USING GENERATIVE ADVERSARIAL NETWORKS TO SIMULATE FLUORODEOXYGLUCOSE UPTAKE FROM COMPUTED TOMOGRAPHY SCANS

FRANCESCA LUIGIA MAROGNA

INTERNSHIP PORTFOLIO SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE

DEPARTMENT OF
COGNITIVE SCIENCE & ARTIFICIAL INTELLIGENCE
SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
TILBURG UNIVERSITY

STUDENT NUMBERS

u4001763
2020140

COMMITTEE

dr. Sharon Ong
Martijn van Leeuwen
dr. Görkem Saygili

LOCATION

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science &
Artificial Intelligence
Tilburg, The Netherlands

DATE

January 22, 2024

USING GENERATIVE ADVERSARIAL NETWORKS TO SIMULATE FLUORODEOXYGLUCOSE UPTAKE FROM COMPUTED TOMOGRAPHY SCANS

FRANCESCA LUIGIA MAROGNA

1 INTRODUCTION

Technological progress in the diagnosis of pathologies, especially tumors, has reached significantly high levels. Clinicians are now able to have at their disposal both the functional and morphological characteristics of pathological lesions. Morphological features define the physical characteristics of a tumor in terms of shape, volume, location, and density (Lee et al., 2017). In contrast, functional features involve altered tissue behavior and impaired physiological processes such as altered blood flow, endocrine dysfunction, and tissue activity (Kumar, Abbas, Fausto, & Aster, 2009).

1.1 *CT & PET Scans*

In the realm of imaging diagnostics that provide morphological features of lesions, one of the most revolutionary advances from the 1970s has been the production of multi-layer computed tomography (CT) with intravenous contrast injection which has enabled the highlighting of even the tiniest pathological lesions in various systems (Cormack, 1980; Hounsfield, 1980). CT scans are constructed according to the amount of X-ray absorption in different types of tissue. The amount of absorption is related to tissue density: bone absorbs the most (and so the skull appears white), cerebrospinal fluid absorbs the least (so the ventricles appear black), and the brain matter is intermediate (and appears gray) (Ward, 2019). Just a decade later, a functional imaging device called positron emission tomography (PET) became

widely used for medical purposes (Phelps et al., 1976). PET scans provide information about the metabolic activity of tissues in the body. PET uses a radioactive tracer injected into the bloodstream. The greater the blood flow in a region, the greater the signal emitted by the tracer in that region (Ward, 2019). The PET scanner detects these signals and creates detailed images of the metabolic activity of the tissues. When combining PET and CT (i.e., PET/CT), diagnostic accuracy approaches near 100%, yielding high sensitivity and specificity (Krause, Schwarzenböck, & Souvatzoglou, 2013). The development of clinical PET/CT, attributed to David Townsend and Ronald Nutt (Jones & Townsend, 2017), occurred in the mid-1990s and won TIME Magazine's Medical Invention of the Year in 2000 (Jaroff, 2000). PET/CT with fluorodeoxyglucose (FDG) finds application in studying numerous neoplastic pathologies by exploiting tumor cells' ability to absorb glucose, coupled with the radioactive marker i.e., FDG (Basu et al., 2014). Indeed, malignant cells have increased glucose consumption as they need plenty of energy to maintain their uncontrolled replication capacity high. Normally, each cell has enzymes that regulate glucose intake; cancer cells do not have this control, and their growth does not come to a halt. Consequently, injection of a glucose-based radionuclide (i.e., a radioactive element that can be detected by equipment such as CT and PET) such as FDG, can be used to identify these abnormal metabolically active tissues (Kwee & Kwee, 2009). The comparison of the two diagnostic methods (PET/FDG + CT) enables the identification of pathological radioactive accumulations in the body while correlating morphological and functional data. As useful and revolutionary as these devices are, they are accompanied by high costs for diagnostic equipment production, usage, and maintenance. As a result, many clinical facilities cannot afford such advanced diagnostics. Moreover, these approaches do not come without radioactive risks for the patients undergoing the procedure (Saif, Tzannou, Makrilia, & Syrigos, 2010).

1.2 *Generating Pseudo PET Scans*

This research aims to generate pseudo PET scans, therefore removing the cons aforementioned, by simulating the FDG uptake from CT scans alone. This work will explore to what extent the data acquired from a non-contrast CT contain sufficient information to differentiate the various regions with different FDG uptake. Radiomic features from CT scans have been shown to vary significantly in regions of differing metabolic activity (Tixier et al., 2014). Radiomics is the post-processing and analysis of large amounts of quantitative imaging features that can be derived from medical images. Radiomics features can reflect the spatial complexity, genomic heterogeneity, and subregional identification of lung cancer (Lee et al., 2017).

Therefore, it is possible that PET scans can be reliably inferred from CT scans. The research was conducted within the Department of Humanities and Digital Sciences (TSHD) of Tilburg University, in collaboration with the nonacademic teaching hospital Elizabeth-TweeSteden Ziekenhuis (ETZ) in Tilburg. The ETZ is a top clinical teaching hospital and trauma center for all residents of the Brabant region (Timmers, Keijsers, Kremer, Janssen, & Smeenk, 2021). The proposed approach is to use a Pix2Pix Generative Adversarial Network (GAN) model to simulate the FDG uptake.

GANs are a class of artificial intelligence algorithms used in unsupervised machine learning. The key idea behind GANs is to train two neural networks, a generator, and a discriminator, simultaneously through adversarial training. The generator aims to create realistic data, while the discriminator's role is to distinguish between real and fake data. As the generator improves, the discriminator needs to become more discerning, creating a competitive and iterative training process that leads to the generation of increasingly realistic data (Goodfellow et al., 2014).

Pix2Pix, or Image-to-Image Translation with Conditional Adversarial Networks, is a specific type of GAN designed for image translation tasks (Isola, Zhu, Zhou, & Efros, 2017). Pix2Pix uses a conditional GAN architecture, where the generator takes an input image from one domain and translates it into an output image in another domain. The conditional GAN framework allows for the generation of images that are conditioned on specific input data. The versatility of Pix2Pix was harnessed for various image translation applications, including the conversion of satellite images to maps, grayscale image colorization, the transformation of sketches into realistic images, and semantic segmentation. This method learns to map input images to output images in a paired dataset during the training phase. For instance, in the context of semantic segmentation, where the goal is to generate a segmented version of an input image, Pix2Pix learns to understand the mapping between images and their corresponding segmented masks (Shukla, Van Gool, & Timofte, 2019). The generator produces images that, when fed into the discriminator, are difficult to distinguish from real examples, resulting in high-quality translations.

The primary goal of this study is to investigate whether, when integrated with a low-dose CT, this simulation method can attain a level of precision in line with that of PET/FDG + CT. A similar study is the one performed by Chandrashekar, Handa, Ward, Grau, and Lee (2022), in which they simulated the FDG uptake in head and neck cancers using non-contrast CT images without the administration of radioactive tracer. The aim here is to expand this research to the whole body by utilizing the dataset collected by Gatidis et al. (2022).

1.3 *Scientific Motivation*

Various inputs and metrics will assess Pix2Pix model performance. Input images cropped to 256x256 and 128x128, explore lesion zooming effects. Adaptive Histogram Equalization (AHE) enhances cropped images, a technique widely used in medical imaging (Pizer et al., 1987). Four unique input scenarios present challenges for the Pix2Pix model. For model evaluation, Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) will be used. SSIM gauges structural similarity, considering luminance, contrast, and structure, providing a holistic assessment (Wang, Bovik, Sheikh, & Simoncelli, 2004). PSNR evaluates image fidelity by measuring the ratio between maximum signal power and introduced noise power, offering a quick assessment (Korhonen & You, 2012). Using both metrics aims for a comprehensive evaluation, considering various aspects of image quality.

1.4 *Societal Motivation*

Additionally, the breakthrough in this research holds the potential to revolutionize medical imaging practices, fostering a more inclusive and efficient healthcare system. Patients with allergies to radioactive markers, who were previously constrained in their diagnostic options, can now benefit from the same high level of accuracy in diagnoses as their non-allergic counterparts (Nyakale, Lockhat, & Sathekge, 2015). Beyond the evident advantages for patients, doctors stand to gain significant benefits as well. With the adoption of non-contrast imaging techniques, medical professionals can substantially reduce the time required for diagnosis. A non-contrast CT scan, taking approximately 15 minutes, represents a remarkable 30-minute decrease compared to the time-consuming process of a contrast CT. This time efficiency not only enhances the overall workflow in medical facilities but also allows healthcare providers to promptly address patient needs, leading to more timely interventions and improved outcomes. As such, the implications of this research extend far beyond the confines of individual cases, promising a transformative impact on the broader landscape of medical diagnostics and patient care

1.5 *Research Questions*

The central issue under investigation in the present study revolves around the feasibility of generating PET scans through the simulation of FDG uptake exclusively from CT scans. The primary focus is on discerning the extent to which PET scans can be accurately synthesized without direct

FDG information. The leading research question that the current paper addresses therefore is:

> *To what extent can PET scans be generated by simulating the FDG uptake from CT scans alone?*

The relevant research questions taken into consideration thus aim at comprehensively addressing the problem statement and are articulated as follows:

RQ1 *How does the size of the input image, namely 256x256 and 128x128, influence the performance of the model?*

RQ2 *How does the enhancement of the input image using AHE affect the performance of the model?*

RQ3 *To what extent do different metrics, specifically SSIM and PSNR, quantify the performance of the model?*

The first research question (RQ1) posits an exploration into the impact of input image size on the model's performance. This investigation seeks to elucidate the relationship between the dimensions of the input CT scans and the efficacy of the ensuing PET scan generation process. By systematically varying the size of input images, the study aims to derive insights into the optimal configuration that maximizes the accuracy of PET scan synthesis.

Additionally, the second research question (RQ2) delves into the impact of enhancing the input image using AHE on the model's performance. This inquiry seeks to understand the influence of AHE, a contrast enhancement technique, on the accuracy and quality of the synthesized PET scans. By investigating the effects of AHE on the input images before the simulation process, the study aims to elucidate whether this enhancement strategy contributes positively to the overall performance of the PET scan generation model. This exploration is crucial in gauging the potential benefits of preprocessing techniques like AHE in enhancing the fidelity of PET scans derived from CT scans alone.

Concurrently, the third research question (RQ3) delves into the nuanced influence of diverse evaluation metrics, specifically SSIM and PSNR, on the model's overall performance. This query aims to unravel the extent to which the choice of assessment criteria impacts the fidelity and reliability of the synthesized PET scans. Through a meticulous examination of SSIM and PSNR metrics, the research endeavors to discern their respective roles and contributions in gauging the authenticity of the generated PET images.

In essence, these refined research questions collectively contribute to a systematic and comprehensive investigation, aiming to provide nuanced

insights into the intricate dynamics of PET scan synthesis through CT-based FDG uptake simulation.

1.6 *Structure of Paper*

The aim and scope of this present research are established in Chapter 1, by stating the problem indication, problem statement, and research questions. Chapter 2 presents a related work of the literature that identifies and defines the current field of PET/CT. The methodology is explained in Chapter 3, specifically the details of the data, the data pre-processing, and the model's implementation determined in Chapter 1. The results of the models are reported in Chapter 4, and the results of said models are discussed in Chapter 5 as well as potential limitations and future research. Additionally, Chapter 6 includes conclusive remarks. Finally, a brief review of the personal development of the author can be found in Chapter 7.

2 RELATED WORK

The simulation of pseudo PET scans from CT scans constitutes a niche yet pivotal area within the domain of medical imaging research, characterized by a scarcity of comprehensive studies. The exploration of simulating PET scans from CT scans has undergone substantial evolution, traversing diverse methodologies and research domains. Within the broader scope of medical image synthesis, researchers have delved into tasks akin to simulating PET scans from CT scans, such as cross-modality image translation. For instance, Nie, Cao, Gao, Wang, and Shen (2016) employed a 3D fully convolutional neural network (FCN) to establish an end-to-end nonlinear mapping from magnetic resonance (MR) images to CT images. Their rationale for choosing FCN lay in its ability to generate structured output, preserving neighborhood information in the predicted CT images. Similarly, Han (2017) endeavored to derive CT images from MR images using a deep convolutional neural network (DCNN), underlining the importance of CT estimation for PET attenuation correction in hybrid PET-MR systems.

In recent years, the paradigm shift induced by deep learning has revolutionized PET image simulation. Pioneering this shift, researchers such as Bi, Kim, Kumar, Feng, and Fulham (2017) and Thanh et al. (2022) employed deep models to synthesize PET images. Bi et al. utilized a multi-channel generative adversarial network to synthesize PET images from CT images with manually annotated lung tumors. Their model learned integration from both CT and annotated labels, demonstrating comparable detection performance to original PET data. Conversely, Thanh et al. introduced a

novel Convolution-mix-Transformer as a Generator network for translating CT images to PET images. A closely aligned study is presented by Ben-Cohen et al. (2019), who devised a system combining an FCN with a conditional GAN to generate simulated PET data from input CT data, specifically targeting false-positive reduction in lesion detection solutions.

This progression toward deep learning techniques underscores an increasing trend in pushing the boundaries of simulation accuracy. Notably, Chandrashekar et al. (2022) simulated FDG uptake in head and neck cancers using non-contrast CT images without the administration of a radioactive tracer. Their use of a Cycle-GAN for CT to PET transformation demonstrated the ability to predict clinical outcomes with similar accuracy to FDG-PET images. However, the specific research gap addressed in this study, focusing on the creation of pseudo-PET scans for whole-body scans, remains unexplored.

As the field advances, a nuanced research gap becomes apparent, necessitating a structured overview to identify contrasts in methodologies and the need for further refinement. The challenge lies in enhancing the accuracy of simulated PET-FDG uptake patterns from CT scans and establishing their robustness in diverse clinical contexts. This related work not only provides a historical context but also sets the stage for identifying the specific research gap that the current study aims to address — a critical step towards advancing the field of PET imaging simulation.

3 METHODOLOGY

In the Methodology chapter of this project, the foundational framework for the research is meticulously outlined, providing a comprehensive understanding of the methodologies employed to investigate the proposed objectives. The chapter delves into the crucial components of the experimental process, beginning with a detailed examination of the raw data utilized for the study. A thorough exploration of the data pre-processing techniques follows, elucidating the steps taken to refine and prepare the dataset for analysis. The chapter then elucidates the intricacies of the train-validation-test split, a pivotal element in ensuring the robustness and generalization of the proposed model. Subsequently, the focus shifts to the methodology employed — specifically, the Pix2Pix architecture — which is expounded upon to convey its relevance and applicability to the research context. This chapter serves as a road map for the reader, guiding them through the intricate details of the experimental design and laying the groundwork for the subsequent analyses and findings. An overview of the entire project's pipeline can be seen in Figure 1

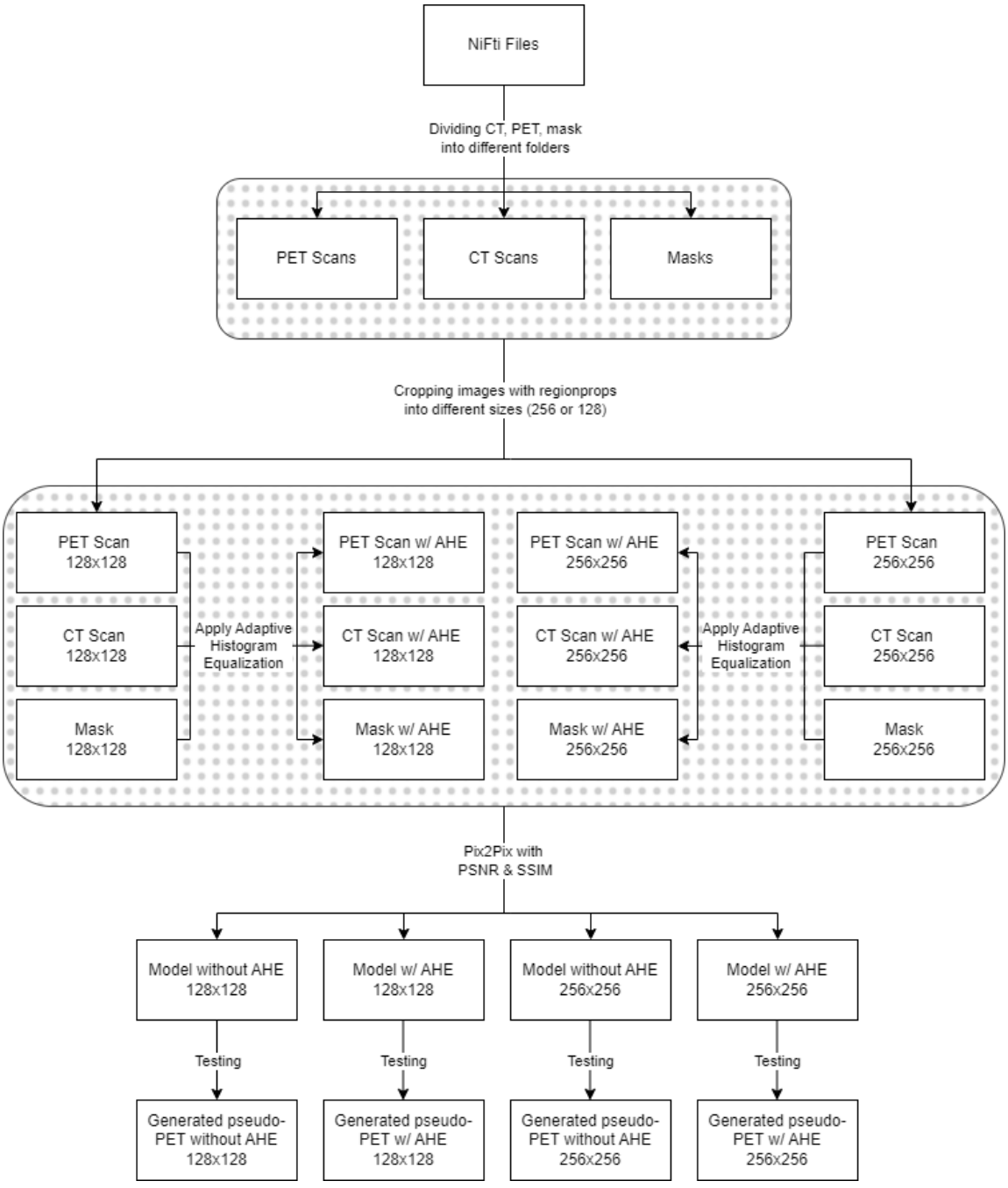

Figure 1: The figure presents a complete view of the project's pipeline.

3.1 *Data*

The dataset utilized for this project is comprised of 1014 whole-body fluorodeoxyglucose (FDG)-PET/CT scans (501 studies of patients with malignant lymphoma, melanoma, and non-small cell lung cancer (NSCLC)) acquired between 2014 and 2018 (Gatidis et al., 2022). All examinations were acquired on a single, state-of-the-art PET/CT scanner. The imaging protocol consisted of a whole-body FDG-PET acquisition and a corresponding diagnostic CT scan. All FDG-avid lesions identified as malignant based on the clinical PET/CT report were manually segmented by radiologists on PET images in a slice-per-slice (3D) manner. These will be referred to as *masks* in the following paragraphs. The creators of the dataset provide the anonymized original DICOM files of all studies as well as the corresponding DICOM segmentation masks (Gatidis et al., 2022).

3.2 *Pre-Processing*

In the realm of medical image analysis, the significance of meticulous data pre-processing cannot be overstated, as it lays the foundation for robust model development and accurate results. This chapter delves into the intricate process of preparing the input data for subsequent analysis, focusing on the Gatidis et al. (2022) dataset stored in NIfTI files (Data Format Working Group, 2003). NIfTI is a very simple, minimalistic format. This format has been widely adopted in neuroimaging research, allowing scientists to mix and match image processing and analysis tools developed by different teams (Li, Morgan, Ashburner, Smith, & Rorden, 2016). The initial step of this pre-processing step involved the extraction and organization of data into distinct folders dedicated to PET scans, CT scans, and corresponding masks. An example of these images can be seen in Figure 2. This segregation was pivotal for streamlined processing and facilitated targeted operations on specific imaging modalities.

Following the organization step, the data underwent a series of transformative procedures to enhance its suitability for subsequent machine learning tasks. The pixel intensity values within the images were normalized through the application of Min-Max normalization (Patro & Sahu, 2015), ensuring a standardized scale across the entire dataset. Subsequently, two resolutions, namely 256 pixels and 128 pixels, were chosen for further analysis. The images were cropped to these dimensions, strategically centered around regions of interest. This cropping process aimed to extract meaningful patches for detailed analysis while managing computational complexity. A key advancement in the pre-processing pipeline was the application of AHE to both the 256 and 128-pixel images. This technique,

tailored to enhance local contrast within images, proved instrumental in revealing subtle features and ensuring that the model could effectively capture intricate patterns present in the medical images (Pizer et al., 1987). The final step in this comprehensive pre-processing pipeline involved the partitioning of the dataset into training, validation, and test sets, maintaining an 80-10-10 ratio. This careful distribution of data aimed to strike a balance between training the model on a sufficiently diverse set of examples and rigorously evaluating its performance on distinct subsets. Crucially, special attention was given to ensuring that individual patients were exclusively represented in only one of these sets, mitigating potential biases.

This chapter provides a comprehensive overview of the intricacies involved in preparing the medical imaging dataset for subsequent analysis. The detailed pre-processing steps, ranging from data extraction to normalization, cropping, and dataset splitting, set the stage for robust model training and evaluation in the pursuit of accurate and reliable outcomes.

An example of the various pre-processing steps performed on the scans from one single patient can be found in Figure 2.

3.2.1 *Normalization*

The images were normalized by using the min-max normalization technique which takes an image as input and performs min-max normalization on its pixel values (Henderi, Wahyuningsih, & Rahwanto, 2021). The function calculates the maximum and minimum pixel values in the input image using `NumPy`'s `np.max` and `np.min` functions, respectively. Subsequently, it scales the pixel values in the image to a normalized range between 0 and 1 using the formula:

$$X_{min\_max} = \frac{(X - min(X)}{(max(X) - min(X)} \tag{1}$$

The normalized image is then returned as the final output. Min-max normalization is commonly employed in image processing before analysis because it scales the pixel values to a standardized range, making it easier to compare and analyze images across different datasets. This normalization technique helps mitigate the impact of varying pixel value ranges in different images, ensuring that the analysis is not biased by the intensity scales of individual pictures. Additionally, it can enhance the performance of machine learning models that rely on consistent input data ranges (Patro & Sahu, 2015).

3.2.2 *Cropping*

The images are then cropped by using a function named `regionprops` designed for processing medical imaging data, specifically CT, PET, and

corresponding masks. The function identifies and analyzes distinct regions within binary segmentation masks derived from label data. For each identified region, the function extracts a region of interest (ROI) centered around a random deviation from the region's centroid (van der Walt et al., 2014). Subsequently, the extracted ROIs, including CT, PET, and mask images, are cropped and saved to designated directories. The rationale behind employing `regionprops` lies in its ability to provide valuable information about individual regions within a segmented image, such as the centroid coordinates, area, and other geometrical properties. In medical image analysis, this information is crucial for precisely isolating and characterizing pathological lesions or structures (Zneit, Alqadi, & Zalata, 2017). By incorporating region-based analysis, the code facilitates the extraction of meaningful image patches centered around identified lesions, enabling subsequent detailed analysis and aiding in tasks such as lesion classification or quantification in the context of medical research and diagnostics.

3.2.3 *Adaptive Histogram Equalization*

To enhance the images by use of AHE, a function designed to perform AHE on an input image was selected. This function utilizes this method with a specified clipping limit of 0.02. The result is a transformed image that has undergone adaptive histogram equalization, and this enhanced image is then returned as the output (van der Walt et al., 2014). Incorporating AHE in image pre-processing is substantiated by its ability to improve the contrast and visibility of local details in an image. Unlike traditional histogram equalization, AHE adapts its enhancement process to smaller, localized regions within the image, mitigating the risk of over-amplifying noise (Pizer et al., 1987). In analytical contexts, particularly in image analysis for research or diagnostic purposes, AHE can be advantageous for revealing subtle features, enhancing the interpretability of images, and aiding in the extraction of meaningful information from diverse image datasets. AHE involves several steps, and there is not a single formula that encapsulates the entire process. However, these are the key steps involved in its computation:

- Divide the Image: the image is divided into non-overlapping or overlapping regions (tiles).
- Compute Local Histograms: for each local region, compute the histogram. The histogram represents the distribution of pixel intensities in that specific region.

- Compute Cumulative Distribution Function (CDF): calculate the cumulative distribution function for each local histogram. The CDF gives the cumulative probability of pixel intensities in the local region.
- Normalize CDF: normalize the CDF to a desired dynamic range (e.g., 0 to 255) to ensure that the pixel values are within a valid intensity range.
- Transform Pixel Values: apply the normalized CDF transformation to the pixel values in the corresponding local region.
- Combine Local Results: combine the results from all the local regions to generate the final enhanced image.

3.2.4 *Train-val-test Split*

Finally, the data was split adhering to the 80-10-10 ratio for training, validation, and test sets. This partitioning ensures a balanced distribution of data for model training, evaluation, and testing. Notably, it was meticulously ensured that patients were exclusively assigned to one of these sets, preventing any bias in the model's performance assessment. The data sets were further categorized into image sizes of 128 and 256, with corresponding quantities specified for each set. For the 128-sized images, the training set comprises 22,906 images, the validation set consists of 2,175 images, and the test set encompasses 4,006 images. Similarly, for the 256-sized images, the training set includes 20,763 images, the validation set contains 2,180 images, and the test set encompasses 1,924 images. This meticulous division and allocation of patients across sets contribute to the robustness and generalizability of the model by ensuring diverse representation and avoiding patient-specific biases during the training and evaluation phases.

3.3 *Pix2Pix GAN*

This study employed the Pix2Pix GAN for image-to-image translation tasks, leveraging the architecture proposed by Isola et al. (2017). Pix2Pix operates on a conditional GAN framework, wherein the generation process is conditioned on input images. In this framework, two key components, the discriminator and the generator, play pivotal roles. The discriminator serves as a critical element in the adversarial training process. Its primary function is to evaluate the authenticity of the generated images by discerning them from the ground truth images in the dataset. Simultaneously, the generator works to produce images that are convincing enough to deceive the discriminator. This adversarial interplay between the discriminator and

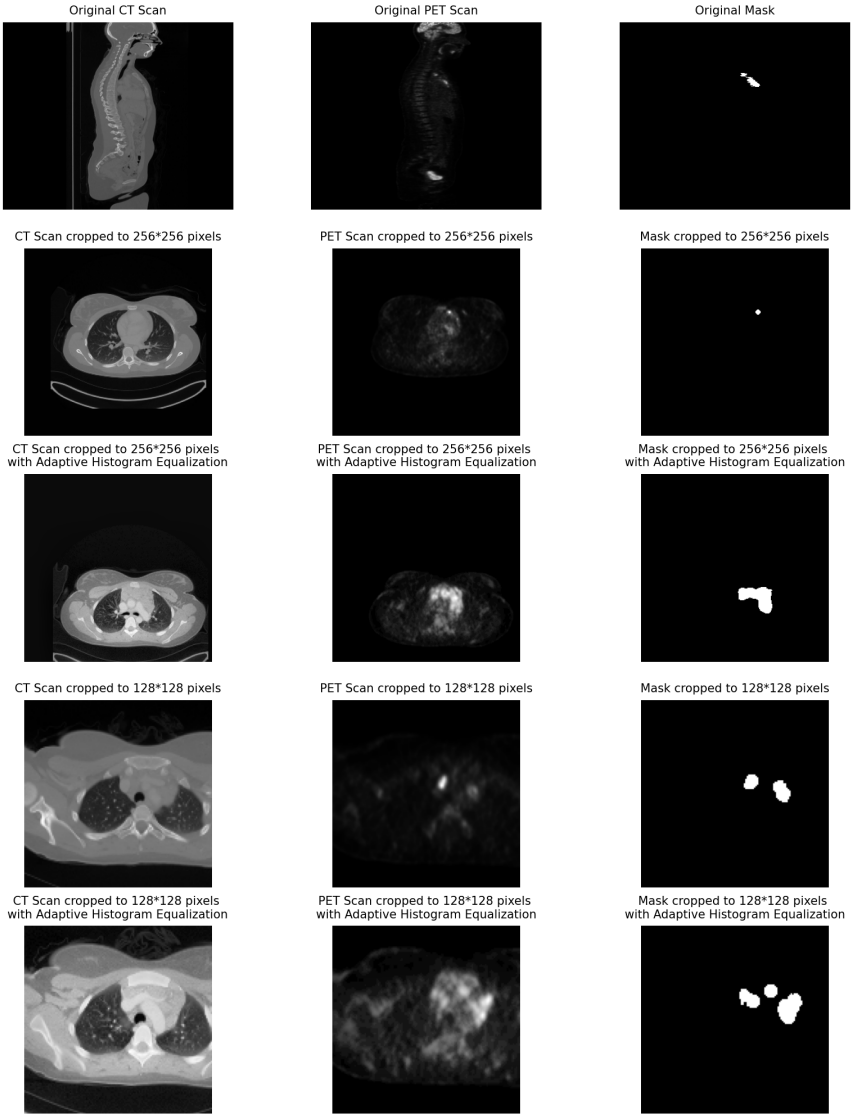

Figure 2: The figure presents a comprehensive overview of image modifications for analysis. The three columns represent distinct image types: the first column showcases CT scans, the second column displays PET scans, and the third column exhibits lesion masks. Each row illustrates a specific stage of image processing. The first row presents the original images extracted from NIfTI files. The second row showcases the images after cropping to a size of 256x256 pixels. The third row displays the same images with AHE enhancement applied. Moving to the fourth row, images are presented after further cropping to a size of 128x128 pixels. Lastly, the fifth row features the corresponding images with AHE enhancement applied. This systematic arrangement facilitates a comprehensive understanding of the sequential modifications applied to the images for analysis.

the generator is instrumental in refining the image translation capabilities of the Pix2Pix model.

Training the model necessitated the use of paired datasets, consisting of corresponding images from source and target domains. This supervised approach enabled the network to learn the mapping between input and output images. The methodology involved meticulous tuning of hyperparameters, with attention to achieving a balance between training efficiency and the avoidance of artifacts in generated images. Limitations such as the need for paired data and computational intensity were acknowledged during the implementation. The Pix2Pix model's effectiveness in producing structured and coherent outputs was evaluated through rigorous experimentation, addressing the specific requirements of image translation tasks within the scope of this study.

3.3.1 *The Discriminator*

The discriminator is designed to assess the authenticity of the generated images produced by a corresponding generator model. The network structure encompasses a series of convolutional layers with increasing depth, from 64 to 512 channels, employing leaky rectified linear unit (LeakyReLU) activations to introduce non-linearity (Nair & Hinton, 2010). Batch normalization is strategically integrated to enhance the stability and efficiency of the model during training (Bjorck, Gomes, Selman, & Weinberger, 2018). The discriminator outputs a binary classification indicating the authenticity of the input image pair. The model is compiled using the Adam optimizer with a learning rate of 0.0002 and a beta parameter of 0.5 while utilizing binary cross-entropy as the loss function (Ruby & Yendapalli, 2020). The overall design of this discriminator adheres to established principles in adversarial training, aiming to discern between real and generated image pairs effectively.

3.3.2 *The Generator*

Comprising encoder and decoder blocks, the generator transforms input images into realistic outputs. The encoder section downsamples the input image through successive convolutional layers, utilizing LeakyReLU for non-linearity (Nair & Hinton, 2010). The encoder gradually reduces the spatial dimensions while increasing the number of filters. The bottleneck layer, without batch normalization and using rectified linear unit (ReLU) activation, captures essential features (Nair & Hinton, 2010). The decoder section then reconstructs the output image by employing transposed convolutional layers, incorporating batch normalization and skip connections to preserve information. The generator is designed to handle images of

dimensions (256, 256, 3) and employs weight initialization with a specified seed for reproducibility. Due to this design flaw, the images with dimensions 128x128 had to be rescaled during loading. The final output is activated using hyperbolic tangent (tanh), ensuring pixel values are within the [-1, 1] range. This architecture adheres to Pix2Pix GAN principles and serves as a pivotal component in conditional image translation tasks.

3.3.3 *Evaluation Metrics: SSIM & PSNR*

In the current project, two prominent image quality evaluation metrics, SSIM and PSNR, were employed to assess the performance of the model. SSIM, introduced by Wang et al. (2004), is designed to measure the structural similarity between the reference and distorted images. It takes into account luminance, contrast, and structure, providing a comprehensive evaluation of perceived image quality. The formula for the SSIM metric is:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_x^2 + c_2)} \tag{2}$$

where:

- μ_x is the average of x;
- μ_y is the average of y;
- σ_x^2 is the variance of x;
- σ_y^2 is the variance of y;
- σ_{xy} is the covariance of x and y;
- $c_1 = (k_1 L)^2, c_2 = (k_2 L)^2$ are the two variables to stabilize the division with weak denominator;
- L is the dynamic range of the pixel-values (typically this is $2^{\#bitsperpixel} - 1$);
- $k_1 = 0.1$ and $k_2 = 0.003$ by default.

On the other hand, PSNR, a well-established metric rooted in signal processing, quantifies the ratio of the maximum possible power of a signal to the power of corrupting noise (Korhonen & You, 2012). PSNR focuses on pixel-wise differences, providing a numerical representation of the distortion between the reference and reconstructed images. Generally, higher PSNR values indicate better image quality. PSNR is expressed

in decibels (dB), and a higher dB value signifies less perceived error or distortion in the reconstructed image. The formula for PSNR is:

$$PSNR = 10log_{10}(\frac{R^2}{\sqrt{MSE}}) \tag{3}$$

where$R$ is the maximal variation in the input image data. If it has an 8-bit unsigned integer data type, R is 255.

Mean Squared Error (MSE) specifies the average difference of the pixels throughout the original ground truth image with the generated image. The higher MSE indicates a greater difference between the original and processed image (Vidya, Veni, & Narayanankutty, 2009).

$$MSE = \frac{\sum_{M,N}[I_1(m,n) - I_2(m,n)]^2}{M,N} \tag{4}$$

where:

- I_1 is the original image;
- I_2 is the generated image;
- m is the height of the image;
- n is the width of the image.

The inclusion of both SSIM and PSNR in the evaluation process allows for a more comprehensive understanding of the model's performance. While SSIM captures perceptual aspects, considering human vision, PSNR provides a traditional and widely used measure of fidelity. Utilizing both metrics ensures a robust evaluation, considering various aspects of image quality and aiding in a more nuanced interpretation of the model's effectiveness (Hore & Ziou, 2010).

3.3.4 *Training the Pix2Pix Model*

The GAN model is utilized to jointly update both the generator and discriminator. The training process involves iterating through epochs and batches, with each batch comprising real and generated samples for both training and validation sets. The discriminator is trained on both real and fake samples, with associated losses. Subsequently, the generator is updated using the GAN model on the real input data, with the overall generator loss and accuracy recorded. The training process includes validation steps, with corresponding metrics evaluated on a separate validation dataset. The training loop also provides periodic summaries of the model's performance, generating plots of real and generated samples for visual inspection. Additionally, plots monitoring the chosen metric and the loss over

time are created, facilitating the monitoring and analysis of the training progress. The loss and metrics plots offer insights into the convergence and performance of the Pix2Pix GAN during the image-to-image translation task.

4 RESULTS

In this pivotal section, we delve into the results of our comprehensive experimentation with four distinct models: those trained with sizes 256 and 128, and their counterparts augmented with AHE. The training process involved the meticulous evaluation of two key metrics, SSIM and PSNR. The approach pursued was rigorous, selecting models based on the lowest loss achieved during validation. Subsequently, the chosen models underwent rigorous testing on an independent test set, where their performance was scrutinized in terms of loss, metric scores, and, most crucially, the visually perceptible quality of generated PET scans. The culmination of this analysis not only sheds light on the efficacy of our trained models but also provides insights into the potential advancements in medical imaging that may result from this groundbreaking research.

4.1 *Training Results*

The presented table (Table 1) encapsulates the outcomes of the training round, providing a comprehensive snapshot of the model's performance across different scenarios and augmentation techniques. The focus here is on the mean PSNR and SSIM metrics for images of varying sizes (128x128 and 256x256) with and without AHE augmentation.

4.1.1 *128x128 Image Size*

Without AHE, the PSNR stands at 52, signifying a high level of fidelity in image reconstruction. The SSIM, at 0.9, indicates substantial structural similarity. These results suggest that the model performs well in preserving image quality for smaller-sized images without the need for additional contrast enhancement.

With AHE, introducing AHE leads to a slight reduction in PSNR to 50, but the SSIM experiences an improvement, reaching 0.94. The AHE augmentation seems to enhance structural similarity, compensating for the minor decrease in PSNR.

4.1.2 *256x256 Image Size*

Without AHE, the PSNR remains consistent at 52, reaffirming the model's ability to maintain high fidelity even for larger-sized images. The SSIM sees an improvement to 0.95, indicating an enhanced preservation of structural details.

With AHE, the PSNR experiences a modest increase to 53.5, suggesting that AHE augmentation contributes to higher fidelity in larger-sized images. The SSIM, remaining at 0.94, indicates that the structural similarity is maintained with the additional contrast enhancement.

	128x128		256x256	
	Without AHE	With AHE	Without AHE	With AHE
PSNR	52	50	52	53.5
SSIM	0.9	0.94	0.95	0.94

Table 1: Validation Results

4.2 *Test Results*

The results from the testing phase are summarized in Table 2, showcasing the model's performance on unseen data. These outcomes provide a critical assessment of the model's generalization capabilities and effectiveness in producing pseudo-PET scans from CT scans under different conditions.

4.2.1 *128x128 Image Size*

Without AHE, the PSNR and SSIM both reach their maximum values (115 and 1, respectively), indicating exceptional fidelity and structural similarity in the generated PET scans. This suggests that the model can accurately simulate PET scans even for smaller-sized input images without the need for AHE.

With AHE, interestingly, the PSNR and SSIM remain constant, emphasizing that the AHE augmentation does not significantly impact the quality or structural similarity in this scenario. The consistently high values reaffirm the model's robustness and proficiency in generating accurate PET-like images.

4.2.2 *256x256 Image Size*

Without AHE, the PSNR reaches 117, showcasing an improvement in fidelity for larger-sized images. However, the SSIM drops to 0.92, indicating

a slight decrease in structural similarity compared to the smaller-sized images without AHE.

With AHE, with AHE augmentation, the PSNR decreases to 114, suggesting a minor trade-off in fidelity, while the SSIM remains at a high value of 1. This implies that the structural similarity is well-preserved even with the addition of AHE for larger-sized images.

In addition to the quantitative evaluation, a visual assessment of the model's performance is crucial. Figures 3, 4, 6, and 5 present images for each tested scenario: the source image (CT scan), the generated PET scan, the expected image (original PET scan from the patient), and the mask of the lesion. This visual evaluation complements the quantitative metrics, providing insights into the model's ability to capture detailed information in the simulated PET scans, including lesion localization.

	128x128		256x256	
	Without AHE	With AHE	Without AHE	With AHE
PSNR	115	115	117	114
SSIM	1	1	0.92	1

Table 2: Test Results

5 DISCUSSION

The training results showcase the model's adaptability across various scenarios, revealing its capability to generate pseudo PET scans from CT scans with consistent high fidelity and structural similarity. The nuanced impact of AHE augmentation on image quality suggests that for smaller-sized images (128x128), the model inherently maintains a high level of accuracy, with AHE contributing subtly to structural enhancement. However, for larger-sized images (256x256), the balance between fidelity and structural similarity sees a nuanced shift with AHE, presenting an interesting area for further investigation.

Moving to the test results, the model's generalization performance is commendable, demonstrating its robustness in producing accurate pseudo-PET scans for previously unseen data. The minor fluctuations in PSNR and SSIM metrics between training and test phases indicate the model's ability to generalize well to different image sizes and augmentation techniques. Notably, the high PSNR values suggest that the model effectively minimizes noise introduced during image processing or compression.

The visual evaluation of model outputs reinforces the quantitative findings, showcasing the model's proficiency in reproducing detailed features,

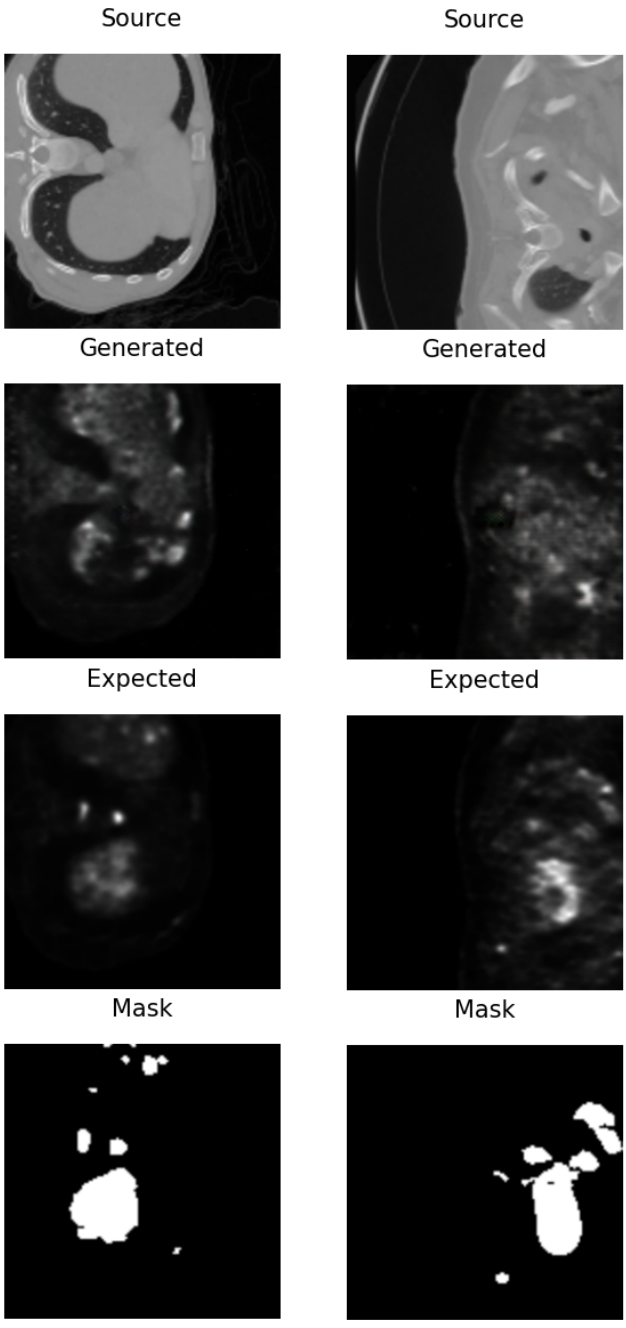

Figure 3: 128x128 Image Size: PSNR (left) and SSIM (right)

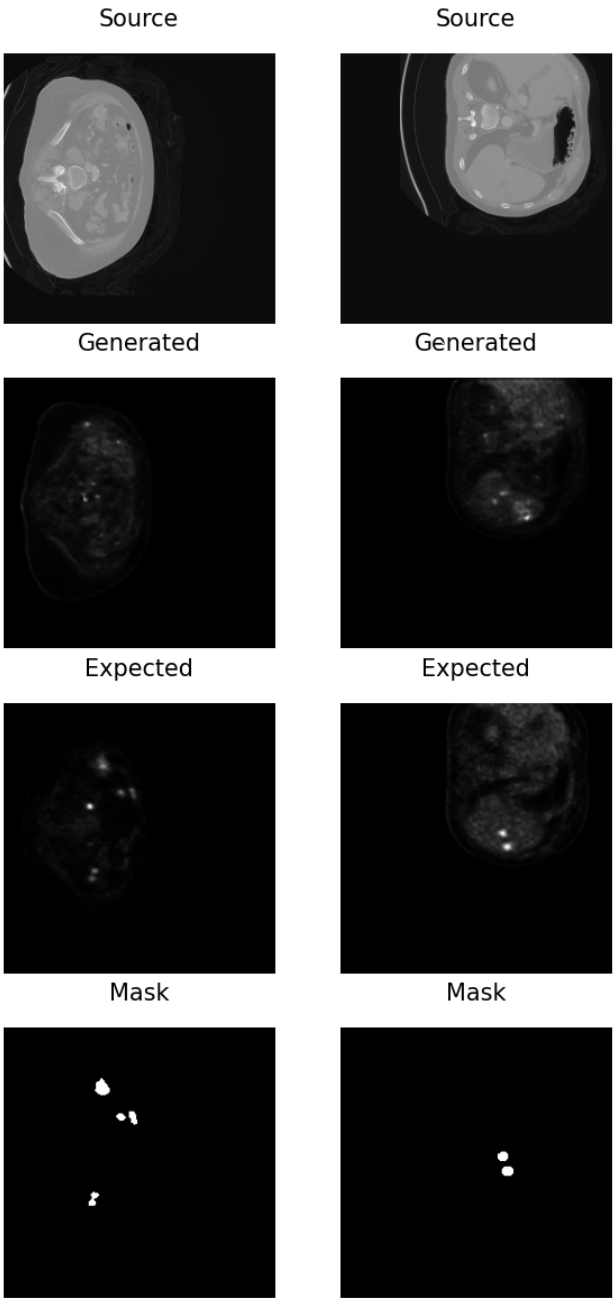

Figure 4: 256x256 Image Size: PSNR (left) and SSIM (right)

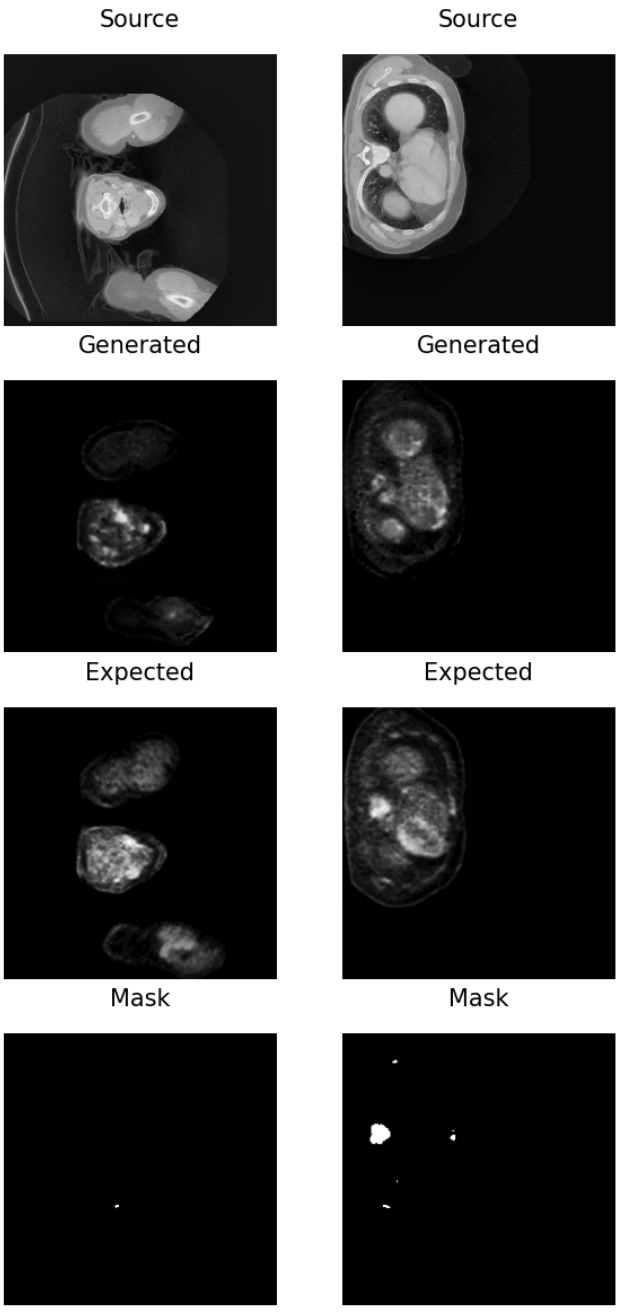

Figure 5: 256x256 Image Size with AHE: PSNR (left) and SSIM (right)

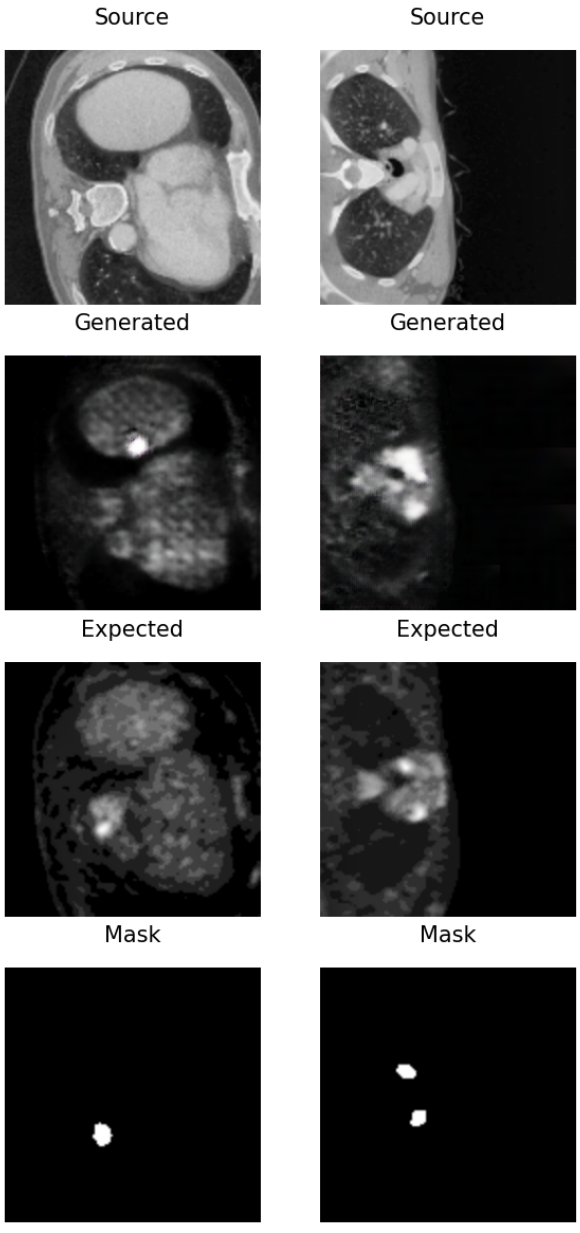

Figure 6: 128x128 Image Size with AHE: PSNR (left) and SSIM (right)

including lesion structures, in the simulated PET scans. The ability to capture clinically relevant information, such as lesion localization, highlights the potential utility of the model in practical medical imaging applications.

In summary, the results suggest that the Pix2Pix model, trained on a diverse dataset and evaluated comprehensively, exhibits promising capabilities in simulating PET scans from CT scans. The discussion points toward avenues for further refinement, such as investigating the impact of AHE on larger-sized images and exploring additional metrics for a more comprehensive evaluation. The model's success in generalization underscores its potential for practical deployment, offering a valuable tool in the synthesis of PET scans for diagnostic purposes.

5.1 *Effect of Cropping*

The difference in results between image sizes 128x128 and 256x256 without AHE reflects the model's performance across varying levels of detail and complexity.

For image size of 128x128, the PSNR remains consistently high at 52 for both scenarios, indicating a high level of fidelity in image reconstruction. This suggests that the model is adept at preserving image quality for smaller-sized images, capturing intricate details with precision. The SSIM stands at 0.9, denoting substantial structural similarity. This aligns with the high PSNR, suggesting that the model maintains both fidelity and structural features effectively for smaller-sized images.

On the other hand, for image size 256x256, the PSNR remains stable at 52, indicating that the model maintains high fidelity even for larger-sized images. This suggests that the model continues to effectively minimize noise and preserve image quality when dealing with higher resolution. Interestingly, the SSIM drops to 0.95 for the larger-sized images. While still indicative of a high level of structural similarity, the decrease suggests a nuanced trade-off between fidelity and structural features in larger images without AHE. This may be attributed to the increased complexity and the need for the model to balance between capturing finer details and maintaining overall image structure.

In summary, without AHE, the Pix2Pix model consistently performs well in terms of PSNR for both image sizes. However, the difference in SSIM for larger-sized images suggests that, without the adaptive contrast enhancement provided by AHE, there is a slight compromise in structural similarity. This finding emphasizes the need for careful consideration of image size and pre-processing techniques in the synthesis of PET scans from CT scans, providing insights for further optimization and refinement of the model.

5.2 *Effect of AHE*

The difference in results between image sizes 128x128 and 256x256 with Adaptive Histogram Equalization (AHE) highlights the impact of this contrast enhancement technique on the model's performance.

For image size 128x128, the PSNR slightly decreases from 52 to 50 when AHE is applied. While this indicates a minor reduction in fidelity, it is worth noting that the PSNR value still remains relatively high, suggesting that the AHE augmentation does not significantly compromise image quality for smaller-sized images. The SSIM, on the other hand, increases from 0.9 to 0.94 with AHE. This improvement indicates that the AHE augmentation contributes to enhanced structural similarity, compensating for the minor decrease in PSNR. The model, with AHE, appears to better preserve the intricate details and overall structure in smaller-sized images.

On the other hand, for image size 256x256 with AHE, the PSNR increases from 52 to 53.5 for larger-sized images. This suggests that AHE plays a more pronounced role in improving fidelity for higher-resolution images. The increase in PSNR indicates that the AHE augmentation positively impacts the preservation of image quality in larger-sized images. The SSIM remains stable at 0.94, indicating that AHE does not compromise the structural similarity for larger-sized images. The model successfully balances fidelity and structural features, providing consistent results with and without AHE in terms of SSIM.

In summary, with AHE, the Pix2Pix model exhibits nuanced differences between image sizes. For smaller-sized images, AHE contributes to improved structural similarity, compensating for a minor decrease in PSNR. For larger-sized images, AHE has a more pronounced positive effect on fidelity without compromising structural similarity. These findings underscore the importance of considering image size and appropriate augmentation techniques, such as AHE, for optimizing the synthesis of PET scans from CT scans.

5.3 *Limitations*

While the AHE augmentation proved beneficial in enhancing structural features, it introduces potential limitations that warrant consideration. AHE, in some instances, can lead to over-amplification of noise, particularly in regions with low contrast. This over-amplification may impact the overall fidelity of the synthesized PET scans, raising concerns about the potential introduction of artifacts in the enhanced images. To mitigate this concern, alternative techniques such as Contrast Limited Adaptive Histogram Equalization (CLAHE) have been developed. CLAHE addresses

the issue by restricting the amplification in regions with low contrast, providing a more controlled enhancement that may be particularly relevant in medical imaging applications (Reza, 2004).

Another noteworthy limitation arises from the prevalence of black pixels in the images, leading to exceptionally high SSIM and PSNR values. As SSIM and PSNR focus on pixel-wise similarities, the dominance of black pixels can skew the accuracy metrics. While these metrics provide valuable quantitative insights, the emphasis on black pixels may not fully reflect the perceptual quality of the synthesized PET scans, especially in scenarios where lesion details are crucial. This emphasizes the need for caution in interpreting extremely high SSIM and PSNR values and underscores the importance of considering additional metrics and visual inspection for a more comprehensive assessment of image quality.

Moreover, perceptual considerations further underscore the limitations of relying solely on PSNR for assessing image quality. PSNR, providing a quantitative measure, may not consistently align with human perception. A high PSNR value does not guarantee visual pleasantness or the indistinguishability of the reconstructed images to the human eye. Therefore, while striving for a higher PSNR is often desirable, it is imperative to complement it with other metrics and visual inspection to ensure that the reconstructed PET scans not only meet quantitative criteria but also align with the intended perceptual quality standards in practical medical applications.

5.4 *Future Work*

The current study has laid a foundation for the synthesis of PET scans from CT scans using Pix2Pix, yet there are avenues for further refinement and enhancement that merit exploration in future research.

5.4.1 *Adoption of Alternative Evaluation Metrics*

While SSIM and PSNR provide valuable quantitative insights, the dominance of black pixels has demonstrated their limitations in certain scenarios. Future work could involve exploring and integrating additional evaluation metrics that offer a more nuanced assessment of image quality such as mean structural similarity index measure (MSSIM) and Fréchet inception distance (FID) (Thanh et al., 2022). Metrics specifically designed to account for lesion details and anatomical structures may provide a more accurate representation of the synthesized PET scans' perceptual quality.

5.4.2 *Region-Specific Assessment*

To address the challenges posed by the abundance of black pixels, a potential solution involves region-specific evaluation. Focusing the assessment on regions of interest, such as cropping images to 64x64 around the lesion, may provide a more targeted evaluation of the model's performance. This approach could offer insights into how well the model captures essential details in specific areas of clinical relevance, facilitating a more granular understanding of its strengths and areas for improvement.

5.4.3 *Implementation of Early Stop*

Introducing mechanisms for early stopping during training could be a valuable addition to the model optimization process. Early stopping, based on criteria such as validation loss or performance plateaus, can prevent overfitting and enhance the model's generalization capabilities (Rice, Wong, & Kolter, 2020). By monitoring the training process and halting it when further improvement is minimal, early stop mechanisms contribute to a more efficient and effective training procedure.

5.4.4 *Incorporation of Advanced Contrast Enhancement Techniques*

Building on the findings related to AHE, future work could explore the integration of more advanced contrast enhancement techniques. Techniques like CLAHE or other state-of-the-art methods may offer improved control over contrast enhancement, addressing potential over-amplification issues and further refining the model's ability to generate high-quality pseudo PET scans.

5.4.5 *Exploration of Multi-Modal Approaches*

Investigating multi-modal approaches that leverage additional information, such as incorporating functional or molecular data, could be a promising avenue. The integration of complementary modalities may enhance the synthesis process, providing the model with richer information to better capture the complexity of PET scans.

In conclusion, these proposed directions for future work aim to refine the model's evaluation, training efficiency, and overall performance. By addressing limitations and exploring innovative techniques, the synthesis of PET scans from CT scans using Pix2Pix can be further optimized, contributing to its robust application in clinical settings.

6 CONCLUSION

In this study, we embarked on a journey to synthesize PET scans from CT scans using the Pix2Pix model, successfully training it on a diverse dataset. Our exploration included the examination of AHE as an augmentation technique, revealing its potential to enhance structural features in the synthesized PET scans. Evaluation metrics, such as SSIM and PSNR, were employed to assess the model's performance. However, observed limitations related to image characteristics, such as potential over-amplification of noise with AHE and challenges posed by black pixels, highlight the need for caution in interpretation. The study recognizes the importance of perceptual considerations and emphasizes the necessity for a holistic approach, encompassing visual inspection and multi-modal exploration to ensure the clinical relevance of the synthesized PET scans. Overall, these findings contribute to ongoing discussions in the field, guiding future endeavors to refine the model, optimize evaluation strategies, and enhance its applicability in clinical settings. This exploration serves as a stepping stone toward advancements in medical imaging, with the ultimate goal of improving diagnostic accuracy and patient care.

7 PERSONAL DEVELOPMENT

Embarking on this project has been a transformative journey that not only expanded my horizons but also catalyzed significant growth, both personally and professionally. This chapter serves as a reflection on the evolution I underwent during the course of this remarkable endeavor, delving into the development of crucial skills and the discovery of my true passion.

Working collaboratively in a team setting was a cornerstone of this project, and through this experience, I honed my ability to function seamlessly within a group. Communication, collaboration, and compromise became second nature, fostering an environment where diverse perspectives merged into a harmonious symphony of ideas. I discovered the strength in unity and the importance of leveraging individual strengths for collective success. The teamwork and cooperation during this project were exceptional, creating a very pleasant and productive atmosphere. Although there were moments of stress, the challenges were navigated successfully, further strengthening my adaptability and resilience. Adaptability and resilience were also essential personal skills that I cultivated. The project's dynamic nature demanded the ability to navigate uncertainties, adapt to changing circumstances, and learn from setbacks. Each challenge became an opportunity for personal development, teaching me to approach difficul-

ties with a positive mindset and an unwavering determination to overcome obstacles.

On the professional front, this project acted as a catalyst for the enhancement of my technical skills, particularly in the realm of server management. Navigating the intricacies of server architecture, troubleshooting, and optimizing performance became integral components of my skill set. I learned how to work more efficiently with online servers, gained proficiency in properly running code on the server, and further developed my skills in image analysis. The hands-on experience allowed me to develop a nuanced understanding of server operations, empowering me to tackle complex technical issues with confidence. In the academic arena, my foray into the image analysis field proved to be intellectually enriching. The exploration of advanced techniques and methodologies in this domain broadened my understanding and proficiency in handling intricate image datasets. I found joy in the meticulous process of extracting meaningful insights from visual information, further fueling my passion for the analytical aspects of my work. Throughout this project, my involvement in the medical field emerged as a revelation. As I delved deeper into image analysis, I realized a profound connection with the applications in the medical domain. The success of the research not only holds the potential to revolutionize medical imaging practices but also aligns with my newfound passion for contributing to advancements that can positively impact lives.

In conclusion, the journey of project development has been more than just a professional milestone; it has been a transformative chapter in my life. The amalgamation of personal and professional growth has shaped me into a more resilient, adaptable, and skilled individual. As I move forward, fueled by a newfound passion for the medical field, I am eager to apply these lessons and skills to future endeavors, contributing meaningfully to both my professional pursuits and the greater good.

REFERENCES

Basu, S., Hess, S., Braad, P., Olsen, B., Inglev, S., & Høilund-Carlsen, P. (2014). The basic principles of FDG-PET/CT imaging. *PET clinics*, *9*(4), 355-370. doi: 10.1016/j.cpet.2014.07.006

Ben-Cohen, A., Klang, E., Raskin, S. P., Soffer, S., Ben-Haim, S., Konen, E., ... Greenspan, H. (2019). Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection. *Engineering Applications of Artificial Intelligence*, *78*, 186-194.

Bi, L., Kim, J., Kumar, A., Feng, D., & Fulham, M. (2017). Synthesis of positron emission tomography (PET) images via multi-channel generative adversarial networks (GANs). In *Molecular imaging, re-*

construction and analysis of moving body organs, and stroke imaging and treatment: Fifth international workshop, CMMI 2017, second international workshop, RAMBO 2017, and first international workshop, SWITCH 2017, held in conjunction with MICCAI 2017 (Vol. 10555, p. 43-51).

Bjorck, N., Gomes, C. P., Selman, B., & Weinberger, K. Q. (2018). Understanding batch normalization. *Advances in neural information processing systems*, *31*.

Chandrashekar, A., Handa, A., Ward, J., Grau, V., & Lee, R. (2022). A deep learning pipeline to simulate fluorodeoxyglucose (FDG) uptake in head and neck cancers using non-contrast CT images without the administration of radioactive tracer. *Insights into Imaging*, *55*(8), 1235-1241. doi: 10.1186/s13244-022-01161-3

Cormack, A. (1980). Early two-dimensional reconstruction (ct scanning) and recent topics stemming from it. *Journal of computer assisted tomography*, *4*(3), 658-664. doi: 10.1097/00004728-198010000-00016

Data Format Working Group. (2003). *NIfTI documentation.* Retrieved from http://nifti.nimh.nih.gov/nifti-1/documentation

Gatidis, S., Hepp, T., Früh, M., La Fougère, C., Nikolaou, K., Pfannenberg, C., ... Rubin, D. (2022). A whole-body FDG-PET/CT dataset with manually annotated tumor lesions. *Scientific Data*, *9*(601), 1-7. doi: 10.1038/s41597-022-01718-3

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Weinberger (Eds.), *Advances in neural information processing systems 27 (NIPS 2014)* (Vol. 27, p. 2672-2680).

Han, X. (2017). MR-based synthetic ct generation using a deep convolutional neural network method. *Medical physics*, *44*(4), 1408-1419.

Henderi, H., Wahyuningsih, T., & Rahwanto, E. (2021). Comparison of min-max normalization and z-score normalization in the k-nearest neighbor (knn) algorithm to test the accuracy of types of breastcancer. *International Journal of Informatics and Information SYstem*, *4*(1), 133-20. doi: 10.47738/ijiis.v4i1.73

Hore, A., & Ziou, D. (2010). Image quality metrics: PSNR vs. SSIM. In *20th international conference on pattern recognition* (p. 2366-2369).

Hounsfield, G. (1980). Computed medical imaging. *Journal of computer assisted tomography*, *4*(5), 665-674. doi: 10.1097/00004728-198010000-00017

Isola, P., Zhu, J., Zhou, T., & Efros, A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the ieee conference on computer vision and pattern recognition (cvpr)* (p. 5967-5976). doi: 10.1109/CVPR.2017.632

Jaroff, L. (2000). A winning combination. *Time*, *156*(23), 72-74.

Jones, T., & Townsend, D. (2017). History and future technical innovation in positron emission tomography. *Journal of Medical Imaging*, *4*(1), 011013-011013. doi: 10.1117/1.JMI.4.1.011013

Korhonen, J., & You, J. (2012). Peak signal-to-noise ratio revisited: Is simple beautiful? In *4th international workshop on quality of multimedia experience* (p. 37-38). doi: 10.1109/QoMEX.2012.6263880

Krause, B., Schwarzenböck, S., & Souvatzoglou, M. (2013). Molecular imaging in oncology. In (Vol. 187, chap. FDG PET and PET/CT). Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-10853-2_12

Kumar, V., Abbas, A., Fausto, N., & Aster, J. (2009). *Robbins cotran pathologic basis of disease* (8th ed.). Elsevier Health Sciences.

Kwee, T. C., & Kwee, R. M. (2009). Combined FDG-PET/CT for the detection of unknown primary tumors: systematic review and meta-analysis. *European radiology*, *19*, 731-744. doi: 10.1007/s00330-008-1194-4

Lee, G., Lee, H. Y., Park, H., Schiebler, M. L., van Beek, E. J., Ohno, Y., ... Leung, A. (2017). Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: State of the art. *European journal of radiology*, *86*, 297-307. doi: 10.1016/j.ejrad.2016.09.005

Li, X., Morgan, P., Ashburner, J., Smith, J., & Rorden, C. (2016). The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *Journal of Neuroscience Methods*, *264*, 47-56. doi: 10.1016/j.jneumeth.2016.03.001

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (p. 807-814).

Nie, D., Cao, X., Gao, Y., Wang, L., & Shen, D. (2016). Estimating CT image from MRI data using 3d fully convolutional networks. In *Deep learning and data labeling for medical applications: First international workshop, LABELS 2016, and second international workshop, DLMIA 2016, held in conjunction with MICCAI 2016* (p. 170-178).

Nyakale, N., Lockhat, Z., & Sathekge, M. (2015). Nuclear medicine-induced allergic reactions. *Current Allergy & Clinical Immunology*, *28*(1), 10-17.

Patro, S., & Sahu, K. (2015). Normalization: A preprocessing stage. *IARJSET*. doi: 10.17148/IARJSET.2015.2305

Phelps, M., Hoffman, E., Coleman, R., Welch, M., Raichle, M., Weiss, E., ... Ter-Pogossian, M. (1976). Tomographic images of blood pool and perfusion in brain and heart. *Journal of Nuclear Medicine*, *17*(7), 603-612.

Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A.,

Greer, T., ... Zuiderveld, K. (1987). Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, *39*(3), 355-368.

Reza, A. M. (2004). Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *Journal of VLSI signal processing systems for signal, image and video technology*, *38*, 35–44.

Rice, L., Wong, E., & Kolter, Z. (2020). Overfitting in adversarially robust deep learning. In *International conference on machine learning* (pp. 8093–8104).

Ruby, U., & Yendapalli, V. (2020). Binary cross entropy with deep learning technique for image classification. *Int. J. Adv. Trends Comput. Sci. Eng*, *9*(10).

Saif, M. W., Tzannou, I., Makrilia, N., & Syrigos, K. (2010). Role and cost effectiveness of PET/CT in management of patients with cancer. *The Yale journal of biology and medicine*, *83*(2), 53-65.

Shukla, S., Van Gool, L., & Timofte, R. (2019). Extremely weak supervised image-to-image translation for semantic segmentation. In *2019 IEEE/CVF international conference on computer vision workshop (ICCVW)* (p. 3368-3377). doi: 10.1109/ICCVW.2019.00419

Thanh, T. N., Le Pham Van, L., Ngoc Anh, T. P., Chi, T. N., Minh, D. C., & Quoc, L. T. (2022). Convolution-mix-transformer generator model to synthesize PET images from CT scans. In *2022 international conference on digital image computing: Techniques and applications (DICTA)* (p. 1-6). doi: 10.1109/DICTA56598.2022.10034560

Timmers, T., Keijsers, M., Kremer, J., Janssen, L., & Smeenk, J. (2021). Supporting women undergoing ivf treatment with timely patient information through an app: Randomized controlled trial. *JMIR Mhealth Uhealth*, *9*(8), e28104. doi: 10.2196/28104

Tixier, F., Hatt, M., Valla, C., Fleury, V., Lamour, C., Ezzouhri, S., ... Le Rest, C. C. (2014). Visual versus quantitative assessment of intratumor F-FDG PET uptake heterogeneity: Prognostic value in non–small cell lung cancer. *Journal of Nuclear Medicine*, *55*(8), 1235-1241. doi: 10.2967/jnumed.113.133389

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., ... the scikit-image contributors (2014). scikit-image: image processing in python. *PeerJ*, *2*, e453. doi: 10.7717/peerj.453

Vidya, P., Veni, S., & Narayanankutty, K. (2009). Performance analysis of edge detection methods on hexagonal sampling grid. *International Journal of Electronic Engineering Research*, *1*(4), 313-328.

Wang, Z., Bovik, A., Sheikh, H., & Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, *13*(4), 600-612. doi: 10.1109/TIP.2003.819861

Ward, J. (2019). *The student's guide to cognitive neuroscience*. Routledge. doi: 10.4324/9781315742397

Zneit, R., Alqadi, Z., & Zalata, D. M. A. (2017). Procedural analysis of rgb color image objects. *International Journal of Computer Science and Mobile Computing*, *6*(1), 197-204.