)

3.4.6 Masked Metrics

Body mask from CT and the threshold value was set 0.03 with 3*3 grow. Masked PSNR/SSIM was computed inside the body and if mask parse it fallbacks to global.

3.5 Hybrid 2.5D Diffusion + Weak Edge-GAN

To cover up the edge-constructing lack of diffusion model, a Hybrid approach was taken where along with a v-prediction diffusion U-Net a spectral-normalized edge PatchGAN was used. Also special data preparation and losses functions were also taken for the smooth training process of the model.

3.5.1 Data Processing for Hybrid

2.5D technique was used in data preparation, meaning there are 3 PET slices for one CT slice, to give more details just not about the functionality, rather about the whole area of that particular PET image, the previous and the next images are also given. For the XY pre-alignment phase correlation technique was used. To focus supervision, 256*256 patches around the body mask are sampled.

PET triplet = [p_{−1}, p_0, p_{+1}]

The final conditioning tensor used during diffusion sampling is as follows

cond = [PET(3)_tanh , Sobel(PET(mid))_tanh]

Here, PET(3)_tanh means that 3 slice PET images that are normalized to tanh space (-1.1)
And sobel(PET(mid))_tanh means the gradient(edge) map of the central pet slice; these things are used to preserve intensity and boundary cues by conditioning the net.

3.5.2 Generator for Hybrid

For the hybrid model, a conditional U-Net was employed as the generator architecture, inspired by the design principles of Latent Diffusion Models [33].Where the parameter was

In channel = 5

Base =64

T_dim = 128

And MinSNR weighting was used in v-loss .

$$v_{tgt} = \sqrt{\overline{\alpha}_t} \cdot \varepsilon - \sqrt{1 - \overline{\alpha}_t} \cdot x_0 \tag{3.8}$$

Here,

v_tgt is the target velocity term

The signal to noise ratio at each timestamp t is calculated as

$$\mathrm{SNR}_t = \frac{\overline{\alpha}_t}{1 - \overline{\alpha}_t} \tag{3.9}$$

The minSNR weighting function is used to further stabilize gradient flow and prevent domination by nearly noise-free (high-SNR) steps [34], [35]. It is stated as

$$w(t) = \min(\gamma, \mathrm{SNR\_t}) / \mathrm{SNR\_t} \tag{3.10}$$

Here γ=5

And the diffusion loss term is calculated as

$$L\_diff = \| w(t) \cdot (\hat{v} - v\_tgt) \|\_2\text{^}2 \tag{3.11}$$

Here,

$\hat{v}$ is the velocity that the model predicts for each step [35].

v_tgt is the ground truth target derived from the input x0 and noise ε

w(t) is used to balance learning across noise levels by down weighting every high SNR step.

This approach of MinSNR is taken to ensure stable gradient flow and to prevent domination by near-clean steps.

3.5.3 Discriminator for Hybrid :

As Discriminator EdgePatchGAN Discriminator was used [7], [36].The parameter was

in channel = 2

Base =64

Inputs = edge(PET mid) and edge (CT),

In all convolution spectral norms were used, and Bias was taken as false for AMP Stability and we get patch logits and intermediate features for feature matching as output of the discriminator.

3.5.4 Loss Functions

Different types of loss functions were used.

1. Diffusion V-loss (MinSNR)

$$L_{\text{diff}} = \left\| w(t) \cdot (\hat{v} - v_{\text{tgt}}) \right\|_2^2 \tag{3.12}$$

This loss calculates the squared error between predicted and target velocity and that is weighted by the MinSNR to maintain the balance across noise levels [34].

2. Reconstruction on input

- Charbonnier loss is a smooth approximation of L1, robust to outliers [37]. The value for weight was taken = 120

$$L_{\text{charb}} = \sqrt{(\hat{x}_0^1 - CT_{01})^2 + \varepsilon^2} \tag{3.13}$$

This loss is used to encourage finer CT intensity recovery and it is robust to outliers .

- L1 Loss. The weight was taken as 40

$$L_{L1} = \left\| \hat{x}_0^1 - CT_{01} \right\|_1 \tag{3.14}$$

Calculate the difference between pixels.

- SSIM loss . The weight for it is used as 0.3

$$L_{SSIM} = 1 - SSIM(\hat{x}_0^1, CT_{01}) \tag{3.15}$$

Calculate the structural difference and helps the model to focus on structural similarity

- Edge-based loss between Sobel-filtered or gradient maps is standard for enforcing anatomical boundary preservation in medical image translation . And the weight was taken as 0.2

$$L_{edge} = \left\| Sobel(\hat{x}_0^1) - Sobel(CT_{01}) \right\|_1 \tag{3.16}$$

This loss is used to preserve anatomical edges and boundaries during reconstruction

3. Adversarial and Feature Map on Edges

- Adversarial (hinge loss on edges)

$$L_{adv} = -\mathbb{E}\left[D_{edge}(edge(\hat{x}_0^1))\right] \tag{3.17}$$

It is used to encourage the generator to produce realistic local edge patterns [38].

- Feature-Matching Loss encourages the generator to match intermediate discriminator activations, improving stability and realism [39].Weight = 0.2

$$L_{fm} = \sum \left\| f_{real} - f_{fake} \right\|_1 \tag{3.18}$$

- Adversarial Warm-up and Scaling gradually increases adversarial loss weight (warm-up) . It is stated as

$$\lambda_{adv}(t) = \min(1, \frac{step}{adv_{warmup}})\ \lambda_{adv}^{max} \tag{3.19}$$

Where,

adv_warmup_steps = 1200

λ_adv_max = 0.02

Total Generator Loss is the weighted loss of all these losses.

Total generator Loss =

$$L_G = L_v + 120L_{charb} + 40L_{L1} + 0.3L_{SSIM} + 0.2L_{edge} + \lambda_{adv}(L_{adv} + 0.2L_{fm}) \tag{3.20}$$

Total Discriminator Loss =

$$L_D = \mathbb{E}[\max(0, 1 - D(x_{real}))) + \max(0, 1 + D(x_{fake}))] + \frac{\gamma}{2}\left\|\nabla_{x_{real}} D\right\|_2^2 \tag{3.21}$$

3.5 5 Optimizers

Adam Optimizer was used for both the Generator and Discriminator

Generator:

Learning Rate = 1e-4

Beta = (0.5,0.999)

Discriminator :

Learning Rate = 3e-4

Beta = (0.5,0.999)

Schedulers:

For both Generator and Discriminator cosine after a short warmup was used

AMP :

To avoid Mixed precision edge cases Discriminator used SN Convolutions with Bias = False

EMA :

For the Generator 0.999 value was used.

3.5.6 Sampling

Run DDIM with 50 steps

CFG on v-pred : unconditional = zeros for conditional and conditional = PET(3)+edge(mid)

3.5.7 Evaluation with affine tone calibration

Perform affine tone calibration per slice alongside computing global and masked PSNR and SSIM.

3.5.8 Training and Validation Protocol:

Generative Adversarial Network (GAN) : Training the model, evaluating it once and saving PET image, Generated CT image and Ground truth CT image, printing the global PSNR and SSIM value.

Diffusion Model : Validating every epoch on EMA, showing the masked PSNR and SSIM value, saving the grids and finding the best SSIM checkpoint.

Hybrid Model: Evaluation reports of PSNR and SSIM on per epoch EMA, and showing visual grids including (Generated CT-Ground Truth CT) error map.

CHAPTER 4

RESULT AND DISCUSSIONS

4.1 Result of Generative Adversarial Network

Quantitative Result:

PSNR : 32.550 dB

SSIM : 0.8799

Qualitative Result:

| Pet | Generated CT | Ground Truth |

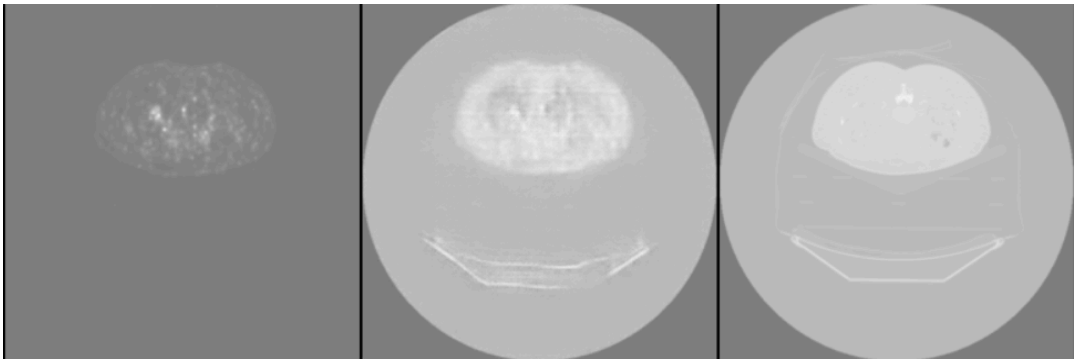


Figure 4.1: Samples of output Generative Adversarial Network (Sample 01)

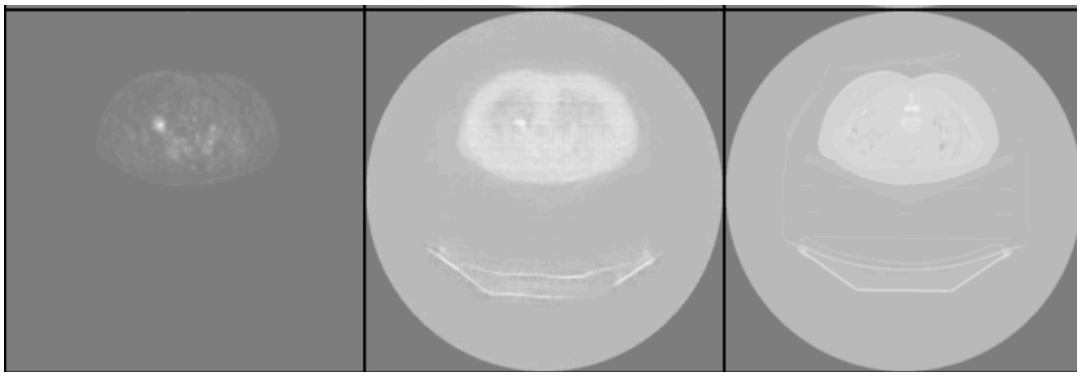


Figure 4.2: Samples of output Generative Adversarial Network (Sample 02)

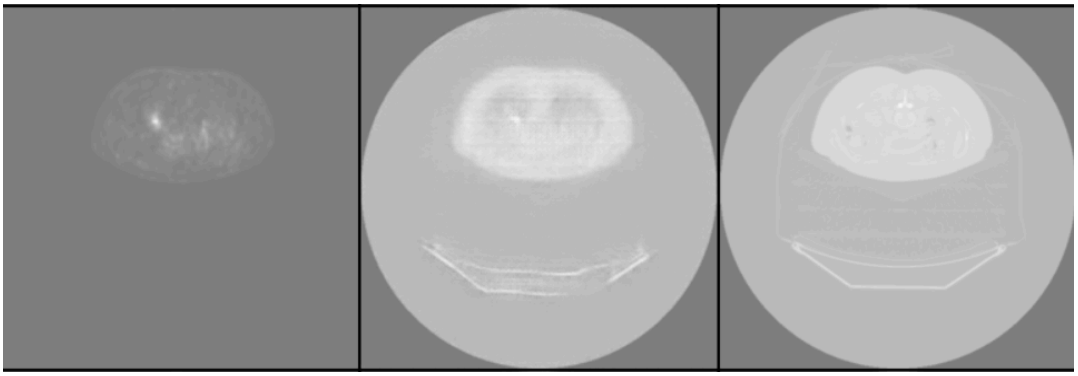


Figure 4.3: Samples of output Generative Adversarial Network (Sample 03)

4.2 Result of Diffusion Model

Quantitative Result:

PSNR : 15.46 dB

SSIM : 0.776

Qualitative Result :

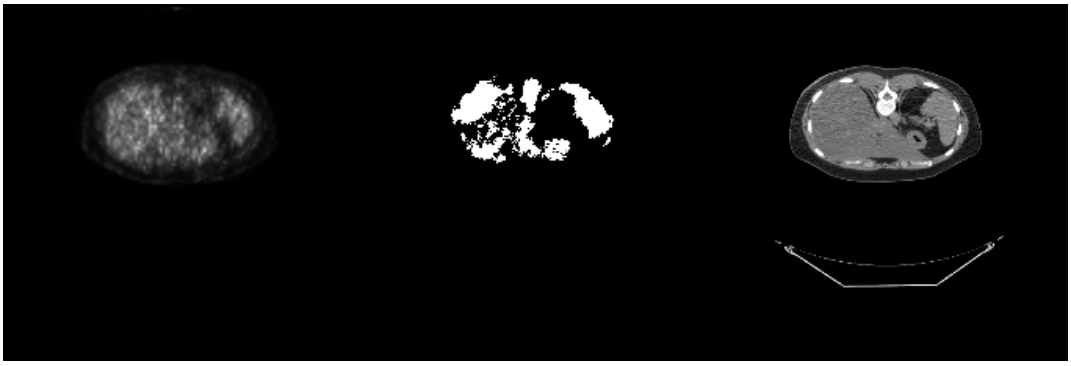


Figure 4.4 : Samples of output of Diffusion Model (Sample 01)

Figure 4.5: Samples of output of Diffusion Model (Sample 02)

Figure 4.6: Samples of output of Diffusion Model(Sample 03)

4.3 Hybrid Model (Diffusion + GAN)

Quantitative Result :

PSNR : 21.48 dB

SSIM : 0.833

Qualitative Result :

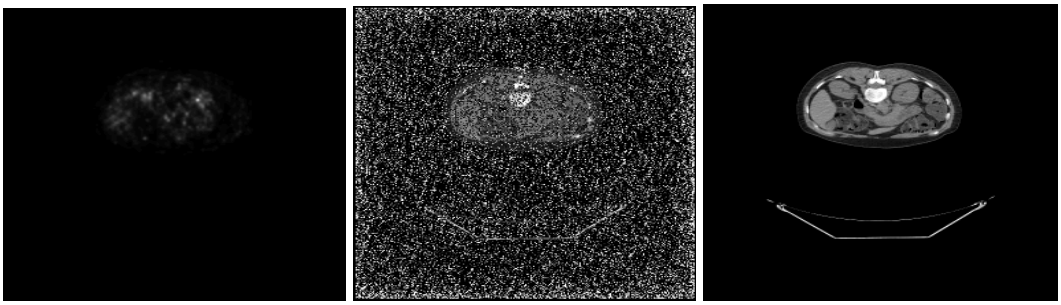


Figure 4.7: Samples of Output of Hybrid Model (Sample 01)

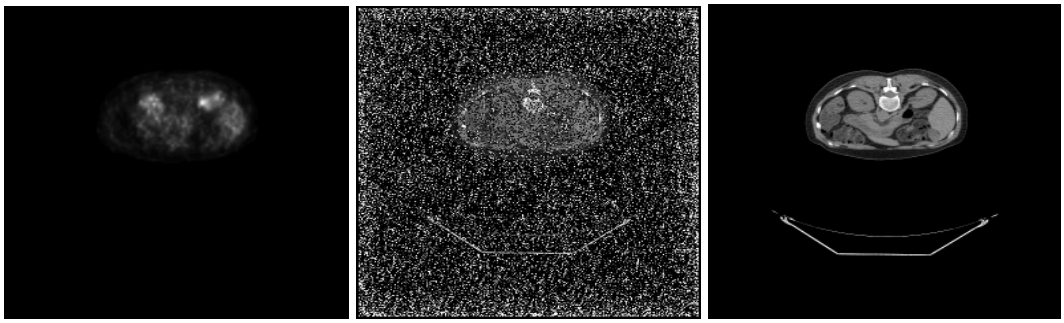


Figure 4.8: Samples of Output of Hybrid Model (Sample 02)

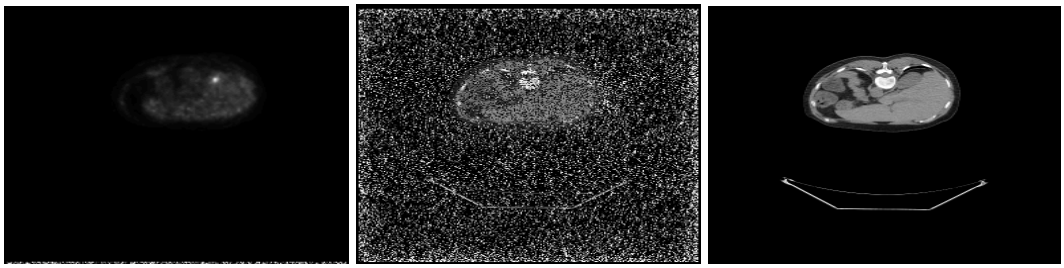


Figure 4.9: Samples of Output of Hybrid Model (Sample 03)

Table 4.1: Comparison of 3 Different Approaches

Model Architecture	PSNR (in dB)	SSIM
Generative Adversarial Network	32.55	0.879
Diffusion Model	15.46	0.776
Hybrid Model	21.48	0.833

4.4 Discussions

4.4.1 GAN Performance

With an average PSNR of 32.55 dB and SSIM of 0.879, the Pix2Pix baseline model produced the best reconstruction quality across all evaluation criteria.

The supervised nature of the training pipeline, which directly guided the generator by several complementary objectives L_1 loss for pixel accuracy, SSIM loss for structural preservation, VGG perceptual loss for high-level texture similarity, and adversarial (GAN) loss for local realism is responsible for this impressive performance.

Instead of promoting global categorization, the PatchGAN discriminator employed in this model promoted patch-wise discrimination.

As a result, it made the generator replicate local contrast patterns and fine-grained information in every receptive field, thus minimizing the over-smoothing issue that frequently arises in conventional convolutional regression networks.

Notwithstanding the overall success, a few minor problems were noted. Around sharp edges, especially in areas of bone or interfaces of high contrast tissue, occasional checkerboard artifacts appeared.

Both bilinear interpolation and kernel-aligned padding can be used in subsequent iterations to reduce these artifacts, which are frequently seen in transposed convolution-based upsampling.

Additionally, there were times when low-contrast tissues were over-sharpening, which somewhat accentuated edge intensity changes.

Even when the matching PET inputs are intrinsically low in resolution, this results from the adversarial pressure giving local contrast priority.After the second epoch, the GAN occasionally showed slight variations in generator loss, suggesting minor mode-collapse tendencies from the standpoint of training stability.

This kind of instability is common in GAN optimization, particularly when learning rates are high or dataset variety is restricted.However, the model consistently generated CT reconstructions that could be interpreted clinically, and the convergence behavior remained controllable.

Key Strengths

- Realistic tissue contrast and density distribution aligned with ground-truth CTs.
- Sharp delineation of anatomical structures (ribs, breast margins, and lungs).
- Best quantitative scores (highest PSNR and SSIM).

Limitations

- Training sensitivity to learning-rate scheduling and batch-size choices.
- Risk of adversarial oscillation if discriminator updates dominate.
- Slight checkerboard artefacts and contrast exaggeration.

In summary, the GAN framework demonstrated that strongly supervised adversarial translation is highly effective for PET→CT synthesis when a sufficient number of paired samples are available. Even under only four epochs, it remained the most faithful and visually convincing model among the compared methods.

4.4.2 Diffusion Model Performance

The conditional diffusion network generated CT images that were structurally consistent and anatomically reasonable; however, the outputs showed noticeably softer edges and lower contrast compared to those produced by the GAN.

This behavior is quite expected, since diffusion models are built around a *noise-prediction* task rather than a direct mapping between intensity domains.

Instead of learning to reproduce a CT image pixel-by-pixel, the model learns to estimate and remove gradually added Gaussian noise, which inherently prioritizes stability and denoising over sharpness.

In this experiment, the diffusion backbone was implemented using a v-prediction formulation with a cosine β-schedule and MinSNR weighting. These design choices ensured that gradients stayed stable throughout the diffusion timeline, preventing exploding or vanishing updates during training. An auxiliary L_1 reconstruction term on x0x_0x0 was also included so that each denoised estimate remained anchored to the true CT domain, further improving convergence and reducing colour drift. Training stability was excellent, no sudden divergence or collapse occurred but only four epochs were completed, far fewer than the typical 100 or more epochs normally required for diffusion convergence.

Consequently, the model learned strong global intensity consistency but failed to capture the finer high-frequency anatomical textures that adversarial frameworks usually recover.

Smoothness versus Sharpness:

There is a trade-off between smoothness and sharpness. Minimizing the expected energy of the added noise in diffusion models promotes spatial smoothness and homogenous textures but as well as it blurs sharp anatomical edges. Though smoothing is beneficial for removing noise artefacts, it washes out subtle tissue boundaries and there is a trade-off.

EMA Averaging Effect:

An exponential moving average (EMA=0.999) was used to stabilize training. This basically improves the consistency of validation metrics, reduces temporal flicker between adjacent slices and produces smoother volumetric continuity. Overall it stabilized the training procedure well.

Masked Evaluation:

PSNR got improved by 0.5dB and SSIM got improved by 0.2 while quantitative evaluation was restricted to the foreground body region basically excluding the background pixels. It actually proves that reconstruction errors mainly arose from mismatched background intensities rather than anatomical inaccuracies, and also indicates that the model's spatial alignment within the body region was accurate.

Interpretation:

Overall, the diffusion network generated *noise-free, globally coherent* CT images with some artefacts.

But in reality it didn't reach the level of GAN in visual sharpness. However, it would require:

1. More training epochs (≥ 20 or more),
2. More number of denoising steps per sample (≥ 200 DDIM iterations), and
3. Integration of explicit edge-preserving or adversarial components.

Though it results in lower PSNR and SSIM, it shows training stability, fewer number of artefacts, and strong theoretical foundation. And that makes it a strong candidate for future work with 3-D volumetric image translation tasks.

4.4.3 Hybrid Model Performance

The Hybrid model is approached to blend the strengths of GAN model and Diffusion model, combining the local edge realism of adversarial learning with reliable intensity reconstruction of diffusion.

Here, the diffusion UNet acted as the primary generator and the lightweight PatchGAN discriminator operated only on *edge maps* rather than to encourage the structural refinement while keeping the training stabilized.

The hybrid approach also ran for only four epochs, same as the diffusion model did, but it achieved higher PSNR and SSIM than diffusion model, which indicates that combining the adversarial feedback helps us to enhance texture detail without harming the underlying

anatomy of the image. And that is the success of Hybrid approach combining strengths of previous two approaches.

Advantages Observed

1. **Sharper Structural Boundaries:** In high contrast regions like ribs, lung wales and soft tissues interfaces, the hybrid approach of edge focused discriminator strengthened gradients. In adversarial loss the common issues happen of hallucinating textures which is avoided here, and that makes it anatomically realistic.
2. **Improved Masked SSIM:** Better preservation of organ morphology and clearer tissue boundaries is noticed after applying the edge based regularization and we got improved masked SSIM by 0.07 over the diffusion baseline.And that is a significant improvement than diffusion model.
3. **Affine Tone Calibration:** Initially it is noticed that there are slight tone inconsistencies between slices while mixing GAN and diffusion gradients. But it was solved by simple affine intensity rescalling during post-processing and it also increased the PSNR by 1 dB. And that indicates that global brightness alignment plays a key role in translation tasks.
4. **2.5-D Conditioning:** By using three consecutive PET slices [p−1,p0,p+1][p_{−1}, p_0, p_{+1}][p−1,p0,p+1] as input channels, the network gets contextual information along the z-axis. This extra spatial context helps to improve the continuity between adjacent CT slices and it also reduces sudden changes in contrast or structure, and that's how it bridges the gap between 2-D and volumetric learning.

Limitations and Practical Observations

- The GAN branch had limited training time and this causes issues to learn deeper adversarial priors.
- Minor tone instabilities are noticed in neighbouring slices if affine calibration was skipped.
- Diffusion sampling still remains computationally heavy because each slice still requires about 50 steps even if we use DDIM acceleration technique.

- Because of the lower discriminator strength outputs are noticed overly smooth occasionally in regions with weaker gradient signals.

Interpretation and Takeaway

The hybrid approach opens the door of integrating multiple deep learning techniques that will combine the strength of each technique and will outperform each of their individual's performance. In our Hybrid framework the adversarial feedback complemented diffusion regularization by reintroducing texture and edge information that were lost during pure denoising.

This demonstrates the potentiality of hybrid architectures for future research particularly for 3-D PET → CT or other multimodal translation tasks where both quantitative accuracy and perceptual realism are essential.

CHAPTER 5

DEMONSTRATION OF OUTCOME BASED EDUCATION (OBE)

5.1 Introduction

While doing the whole work, we have followed all the Outcome-Based Education principles, incorporating Course outcomes, Programme outcomes and knowledge profiles. And We have also taken all the ethical consideration while accessing dataset and sustainability was also checked while implementing the idea. Also throughout our task, we had to go through various complex engineering problems to reach the final outcome.

5.2 Course Outcome (CO) Addressed :

Table 5.1: Course Outcome Addressed

Cos	CO Statement	POs	Put Tick
CO1	Apply knowledge of electrical and electronic engineering to the solution of complex engineering problems.	PO1	√
CO2	Identify a contemporary real life problem related to electrical and electronic engineering by reviewing and analyzing existing research works.	PO2	√
CO3	Determine functional requirements of the problem considering feasibility and efficiency through analysis and synthesis of	PO4	√

	information.		
CO4	Select a suitable solution and determine its method considering professional ethics, codes and standards.	PO8	√
CO5	Adopt modern engineering resources and tools for the solution of the problem.	PO5	√
CO6	Prepare management plan and budgetary implications for the solution of the problem.	PO11	
CO7	Analyze the impact of the proposed solution on health, safety, culture and society.	PO6	√
CO8	Analyze the impact of the proposed solution on environment and sustainability.	PO7	
CO9	Develop a viable solution considering health, safety, cultural, societal and environmental aspects.	PO3	√
CO10	Work effectively as an individual and as a team member for the accomplishment of the solution.	PO9	√
CO11	Prepare various technical reports, design documentation, and deliver effective presentations for demonstration of the solution.	PO10	√

CO12	Recognize the need for continuing education and participation in professional societies and meetings.	PO12	√

5.3 Aspects of Program Outcomes (POs) Addressed

Table 5.2: Program Outcome Addressed

	Statement	Different Aspects	Put tick
PO3	**Design/development of solutions:** Design solutions for complex electrical and electronic engineering problems and design systems, components or processes that meet specified needs with appropriate consideration for public health and safety, cultural, societal, and environmental considerations.	Public Health	√
		Safety	√
		Cultural	
		Societal	
		Environmental	
PO4	**Investigation:** Conduct investigations of complex electrical and electronic engineering problems using research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of information to provide valid conclusions.	Design of Experiments	√
		Analysis and interpretation of data	√
		Synthesis of	√

		Information	
PO6	**The engineer and society:** Apply reasoning informed by contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to professional engineering practice and solutions to complex electrical and electronic engineering problems.	Societal	
		Health	
		Safety	√
		Legal	√
		Cultural	
PO7	**Environment and sustainability:** Understand and evaluate the sustainability and impact of professional engineering work in the solution of complex electrical and electronic engineering problems in societal and environmental contexts.	Societal	
		Environmental	
PO8	**Ethics:** Apply ethical principles embedded with religious values, professional ethics and responsibilities, and norms of electrical and electronic engineering practice.	Religious Values	
		Professional Ethics and responsibilities	√
		Norms	√
PO10	**Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.	Comprehend and write effective reports	√
		Design Documenta	√

		tion	
		Make Effective presentations	√
		Give and receive clear instructions	√
PO11	**Project management and finance:** Demonstrate knowledge and understanding of engineering management principles and economic decision-making and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.	Engineering Management principle	√
		Economic Decision-making	
		Manage projects	√
		Multidisciplinary environments	

5.4 Knowledge Profiles (K3 – K8) Addressed

Table 5.3: Knowledge Profile Addressed

K	Knowledge Profile (Attribute)	Put

		Tick (√)
K3	A systematic, theory-based formulation of engineering fundamentals required in the engineering discipline	√
K4	Engineering specialist knowledge that provides theoretical frameworks and bodies of knowledge for the accepted practice areas in the engineering discipline; much is at the forefront of the discipline	√
K5	Knowledge that supports engineering design in a practice area	√
K6	Knowledge of engineering practice (technology) in the practice areas in the engineering discipline	√
K7	Comprehension of the role of engineering in society and identified issues in engineering practice in the discipline: ethics and the engineer's professional responsibility to public safety; the impacts of engineering activity; economic, social, cultural, environmental and sustainability.	
K8	Engagement with selected knowledge in the research literature of the discipline	√

5.5 Use of Complex Engineering Problems

Throughout our whole project we had to face a lot of complex engineering problems and had to solve it to reach the ultimate goal. From dataset preprocessing, handling dicom

data,resizing it transforming into usable pixel array, engineering knowledge of dataset handling was required. In the next portion, deep learning model was implemented, for that deep understanding of deep learning architecture, basics of neural network, CNN, Generative Adversarial Network,U-Net Generator, PatchGAN discriminator,Diffusion model architecture, Hybrid approach combining two was used which requires complex engineering knowledge, and one of the key challenges was to keep image shape accurate which requires problem solving,debugging ability. And lots of loss functions are also used to calculate the difference between images accurately. Implementing those successfully needed complex engineering problem solving ability.

5.6 Socio-Cultural, Environmental, And Ethical Impact

This work advances the current state of art to implement a real life applicable medical technology to translate images across different modalities. Once the result will be satisfactory enough to apply in real life it would be really major advancement in the medical sector and lots of patients will be secured from extra radiation of multiple medical tests and also the cost that they had to bear . This approach has no negative environmental impact. And about Ethical consideration, the dataset we have used QIN-Breast is licensed for public use of medical research.

5.7 Attributes of Ranges of Complex Engineering Problem

Solving (P1 – P7) Addressed

Table 5.4: Attributes of Ranges of Complex Engineering Problem

P	Range of Complex Engineering Problem Solving	Put Tick (√)
Attribute	Complex Engineering Problems have characteristic P1 and some or all of P2 to P7:	
Depth of	P1: Cannot be resolved without in-depth engineering	√

knowledge required	knowledge at the level of one or more of K3, K4, K5, K6 or K8 which allows a fundamentals-based, first principles analytical approach	
Range of conflicting requirements	P2: Involve wide-ranging or conflicting technical, engineering and other issues	√
Depth of analysis required	P3: Have no obvious solution and require abstract thinking, originality in analysis to formulate suitable models	√
Familiarity of issues	P4: Involve infrequently encountered issues	√
Extent of applicable codes	P5: Are outside problems encompassed by standards and codes of practice for professional engineering	√
Extent of stakeholder involvement and conflicting requirements	P6: Involve diverse groups of stakeholders with widely varying needs	
Interdependence	P7: Are high level problems including many component parts or sub-problems	√

5.8 Attributes of Ranges of Complex Engineering Activities (A1 –A5) Addressed

Table 5.5: Attributes of Ranges of Complex Engineering Activities Addressed

	Range of Complex Engineering Activities	Put

Attribute	Complex activities means (engineering) activities or projects that have some or all of the following characteristics:	Tick (√)
Range of resources	A1: Involve the use of diverse resources (and for this purpose resources include people, money, equipment, materials, information and technologies)	√
Level of interaction	A2: Require resolution of significant problems arising from interactions between wide-ranging or conflicting technical, engineering or other issues	√
Innovation	A3: Involve creative use of engineering principles and research-based knowledge in novel ways	√
Consequences for society and the environment	A4: Have significant consequences in a range of contexts, characterized by difficulty of prediction and mitigation	
Familiarity	A5: Can extend beyond previous experiences by applying principles-based approaches	√

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

This study presents a detailed exploration of Image translation across different modalities from PET to CT using advanced deep learning techniques. The work mainly focuses on three architectures, a Pix2Pix GAN, a Conditional Diffusion Model and a Hybrid diffusion + Edge-GAN using the PET/CT pair of QIN-Breast dataset from The Cancer Imaging Archive.

The Pix2Pix GAN acted as the baseline of all, with U-Net as generator and PatchGan as discriminator it achieved highest reconstruction fidelity with the result of SSIM 0.8799 and PSNR around 32.55 dB. In the qualitative portion it has been shown that it produced sharp, realistic CT images with well-defined organ boundaries as it has been gone through adversarial training and the inclusion of perceptual and SSIM based loss functions really helped the neural network to capture the difference between the target image and the generating image and updating the weight accordingly.

The Diffusion model, implemented with v-prediction, cosine β-scheduling, and MinSNR weighting, delivered anatomically coherent and noise-free outputs. Although its PSNR and SSIM were slightly lower than those of the GAN, the results were impressively stable and completely free of adversarial artefacts. This confirmed that diffusion-based frameworks can reconstruct the global intensity structure of CT volumes in a theoretically sound and probabilistic way. The main limitation was computational cost and relatively soft detail, which stemmed from limited training epochs and a small number of denoising steps.

To address these weaknesses, the Hybrid Diffusion + Edge-GAN was proposed. Basically it combines the strength of both previous approaches, the smooth intensity reconstruction of the diffusion model with the localized edge enhancement of an adversarial discriminator. And the model generated images that are visually sharper and more anatomically faithful than previous diffusion models.

And preprocessing the date to make it 2.5D PET conditioning improved spatial coherence across adjacent slices and reduced abrupt intensity transitions..

Overall, the comparative study highlighted that:

- GAN-based models perform better when the computational resources are limited and the dataset is paired.
- Diffusion models give us unmatched stability and theoretical rigor for large-scale or 3-D medical reconstruction tasks and it requires huge computational resources.
- Hybrid models are the promising middle ground, combining the strengths of both types of models ,specifically the sharpness of adversarial learning in GAN models and the reliability of diffusion inference in DDPM.

Throughout our work, it has been shown that deep learning can indeed reconstruct clinically meaningful CT representations from PET images,which will open the door to reduce the need for dual-modality scanning and will minimize the patient's radiation exposure. Though the current implementation was in 2-D slice translation, the results give us a solid foundation for future expansion into volumetric or cross-modality reconstruction pipelines.

6.2 Future Work

As the outcomes are promising there are several directions opened for further improvement and make it clinically applicable.

6.2.1. 3-D and Volumetric Extensions

Future work should be based on 3-D volumetric translation instead of 2-D slice wise training. Thus it will contain spatial context across all dimensions and the continuity between the slices will improve and the model will be able to capture anatomical variations which our 2-D training couldn't capture. An advanced GPU platform will make it possible to build a model like this and memory efficient approaches should be taken to keep the training more feasible.

6.2.2. Enhanced Dataset and Preprocessing

The dataset used here contains limited numbers of paired data and that's why the number of required data for making the model more robust is needed. Any public dataset containing more numbers of paired or unpaired PET/CT images will help the model to capture more details. Moreover, implementing automated SUV normalization, HU calibration, and cross-patient intensity standardization may improve the model performance by reducing noise.

6.2.3. Improved Diffusion Schedules

Some of the noise scheduling techniques like learned variance schedules,EDM (Elucidated Diffusion Model) , even DDPM++ framework are unexplored here which should be approached in future work that could benefit diffusion models. Basically these techniques are used to accelerate convergence and recover sharper details with fewer denoising steps.

6.2.4. Adaptive Hybridization

Introducing dynamic adversarial weighting could be a big step where the discriminator's contribution will be gradually increased as diffusion training stabilizes. Adaptive loss balancing would help to ensure that fine texture enhancement won't compromise with PSNR based stability.

6.2.5. Integration with Clinical Workflows

Depending only quantitative values won't be enough for making the model clinically applicable. In a practical scenario the generated CT image might be used in attenuation correction, dose planning or radiation-free follow-up imaging. Future work should collaborate with radiologists to have the proper qualitative evaluations as well.

6.2.6. Exploration of Cross-Modality and Multi-Task Learning

Another promising approach is combining images from multiple modalities, for example (PET,CT,MRI), which will help the model to have more information about the organ of particular image and it will have more versatility to translate image in multiple modalities. Combining multiple modalities will increase computational cost which must be checked.

6.2.7. Explainability and Uncertainty Estimation

While evaluating, rather than just depending on accuracy, approach should emphasize on interpretability as well. Uncertainty quantification and attention visualization techniques like Grad-CAM, diffusion time attention maps will help us to identify which regions are influencing the predictions most.

Closing Remark

In conclusion , our approach shows that we can translate PET images into corresponding CT images by using deep learning techniques. All the approaches including the GAN, Diffusion and Hybrid offered unique advantages and that shows a clear trajectory toward a more accurate, stable and practically applicable image translation system in healthcare.
The proposed hybrid pipeline particularly highlighted how combining the principles of probabilistic diffusion with targeted adversarial supervision can bridge the gap between stability and realism and can lead us to a more advanced model.

Larger dataset with 3-D integration of those deep learning techniques may lead us to more clinically applicable solutions contributing to reduce the radiation exposure of patients and the treatment cost as well.

References

[1] G. Antoch *et al.*, "Accuracy of Whole-Body Dual-Modality Fluorine-18–2-Fluoro-2-Deoxy- D -Glucose Positron Emission Tomography and Computed Tomography (FDG-PET/CT) for Tumor Staging in Solid Tumors: Comparison With CT and PET," *J. Clin. Oncol.*, vol. 22, no. 21, pp. 4357–4368, Nov. 2004, doi: 10.1200/JCO.2004.08.120.

[2] B. Huang, M. W.-M. Law, and P.-L. Khong, "Whole-Body PET/CT Scanning: Estimation of Radiation Dose and Cancer Risk," *Radiology*, vol. 251, no. 1, pp. 166–174, Apr. 2009, doi: 10.1148/radiol.2511081300.

[3] A. Shokraei Fard, D. C. Reutens, and V. Vegh, "From CNNs to GANs for cross-modality medical image estimation," *Comput. Biol. Med.*, vol. 146, p. 105556, July 2022, doi: 10.1016/j.compbiomed.2022.105556.

[4] I. J. Goodfellow *et al.*, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, Eds., Curran Associates, Inc., 2014. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2014/file/f033ed80deb0234979a61f95710dbe25-Paper.pdf

[5] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," 2014, *arXiv*. doi: 10.48550/ARXIV.1411.1784.

[6] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2016, *arXiv*. doi: 10.48550/ARXIV.1611.07004.

[7] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI: IEEE, July 2017, pp. 5967–5976. doi: 10.1109/CVPR.2017.632.

[8] K. Armanious *et al.*, "MedGAN: Medical image translation using GANs," *Comput. Med. Imaging Graph.*, vol. 79, p. 101684, Jan. 2020, doi: 10.1016/j.compmedimag.2019.101684.

[9] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds., Curran Associates, Inc., 2020, pp. 6840–6851. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf

[10] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," presented at the International conference on machine learning, PMLR, 2021, pp. 8162–8171.

[11] S. Zhang, J. Liu, B. Hu, and Z. Mao, "GH-DDM: the generalized hybrid denoising diffusion model for medical image generation," *Multimed. Syst.*, vol. 29, no. 3, pp. 1335–1345, June 2023, doi: 10.1007/s00530-023-01059-0.

[12] W. A. Weber, A. L. Grosu, and J. Czernin, "Technology Insight: advances in molecular imaging and an appraisal of PET/CT scanning," *Nat. Clin. Pract. Oncol.*, vol. 5, no. 3, pp. 160–170, 2008.

[13] S. K. Yang, N. Cho, and W. K. Moon, “The role of PET/CT for evaluating breast cancer,” *Korean J. Radiol.*, vol. 8, no. 5, pp. 429–437, 2007.

[14] S. Dayarathna, K. T. Islam, S. Uribe, G. Yang, M. Hayat, and Z. Chen, “Deep learning based synthesis of MRI, CT and PET: Review and analysis,” *Med. Image Anal.*, vol. 92, p. 103046, Feb. 2024, doi: 10.1016/j.media.2023.103046.

[15] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., in Lecture Notes in Computer Science, vol. 9351. , Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[16] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks,” 2017, *arXiv*. doi: 10.48550/ARXIV.1703.10593.

[17] Z. Xiao, K. Kreis, and A. Vahdat, “Tackling the generative learning trilemma with denoising diffusion GANs,” *ArXiv Prepr. ArXiv211207804*, 2021.

[18] H. Jiang *et al.*, “Fast-DDPM: Fast denoising diffusion probabilistic models for medical image-to-image generation,” *IEEE J. Biomed. Health Inform.*, 2025.

[19] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, “Score-Based Generative Modeling through Stochastic Differential Equations,” 2020, *arXiv*. doi: 10.48550/ARXIV.2011.13456.

[20] Y. Lu, “The level weighted structural similarity loss: A step away from MSE,” presented at the Proceedings of the AAAI conference on artificial intelligence, 2019, pp. 9989–9990.

[21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[22] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multiscale structural similarity for image quality assessment,” presented at the The thirty-seventh asilomar conference on signals, systems & computers, 2003, Ieee, 2003, pp. 1398–1402.

[23] A. Hore and D. Ziou, “Image quality metrics: PSNR vs. SSIM,” presented at the 2010 20th international conference on pattern recognition, IEEE, 2010, pp. 2366–2369.

[24] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” presented at the European conference on computer vision, Springer, 2016, pp. 694–711.

[25] X. Li *et al.*, “Data From QIN-Breast.” The Cancer Imaging Archive, 2016. doi: 10.7937/K9/TCIA.2016.21JUEBH0.

[26] “Generative Adversarial Network (GAN) - GeeksforGeeks.” Accessed: Oct. 29, 2025. [Online]. Available: https://www.geeksforgeeks.org/deep-learning/generative-adversarial-network-gan/

[27] “What are Diffusion Models? - GeeksforGeeks.” Accessed: Oct. 29, 2025. [Online]. Available: https://www.geeksforgeeks.org/artificial-intelligence/what-are-diffusion-models/

[28] “Kaggle: Your Home for Data Science.” Accessed: Oct. 29, 2025. [Online]. Available: https://www.kaggle.com/

[29] G. N. Hounsfield, “Computerized transverse axial scanning (tomography): Part 1. Description of system,” *Br. J. Radiol.*, vol. 46, no. 552, pp. 1016–1022, 1973.
[30] J. W. Keyes, “SUV: standard uptake or silly useless value?,” *J. Nucl. Med.*, vol. 36, no. 10, pp. 1836–1839, 1995.
[31] “U-Net Architecture Explained,” GeeksforGeeks. Accessed: Oct. 29, 2025. [Online]. Available: https://www.geeksforgeeks.org/machine-learning/u-net-architecture-explained/
[32] “The PatchGAN structure in the discriminator architecture. | Download Scientific Diagram.” Accessed: Oct. 29, 2025. [Online]. Available: https://www.researchgate.net/figure/The-PatchGAN-structure-in-the-discriminator-architecture_fig5_339832261
[33] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, “High-resolution image synthesis with latent diffusion models,” presented at the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 10684–10695.
[34] T. Salimans and J. Ho, “Progressive distillation for fast sampling of diffusion models,” *ArXiv Prepr. ArXiv220200512*, 2022.
[35] J. Ho and T. Salimans, “Classifier-Free Diffusion Guidance,” 2022, *arXiv*. doi: 10.48550/ARXIV.2207.12598.
[36] C. Li and M. Wand, “Precomputed real-time texture synthesis with markovian generative adversarial networks,” presented at the European conference on computer vision, Springer, 2016, pp. 702–716.
[37] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, “Deterministic edge-preserving regularization in computed imaging,” *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 298–311, 1997.
[38] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, “Spectral normalization for generative adversarial networks,” *ArXiv Prepr. ArXiv180205957*, 2018.
[39] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, “Improved techniques for training gans,” *Adv. Neural Inf. Process. Syst.*, vol. 29, 2016.

Appendices

Code

All experiments, model training, and evaluations were conducted in a Kaggle GPU environment (NVIDIA Tesla T4, 16 GB VRAM).

The complete implementation, including dataset preparation, model training, and evaluation scripts, is available in the following Kaggle notebook:

Kaggle Repository:

https://www.kaggle.com/code/almaheekhan/thesis-defence