Structure-Enhanced Translation from PET to CT Modality with Paired GANs

Tasnim Ahmed
tasnimahmed@iut-dhaka.edu
Islamic University of Technology
Gazipur, Dhaka, Bangladesh

Ahnaf Munir
ahnaf@iut-dhaka.edu
Islamic University of Technology
Gazipur, Dhaka, Bangladesh

Sabbir Ahmed
sabbirahmed@iut-dhaka.edu
Islamic University of Technology
Gazipur, Dhaka, Bangladesh

Md. Bakhtiar Hasan
bakhtiarhasan@iut-dhaka.edu
Islamic University of Technology
Gazipur, Dhaka, Bangladesh

Md. Taslim Reza
taslim@iut-dhaka.edu
Islamic University of Technology
Gazipur, Dhaka, Bangladesh

Md. Hasanul Kabir
hasanul@iut-dhaka.edu
Islamic University of Technology
Gazipur, Dhaka, Bangladesh

ABSTRACT

Computed Tomography (CT) images play a crucial role in medical diagnosis and treatment planning. However, acquiring CT images can be difficult in certain scenarios, such as patients inability to undergo radiation exposure or unavailability of CT scanner. An alternative solution can be generating CT images from other imaging modalities. In this work, we propose a medical image translation pipeline for generating high-quality CT images from Positron Emission Tomography (PET) images using a Pix2Pix Generative Adversarial Network (GAN), which are effective in image translation tasks. However, traditional GAN loss functions often fail to capture the structural similarity between generated and target image. To alleviate this issue, we introduce a Multi-Scale Structural Similarity Index Measure (MS-SSIM) loss in addition to the GAN loss to ensure that the generated images preserve the anatomical structures and patterns present in the real CT images. Experiments on the 'QIN-Breast' dataset demonstrate that our proposed architecture achieves a Peak Signal-to-Noise Ratio (PSNR) of 17.70 dB and a Structural Similarity Index Measure (SSIM) of 42.51% in the region of interest.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision**.

KEYWORDS

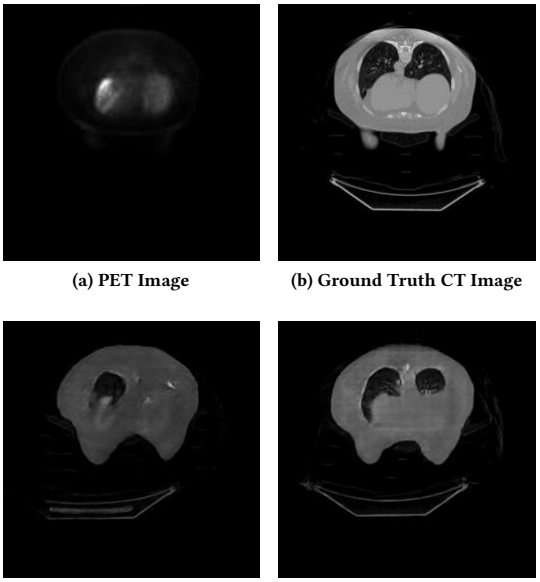
Medical Image Translation, GAN, PET to CT, Medical Imaging, QIN-Breast, Breast Cancer Treatment

ACM Reference Format:
Tasnim Ahmed, Ahnaf Munir, Sabbir Ahmed, Md. Bakhtiar Hasan, Md. Taslim Reza, and Md. Hasanul Kabir. 2023. Structure-Enhanced Translation from PET to CT Modality with Paired GANs. In *Proceedings of 6th International Conference on Machine Vision and Applications (ICMVA '23).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

ICMVA '23, March 10–12, 2023, Singapore

ACM ISBN 978-1-4503-9953-1/23/03…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION



(a) PET Image **(b) Ground Truth CT Image**

(c) Generated CT Image using baseline Pix2Pix **(d) Generated CT Image using our proposed architecture**

Figure 1: Our proposed architecture is able to better generate CT Images from PET images. This is a representative example of the performance of our model in a challenging dataset.

The combination of PET and CT imaging is now a regular part of oncology diagnosis and staging [1]. These imaging techniques carry useful markers for cancerous cells which facilitate the detection of malignant lesions [2]. In addition, PET/CT imaging is becoming a crucial technique for evaluating prospective drug therapies [3]. Although both of these imaging techniques are useful to detect cancer at an early stage, CT images are often considered superior to PET images in medical diagnosis for several reasons [4]. CT scans have a higher spatial resolution containing greater detail and clarity compared to that of PET. Moreover, CT scans provide detailed anatomic information and are better at localizing abnormalities within the body. While differentiating tissue, such as bone, muscle, fat, etc., CT imaging provides greater accuracy than PET scans

[5]. Hence, an image translation pipeline to produce accurate CT images from the available PET image can be a vital tool in places where CT scans are difficult to avail. To facilitate this, medical image translation can be considered as one of the potential solutions.

Medical image translation focuses on the transformation of medical images from one imaging modality to another, or from one representation to another. This can be useful for improving image quality, facilitating image comparison across modalities, etc [6]. In this regard, the remarkable ability of learning complex patterns from highly diversified datasets of the Deep learning-based algorithms can be utilized in this regard, which has been found to be extremely useful in a wide variety of detection [7–9], recognition [10–13] and classification [14–16] tasks in recent times. More specifically, in medical image translation, the architectures like Generative Adversarial Networks (GANs), Convolutional Neural Networks (CNNs), Attention-based models, etc., have enabled generation of highly accurate images that are almost impossible to differentiate from real ones, even for the domain experts [17–19]. Despite there being a number of works focusing on CT image generation from Magnetic Resonance (MR) image [20, 21], no works have been proposed that focus on translating PET images to CT imaging modality.

This work proposes a pipeline that can translate PET images to CT imaging modality with high precision. We utilized the Pix2Pix GAN and introduced a Multi-Scale Structural Similarity loss to alleviate the limitations of traditional GANs in capturing structural similarity between generated and target images. The pipeline focuses on structural information to generate images having similar anatomical features as the original image. Experiments on the publicly available 'QIN-Breast' dataset have provided promising results with significant PSNR and SSIM values in the region of interest.

2 METHODOLOGY

Our pipeline adopts the Pix2Pix architecture [22] that exploits the Conditional GAN architecture. As shown in Figure 2, for every pair of ground truth PET and CT images, the PET image is used as input to the generator to produce the corresponding CT image. The generator of Pix2Pix uses the U-Net architecture [23] while PatchGAN network is proposed for the discriminator. We replace the U-Net architecture with ResNet [24] and incorporate an Image Quality Metric (IQM) based loss term. The generator-produced and the ground truth CT image are then used to train the discriminator.

Generally, two types of IQM, subjective and objective, are used for comparing image qualities in medical image analysis [25]. The subjective measurement is carried out by a group of experts that inspect and evaluate the images. Although this method yields the most consistent results, the underlying process is too slow and expensive. Objective measurements aim to overcome these limitations albeit being less sensitive to the Human Visual System (HVS) compared to the subjective techniques [26]. In this work, we utilize one of the most popular objective measurement methods called Structural Similarity Index Measure (SSIM).

2.1 Dataset

To effectively translate the PET images to CT modality, ground truth PET-CT pairs are required for the same subjects so that the model

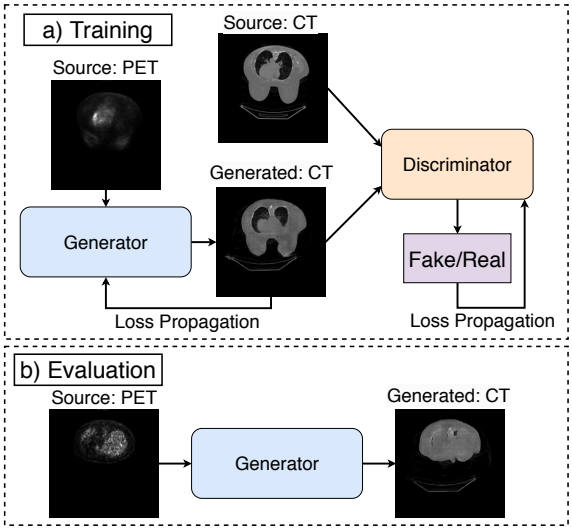


Figure 2: Overview of the proposed pipeline.

can understand the co-relation and the probability distribution of both modalities. In this regard, the 'QIN-Brest' [27] is the largest publicly available dataset containing around 9000 pairs of PET-CT scans for the same subjects in different stages of treatment. Hence we have chosen this dataset for conducting all the experiments. The dataset contains images of three modalities- MR, PET, and CT of 68 patients over 214 studies. In our experiments, we could only utilize the subjects for which both of the PET/CT images were available. The samples represent three phases — prior to the treatment, after one pass of treatment, and after the second pass or prior to surgery.

2.2 Structural Similarity Index Measure (SSIM)

SSIM improves upon existing subjective measurements and better represent the HVS [28]. It evaluates the visual impact of three different characteristics of an image: luminance (l), contrast (c), and structure (s). For each pixel, a sliding window is placed to calculate the SSIM of the region under the window, ensuring local assessment of the image. For any pair of corresponding pixels x and y in images I and I' respectively, the SSIM is calculated using:

$$SSIM(x, y) = [l(x, y)^{\alpha} \cdot c(x, y)^{\beta} \cdot s(x, y)^{\gamma}] \tag{1}$$

Here, α, β, and γ are parameters to control the relative importance of the three comparison functions. These three functions can be defined as:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{2}$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{3}$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{4}$$

Here, μ_x and μ_y are the mean intensities of the regions under the sliding windows centered at the pixels x and y respectively. The

variances of the two regions are represented as σ_x^2 and σ_y^2 while σ_{xy} denote the covariance between them. The constants C_1, C_2, and C_3 are used in the equations to stabilize the division operation if the denominator is too small. By setting $C_3 = C_2/2$ and $\alpha = \beta = \gamma = 1$, we can reduce Equation 1 to:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_1)} \tag{5}$$

Multi-scale SSIM (MS-SSIM) makes use of image details from different scales [29]. A series of low-pass filters followed by downsampling is applied to the original image before computing the SSIM measurements at each of these new lower resolutions:

$$SSIM(x, y) = [l_M(x, y)]^{\alpha_M} \cdot \prod_{j=1}^{M} [c_j(x, y)]^{\beta_j} . \cdot [s_j(x, y)]^{\gamma_j} \tag{6}$$

Here, M corresponds to the lowest resolution while $j = 1$ corresponds to the original resolution of the image.

2.3 Loss Functions

The vanilla GAN framework simultaneously trains a generator model G and a discriminator model D. Here the generator is responsible for creating new data instances Y that resemble the training data X. The discriminator is used to determine whether the new data instances are from the training data or the generator.

$$\min_G \max_D \mathcal{L}_{Adv}(G, D) = \mathbb{E}_y[log(D(y))] + \mathbb{E}_x[log(1 - D(G(x)))] \tag{7}$$

CGAN modifies the generator ($G : X \rightarrow Y$) by minimizing pixel-level L1 loss between the source (x) and target (y) image.

$$\min_G \mathcal{L}_{L1}(G) = \mathbb{E}_{x,y}[||y - G(x)||_1] \tag{8}$$

In order to ensure superior perceptual quality in the generated images, we introduce the MS-SSIM term along with this loss function. For a generated image I, the MS-SSIM loss is computed using:

$$\mathcal{L}_{MS\text{-}SSIM}(I) = 1 - MS\text{-}SSIM(I) \tag{9}$$

Equation 10 illustrates the combined loss function, where C dictates the contribution of the MS-SSIM loss:

$$\mathcal{L}(I) = \mathcal{L}_{L1}(G) + C \cdot \mathcal{L}_{MS\text{-}SSIM}(I) \tag{10}$$

3 RESULTS ANALYSIS

3.1 Evaluation Metrics

The performance of our proposed architecture was evaluated by demonstrating that the CT images generated by the proposed pipeline are superior in both their practicability and quality. The primary objective behind the implementation of our pipeline was to achieve a higher level of structural similarity between the ground truth image and the generated one. The Normalized Mean Absolute Error (NMAE), Peak Signal to Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM) were used to evaluate the performances of trained models. We excluded the background while calculating the metrics to prevent misleading results. The majority of the samples in our dataset consist mostly of black pixels, with the center of the image containing important clinical information. The real CT image and the generated image both have layers of black pixels around the relevant pixels, which results in exaggerated performance when calculated for the entire image. To obtain a more accurate measure of the performance, we first applied thresholding to the real CT image to convert it to a binary image. Then, a bounding box was generated around the relevant regions of the image by excluding the surrounding black pixels. The performance metrics were calculated only within this bounding box, which penalized any differences in contrast, luminance, or structure between the real and generated images of those regions.

Table 1: Performance comparison of the various Pix2Pix GAN modes on the QIN-Breast dataset with our proposed pipeline.

GAN Setting	Generator Architecture	Learning Rate Decay	Loss	NMAE ↓	PSNR ↑	SSIM ↑
Pix2Pix	ResNet	Linear	L_{GAN}	0.33	15.76	39.33
		Linear	$L_{GAN} + L_{MS-SSIM}$	**0.35**	**17.70**	**42.51**
		Plateau	$L_{GAN} + L_{MS-SSIM}$	0.42	16.89	38.11

3.2 Quantitative Results

A summary of the quantitative results are shown in Table 1. It is evident that the Pix2Pix GAN with $L_{MS-SSIM}$ achieved significantly better performance although NMAE is slightly higher than the baseline Pix2Pix GAN. Incorporating $L_{MS-SSIM}$ improved the SSIM of our proposed pipeline by 3.18% and increased PSNR by almost 2 dB. Images are typically evaluated using one of two traditional quality metrics: NMAE or PSNR because both have a physical meaning and are straightforward to calculate. These quality measures can give the same value of quality to two different distorted images, even if one of them is more perceivable than the other, which is in direct contradiction to the HVS [30]. On the other hand, SSIM considers the structural and contextual information in the image, making it a more reliable metric for measuring perceptual quality compared to PSNR and NMAE. Therefore, it is reasonable to expect that a model with better SSIM performance will produce accurate and realistic-looking images.

3.3 Ablation Study

An ablation study was conducted to understand the contribution of different components of the proposed pipeline to the overall performance. We considered several combinations of the design choices, such as the number of epochs, batch size, learning rate (LR), and learning rate decay (lrDecay) to analyze their effects.

Appropriate selection of lrDecay enables a model to discover globally optimal weights, which improves the optimization of the objective function [15]. We experimented with two lrDecay strategies: Linear and Plateau. In plateau, the LR is reduced when the performance metric plateaus. In linear lrDecay, the LR is reduced linearly over time, based on a predefined schedule which can help the model converge gradually and avoid overshooting a minima in the loss function. The choice between these two strategies is highly

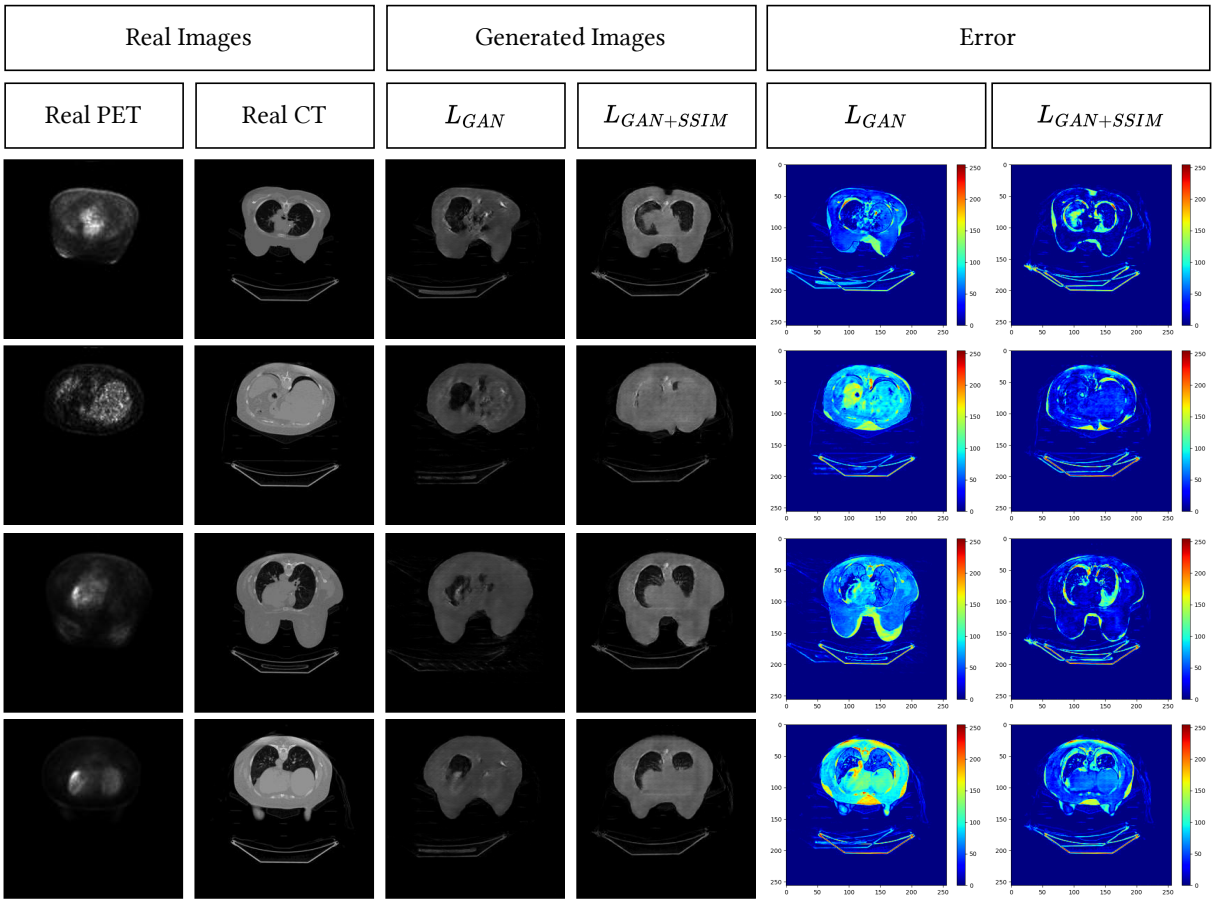


Figure 3: Generated images compared with real CT images.

dependent on the specific problem domain and desired outcome. In our experimental setup, we noticed a significant improvement in performance while using linear decay over plateau as shown in Table 1. Since the LR decreased in a predictable manner, it was easier to control the training process using linear lrDecay. Additionally, using plateau lrDecay, we noticed oscillations in the LR when performance metrics were oscillating around a minimum, which in turn caused the performance drop. An initial LR of 0.0002 for the Adam optimizer provided the optimal results after training for 100 epochs. While most generative models are stochastically designed to sample from an acquired distribution, GAN models use input images to generate predetermined output images. In this case, we did not find any statistically significant variation in target samples taken independently from the same source.

3.4 Qualitative Analysis

The results of the proposed method were compared to the ground truth CT images using visual inspection. The proposed method was able to produce CT images with high structural similarity to the ground truth images. The generated images were able to capture clinically relevant information, such as the shape and location of anatomical structures, as illustrated in Figure 3. The output of the baseline Pix2Pix showed significant structural deformities in the generated images, which were largely corrected after incorporating the $L_{MS-SSIM}$ loss. Another intriguing finding is that the sample images generated by the baseline Pix2Pix significantly under-performed in matching the luminance and contrast with the ground truth image. Incorporating $L_{MS-SSIM}$ significantly reduces the difference in luminance and contrast as it penalizes the generator for luminance and contrast difference from the ground truth image by following the equation 4. We also discovered a few instances of misalignment between image pairs (PET-CT) in our dataset, which we believe contributed to the overall performance drop.

4 CONCLUSION

The results of our study demonstrate the feasibility of generating high-quality CT images from PET images using Pix2Pix GAN combined with multi-scale structural similarity loss on a publicly available dataset, QIN-Breast. The generated CT images closely resemble the real CT images, with similar anatomical structures and patterns, and have high structural similarity and low error

compared to the real CT images. This approach has the potential to provide valuable information for medical diagnosis and treatment planning and could have significant implications for clinical practice. The proposed method has the potential to provide an alternative solution for obtaining CT images in scenarios where the acquisition of CT images is infeasible.

REFERENCES

[1] Wolfgang A. Weber, Anca L. Grosu, and Johannes Czernin. "Technology Insight: advances in molecular imaging and an appraisal of PET/CT scanning". In: *Nature Clinical Practice Oncology* 5.3 (2008), pp. 160–170.

[2] Judith E. Kalinyak et al. "Breast cancer detection using high-resolution breast PET compared to whole-body PET or PET/CT". In: *European Journal of Nuclear Medicine and Molecular Imaging* 41.2 (2014), pp. 260–275.

[3] David A. Mankoff and Sharyn I. Katz. "PET imaging for assessing tumor response to therapy". In: *Journal of Surgical Oncology* 118.2 (2018), pp. 362–373.

[4] Moon Woo Kyung Yang Sang Kyu Cho Nariya. "The Role of PET/CT for Evaluating Breast Cancer". In: *kjr* 8.5 (2007), pp. 429–437.

[5] Marcus D Seemann. "PET/CT: fundamental principles". In: *European journal of medical research* 9.5 (2004), pp. 241–246.

[6] Aziz Alotaibi. "Deep Generative Adversarial Networks for Image-to-Image Translation: A Review". In: *Symmetry* 12.10 (2020).

[7] Raian Rahman, Zadid Bin Azad, and Md. Bakhtiar Hasan. "Densely-Populated Traffic Detection Using YOLOv5 and Non-maximum Suppression Ensembling". In: *Proceedings of the International Conference on Big Data, IoT, and Machine Learning*. Springer Singapore, 2022, pp. 567–578.

[8] Shafkat Farabi et al. "Improving Action Quality Assessment Using Weighted Aggregation". In: *Pattern Recognition and Image Analysis*. Springer, 2022, pp. 576–587.

[9] Mohsinul Kabir et al. "DEPTWEET: A typology for social media texts to detect depression severities". In: *Computers in Human Behavior* 139 (2023), p. 107503.

[10] A. B. M. Ashikur Rahman et al. "Two Decades of Bengali Handwritten Digit Recognition: A Survey". In: *IEEE Access* 10 (2022), pp. 92597–92632.

[11] Md. Bakhtiar Hasan, Tasnim Ahmed, and Md. Hasanul Kabir. "HEATGait: Hop-Extracted Adjacency Technique in Graph Convolution based Gait Recognition". In: *4th International Conference on Advances in Computer Technology, Information Science and Communications (CTISC)*. 2022, pp. 1–6.

[12] Arowa Yasmeen et al. "CSVC-Net: Code-Switched Voice Command Classification using Deep CNN-LSTM Network". In: *10th International Conference on Informatics, Electronics & Vision (ICIEV)*. 2021, pp. 1–8.

[13] Ayesha Khatun et al. "A Systematic Review on the Chronological Development of Bangla Sign Language Recognition Systems". In: *10th International Conference on Informatics, Electronics & Vision (ICIEV)*. 2021, pp. 1–9.

[14] Tasnim Ahmed et al. "A Novel Approach to Classify Electrocardiogram Signals Using Deep Neural Networks". In: *2nd International Conference on Computer and Information Sciences (ICCIS)*. 2020, pp. 1–6.

[15] Sabbir Ahmed et al. "Less is More: Lighter and Faster Deep Neural Architecture for Tomato Leaf Disease Classification". In: *IEEE Access* 10 (2022), pp. 68868–68884.

[16] Tasnim Ahmed et al. "A Complete Bangla Optical Character Recognition System: An Effective Approach". In: *22nd International Conference on Computer and Information Technology (ICCIT)*. 2019, pp. 1–7.

[17] Salome Kazeminia et al. "GANs for medical image analysis". In: *Artificial Intelligence in Medicine* 109 (2020), p. 101938.

[18] Shizuo Kaji and Satoshi Kida. "Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging". In: *Radiological Physics and Technology* 12.3 (2019), pp. 235–248.

[19] Fahad Shamshad et al. *Transformers in Medical Imaging: A Survey*. 2022. arXiv: 2201.09873 [eess.IV].

[20] Yang Lei et al. "MRI-based synthetic CT generation using deep convolutional neural network ". In: *Medical Imaging 2019: Image Processing*. Vol. 10949. SPIE, 2019, 109492T.

[21] Vasant Kearney et al. "Attention-Aware Discrimination for MR-to-CT Image Translation Using Cycle-Consistent Generative Adversarial Networks". In: *Radiology: Artificial Intelligence* 2.2 (2020), e190027.

[22] Phillip Isola et al. "Image-To-Image Translation With Conditional Adversarial Networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.

[23] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention*. Springer International Publishing, 2015, pp. 234–241.

[24] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778.

[25] Anuja khodaskar and Siddarth Ladhake. "Semantic Image Analysis for Intelligent Image Retrieval". In: *Procedia Computer Science* 48 (2015), pp. 192–197.

[26] Yingjing Lu. "The Level Weighted Structural Similarity Loss: A Step Away from MSE". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 33.01 (2019), pp. 9989–9990.

[27] X Li et al. "Data from QIN-breast". In: *The Cancer Imaging Archive* (2016).

[28] Zhou Wang et al. "Image quality assessment: from error visibility to structural similarity". In: *IEEE Transactions on Image Processing* 13.4 (2004), pp. 600–612.

[29] Z. Wang, E.P. Simoncelli, and A.C. Bovik. "Multiscale structural similarity for image quality assessment". In: *37th Asilomar Conference on Signals, Systems & Computers*. Vol. 2. 2003, 1398–1402 Vol.2.

[30] Alain Hore and Djemel Ziou. "Image quality metrics: PSNR vs. SSIM". In: *20th international conference on pattern recognition*. IEEE. 2010, pp. 2366–2369.