ARTIFICIAL COUNT ENHANCEMENT OF LUNG SCINTIGRAPHIC IMAGES USING DEEP LEARNING TECHNIQUES

SIRAJ-EDDIN GHASSEL

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Master of Computer Science

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Abstract

This thesis examines the use of artificial intelligence (AI) for enhancing photon counts in ventilation/perfusion (V/Q) scans. V/Q scintigraphy is essential in the nuclear medicine field for diagnosing pulmonary embolism (PE), a severe and potentially life-threatening condition. Traditional V/Q scanning methods are often lengthy, which not only cause discomfort for patients, and thus possibly affect the quality of the images obtained, but also has contributed to their declining use and the uprise of computed tomography pulmonary angiography (CTPA). This thesis aims to address these issues by proposing AI-based techniques that enhance low-count images in V/Q scans, with the aim of reducing scan times while maintaining accurate diagnoses.

The research began with a systematic review to assess the role of AI in V/Q scans. This review identified several promising areas for further research, including image enhancement, artifact removal, and the creation of pseudo-planar images from single-photon emission computed tomography (SPECT). This latter aspect is relevant since there has been research to support the transition from traditional planar imaging towards SPECT; thus, alternatives are needed that help physicians adapt.

A critical part of the research was evaluating the impact of various image resizing techniques on the noise characteristics inherent in V/Q images. This involved a comparative analysis of standard upsampling and downsampling methods, such as linear interpolation, against techniques designed to preserve Poisson counting statistics. These included linear interpolation followed by Poisson resampling for upsampling and sliding window summation for downsampling. Image quality was assessed using the structured similarity index (SSIM) and the logarithm of mean squared error (MSE), to compare resized images with those at the target resolution. The findings revealed that upsampling with Poisson resampling after linear interpolation yielded images consistent with correct photon count properties at the target resolution. In downsampling, while linear interpolation and sliding window summation were both effective at a 2x reduction, the latter method produced realistic images at a 4x reduction.

Furthermore, the thesis showcases the development, training, and validation of a deep learning model tailored for count enhancement using scintigraphic images. This model successfully created diagnostic quality images from simulated low-count images, which were derived from diagnostic-quality images at 10% counts, with full-count images serving as the ground truth. The effectiveness of various loss functions was determined through both qualitative and quantitative analyses. It was observed that a combination of L_1, perceptual, and adversarial losses yielded images most comparable to the ground truth. These AI-enhanced images underwent clinical evaluation by both experienced and novice V/Q scan readers. The assessment was based on three categories: low, moderate, and high similarity. The general consensus among the readers indicated a moderate to high similarity, suggesting that the AI-enhanced images retained essential diagnostic features.

Table of Contents

List of Figures

List of Tables

Acronyms

AI	Artificial Intelligence
ANNs	Artificial Neural Networks
ANT	Anterior
CAD	Computer-Aided Diagnosis
cGAN	Conditional Generative Adversarial Network
CNN	Convolutional Neural Network
COPD	Chronic Obstructive Pulmonary Disease
CT	Computed Tomography
CTEPH	Chronic Thromboembolic Pulmonary Hypertension
CTPA	Computed Tomography Pulmonary Angiography
CXR	Chest X-Ray
DOACs	Direct-Acting Oral Anticoagulants
DTPA	Diethylenetriaminepentaacetic Acid
DVT	Deep Vein Thrombosis
EANM	European Association of Nuclear Medicine
FDG	Fluorodeoxyglucose
FDR	False Discovery Rate
GANs	Generative Adversarial Networks
IQR	Interquartile Range
LAD	Least Absolute Deviations
LAO	Left Anterior Oblique
LPO	Left Posterior Oblique
MAA	Macroaggregated Albumin
MBq	Mega Becquerel
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
PA	Pulmonary Angiography
PE	Pulmonary Embolism
PESI	Pulmonary Embolism Severity Index
PET	Positron Emission Tomography
PIOPED	Prospective Investigation of Pulmonary Embolism Diagnosis

PMTs	Photomultiplier Tubes
POST	Posterior
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RAO	Right Anterior Oblique
RELIANCE	Recommendations for Evaluation of AI for Nuclear Medicine
ReLU	Rectified Linear Unit
RPO	Right Posterior Oblique
SPECT	Single Photon Emission Computed Tomography
SUV	Standardized Uptake Value
VTE	Venous Thromboembolism

Chapter 1: Introduction

1.1 Problem Statement

In nuclear medicine, ventilation/perfusion (V/Q) scintigraphy is a critical diagnostic tool for the detection of pulmonary embolism (PE), a potentially life-threatening condition. Despite its clinical importance, a fundamental challenge persists in the form of prolonged acquisition times for scintigraphic images. Such extended durations often lead to patient discomfort and, critically, can degrade image quality due to patient movement. Importantly, long acquisition times also constitute one of the major limitations to V/Q scans with respect to other modalities for diagnosing PE, such as computed tomography pulmonary angiography (CTPA). However, shortening image acquisition times, while desirable, can result in low count scintigraphic images, where the clarity and reliability of the images are inherently compromised. Consequently, there is a pressing need for innovative solutions to enhance these low count images to increase diagnostic precision. In recent years, nuclear medicine studies have been transitioning from two-dimensional planar imaging to three-dimensional SPECT imaging by rotating the camera around the patient and reconstructing a 3D tomographic volume of ventilation and perfusion from the acquired projection images. The development of techniques for generating pseudo-planar images from raw SPECT projection data is desirable as a fallback in the presence of SPECT image artifacts such as those associated with patient motion. This advancement would not only assist less experienced physicians in interpreting complex SPECT data but also improve the overall efficiency and accuracy of diagnoses in nuclear medicine, particularly for V/Q scans.

1.2 Thesis Statement

This thesis explores the application of deep learning techniques to enhance V/Q scintigraphic images, specifically focusing on augmenting the quality of low-count images and generating pseudo-planar images from raw SPECT projection data. The primary goal is to leverage deep learning models to enhance the clarity and diagnostic value of scintigraphic images characterized by low photon counts. Raw SPECT data typically contains lower count and lower grid size compared to planar acquisition data. Thus, a critical preprocessing step in preparing these images for neural network training involves resizing them to a uniform grid size. This resizing is done using Poisson resampling correction for upsampling and sliding window summation for downsampling. By developing and implementing innovative deep learning approaches, this research aims to significantly improve the interpretation of V/Q scans, thereby enhancing the diagnostic process in nuclear medicine.

1.3 Main Contributions

This research is situated at the intersection of computer science and nuclear medicine, aiming to innovate in several key areas:

1- The Role of Artificial Intelligence (AI) in V/Q Scintigraphy: This systematic review contextualizes and critiques the current standing of AI applications in V/Q scintigraphy. I was the 2nd author on this publication. My contributions included carrying out a jointly reviewed literature search, undertaking data extraction, and co-writing portions of the manuscript.

> Jabbarpour A, **Ghassel S**, Lang J, Klein R, Moulton E. "The Past, Present, and Future Role of Artificial Intelligence in Ventilation/Perfusion Scintigraphy: A Systematic Review." *Seminars in Nuclear Medicine*, vol. 53, no. 6, 2023, pp. 752–65, https://doi.org/10.1053/j.semnuclmed.2023.03.002.

2- Techniques for Resampling Scintigraphic Images: We developed robust methodologies for resizing scintigraphy images, in a manner that preserves photon counting statistics and associated noise properties. These methods provide a robust tool for practitioners, ensuring that the integrity and quality of resized images are maintained, a crucial aspect for accurate image analysis, particularly in the context of machine learning algorithms where image size standardization is common practice. I was the first author of this publication. This work has been submitted to Physics in Medicine and Biology and a provisional patent has been submitted.

> **Ghassel S**, Jabbarpour A, Lang J, Moulton E, Klein R. "The effect of resizing on the natural appearance of scintigraphic images: an image similarity analysis," 2023. *(Submitted)*

> Eric Moulton (Jubilant), **Siraj Ghassel** (Ottawa Hospital Research Institute), Jochen Lang (University of Ottawa), Ran Klein (Ottawa Hospital Research Institute), Indranil Nandi (Jubilant); Jubilant-DraxImage, Methods and Techniques of Resizing of Scintigraphy Images. Provisional Patent Application: United States of America Patent Office, 63/584,367.

3- Deep-Learning Model for Count Enhancement: At the heart of this thesis is the development and validation of a deep-learning model aimed at augmenting counts in scintigraphic images. The following abstracts have been presented at the annual conference of the Society of Nuclear medicine and Molecular Imaging (SNMMI); University of Ottawa, Faculty of Medicine research day; and University of Ottawa, Faculty of Engineering poster day.

> **Ghassel S**, Jabbarpour A, Lang J, Moulton E, Klein R. "Count Enhancement of Perfusion Images in Lung Scintigraphy using Artificial Intelligence," *J. Nucl. Med.*, vol. 64, no. supplement 1, p. P1240, Jun.2023 *(Abstract)*

A manuscript is in preparation to develop a deep learning model for the count enhancement of scintigraphic images.

1.4 Thesis Structure

This thesis represents an exploration into the application of artificial intelligence in V/Q scintigraphy. The research initiated with approval from the Ottawa Hospital Research Ethics Board, enabling the collation of patient data from The Ottawa Hospital. In adherence to ethical standards, all personally identifiable information were anonymized, ensuring patient confidentiality.

The thesis begins with a background chapter that lays the groundwork for understanding the context and technical aspects of the research. This chapter explains the fundamental principles of nuclear medicine, the intricacies of V/Q scintigraphy, and the role of AI in enhancing medical imaging. The third chapter presents a systematic review on the current state of AI in V/Q scans. This review serves a dual purpose: (1) it confirms the unmet need for accelerating V/Q scan image acquisition and (2) establishes the relevance and potential of AI in this domain.

In response to the identified need for accelerating image acquisition and the potential of AI, the fourth chapter focuses on image spatial resampling techniques, an essential preprocessing step for neural network training. This involves standardizing the spatial grid sizes of image inputs to ensure count and noise characteristics are preserved in the resizing process. The chapter presents a novel method developed to resize images while preserving the inherent noise characteristics of scintigraphic images.

Chapter five forms the core of the thesis and builds on the previous chapters. It details the formulation, training, and validation of a deep-learning model aimed at count enhancement of scintigraphic images. This chapter represents the practical application of the concepts and methodologies discussed earlier, showcasing the actual development and testing of the AI model.

The sixth and final chapter synthesizes the key findings of the research, examines the limitations encountered and proposes future research directions, underlining areas where further advancements can be made.

Chapter 2: Background and Related Works

2.1 Pulmonary Embolism

Lungs perform the critical function of exchanging gasses between the ambient air and the blood that delivers these gasses to and from all the body organs. To perform this gas exchange, blood is collected from the body via the venous system. The smaller venous tributaries come together to form the largest veins (the superior and inferior vena cava) which terminate in the right chambers of the heart. The right heart pumps the blood via the pulmonary arteries to the two lungs. This blood is low in oxygen and high in carbon-dioxide. In the lungs, the pulmonary veins bifurcate to smaller and smaller vessels terminating in the capillaries.

These capillaries encase small air sacks called alveoli producing a large surface area where gas exchange can readily occur. Concentrated carbon dioxide in the blood is diffused to the air and concentrated oxygen in the air is diffused to the blood (where is binds to red blood cells using the hemoglobin molecule).

The oxygenated blood flows through the capillaries which gradually merge to form the pulmonary veins. These pass blood to the left chambers of the heart that pumps the blood to the aorta and from there to the rest of the body. In doing so, the closed-loop circulatory system is complete with the blood delivering essential oxygen to the body and carrying away its waste carbon dioxide.

The lungs are enclosed in the thoracic cavity, between the rib cage and diaphragm. As the diaphragm, a large muscle above the liver and stomach, contracts and relaxes it expands and contracts the lungs respectively. Correspondingly, the lungs inhale and exhale ambient air via the bronchi, tracheal tube, mouth, and nose to exchange the air in the alveoli with the atmosphere around us.

The typical human has two lungs. The right lung is comprised of three main lobes, while the left has two (Figure 2.1). These are further divisible into segments and subsegments based on the branching of the bronchi and blood vessels that make up the lung. With the gradual decrease of blood vessels to the lungs and the eventual restriction by the capillaries (6 μm diameter), even cells in the blood can struggle to pass through them. In fact, some of the larger cells in the blood (as large as 8 μm in diameter) must deform to pass through these capillaries. Larger and non-deformable objects can become lodged in these blood vessels leading to blockages. These blockages, depending on their location, can restrict perfusion to a lung subsegment, segment, or in extreme cases even entire lobes.

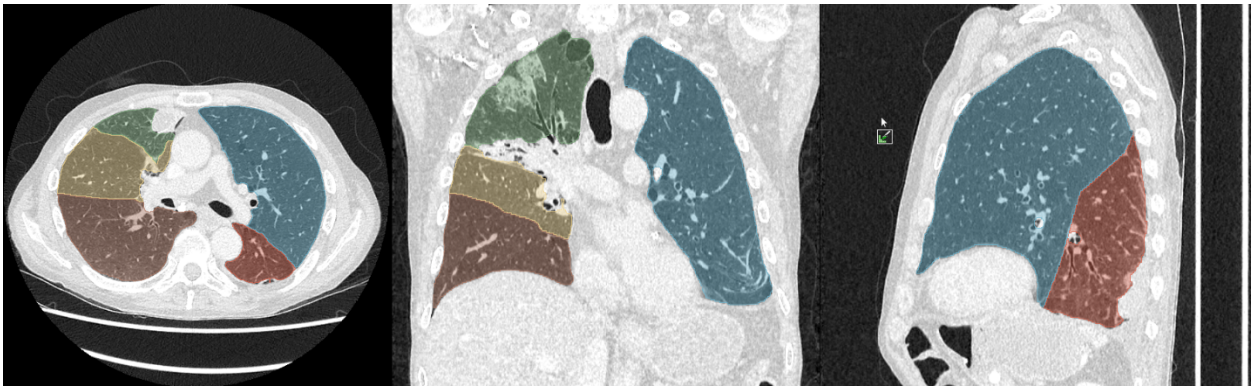

Figure 2.1- Segmented CT lungs. From left to right, the lung orientations are axial, anterior, and left side with each lung segmented into their respective lobes.

2.1.1 Understanding Pulmonary Embolism

Pulmonary embolism is a life-threatening medical condition that requires immediate attention and intervention [1]. It occurs when one or more arteries in the lungs become blocked by blood clots. This obstruction disrupts the normal blood flow to the lungs and can result in various complications, including chest pain, shortness of breath, and, in severe cases, sudden death [1].

The clinical presentation of PE can vary from one patient to the other, depending on factors such as clot size, lung involvement, and overall health. PE is part of a broader spectrum of conditions known as venous thromboembolism (VTE) [2], which also encompasses deep vein thrombosis (DVT). DVT involves the formation of blood clots in the deep veins, often preceding the occurrence of PE. Typically, PE arises when these DVT clots dislodge and travel through the bloodstream to the lungs.

Diagnosing PE can be challenging due to its variable and non-specific symptoms. A high level of suspicion, thorough assessment of risk factors, and appropriate use of diagnostic tests are crucial. These tests may include D-Dimer blood tests [3], chest x-rays [4], (CTPA) [5], and (V/Q) scans [6].

The severity of PE and the potential for life-threatening complications, such as right heart failure and recurrent embolism, underscore the urgent need for treatment. Early interventions aim to prevent clot growth, dissolve existing clots through thrombolytic therapy, and restore normal blood flow in the lungs [7]. Anticoagulant medication is commonly prescribed, while surgical procedures may be necessary in severe cases [2].

2.1.2 Physiology of PE

DVT consists of three factors: blood stasis, endothelial damage, and hypercoagulability [8]. Any disruption to these elements can disrupt the balance of the blood's natural anticoagulant properties and promote clot formation.

Once formed, the DVT clot may either remain in place, adhering to the vein wall, or become an embolus that travels through the bloodstream. Eventually, the embolus may reach a point where the vessel diameter becomes narrower than the embolus itself. In the case of PE, these emboli

traverse the right heart chambers and lodge in the pulmonary arteries, obstructing the blood flow to the lungs.

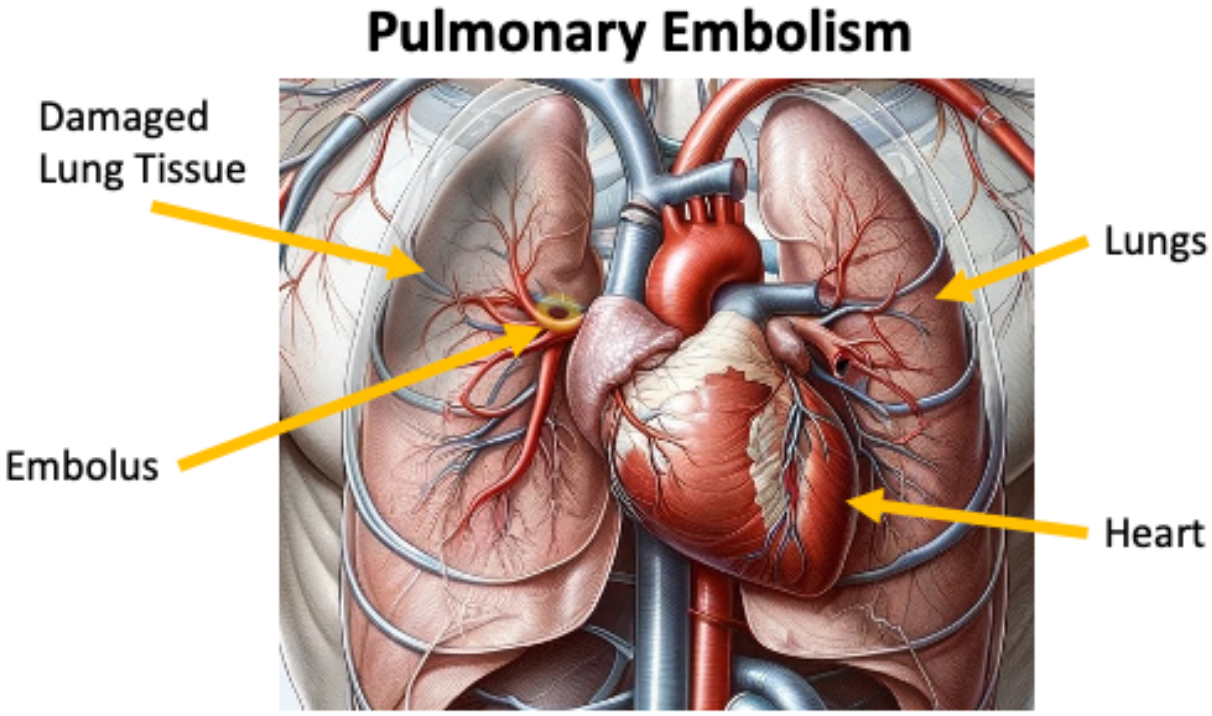

Figure 2.2- Embolus forming in the right lung and lodging in the pulmonary arteries.Image generated from OpenAI DALLE-2.

The consequences of this blockage are immediate and varied. Firstly, the impaired blood flow creates a ventilation-perfusion (V/Q) mismatch in the lungs [1]. Although the affected areas continue to receive air, the compromised blood supply impairs gas exchange, resulting in reduced oxygen levels in the blood, known as hypoxemia [9].

PE can also lead to chronic complications such as post-embolic syndrome and chronic thromboembolic pulmonary hypertension (CTEPH), both of which profoundly affect patients' quality of life [10]. Post-embolic syndrome, characterized by persistent shortness of breath, reduced exercise capacity, and leg discomfort, can manifest in up to half of PE patients [7]. On the other hand, although less common, CTEPH represents a severe form of pulmonary hypertension resulting from unresolved thromboembolic occlusion of pulmonary arteries.

The obstruction of pulmonary arteries may cause increased pressure on the right side of the heart as it tries to pump blood against the resistance [11]. Over time, this strain can lead to serious complications such as right heart failure or pulmonary hypertension, particularly in cases of untreated or extensive PE.

2.1.3 The Prevalence and Impact of PE

PE is far from being an uncommon medical event that significantly contributes to global morbidity and mortality [12]. Estimates indicate that developed countries experience approximately 60-70 cases of PE per 100,000 population annually [13], although the actual incidence might be higher due to underdiagnosis or misdiagnosis [14]. This prevalence makes PE a pressing concern for healthcare providers, necessitating continuous advancements in prevention, diagnosis, and management.

Mortality rates associated with PE are alarmingly high, particularly when the condition remains untreated or is not promptly diagnosed [12]. Acute PE ranks as the third leading cause of cardiovascular death, following heart disease and stroke, further underscoring the urgency of early detection [15].

The economic impact of PE should not be underestimated. It contributes to prolonged hospital stays, readmissions, and long-term treatments, resulting in substantial healthcare costs [16]. Indirect costs also arise from lost productivity due to PE-related disability or death. Therefore, implementing efficient diagnostic and management strategies for PE not only saves lives but also demonstrates economic prudence.

2.1.4 Patient Management: Strategies and Approaches to Dealing with PE

Over the years, considerable progress has been made to improve the diagnosis and treatment of PE, with patient outcomes seeing significant improvement. One such key development, as highlighted by Wells *et al.* in the article "Diagnosis of Venous Thromboembolism: 20 Years of Progress," [2] is the implementation of risk stratification to guide management strategies for patients with PE.

The first step in the management of a patient suspected of PE involves risk stratification, where patients are classified as high-risk, intermediate-risk, or low-risk based on their clinical condition and diagnostic test results. High-risk patients often exhibit hemodynamic instability or cardiac arrest, while low-risk patients demonstrate normal blood pressure and have a low score on the Pulmonary Embolism Severity Index (PESI) [17]. Those falling in between are categorized as intermediate-risk patients.

Immediate therapeutic intervention is required for high-risk PE patients. The preferred strategy typically involves systemic thrombolysis, where medication is given to dissolve the blood clot. However, in some cases, surgical or catheter-directed thrombolysis may be considered. These are more invasive procedures and are typically reserved for when the patient does not respond to systemic thrombolysis.

The management of intermediate-risk and low-risk patients involves the use of anticoagulation therapy. The options include Vitamin K antagonists, direct oral anticoagulants (DOACs), or heparin, which effectively prevent the clot from growing and new clots from forming. The choice of agent depends on various factors including the patient's clinical characteristics, preference, and potential for drug interactions.

Moreover, the duration of anticoagulation therapy is another crucial consideration in the management of PE. Wells' article underlines that the decision should be made based on a balanced

assessment of the risk of recurrent VTE and the risk of bleeding associated with prolonged anticoagulation. For most patients, a minimum duration of anticoagulation of three months is recommended.

An integral part of patient management is the prevention of PE recurrence. This means lifestyle changes such as increased physical activity and weight loss in overweight individuals, as well as addressing modifiable risk factors like smoking and high blood pressure. Long term patient monitoring may include follow-up imaging to assess the resolution of PE with therapy.

2.2 Nuclear Medicine Principles and Image Acquisition

The field of nuclear medicine began with the ground-breaking discoveries of X-rays by Wilhelm Conrad Roentgen in 1895 and natural radioactivity by Henri Becquerel in 1896. These seminal findings paved the way for the development of radiotracers—substances emitting radiation that can be tracked within the body. Radiotracers allowed physicians to visualize the functioning of organs and systems in real-time, thereby facilitating the diagnosis and treatment of diseases.

Beyond diagnosis and treatment, nuclear medicine contributes to research and drug development. By visualizing and quantifying physiological processes, it offers researchers valuable insights into the pathophysiology of diseases and the efficacy of novel treatments.

2.2.1 Radiotracers: The Core of Nuclear Medicine

All nuclear medicine depends on the use of radiopharmaceuticals [18]. Radiopharmaceuticals, commonly known as radiotracers, consist of a radioactive atom chosen for its detectable radiation emissions, combined with a molecule or compound that exhibits a specific affinity for tissues. This property allows radiotracers to serve as biological markers, tracing the movement and concentration of the associated molecule within the body.

Various types of radioisotopes exist, each possessing distinct properties chosen based on the specific diagnostic objective. For example, technetium-99m (^{99m}Tc) [19], the most used radioisotope in lung imaging, is favored for its ideal physical characteristics, including a short half-life and the emission of low-energy gamma rays that are easily detectable and cause minimal tissue damage.

Upon administration, the radiotracer circulates throughout the body and accumulates in specific organs, bones, or tissues, depending on its ligand. As it undergoes radioactive decay, the isotope emits gamma radiation, which is detected by a gamma camera [20]. This information is then used to generate images that unveil the functional aspects of organs and tissues.

2.2.2 The Gamma Camera: Transforming Gamma Rays into Images

At the core of the gamma camera lies the scintillation crystal, typically composed of sodium iodide doped with thallium [21]. When a gamma photon collides with the crystal, it generates a brief flash of visible light, known as scintillation. An array of photomultiplier tubes (PMTs) detects these flashes, amplifies the signal, and converts it into an electrical pulse. Processing of the electronic

signals from all PMTs by a computer (using Anger logic) decodes the location of the photon absorption in the crystal and measures the energy of the photon. This series of transformations enables the gamma camera to render the previously invisible gamma rays into a visible format. A good quality image requires the detection of millions of individual photons by the gamma camera.

However, this process would not be possible without collimators, which consist of lead honeycombs placed in front of the scintillation crystal. Collimators play a crucial role in filtering only the incoming gamma photons along straight paths of the collimator (typically perpendicular to the scintillating crystal), ensuring that the resulting image accurately represents the distribution of the radiotracer within the body. By filtering out photons from oblique angles, collimators significantly contribute to the sharpness and resolution of the resulting image, but at the cost of very low counting efficiency (only a small fraction of the photons hitting the collimator penetrate it).

As the gamma camera captures and records gamma photon events, a computer processes this information to generate an image. This image provides a snapshot of the spatial distribution of the radiotracer within the body at a specific moment, offering insights into the body's physiology.

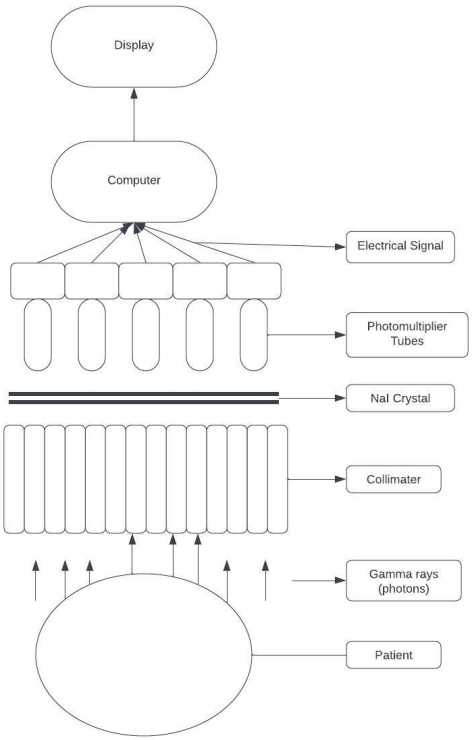

Figure 2.3 -The composition of a gamma camera.

Since its invention in the 1950s, the gamma camera has undergone significant advancements [22]. Modern gamma cameras often feature multiple detector heads that can rotate around the patient, capturing different views and enabling the generation of 3D images. Moreover, improvements in detector technology, data acquisition systems, and reconstruction algorithms have greatly enhanced the image quality, resolution, and diagnostic value of scans.

2.2.3 Planar Imaging

In the mid-20th century, advancements in detector technology and image reconstruction algorithms [23]–[25] pushed the emergence of planar imaging, the earliest form of nuclear imaging. Namely, Hal Anger in the 1950's developed the first gamma cameras which could efficiently detect the gamma-ray emissions from radiotracers in the patient to generate planar (i.e., two dimensional) images of the radiotracer distribution within the patient [26]. The image is a projection of the structures within the (three-dimensional) patient – like a shadow of a 3D object. The camera can be positioned to capture planar images from various angles, such as anterior (front), posterior (back), and lateral (side).

While planar imaging has its strengths, it also has limitations. On one hand, it is straightforward, efficient, and suitable for many diagnostic purposes, including bone scans [27], thyroid scans [28], and renal scans [29]. On the other hand, planar imaging lacks depth information [30], making it challenging to differentiate between overlapping structures in the two-dimensional image. This limitation can potentially obscure certain areas, leading to potential misinterpretations. Furthermore, the radiation emitted from deeper regions of the body, must traverse more tissue than nearer regions, resulting in signal attenuation and scatter. Consequently, nearer objects present a clearer and more intensely in the image than far objects.

2.2.4 Single Photon Emission Computed Tomography (SPECT)

The introduction of Single Photon Emission Computed Tomography (SPECT) in the field of nuclear medicine brought about an advancement from two-dimensional planar imaging to the three-dimensional representation of biological processes [31]. SPECT was developed to overcome the limitations of planar imaging, particularly the lack of depth information and challenges in visualizing overlapping structures. With SPECT, medical professionals and researchers gained the ability to explore the human body in three dimensions, obtaining comprehensive, multi-angular images that provide greater accuracy in diagnosis and insights into function.

SPECT operates on the same fundamental principle as planar imaging, detecting the gamma radiation emitted by radiotracers administered to the body. However, SPECT uses a robotic gamma camera that rotates around the patient and captures a series of two-dimensional images from various angles around the patient's body [32]. These images, called projections, are then processed by a computer algorithm such as filtered back-projection [33] or iterative reconstruction [34]. The outcome is a three-dimensional tomographic image that can viewed as slices through the patient, offering depth perception beyond what planar imaging can achieve. Whereas in planar images three dimensional (3D) structures overlap orthogonal to the projection plane, in 3D tomographic images these 3D structures are spatially separated.

2.2.5 Count Statistics and Image Noise

Nuclear medicine as a discipline relies on counts of radioactive decay events to visualize and interpret functional processes within the body [35]. Radioactive decay is inherently a random process [35]. The radioactive atoms in a radiopharmaceutical, decay at random intervals [36]. Yet,

given a sufficiently large number of these atoms, their average decay rate becomes predictable, following the well-documented laws of radioactive decay [35]. The number of disintegrations detected over a particular time are referred to as counts [37]. Count statistics, therefore, delve into the probabilities and distributions associated with these counts [37].

The number of detected events in a nuclear medicine study follows a Poisson distribution, a fundamental concept in count statistics [38]. Given a mean number of detected events, λ, the variance of the number of detected events is also λ. This means that the relative fluctuation or noise, often denoted as the standard deviation divided by the mean, decreases as the square root of λ [38]. Practically, this implies that acquiring more counts (higher λ) results in a better signal-to-noise ratio, leading to clearer and more reliable images [37].

$$P(X = k) = \left(e^{(-\lambda)}\lambda^{k}\right)/k! \qquad \textit{Equation 1}$$

$$\text{with } Var(X) = \lambda \text{ and } \sigma/\lambda = 1/\sqrt{\lambda}$$

However, acquiring images in nuclear medicine is not merely about maximizing counts. There is a delicate balance to maintain. More particularly, increasing counts often involves administering higher doses of radiopharmaceuticals or increasing the scan duration [36]. Both have implications: higher radiation doses pose risks to patients, and longer scans can lead to patient discomfort, increased potential of motion artifacts, and decreased clinical throughput [39].

Noise in nuclear medicine does not simply arise from the inherent randomness of radioactive decay. The detection process itself introduces noise. Detectors, while designed to be efficient, are not perfect. Not every gamma photon emitted from a decaying atom in the body gets detected. Those that do interact may not do so perfectly, leading to scattering or partial energy deposition [39]. These imperfect interactions, along with electronic noise in the imaging system, contribute to the overall noise in a nuclear medicine image.

2.3 Lung Scintigraphy: A Key Diagnostic Tool in Pulmonary Embolism

2.3.1 Defining Lung Scintigraphy: What It Is and Its Importance in Nuclear Medicine

Lung scintigraphy, also known as a lung scan, is a nuclear medicine imaging test widely used to investigate pulmonary function and blood flow, making it a crucial tool for diagnosing and managing various lung conditions [40]. In the context of PE, it can easily visualize the absence of blood flow due to blood clots that obstruct blood vessels in the lungs which are otherwise much smaller than the spatial resolution of the gamma camera.

A lung V/Q scan consists of two complementary scans: a ventilation scan (V) and a perfusion scan (Q). While protocols can vary somewhat between clinics, herein we focus on one such protocol as used at The Ottawa Hospital, and to capture the data used in this thesis.

The ventilation scan evaluates airflow to the lungs. 370-555 MBq (10-15 mCi) of ^{99m}Tc-Pertechnetate is added to a Technegas machine [41], where it is heated and burned in a carbon crucible in an oxygen free environment (under argon gas). This produces a ^{99m}Tc-labelled fine particle soot, which the patient inhales. The soot penetrates deep into the smallest airways in the lungs (alveoli) where it becomes lodged. The patient inhales the Technegas, then lays under the camera with the technologist monitoring the photon count rate to achieve a pre-defined target range. The gamma camera then captures images of ventilated regions of the lung, enabling clinicians to identify areas of ventilation deficits (defects), Figure 2.4.

The perfusion scan, on the other hand, assesses blood flow to the lungs. 148-185 MBq (4-5 mCi) of ^{99m}Tc-macroaggregated albumin (MAA) or a half-dose of (74-93 (2-2.5 mCi)) for pregnant or pulmonary hypertension patients is injected into the patient's bloodstream. As the radiotracer travels through the lungs with the blood, the MAA particles lodge in the capillaries of the lung. The gamma camera records the activity distribution, highlighting perfused regions. Obstructed regions, where blood flow is compromised, present as low count regions, and are referred to as perfusion defects, Figure 2.4.

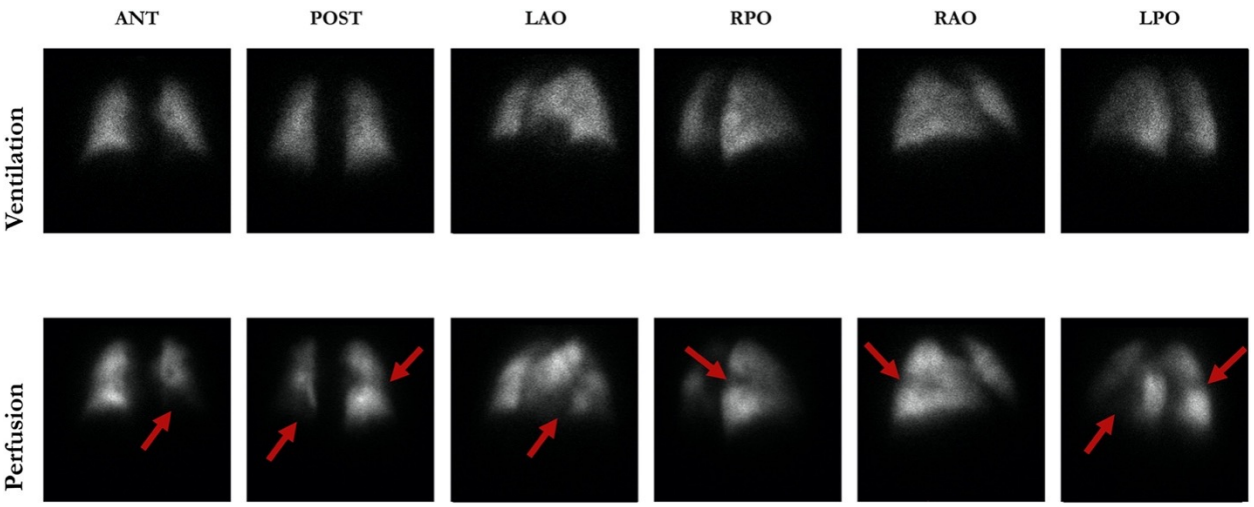

Figure 2.4- Ventilation-perfusion (V/Q) planar images for pulmonary embolism (PE) detection in six views. Red arrows indicate mismatched perfusion defects (V/Q mismatches) corresponding to a positive diagnosis of PE. ANT, anterior; LAO, left anterior oblique; LPO, left posterior oblique; POST, posterior; RAO, right anterior oblique; RPO, right posterior oblique.

Because perfusion imaging is performed immediately after the ventilation imaging and using the same radioisotope (^{99m}Tc) for both scans, the perfusion image is contaminated by the ventilation signal. To overcome this, the administered activity of MAA is intentional to achieve more than four times the counts than in the ventilation image. Thus, the perfusion signal “swamps out” the ventilation signal.

SPECT acquisitions are captured using a 128×128 matrix, with energy windows set at 140±7.5% (129-150) and a lower scatter range of (108-129). The dual headed gamma camera is programmed to acquire 64 views per head over 180° rotation, with a duration of 15 seconds for ventilation and

8 seconds for perfusion. Planar acquisitions are captured on a 256×256 matrix for six standard projections. Ventilation images are acquired until either 150,000 counts per image are achieved or 300 seconds elapse, whichever comes first. For perfusion images 600,000 counts per image or 300 seconds are targeted. Figure 2.5 demonstrates a typical lung V/Q protocol at The Ottawa Hospital.

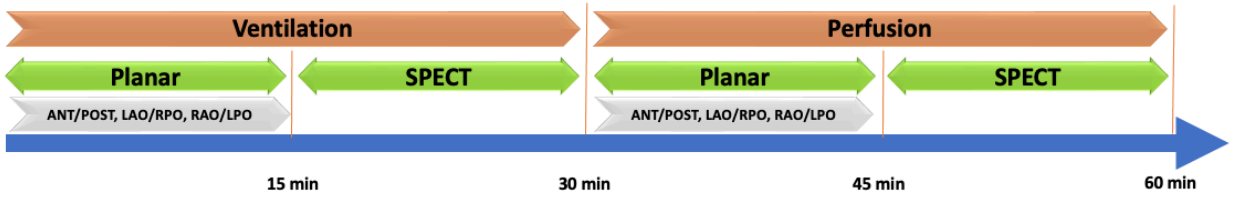

Figure 2.5- Timeline of a lung V/Q protocol at The Ottawa Hospital. Ventilation and Perfusion are performed for both planar and SPECT. ANT, anterior; LAO, left anterior oblique; LPO, left posterior oblique; POST, posterior; RAO, right anterior oblique; RPO, right posterior oblique.

2.3.2 Interpreting Lung Scintigraphy: A Guided Approach Based on PIOPED and EANM Guidelines

The Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED) [42] guidelines offer a systematic approach to interpreting V/Q scans. Based on studies of PE validated by invasive angiography, these guidelines lean heavily on the segmental and lobar structure of the lungs for their categorizations. They classify lung V/Q image patterns based on the observation of these segments and lobes into:

1. Normal: There is a homogeneous distribution of activity across all lung segments and lobes in both ventilation and perfusion.
2. Mismatch: A situation where a perfusion defect is evident within specific segments or lobes, even when ventilation remains active. This is indicative of PE.
3. Reverse Mismatch: This reveals a ventilation defect in certain segments or lobes while preserving perfusion, which is often a sign of airway obstruction.
4. Matched Defect: A defect is observed in both ventilation and perfusion within particular segments or lobes, suggesting chronic lung conditions.

From the PIOPED's standpoint, a high-probability scan, which often involves two segmental mismatches, indicates the presence of PE, and prompts immediate therapeutic actions. In contrast, a scan that demonstrates low or intermediate probability (< two segmental matches) often necessitates a more in-depth diagnostic review. In parallel, the European Association of Nuclear Medicine (EANM) guidelines [43] considers very high probability of PE if there is 1 segment or ≥ 2 subsegments mismatches. It's considered non-PE diagnosis if there are ≤ 1 subsegment mismatches.

2.3.3 Planar and SPECT Imaging in Lung Scintigraphy

Planar imaging was pivotal in establishing lung scintigraphy. However, its limitations in portraying depth and detailed pathologies were evident from the beginning. To overcome this limitation, planar projections are acquired for multiple angles and the interpreting physician must piece together in their mind the presence, location, size, and severity of defects based on consistent

patterns across multiple projections. SPECT seeks to resolve this shortcoming of planar lung scintigraphy by adding three-dimensional imaging to enhance the visualization of pulmonary structures, by reducing overlap between structures, improved localization, and increased defect-to-tissue image contrast. Research, including studies by Zuckier [44], pinpointed SPECT's heightened sensitivity in detecting subtle, peripheral lung defects.

SPECT imaging is not without its pitfalls. For one, the projection data (acquired prior to image reconstruction) is not all captured simultaneously. As the camera rotates around the patient, positioning is assumed consistent, but this assumption may be violated in the presence of patient motion. Accurate motion correction can be tedious and is not always reliable, leaving image artifacts in the reconstructed SPECT that cannot always be identified as motion artifacts. Furthermore, there has been concern that the higher sensitivity of SPECT (over planar) for visualizing small defects, may result in overcalling disease and overtreating patients. For those practitioners who wanted to fall back on true and tested planar imaging during SPECT, pseudo-planars were introduced to SPECT processing.

Two methods were developed to generate pseudo-planar images. The first combined sequential SPECT projections – typically three – in a summed angular technique [45]. This approach produces higher count images (less noise) than a single projection, but still significantly less than a full quality planar acquisition due to the shorter acquisition times of individual SPECT projections. A second, more sophisticated technique, re-projects SPECT images by modeling the associated physics to produce a pseudo-planar using all the image data [46]. This approach has not been widely accepted as the images appear over-smoothed and can have spontaneous artifacts associated with inaccurate estimation of photon attenuation maps, which are crudely estimated from the scintigraphy data.

Many clinicians that were trained to read planar lung scintigraphy, are apprehensive to transition to SPECT imaging. Thus, a dichotomy in practices exists [47]. A need still exists for robust pseudo-planar generation from SPECT data. Such a solution must generate realistic looking projections that reliably preserve disease patterns in the imaging data. Deep-learning, generative algorithms may be uniquely suitable for this task. Essentially, they could follow either of the above-mentioned approaches: (1) either enhance the image quality of individual SPECT projections to artificially increase the imaging count statistics and/or image spatial resolutions, or (2) they could use all the imaging data (before or after SPECT reconstruction) to generate few projection images. Regardless, deep-learning approaches have shown ability to generate realistic images surpassing conventional image filtering methods.

2.4 Deep Learning

Deep Learning (DL), as defined by LeCun et al. [48], is a branch of machine learning that employs a hierarchical stacking of multiple levels of representation, each interlaced with non-linear functions. This architecture allows the exploration of a broader range of representations, particularly advantageous when the application requires invariance to certain variations in the input data. In essence, these layers of representation are often a weighted aggregate of the inputs, leading to an incremental rise in the level of abstraction achieved with each added layer. Notably, in image processing, initial layers typically detect simple elements like edges, progressing to more

complex patterns and eventually recognizing complete objects. This layered learning process bypasses the need for manual feature engineering, streamlining the model training process.

A key learning technique in DL is Stochastic Gradient Descent (SGD) [48]. This method involves randomly selecting input vectors and averaging the error between the model's predictions and actual outputs for each vector. This error is then backpropagated through the network, allowing for the adjustment of each learned weight by calculating partial derivatives [49]. Consequently, each weight's contribution to the total error can be quantified, facilitating precise adjustments. An important aspect of backpropagation is the incorporation of differentiable non-linear activation layers between representation layers [49]. The weight vector w is updated in successive training steps as per the following equation, where E represents the prediction error and η is a tunable hyperparameter of the gradient decent rate:

$$w_t = w_{t-1} - \eta \frac{\partial E}{\partial w} \qquad \textit{Equation 2}$$

Several optimization strategies enhance the efficiency of SGD. Weight decay optimization, for instance, prevents weights from growing excessively large by adding a scaled L2 norm of the weights to the update equation [50]. Momentum-based optimization, on the other hand, aims to stabilize SGD by incorporating a momentum variable that considers previous gradients to influence current updates [51]. The Adam optimizer, a more recent development by Kingma and Ba [52], combines the benefits of AdaGrad and RMSProp, utilizing first and second moment variables with distinct decay rates.

Selecting the appropriate non-linear activation function is a critical decision in the model's architecture. The sigmoid or logistic function is often favored for its interpretability as a probability measure, mapping real numbers to the [0,1] range [48]. The softmax function, useful for multi-class outputs, is another variant. However, these functions can cause the vanishing gradient problem in deep networks due to minimal derivatives at extreme values. The Rectified Linear Unit (ReLU) function, which returns the positive part of its input, has been shown to be more effective for deep networks compared to sigmoid-type functions [53]. Commonly, DL models use the sigmoid or softmax function only in the final layer for probabilistic output, while employing ReLU or similar functions in all other layers for enhanced learning efficiency.

2.4.1 Convolutional Neural Networks (CNNs)

In Fully Connected Neural Networks (FCNNs), the architecture is such that the weights are integrally connected to every position within the input vector or matrix across each layer. This configuration processes each input element as a separate, independent dimension. A significant limitation of this structure is its inability to independently recognize patterns irrespective of their spatial location. To circumvent this issue and facilitate the extraction of high-level features for an

FCNN, one effective strategy is the utilization of filters. These filters are applied to the input through a convolution process, thereby extracting the requisite information.

The ability of FCNNs to independently learn new representation spaces is paralleled in Convolutional Neural Networks (CNNs), where the learning of filters is automated through the backpropagation [49]. This capability has been instrumental in the rapid adoption of CNNs, especially for deep learning tasks involving image processing. The widespread acceptance and utilization of CNNs in the field were bolstered by the groundbreaking achievements of Krizhevsky [54] and Szegedy [55] whose models delivered state-of-the-art results in various image-based applications.

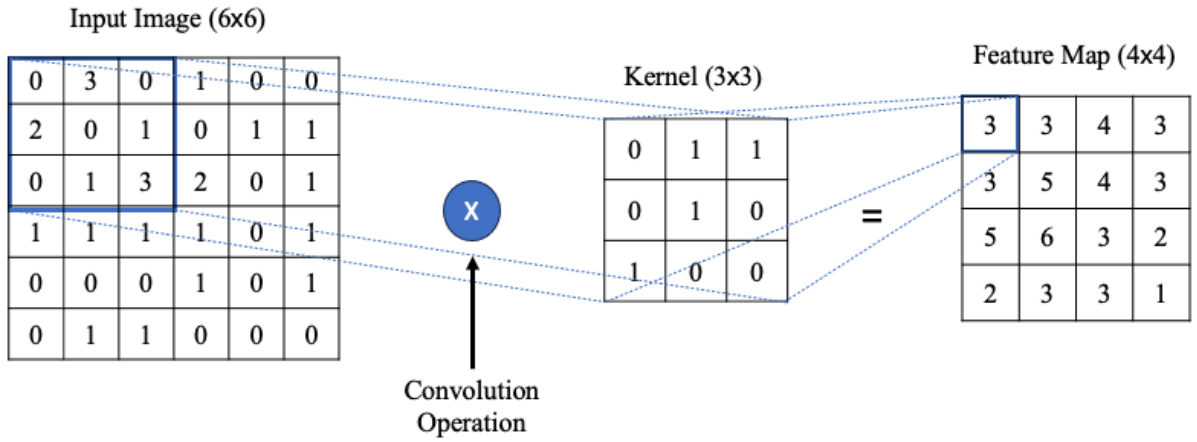

Figure 2.6- Illustration of a convolutional operation applied to a 6x6 input image using a 3x3 kernel. The operation produces a 4x4 feature map, effectively capturing spatial hierarchies in the input image.

2.4.1.1. ResNet

He et al. [56], introduced the ResNet architecture and the concept of residual blocks, specifically designed to train deeper (CNNs) more effectively. This architecture proved to be particularly effective in image classification and detection, and its applicability has since expanded to diverse areas such as image denoising [57]. A notable contribution of their work was the mitigation of the vanishing or exploding gradient problem, a common challenge in training deep neural networks. This was achieved through normalized layer initialization and the use of intermediate normalization layers like batch normalization. Despite these advancements, He et al. [56] identified a degradation issue arising from adding excessive layers to a model, which resulted in higher training and test errors for deeper networks.

To address the degradation problem, they introduced residual learning blocks for computer vision, which can be stacked to form a network. Each block consists of two paths: the first path containing convolutional layers with ReLU activation, and the second being a shortcut connection that adjusts the input size if necessary. ResNet-34 uses two convolutional layers, whereas ResNet-50/101/152 incorporate a bottleneck layer with a 1x1 kernel to transform the input into the required shape for the shortcut connection. The outputs from both paths are combined and processed through a ReLU activation layer. The residual architecture enables models to include more layers without an

excessive increase in parameters, thereby reducing the risk of overfitting. This is particularly beneficial as overfitting leads to poor model generalization on data that is not similar to the training set.

2.4.1.2. U-Net

The U-Net, introduced by Ronneberger et al. in 2015 [58], was designed initially for biomedical image segmentation, addressing challenges with sparse yet high-resolution data. The U-shaped architecture consists of two parts: a contracting path and an expansive path.

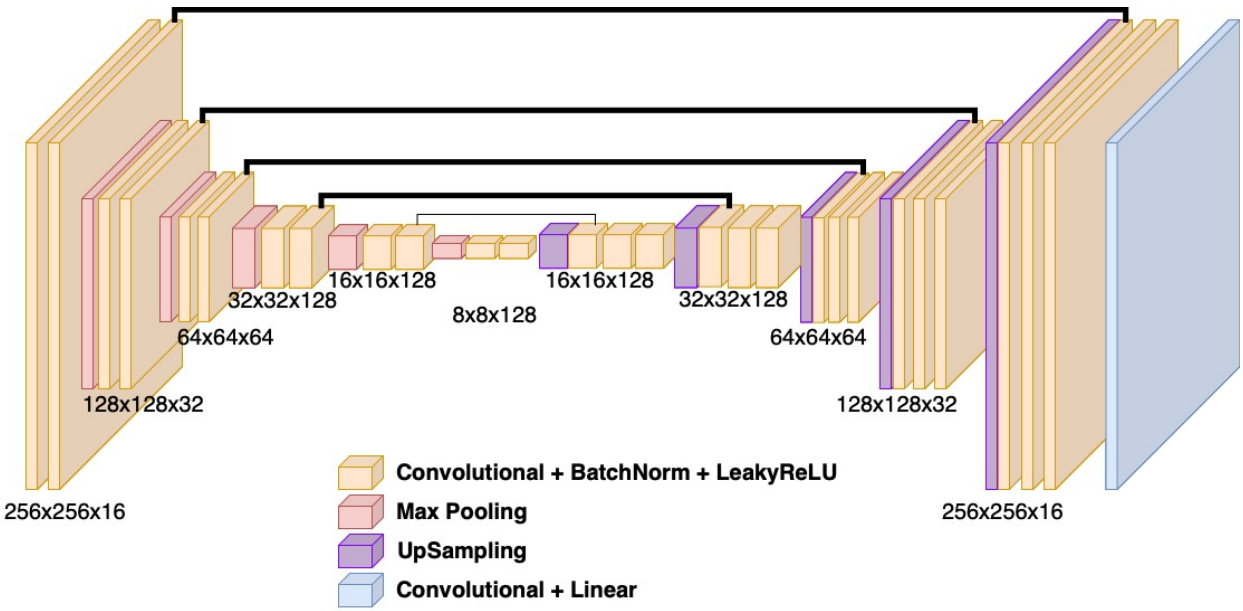

Figure 2.7 - Illustration of a U-Net architecture [58] for a 256x256 input image. This model employs successive convolutional operations followed by Leaky ReLU activation. Max pooling operations are utilized for spatial downsampling, while upsampling layers are paired with linear-activated convolutional layers to refine features. Skip connections link layers with matching spatial dimensions, ensuring the integration of features from various depths of the network for effective image segmentation.

Contracting Path:

In the contracting path of the U-Net architecture, the input image is processed through several convolutional layers coupled with ReLU activations. This combination increases the depth of feature representation while reducing the spatial dimensions of the image. Max-pooling layers are integrated into this pathway, to serve two purposes: firstly, they contribute to the reduction in image size, and secondly, they promote spatial invariance. This design enables the network to expand its receptive field at each hierarchical level, allowing it to capture and integrate broader contextual information from the input image.

Expanding Path:

Next, upsampling is performed followed by 2×2 "up-convolutions". Each upsampled map is then concatenated with a corresponding feature map from the contracting path, combining detailed and contextual information. At the end of this path, a final convolution layer outputs pixel-wise class probability.

2.4.2 Generative Adversarial Networks (GANs)

In 2014, Goodfellow et al. [59] introduced GANs and paved the way for a new paradigm in unsupervised deep learning. At its core, a GAN consists of two neural networks – the generator and the discriminator – that are trained concurrently. The generator creates synthetic samples, while the discriminator discerns the samples, distinguishing between real samples from the dataset and fake samples generated by the generator. Mathematically, GANs are modeled as a min-max two-player game, with the objective function being:

$$min_G max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[logD(x)] + \mathbb{E}_{z\sim p_z(z)}\left[\log\left(1 - D\left(G(z)\right)\right)\right]$$

Here, G seeks to minimize the log probability of the discriminator being correct, while D aims to maximize its correct classification.

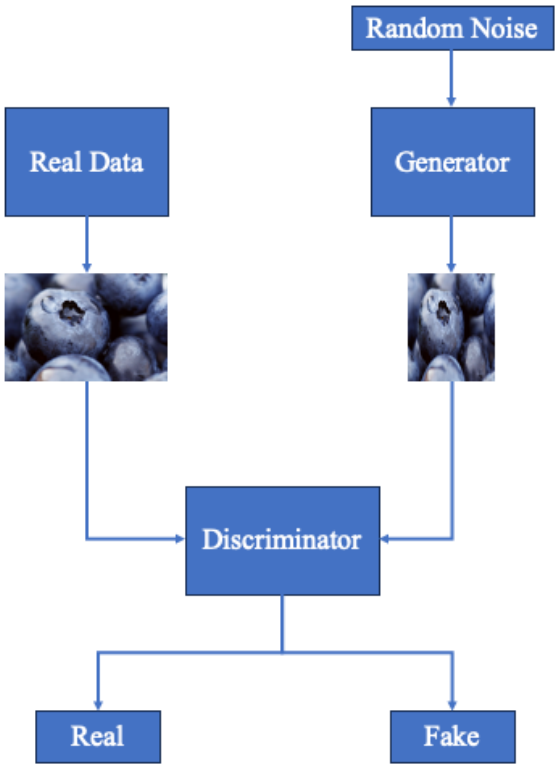

Figure 2.8- A representation of a GAN. Beginning with random noise input, the generator produces synthetic images. The discriminator evaluates these generated images alongside real data, as wither real or fake. The iterative process between the generator and discriminator facilitates the production of similar images as training progresses.

2.4.2.1. Conditional GANs

Expanding on the foundational GAN structure, Conditional GANs (cGANs) were proposed by Mirza and Osindero in 2014 [60]. In cGANs, both the generator and discriminator are conditioned on some auxiliary information, this can be labels or other data. This modification allows for a more targeted generation process. For cGANs, the objective function is modified to condition on auxiliary information *y*:

$$min_G max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[logD(x|y)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z|y))] \quad \textit{Equation 3}$$

2.4.2.2. Image to Image Translation GANs (Pix2Pix)

In 2017, Isola et al. [61], developed the Pix2Pix framework, a method for image-to-image translation tasks using conditional GANs. The idea is to translate an image from the source domain to the target domain. The network employs paired data, where the source and target images are aligned at the pixel level.

Architecturally, Pix2Pix employs a U-Net based generator for mapping input images to output images and employs a "PatchGAN" discriminator that classifies patches of images as real or fake. This allows the discriminator to focus on localized image structures.

2.4.2.3. Cycle GANs

In 2017, Zhu et al. [62], introduced Cycle GANs, a method for unpaired image-to-image translation. Unlike Pix2Pix, Cycle GANs don't necessitate aligned paired datasets. This was achieved by introducing a cycle consistency loss, ensuring that an image translated from one domain to another and then back again remains unchanged. The model is made up of two generators and two discriminators. If *G* translates from domain *X* to *Y* and *F* does the reverse, the cycle consistency loss can be formulated as:

$$\mathbb{L}_{cycle}(G,F) = \mathbb{E}_{x\sim p_{data}(x)}[\|F(G(x)) - x\|1] + \mathbb{E}_{y\sim p_z(y)}\big[\big\|G\big(F(y)\big) - y\big\|1\big] \quad \textit{Equation 4}$$

2.4.3 Loss Functions

Loss functions, or cost functions, are mathematical formulations that get the difference between the predicted output and the actual target in a learning model. Their primary purpose is to provide a metric, based on an optimization algorithm to adjust model parameters and guide the training process. This section delves into specific loss functions used in deep learning.

2.4.3.1. Mean Squared Error (MSE) Loss

The MSE is a commonly used loss functions, especially in regression problems. It calculates the average of the squared differences between predicted and actual values. Mathematically, for a dataset with *n* samples:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2 \qquad \textit{Equation 5}$$

Where y_i and x_i represent the actual value and predicted value for the i^{th} samples respectively. MSE is sensitive to outliers due to the squaring the term, it remains differentiable, making it suitable for optimization using gradient descent.

2.4.3.2. Mean Absolute Error (L1) Loss

The L1 loss computes the average of the absolute differences between the target and predicated values. Mathematically:

$$L1 = \frac{1}{n}\sum_{i=1}^{n}\|y_i - x_i\| \qquad \textit{Equation 6}$$

It is less sensitive to outliers compared to MSE and is used when the model should be robust against large deviations.

2.4.3.3. Perceptual Loss

Perceptual loss was introduced by Johnson et al. [63], as a concept that considers the approach to enhancing the visual fidelity in tasks like style transfer and super-resolution imaging. This method diverges from traditional loss functions such as Mean Squared Error (MSE) or L1, which are based on pixel-wise differences. Instead, perceptual loss delves into the realm of feature-level comparison, to capture more nuanced differences between images.

Perceptual loss uses pre-trained deep neural networks, such as the VGG network [64]. These networks serve as feature extractors that analyze both predicted and target images. The core objective is to minimize the differences in the feature representations extracted from these images. In this context, the content loss is quantified as the Euclidean distance between these feature representations, typically sourced from deeper network layers known for retaining content and spatial structure of the images. This approach ensures that the predicted image, while not required to be an exact pixel-level replica, is perceptually akin to the target image.

Furthermore, to address variations in style—encompassing elements like colors, textures, and patterns—the style loss is incorporated into the perceptual loss framework [65][66]. This aspect of loss is computed using the Gram matrix of both predicted and target images, and the squared Frobenius norm of the difference between these matrices represents the style loss. This component

of perceptual loss thus penalizes disparities in stylistic elements between the two images, ensuring a more comprehensive measure of similarity beyond mere content alignment.

2.4.3.4. Binary Cross Entropy (BCE) Loss

BCE loss is commonly used in classification tasks. Given a predicted probability p and a binary target y, BCE is defined as:

$$BCE = -ylog(p) - (1 - y)\log(1 - p) \quad \text{Equation 7}$$

This loss quantifies the divergence between the actual and the predicted probabilities. In essence, it penalizes the predicted probabilities that deviate from the actual class labels. Its magnitude grows substantially as the predicted probability approaches the wrong label.

Notably, the binary cross-entropy loss has found significance in the training of GANs. During the training process of GANs, the discriminator is often trained using the BCE loss. When the discriminator examines a real sample, it should output a value close to one, and for a generated sample, a value near zero. The BCE loss ensures that the discriminator's predicted probabilities align with these expectations. As the GAN training proceeds, the generator evolves to produce samples that increasingly deceive the discriminator, pushing its output probabilities closer to the uncertainty boundary of 0.5 for generated samples.

2.5 Image Quality Assessment

2.5.1 Structural Similarity Index (SSIM)

SSIM is a metric used to measure the similarity between two images, often used for assessing the quality of image restoration in tasks such as denoising, compression, and super-resolution. As a quality indicator, it requires paired reference images to which the test images can be compared. Unlike traditional error summation methods like MSE, SSIM is designed to look at changes in structural information, luminance, and texture. Wang et al. [67] introduced SSIM, emphasizing its effectiveness over traditional means for its closer alignment with human visual perception [68]. The SSIM index between two images x and y is defined as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad \text{Equation 8}$$

Where:

- μ_x and μ_y are the average of x and y respectively.

- μ_x^2 and μ_y^2 are the variances of *x* and *y* respectively.
- $\sigma_{xy}$is the covariances of *x* and *y.*
- c_1 and c_2 are scalars to prevent division by zero. These are typically scaled to the dynamic range of the pixel values, L, such that c_1=0.01L and c_2=0.03L,but can vary in implementation.

SSIM provides a score between -1 and 1, where a value of 1 indicates that the test image and reference image are identical in terms of structural information.

2.5.2 Peak Signal-To-Noise Ratio (PSNR)

PSNR is a widely used metric in the field of image and video processing to quantify the quality of reconstructed images [69]. It represents the ratio between the maximum possible power of an image (signal) and the background noise in the image.

$$PSNR = 10 * log_{10}\left(\frac{MAX_y^2}{MSE(x,y)}\right) \quad \textit{Equation 9}$$

Where:

- MAX_y^2 is the max possible pixel value of the reference image. For example, for a typical 8-bit image is 255.
- $MSE(x,y)$ between the predicted and reference images.

The typical value for PSNR in images and video compression is between 30 and 50 dB, where higher is better. If the value is over 40 dB it is considered very good and below 20 dB are normally unacceptable.

2.5.3 Task Based Assessment

In the diagnosis of PE, physicians often rely on standardized interpretation criteria, notably PIOPED and EANM (2.3.2). These criteria provide guidance on determining the mismatch size for accurately diagnosing or dismissing PE. However, an element of subjectivity persists in the diagnostic process. Specifically, physicians make personal judgments about the severity of a perfusion defect, its alignment with the pulmonary vascular anatomy, and its size in relation to the affected segment [70]. Such subjective estimations can lead to varying diagnostic outcomes. To illustrate, within the framework of PIOPED criteria, if one physician estimates the size of multiple small mismatches to be less than 25% of the affected segments, a low PE probability might be inferred. However, another physician estimating the same mismatches to be larger than 25% could interpret it as a high probability for PE [70]. Although many physicians strictly adhere to these standardized guidelines, some incorporate their accumulated expertise and experiential insights, referred to as the gestalt interpretation, to make their final PE diagnosis [71].

To assess the quality of an image on a task-based merit various observer experiments may be performed. For example, the agreement between different observers (inter-observer) or between repeat interoperations by the same observer (intra-observer) may be assessed. In other cases, the

interpretation of an image may be compared to another, established reliable ground truth reference such as an objective test, patient outcome, or interpretation of an expert panel. Regardless, recruiting human observers is typically expensive, time consuming, produces few data points, and fraught with human errors. Quantitative and automated quality assurance measures are, therefore, preferable when possible and sufficient.

2.6 Enhancement and Denoising in Nuclear Medicine: A Review of Related Literature

2.6.1 Classical Filtering Approaches

Classical filtering approaches for image enhancement and denoising have been used in the field of medical imaging [69]. These methods, grounded in spatial or frequency domain operations, aim to improve image quality by reducing the noise and augmenting prominent features [72]. Some of the most common filtering techniques include mean filtering, median filtering, and Gaussian filtering, where the value of a pixel is determined based on a weighted average of its neighbors or by filtering in the frequency domain to attenuate unwanted frequencies. Such filters operate under the assumption that noise can be effectively differentiated from the image's true underlying structures, either spatially or in the frequency domain [73].

A prevalent issue is the inadvertent blurring of essential image details during the denoising process [74]. This compromise between noise reduction and preserving detail is particularly pronounced in methods like mean and Gaussian filtering [75], where the averaging process inherently smooths sharp intensity transitions, which could be critical in applications like medical imaging. Another limitation lies in the assumption of uniform noise characteristics throughout the image [76]. Many traditional filters have fixed structures and parameters, making them potentially not fit to handle non-uniform, adaptive, or complex noise distributions present in real-world images.

Further complicating matters is the challenge of preserving edges. Edges, which represent abrupt intensity changes, are fundamental features in images [77]. Yet, many classical filtering approaches, in their bid to reduce noise, often end up blurring or distorting these edges. Techniques like the bilateral filter were introduced to address this by preserving edges while denoising, but they too can sometimes fail, especially in images with intricate texture or in scenarios where noise and image details have similar frequencies [78]. As imaging technology has advanced and the need for high-quality, denoised images has grown, it has become evident that while classical filtering methods provide an initial first step, they may not always be the optimal choice for every image enhancement and denoising application [79]. As such, we've turned to deep learning as an alternative to enhance scintigraphic images [48].

2.6.2 Image Enhancement in PET Imaging

In the field of medical imaging, Positron Emission Tomography (PET) plays a vital role in detecting, diagnosing, and treating various diseases, including cancer, heart disease, and

neurological disorders [80]. However, traditional PET imaging faces limitations such as noise and low resolution, primarily due to the need to minimize patient exposure to radiation.

CNNs and transformers have been utilized to improve the quality of PET images and enhance diagnostic accuracy [81][82]. They effectively enhance the resolution, reduce noise, and reconstruct low-dose images into full-dose equivalents, bridging the gap between low-dose and full-dose images.

Xu et al. conducted a study employing a U-Net model with residual learning to reconstruct full-dose images from a single low-dose image [83]. Similarly, Xiang et al. developed a deep auto-context CNN model that synthesized full-dose images from low-dose images, demonstrating comparable image quality with significantly improved speed compared to conventional methods [84].

Another approach to enhance PET images involves combining PET with other imaging modalities, such as Magnetic Resonance (MR) imaging. By incorporating high-resolution anatomical information from MR imaging, models can generate more accurate and robust images. Liu et al. applied three modified U-Net architectures to enhance PET noise using concurrent MR images without the need for full-dose PET images, resulting in a higher signal-to-noise ratio [85].

Recent advancements have also explored the application of GANs for image-to-image transformation tasks in PET imaging [86]. Zhao et al. utilized a 3D cGAN to denoise low-dose brain PET images [87], while Lei et al. employed Cycle-GAN for noise reduction and predicting whole-body full-dose 18F-FDG PET images from low-dose ones [88].

2.6.3 Image Enhancement in SPECT Imaging

SPECT imaging faces challenges related to the low signal-to-noise ratio, long acquisition times, and suboptimal resolution [89]. Deep learning techniques, have emerged as effective solutions to overcome these hurdles and enhance the quality of SPECT imaging [90].

One application of deep learning in SPECT involves reducing the acquisition time, which is especially crucial in pediatric studies due to children's limited compliance during lengthy scans. Lin et al. conducted a study using a 3D U-Net model to estimate full-time SPECT acquisitions from half-time acquisitions in pediatric renal SPECT scans [91]. The model successfully synthesized SPECT images with high accuracy (91.7%), sensitivity (83.3%), and specificity (100%), demonstrating its potential to reduce acquisition time while maintaining image quality.

The U-Net architecture has been used to reduce SPECT imaging acquisition time. By generating synthetic intermediate projections (SIPs) from fewer actual projections, this method effectively reduces the total number of required projections without sacrificing image quality [92]. It was trained on every fourth out of 120 acquired SPECT projections, the model synthesized 30 SIPs per case, thereby reducing the original acquisition from 120 to 30 projections, supplemented by 90 deep learning-generated SIPs. The generated SIPs achieved a mean SSIM value of 0.926, and for attenuation-corrected ordered-subset expectation maximization (AC-OSEM), the reconstruction with synthesized projections resulted in higher SSIM (0.993 vs. 0.989) and PSNR (49.5 vs. 47.2) values compared to using 30 projections.

Pan et al. developed a deep learning-based method, specifically a lesion-attention weighted U^2-Net model, to enhance the quality of SPECT bone scan images [93]. The model was trained with SPECT images and corresponding 1/7 scan time SPECT images. Quantitative evaluation demonstrated synthesized images with PSNR of 40.8 and SSIM of 0.788 compared to baseline 1/7 SPECT of 37.7 and 0.765 for PSNR and SSIM respectively.

Chapter 3: State of Artificial Intelligence in Nuclear Medicine

I am a co-author on the following article:

> Jabbarpour A, **Ghassel S**, Lang J, Klein R, Moulton E. "The Past, Present, and Future Role of Artificial Intelligence in Ventilation/Perfusion Scintigraphy: A Systematic Review." *Seminars in Nuclear Medicine*, vol. 53, no. 6, 2023, pp. 752–65, https://doi.org/10.1053/j.semnuclmed.2023.03.002. [94]

The article is a systematic literature review that traces the journey of artificial intelligence in the context of V/Q scintigraphy. Through a systematic exploration of various AI methodologies, the paper identifies significant opportunities for innovation and development in the realm of V/Q scintigraphy. The content was slightly modified from the manuscript to fit within the context of this thesis.

3.1 Background

V/Q scans, tracing back to the 1960s, [95][96] are used for diagnosing PE. V/Q studies have been instrumental in both acute and chronic settings, helping to diagnose new instances of PE and track the progression of the disease over time. The early 1990s saw a standardization in the interpretation of acute PE on V/Q scintigraphy, and this period also marked the onset of applying AI algorithms for automating acute PE diagnosis [97]. Interestingly, some early AI models even outperformed experienced radiologists in diagnosing PE [98], [99].

However, the advent of CTPA saw a shift in the diagnostic approach for PE, leading to a decline in the use of AI for V/Q scintigraphy. By the mid-2000s, CTPA was already outpacing V/Q scintigraphy. Presently, V/Q scintigraphy makes up only 2%-11% of imaging studies for PE diagnosis, [100], [101].

The dynamics of V/Q scintigraphy usage, coupled with advances in medical image processing, have impacted the potential of AI in lung scintigraphy and its clinical value. Therefore, it is crucial to understand the past, present, and potential future of research in lung scintigraphy and AI. In this context, we undertook a systematic review to not only assess the use of AI in V/Q scintigraphy for PE detection but to explore all potential applications, with the goal of identifying key focus areas and gaps that could shape future research and development initiatives.

3.2 Methods

This review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines [102].

3.2.1 Search Strategy

We conducted a literature search on the Ovid MEDLINE database, covering the period from January 1, 1946, to August 5, 2022. Their focus was on studies that employed artificial intelligence

models, including machine learning and deep learning, within the context of ventilation and perfusion imaging. Search terms included "artificial intelligence", "machine learning", "deep learning", and "ventilation / perfusion".

Criteria for inclusion in the review were that the articles had to be written in English, had to be peer-reviewed, and had to be original publications or full conference proceedings in manuscript format. We excluded any articles not relevant to computer vision, without AI techniques, non-original works, or were merely abstracts and poster presentations. They also excluded any studies that used V/Q images for subsequent analyses after applying AI to other types of images.

Alongside the database search, we also manually searched for articles to ensure no relevant studies were overlooked. They eliminated any duplicate findings. We independently reviewed the titles and abstracts of the articles to determine if they met the criteria.

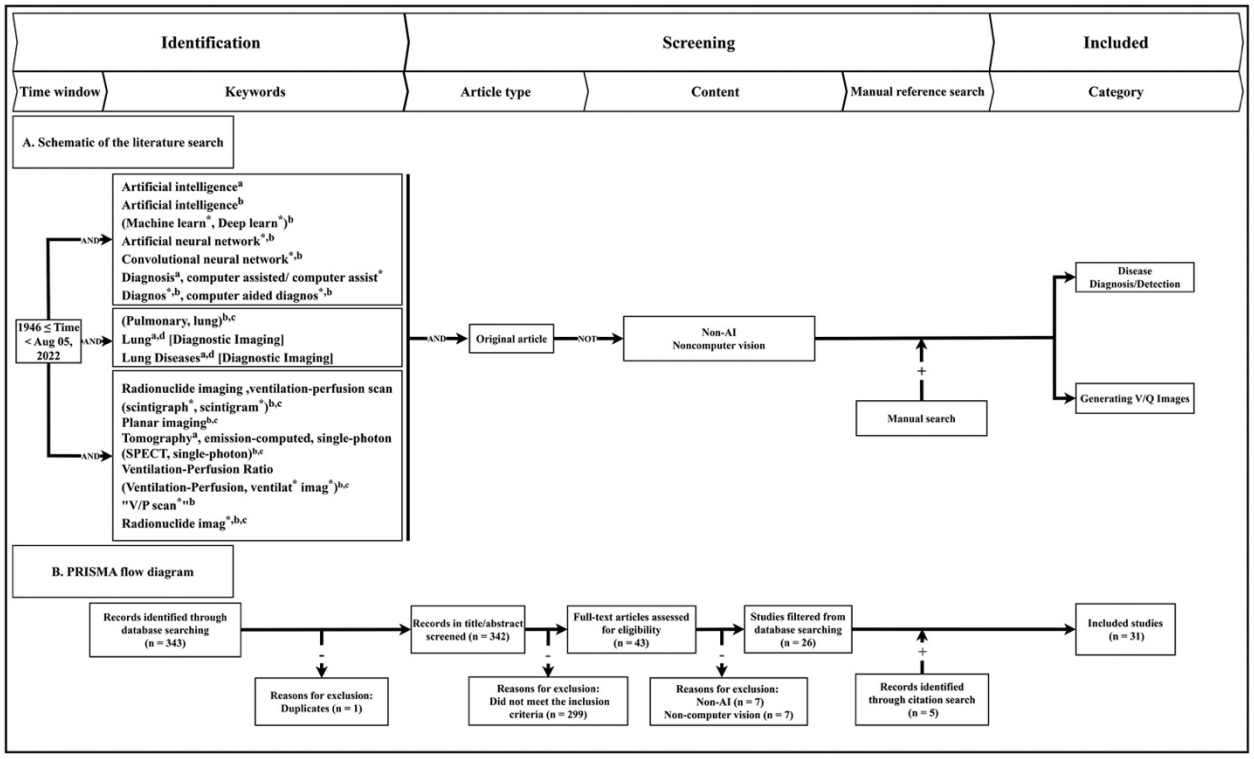

*Figure 3.1- (A) A schematic representation of our search strategy including the time window, keywords, screening criteria, and the final application category. (B) PRISMA flow diagram of systematic literature review process corresponding to headers in (A). a: Medical Subject Headings (MeSH) and all Subheadings as used in Medline/PubMed. b: word of phrase appearing in titles and abstracts. c: keyword supplied by the author. d: MeSH topical qualifier for Diagnostic Imaging. *: wildcard for all words beginning with given characters. /: specific MeSH Subject heading.*

3.2.2 Data Extraction

Using a standardized data extraction form, we separately extracted the necessary information from each study. The data gathered included, among other things, the year of publication, imaging techniques used, the stage of the disease being studied (acute or chronic), acquisition protocols, ventilation and perfusion agents, and the size of the study population for training, validation, and

testing. We also noted the outcome label, incidence rate if applicable, model tasks, input features, validation methods, algorithms used, the source of ground truth for output labels, and performance metrics of the most effective model reported.

3.2.3 Statistics

Descriptive statistics are reported as the median and the interquartile range (IQR) for continuous values or as a ratio and percentage for categorical data. Differences in proportions with the two-proportion z-test.

3.3 Results

3.3.1 Study Selection and Grouping

The search process, along with the PRISMA flow diagrams, are represented in Figure 3.1. The initial database search resulted in 343 articles, which reduced to 342 after one duplicate was eliminated. After screening the titles and abstracts, 299 articles were discarded. Following a comprehensive text review, 26 papers met the eligibility criteria. Five additional articles from the manual search were included in the analysis.

Hence, a total of 31 publications fulfilled the inclusion criteria. These studies were then categorized into two distinct groups: (1) disease diagnosis/detection (22 papers, constituting 71% of the total), and (2) generating V/Q images from non-scintigraphic data (9 papers, or 29% of the total). A summarized version of the studies in each category can be viewed in Figure 3.1 and Figure 3.2.

Publications focusing on AI for disease diagnosis/detection were published from 1993 to 2021, with a majority occurring between the mid-90s and early-2000s (the median year of publication being 2001, with an interquartile range of 1996-2003). In contrast, all the studies centering on generating V/Q images were published within a narrow period from 2019 to 2022. A visual display of the breakdown of publications per year and application category can be seen in Figure 3.2.

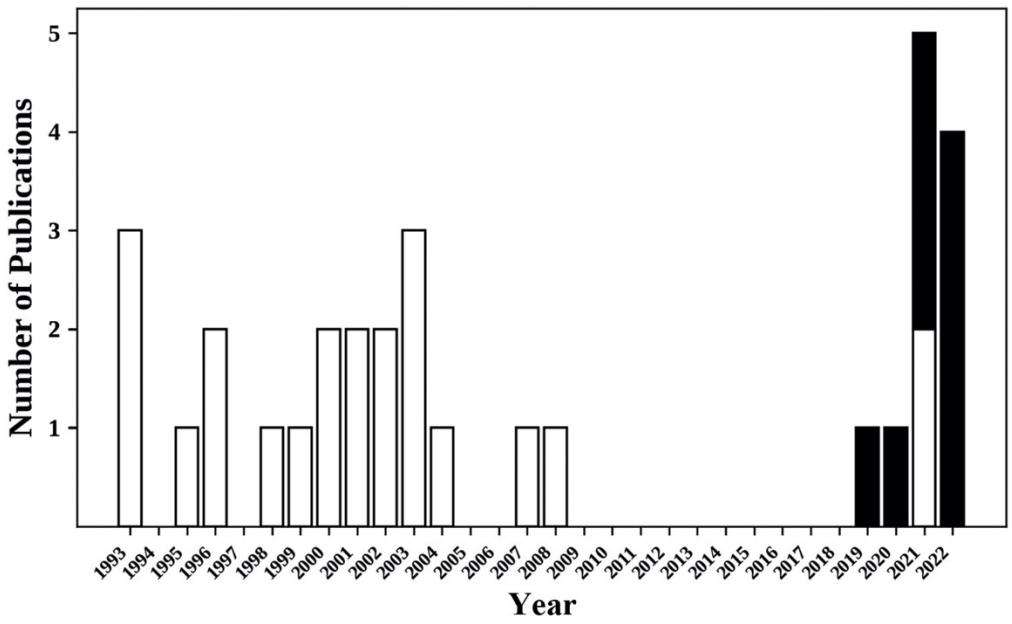

Figure 3.2- Breakdown of publications by year and application category. White – Disease Diagnosis/Detection. Black – V/Q Image Generation.

3.3.2 AI for Disease Diagnosis/Detection

The study found PE was the primary disease of interest in most of the research papers, accounting for 95% of the total, with only a single paper studying the diagnosis of pulmonary arterial hypertension. No papers were identified that pertained to other indications of V/Q scintigraphy, such as assessments of pulmonary function before surgery, COPD, or CTEPH. Most of the studies (82%) utilized both ventilation and perfusion images, with a minority (18%) using only perfusion images.

In terms of the radiotracers used, all the studies that disclosed this information used ^{99m}Tc-MAA as a perfusion agent. However, the ventilation agents varied, with 55% using ^{133}Xe, 20% using ^{99m}Tc-DTPA, and 20% not using ventilation images at all. The study noted that ^{133}Xe was more commonly used in earlier studies.

Planar imaging was the preferred technique used in 95% of all the papers, with only a single study using SPECT for perfusion images. None of the studies used ventilation SPECT as a model input. In the context of disease stages, the studies were either purely acute (50%) or a mix of acute and chronic (23%) stages, with some (27%) not reporting such information.

When looking at the model types, shallow artificial neural networks (ANNs) such as multi-layer perceptrons were the most studied (91%) and were used significantly more than other machine learning models. Feature extraction was performed manually in 41% of the studies and automatically in the remaining 59%.

Ground truth labeling for training and testing sets was determined with invasive pulmonary angiography in 48% of the papers, consensus of physicians in 19%, a single physician in 5%, not mentioned in 5%, and a mix of physician consensus for training and PA for testing in 24%.

Interestingly, most of the AI research in V/Q scintigraphy for disease diagnosis and detection was led by three independent groups from Massachusetts General Hospital, Duke University Medical Center, and Lund University, contributing to 32%, 18%, and 18% of the research papers, respectively.

3.3.3 AI for Generating V/Q Images

All the studies identified in this literature review aimed to generate synthetic V/Q images from CT scans. The goal was to protect functional areas of the lung during radiation therapy. As such, the classification of the patients into acute or chronic disease stages was not relevant in these studies.

This emerging research field within the intersection of medical imaging and artificial intelligence started with a solitary study in 2019 and rapidly expanded to include eight additional studies in the ensuing years. Notably, none of the studies attempted to convert V/Q images into CT scans or any other imaging modality.

In terms of the type of images generated, there was an almost even split between studies generating ventilation images (44%) and those generating perfusion images (56%). When creating perfusion images, all studies used 99mTc-MAA SPECT as the ground truth modality. On the other hand, for ventilation images, 75% used 99mTc-labelled carbon particles SPECT, and 25% used Gallium-68 PET. None of the studies aimed to generate ventilation or perfusion planar images.

Most studies used Dynamic CT (4DCT) to generate ventilation SPECT images (75%), while static CT was primarily used for perfusion SPECT images (80%). All these studies employed CNNs, with the U-Net architecture inspiring 78% of them. One study (11%) used a cGAN, and another one (11%) employed a generic fully CNN.

3.4 Discussion

3.4.1 Early Research on AI in V/Q Scintigraphy for PE: Ahead of Its Time and a Missed Opportunity

The period from the mid-1990s to the early 2000s saw a surge in scientific studies exploring the use of AI in detecting acute PE using V/Q scintigraphy. V/Q scintigraphy was among the first applications in medical imaging to be investigated with AI. A bibliometric study shows that until 1993, AI-related research in medicine was rather rare, constituting only 0.64% of all literature published between 1977 and 2018.

During this period, AI models were trained using manually extracted clinical and imaging features related to PIOPED criteria, such as size and morphology of V/Q mismatches. These models mainly used Pulmonary Angiography (PA) as a ground truth label for training AI models. However, as technology advanced, research transitioned from manual to automated feature extraction from both ventilation and perfusion images.

By the early 2010s, when technology had evolved to effectively utilize CNNs for computer vision, CTPA had already surpassed V/Q scintigraphy for detecting acute PE. Consequently, the research focus shifted towards CTPA.

The early to mid-2010s witnessed the rise of CNNs and deep learning, opening a new chapter in medical image processing for workflow acceleration and computer-aided diagnosis (CAD). However, with CTPA overshadowing V/Q scintigraphy, the research on automatic PE detection or diagnosis with CNNs was primarily focused on CTPA. The number of studies on CTPA for image-level classification and object detection for PE has grown significantly in the past three years, allowing for dedicated meta-analyses, systematic reviews, and public datasets. In contrast, the research on applying state-of-the-art AI techniques, like CNNs, to V/Q scintigraphy remains limited.

The shift in ground truth labeling for PE is also noteworthy. PA, an invasive procedure, was considered the gold standard of PE diagnosis and was frequently performed in the early 90s to validate diagnosis on V/Q scintigraphy. Several studies have shown that ANNs can diagnose PE as effectively as, or even better than, experienced physicians. However, as non-invasive imaging for diagnosis of acute PE became more established, the use of PA declined and is rarely performed today. Consequently, only physician interpretations are now available for training AI models for PE diagnosis.

This transition has significant implications for AI's role in PE diagnosis. For instance, given the lack of invasive PA, new AI models will require innovative approaches to surpass expert radiologists or nuclear medicine physicians. Moreover, combining data from several sources will require careful harmonization of imaging interpretations.

V/Q SPECT, a newer technology, has been severely understudied in the context of AI for PE diagnosis. V/Q SPECT only began to replace planar acquisitions in the 2000s and never received the same attention as the pioneering planar studies in the early 90s. Our review only identified a single recent study of AI for SPECT, which showed promising results, indicating that CNNs can likely perform well at PE diagnosis and detection.

3.4.2 Current Trends in AI for V/Q Scintigraphy: Image Translation to V/Q Images

This review revealed a surprising fact: despite the growing use of AI in diagnosing and detecting a range of medical conditions, there is a noticeable lack of current research focused on using V/Q scintigraphy for detecting PE. It seems that the research landscape has evolved, with a new focus on creating synthetic V/Q images from static 3D or dynamic 4D CT scans. This shift is mainly aimed at preventing damage to functional lung tissue during radiotherapy, akin to SPECT-guided radiotherapy treatment planning.

In a fascinating twist, instead of using V/Q scans as input for PE detection, as was the case in the past, these scans are now being used to train and validate the output of image generation models. This shift in research focus began in 2019, with encouraging results both in terms of generating the synthetic image and manually segmenting functional lung regions based on the synthetic image. This cutting-edge work has been spearheaded by two research groups: one based at the Cancer Hospital Chinese Academy of Medical Sciences and Peking Union Medical College, and the other at The Hong Kong Polytechnic University.

However, it's worth noting that to date, this research hasn't transitioned into clinical applications. Furthermore, it appears that no one has attempted to use these synthetic V/Q images for disease detection, such as acute PE detection, which traditionally has been the primary use of V/Q scintigraphy.

3.4.3 Future Opportunities in AI for V/Q Scintigraphy

Just as crucial as the existing literature this review has brought to light are the identified gaps within it. The application of AI in nuclear imaging covers a broad spectrum of areas such as instrumentation and image acquisition, image reconstruction and low-dose/fast scanning, quantitative imaging, cross-modality image-to-image translation, image registration, internal dosimetry, and image interpretation, which includes computer-aided detection, diagnosis, or prognosis. What follows are some potential directions for integrating these innovations into V/Q scintigraphy. These suggestions could serve to rekindle the forward momentum that V/Q scintigraphy experienced around the dawn of the new millennium.

3.4.4 Workflow

AI has been utilized to enhance image quality in several ways, such as removing noise and artifacts, correcting motion, and reducing dosage. Generally, there are four proposed methods within the image generation pipeline: (1) improving raw or sinogram data, (2) image reconstruction, (3) enhancing the reconstructed image, and (4) employing a combination of analytical reconstruction and AI models, a method known as hybrid image reconstruction.

Low-dose and fast scan approaches are two aspects of the same concept. As the activity level or acquisition time decreases, there's an unavoidable increase in Poisson noise, which subsequently affects image quality, the ability to detect lesions, and the quantitative accuracy of nuclear medicine imaging. Reducing the activity of the administered radiopharmaceutical is desirable due to its associated radiation risk. Additionally, shortening the acquisition time improves clinical throughput, decreases patient discomfort, reduces the likelihood of patient movement, thereby increasing profitability and enhancing image quality. Numerous attempts have been made to suppress noise in nuclear medicine images, from post-reconstruction processing/filtering algorithms and anatomically guided reconstruction algorithms to deep learning-based denoising and statistical modeling during iterative reconstruction. Recently, deep learning enhancements in PET imaging have significantly outpaced those in SPECT, with a few exceptions in image reconstruction, count enhancement, and fast scanning. However, no AI count enhancement study was found for planar scintigraphy, despite several non-AI attempts.

Deep learning has previously delivered promising results for artifact removal in various medical imaging modalities like CT and PET. In the case of ventilation scintigraphy that uses aerosols, hot spots resulting from central deposition in large airways are particularly noteworthy. While traditional image processing algorithms like image intensity thresholding and erosion techniques have been used to address this issue, these methods often require manual intervention and are unreliable for widespread automation. AI stands as a prime candidate for automatically eliminating these artifacts to improve diagnostic quality.

Even though SPECT has several benefits, some centers, especially in the United States, prefer to capture planar images due to the ease of visualizing the lung through the traditional 6 or 8 views. To encourage clinics to transition to the more sensitive SPECT while maintaining familiar planar imaging, pseudo-planar images from tomographic reconstructions offer a promising solution. However, the blurry appearance of these images has been criticized, and some researchers question their ability to diagnose PE as effectively as true planar images. Deep learning solutions might be

the answer where analytical approaches have fallen short, enabling the benefits of pseudo-planars without compromising diagnostic or imaging quality. This could be achieved either through count enhancement of SPECT projection data or by generating planars from the reconstructed SPECT.

3.4.5 Anatomical Information

Historical research, dating back to the year 2000, has demonstrated that AI struggled most with cases that involved abnormal Chest X-Ray (CXR) or sub-segmental PE within the test set. These challenging cases are also a reflection of the inherent limitations of V/Q scintigraphy as encountered in clinical practice. For instance, it's been documented that V/Q planar has a low sensitivity for identifying perfusion defects in medial segments. Furthermore, other medical conditions like pneumonia, pleural or pericardial effusions, cardiovascular disease, pneumothorax, or malignancies can lead to V/Q mismatches, resulting in false positives. If the data from CXR is not included as input for ANNs, abnormal CXR cases can mistakenly be interpreted by ANNs as PE. Research by Bailey and his team has already established the value of SPECT/CT images where functional data is combined with anatomical structures. This not only improves the diagnostic accuracy of V/Q scintigraphy for PE, but also helps clinicians with differential diagnoses as mentioned earlier. Therefore, there's scope for future CNNs to incorporate CXR or CT data along with V/Q scintigraphy to enhance prediction accuracy.

3.4.6 Other Pathologies

Apart from their conventional applications, like surgical planning, V/Q scans have been recently showing promise in identifying new indications in other diseases, such as CTEPH and COPD. These are conditions where AI could potentially play a significant role. For instance, CTEPH, often resulting from residual PE, leads to an increase in pulmonary arterial pressure over time, which ultimately results in right ventricle failure and high mortality rates. A recent study showed that applying PIOPED criteria with a threshold of 2.5 segmental mismatched perfusion defects for diagnosing CTEPH resulted in a sensitivity of 100% and a specificity of 94.7%. This suggests that AI, already employed in detecting segmental and subsegmental mismatches on V/Q scintigraphy, can be repurposed with ease to diagnose CTEPH.

COPD, another lung disease with growing prevalence, is associated with sustained respiratory airflow limitation, often due to exposure to harmful aerosols that lead to obstructed airways. Moreover, patients suspected of PE tend to be more susceptible to COPD. One common method for assessing COPD severity is the Global Initiative for Chronic Obstructive Lung Disease (GOLD) criteria, which is based on the forced expiratory volume in 1 second and forced vital capacity measured using a spirometer. Depending on the severity of the airflow limitation, COPD is categorized into four GOLD stages. Higher stages of COPD result in more severe inhomogeneity of ventilation, which can be visually identified on V/Q scintigraphy. In fact, one recent study found a strong correlation between the penetration grade of ventilation agents and total preserved lung function as measured with V SPECT to GOLD stages. Luckily, CNNs are proficient at texture analysis and can therefore extract features related to the heterogeneity, penetration grade of

ventilation agents, or lung function found in V/Q images and ultimately link them to the severity of COPD.

3.4.7 Challenges and Limitations

In the process of developing and testing AI, it's essential to have definitive and objective assertions about the AI algorithms to confirm their effectiveness in medical scenarios. The Society of Nuclear Medicine and Molecular Imaging has recently proposed best practices through the RELIANCE (Recommendations for EvaLuation of AI for NuClear medicinE) guideline to boost faith in the usage of AI in clinical contexts. A proper claim should specify the clinical task, the patient demographics, the imaging process in use, techniques to gather specific information related to the task, and the performance metrics of the algorithm on the given clinical task.

For instance, a count-enhancement algorithm's primary aim during training would be to reduce the mean square error between the model's predictions and the reference images. However, the most crucial aspect is ensuring that the model's predictions do not modify the diagnosis or patient treatment based on the original full-dose or full-length acquisition. This could occur, for example, by introducing false positives or by eliminating perfusion defects linked to PE. Sticking to these evaluation best practices, such as carrying out further clinical evaluation studies, will assist in determining if these algorithms are suitable for clinical application, give healthcare professionals the confidence to use them, and ultimately raise the standard of care.

It's critical for AI developers to be aware of the need to back up clinically relevant claims, as these challenges may at times surpass the complexity of developing the AI itself. A possible criticism of this work could be that we might have overlooked important contributions in this field. To tackle this concern, we carried out a systematic review of a leading database in the domain of medical imaging, Ovid MEDLINE. Nonetheless, we recognize that there is no absolute assurance.

3.5 Conclusion

Some of the initial efforts in applying AI to medical imaging were centered around lung scintigraphy, and these efforts were quite promising in terms of potential clinical applications. However, this area was largely overlooked later, possibly because of the concurrent emergence and dominance of CTPA as an alternative technique. Given the advancements in AI technologies, the continued use of scintigraphy for diagnosing PE in certain patient groups, and the discovery of new applications such as CTEPH and COPD, there are compelling reasons to revisit and reignite research into the use of AI in lung V/Q scintigraphy.

Chapter 4: Resizing of Scintigraphic Images

I am the first author of the article:

> **Ghassel S**, Jabbarpour A, Lang J, Moulton E, Klein R. "The effect of resizing on the natural appearance of scintigraphic images: an image similarity analysis," *Phys Med Biol*, 2023. *(In Review)*

The article informs about the proper methodology of resizing scintigraphic images so as to preserve image noise properties. The content was slightly modified from the manuscript to fit within the context of this thesis.

4.1 Introduction

Image resizing is a common image processing operation for resampling an image from one grid size to another [103]. When the image is upsampled (the pixel density increases), a choice of interpolation methods may be applied, the most common of which are nearest neighbor, bilinear, bicubic, and b-spline. When the image is downsampled (the pixel density decreases), standard procedure recommends applying a low pass filter to prevent aliasing [69]. While each interpolation method tries to maximize the similarity of the destination to the source image [104], they differ in how the pixel values in the neighborhood of the source coordinate are combined to calculate the final value at the destination coordinate.

In almost all imaging modalities, the process of resizing may not substantially alter the semantic nature of the image [105]. However, in the case of nuclear medicine scintigraphy, where the native image unit is the number of detected events (i.e., photon counts) [106], Poisson counting statistics play a visually perceivable and mathematically significant role in the image noise [107]. As dictated by Poisson counting statistics, the variance in the signal is equal to the mean (expected true counts) of the sample; hence, the relative noise decreases as one over the square root of the mean counts as shown in Equation 10.

$$relative\ noise = \frac{standard\ deviation}{mean} = \frac{\sqrt{\mu}}{\mu} = \frac{1}{\sqrt{\mu}} \qquad \textit{Equation 10}$$

Total photon counts, and thus photon density, need to be conserved if downstream operations are dependent on accurate photon counts or noise modeling. Since resizing intrinsically modifies the number of pixels and pixel spacing [108], the resulting resized image should reflect the splitting or joining of counts from the original image in the target image when upsampling or downsampling, respectively. Accurate accounting of counts models what the image would have looked like had the image been acquired at the desired spatial grid and corresponding pixel spacing, including the magnitude of the noise in each pixel [109].

On one hand, naïve interpolation during upsampling naturally gives rise to a larger sum of counts – due to an increased number of pixels – approximately by a factor of the upsampling ratio. It is

therefore essential to reduce the excess photon counts while maintaining the natural Poisson noise of scintigraphic images. In fact, White and Lawson [110] have demonstrated that Poisson resampling is the appropriate technique for artificially reducing counts in scintigraphic images. While this technique was originally intended to generate synthetic low-count scintigraphic images from high-count ones, this method can be reasonably repurposed to correct for the added collateral counts introduced during upsampling.

Downsampling, on the other hand, is analogous to acquiring an image on a smaller spatial grid with a larger pixel size. Properly implemented, this operation effectively corresponds to the summation of photon counts within a sliding window whose size is given by the downsampling ratio. This method will necessarily conserve count statistics and Poisson noise of the resulting image mimicking how a gamma camera would have aggregated photon counts within a larger pixel size. In more concrete terms, if we use the example of a 256×256 image to be resized to 128×128, the resulting image should be the same as if the image had been natively acquired on a 128×128 imaging grid (within the acceptable limits of random noise associated with two independent image samples). Each resulting pixel is thus expected to have 4 times more counts (the sum of 4 pixels sampling the same corresponding image space) on average than the original image pixel. Examples of upsampling and downsampling effects are demonstrated in Figure 4.1 with simple examples. By performing simple linear interpolation, the total number of counts in the image are increased or decreased respectively by a factor of number of pixels (or their density). This can be corrected by factoring the pixel counts to generate count-preserved images. In the case of upsampling, the image appears smoothed, without local noise pattern. The noise and count-preserved images, apply Poisson resampling of each pixel count to reproduce the statistical noise for each pixel.

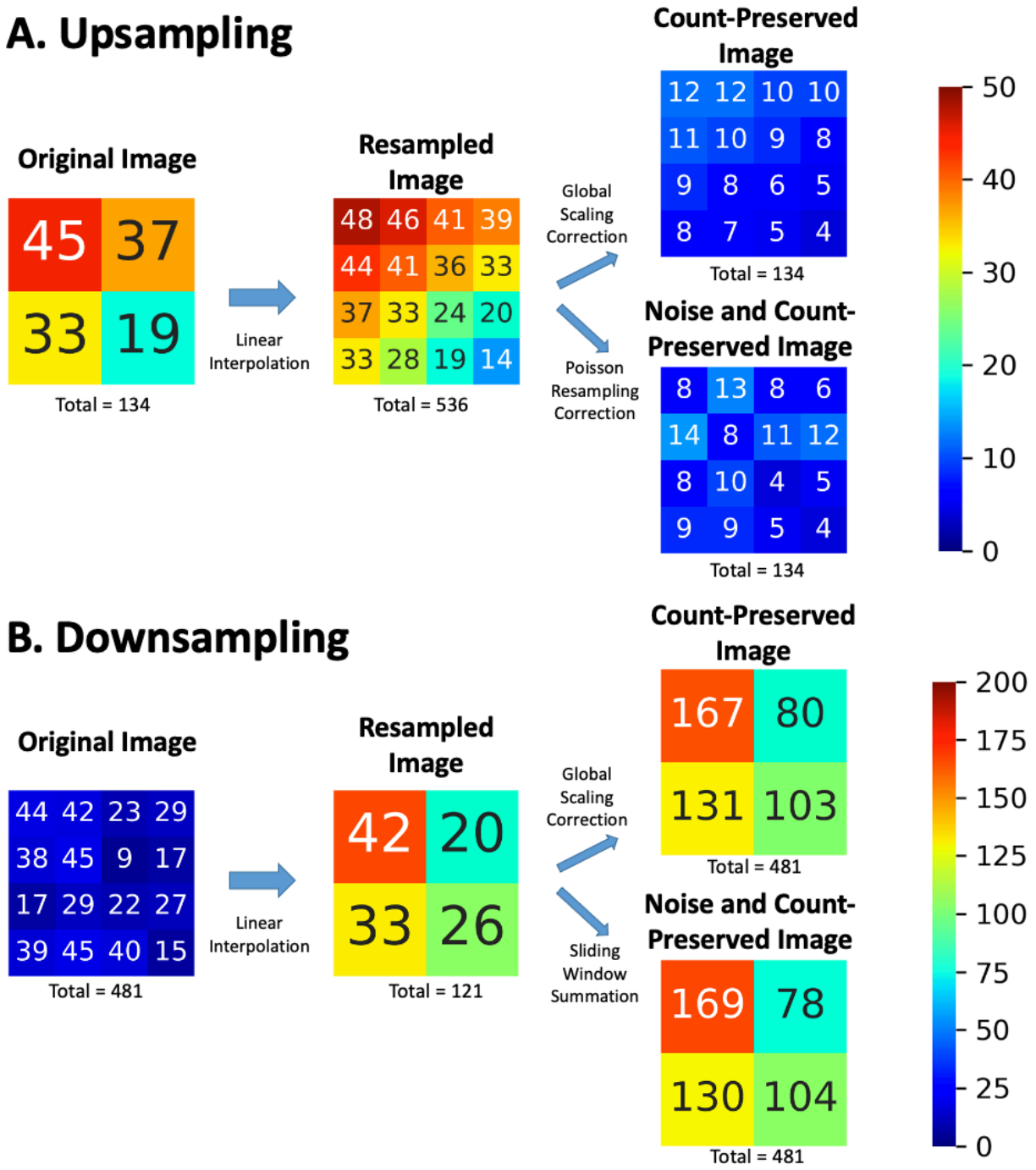

Figure 4.1- Example of (A) upsampling and (B) downsampling operations on nuclear scintigraphy images (original) with linear interpolation, showing the resulting images and incorrect total events. "Count-preserved images" obtained with a global scaling correction are contrasted with an ideal "Noise and Count-Preserved images" as if it had been acquired by an imaging system with the target pixel sizes and local noise associated with the increased or decreased counts per pixel. Numbers represent pixel intensity as event counts in the corresponding pixel and rounded to the nearest integer.

Joint count and noise preservation are paramount for many image processing investigations in nuclear medicine [10][11]. Currently, considering the immense work in AI model development in medical imaging [113][114], AI developers may circumvent the high variability in image sizes in real-life clinical settings by forcing a model's input images to a fixed spatial grid under the

assumption that these resized images reflect natively acquired ones on the destination spatial grid. This does not pose a problem for modalities such as MR or CT, where the content of the image does not change after resizing; however, such is not the case in nuclear medicine where trivial resizing operations introduce false pixel count and/or noise representations in images.

Given the above, we aimed to compare traditional resizing with naïve linear interpolation [103] against upsampling with Poisson resampling corrections or downsampling with sliding window summations. In particular, we sought to evaluate the effect of each of these resizing techniques on the similarity of the resulting images to real images acquired at the target spatial grid using real phantom data. Through this study, we seek to demonstrate inaccuracies resulting from naively applying traditional image resizing methods in nuclear scintigraphy and to establish a robust standard for scintigraphic image resizing for future research and developments.

4.2 Methods

4.2.1 Establishing reference similarity curves as a function of count level

In order to evaluate and compare resizing strategies, we postulated the following: (1) a successfully resized image of an object should exhibit the same content and noise characteristics as if the image had been acquired on the target spatial resolution, and (2) two images of a given object acquired with the same imaging protocol (i.e., at the same spatial resolution) will both have some intrinsic noise and therefore be similar up to a certain point; in other words, the images are not identical. Therefore, it follows that the similarity between a resized image of an object and another image of the same object at the target resolution should be equal to that between two independently acquired images of the object at the target resolution (Figure 4.2).

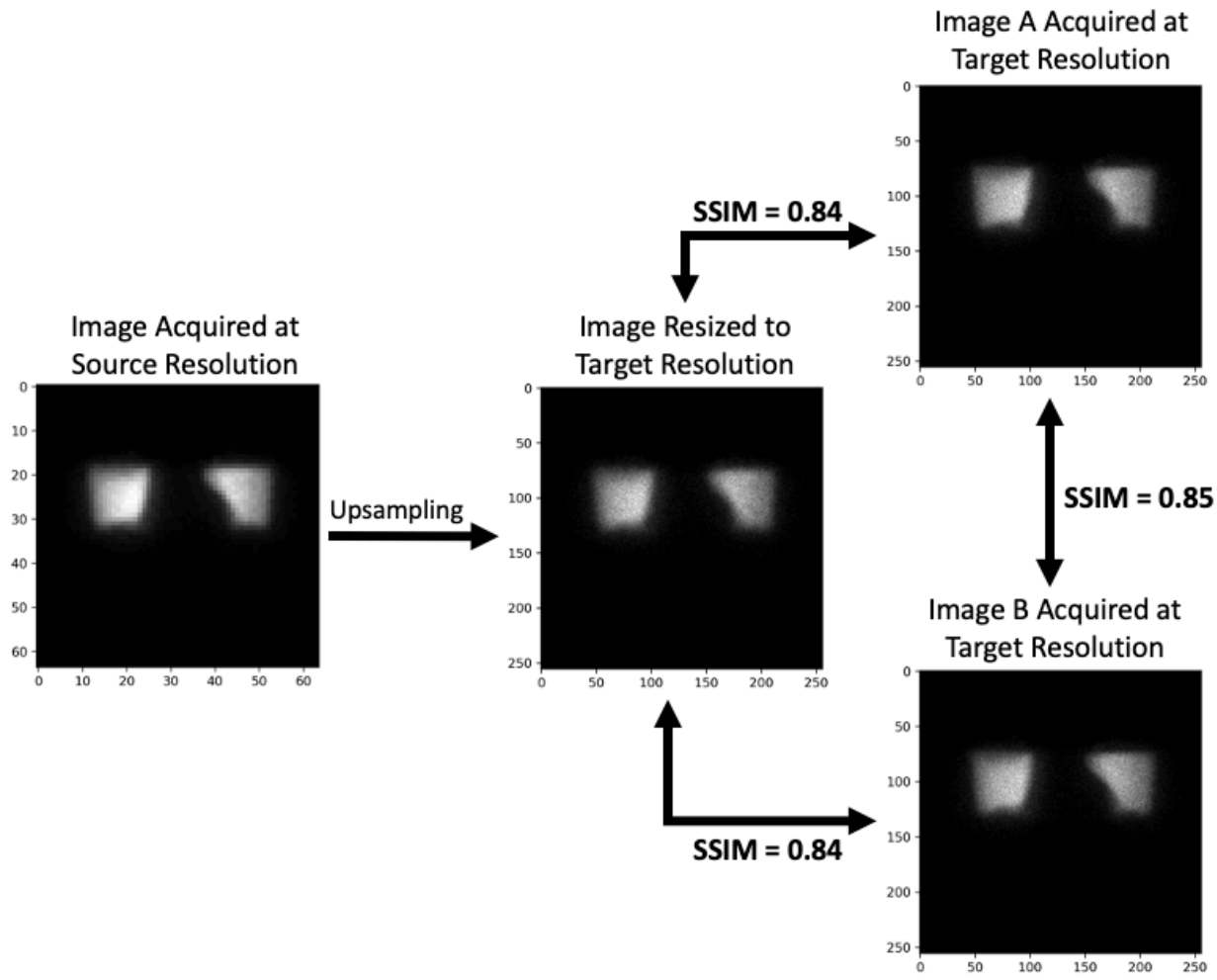

Figure 4.2- Postulate behind successful resizing of scintigraphic images. The similarity between the resized image and either image A or image B at the target resolution is approximately the same as the similarity between images A and B. If this is the case, then the image has successfully been resized. SSIM = Structural Similarity Index.

It is worth noting that no one image acquired at the target spatial resolution may serve as an absolute reference truth, as each image has some degree of inherent noise. Hence, our analysis rests upon measures of similarity between pairs of images throughout this work. To this end, we sought to derive curves establishing reference measures of similarity achievable between two independent, natively acquired images of the phantom for various count levels and spatial resolutions. Against these reference curves we could then compare similarity curves between resized images and those acquired at the target spatial resolution. Consequently, the best resizing method would yield curves that most closely overlaps the reference curves.

Time frames from the dynamic image series were randomly split and the images were summed to generate statistically independent images ranging from 11 to 550 kcnts (Figure 4.3). Various similarity metrics between the "full count" (550 kcnts) image and the reduced count (11-550 kcnts) images were calculated for: (1) two images acquired at the same spatial resolution, and (2) a resized image at a target spatial resolution and another natively acquired image at the same resolution. By summing combinations of randomly selected dynamic frames, we generated 1000 independent

permutations of images with similar number of counts and thus could bootstrap the reference similarity metrics to generate averages and confidence intervals. Reference similarity curves were generated as a function of count level between two images acquired at the same spatial resolution.

To evaluate the proposed method with the highest achievable similarity of the target image, we calculated the structured similarity index (SSIM) [67] and the logarithm of the mean squared error (MSE). Prior to computing similarity metrics between pairs of images, each image was normalized to the maximum intensity pixel so as to (1) be more sensitive to the statistics of the image (i.e., contrast, noise, texture, etc.) instead of the actual count values of the image and (2) be able to compare results between spatial resolutions [115].

4.2.2 Establishing reference similarity curves with synthetic low-count images

We also derived a second reference curve by first generating pairs of independent scintigraphic images with the highest possible count value given our dynamic acquisition (here, 550kcnts) and then synthesizing low-count versions of them with Poisson resampling (Figure 4.3). While it has already been demonstrated that Poisson resampling effectively yields low-count versions of high-count images that preserve the natural noise characteristics of single photon emission scintigraphy [110], we decided to verify that two low-count Poisson resampled images would also be as similar to each other as two native low-count images. By demonstrating this, we can provide confidence that Poisson resampling is indeed a reliable technique to simulate low-count images when dynamic acquisitions are not available – which is the case in most clinical settings – and that they can be combined with appropriately resized images from other spatial resolutions.

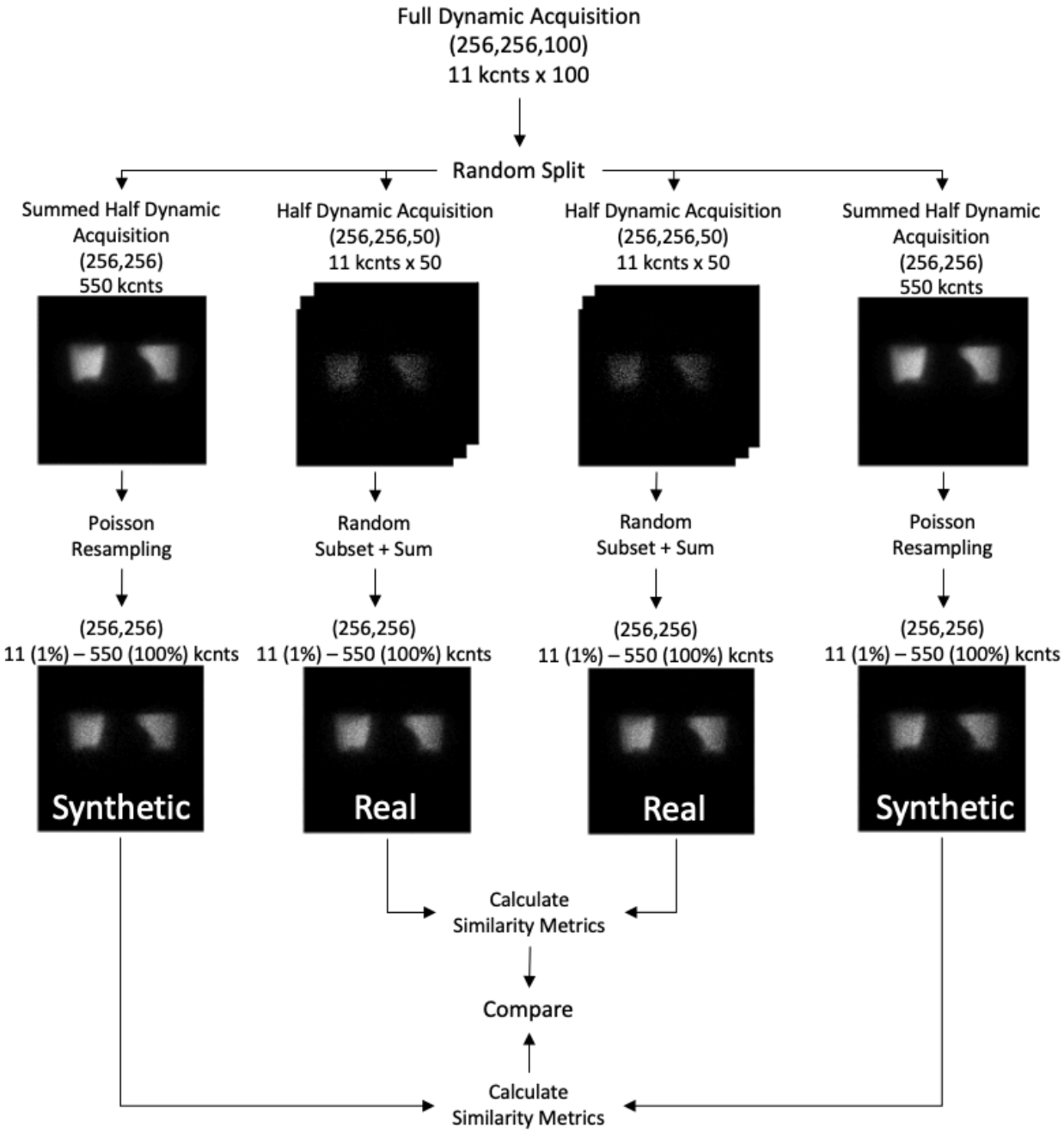

Figure 4.3- A simulated full dynamic acquisition conducted on a 256x256 grid size to create real and synthetic low-count images.

4.2.3 Upsampling Method

Next, we compared the similarity of upsampled images with and without Poisson resampling corrections with the images of matched count level natively acquired on the target spatial resolution (Figure 4.4). First, we reconstructed an image at a given count level at the lower spatial resolution by randomly selecting and summing the appropriate number of frames from the dynamic sequence. Second, we resized the image at the lower spatial resolution using linear interpolation to attain the higher target spatial resolution. The resized image was then rescaled by a factor

corresponding to the increased pixel density (e.g., for 64×64 to 256×256 upsampling the rescaling factor was $(64/256)^2 = 1/16$) so as to preserve the total number of counts in both images. Third, we applied a per pixel Poisson resampling correction to the resized image to simulate the counting statistical noise that would have been present at the per pixel target count level. Briefly speaking, a Poisson resampling correction comprises of resampling all pixels of the image using a binomial distribution where the initial pixel value constitutes the number of trials and the probability of success in our case is given the rescaling factor. The method is detailed in [110]. Fourth, for each upsampled image, we computed the image similarity metrics between it and another randomly reconstructed image natively acquired at the target spatial resolution. This process was repeated 1000 times with the dynamic acquisition images of the lower spatial grid to bootstrap confidence intervals of the image similarity curves as a function of count level.

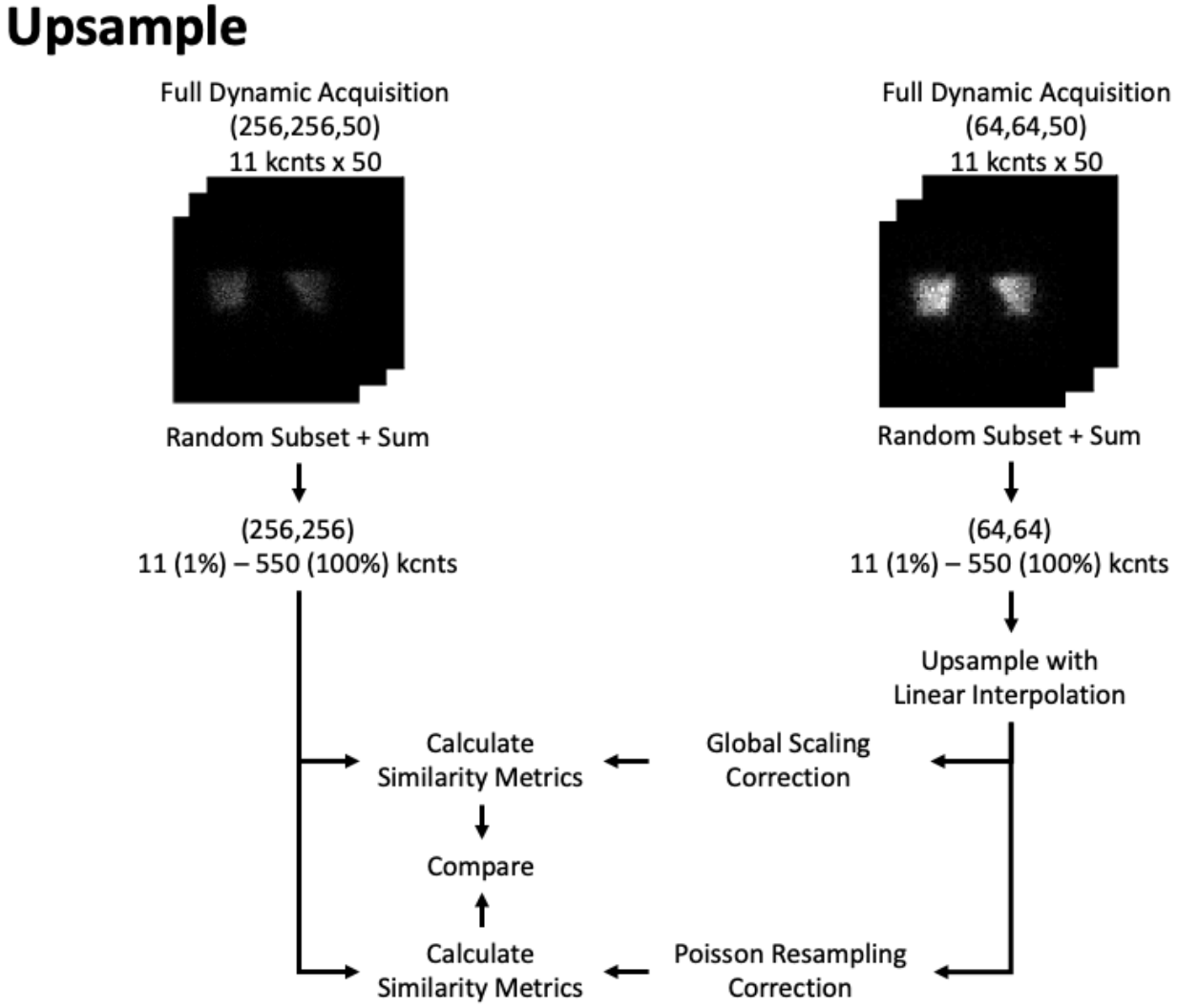

Figure 4.4- Upsampling method. Example workflow resizing from a 64×64 to a 256×256 grid size using 2D phantom planar dynamic acquisition. The similarity metrics of both methods are compared against pre-determined reference similarity curves at the target (256×256) resolution.

4.2.4 Downsampling Method

Similar to the upsampling experiment, we compared the similarity of downsampled images with either linear interpolation or a sliding window summation method with images of the matched count level natively acquired on the target spatial resolution (Figure 4.5). We started by producing an image of a given count level by randomly selecting and summing the appropriate number of frames of the dynamic acquisition. The first method was to downsample the images by the appropriate factor with linear interpolation to the lower spatial grid. The second method was a sliding window summation which consisted of applying a 2×2 or a 4×4 non-overlapping sliding window across the higher resolution image and computing the sum of counts therein. The image similarity metrics were calculated between each downsampled image and another randomly reconstructed native image at the target spatial resolution method. This process was also repeated with 1000 bootstrap samples to generate mean and confidence intervals as a function of count level.

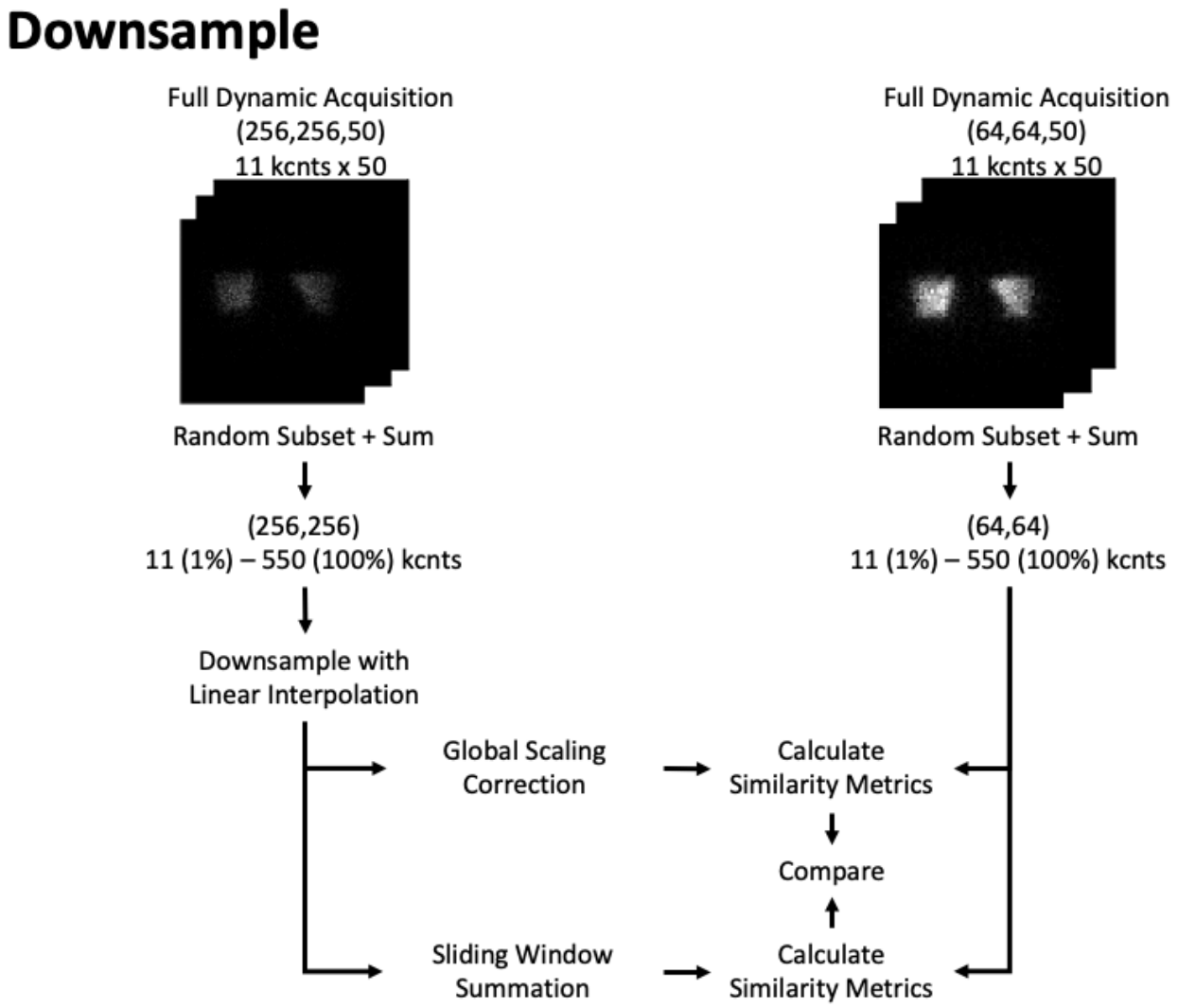

Figure 4.5- Downsampling method: Example workflow resizing from a 256×256 to a 64×64 grid size using 2D phantom planar dynamic acquisition. The similarity metrics of both methods are compared against pre-determined reference similarity curves at the target (64×64) resolution.

4.3 Experiments

4.3.1 Phantom

This study used real planar scintigraphic images acquired using a physical phantom and was performed by Ran Klein and myself. The acquisition protocol was designed to emulate a pulmonary ventilation-perfusion (V/Q) scintigraphy exam using a specially designed Data Spectrum Anthropomorphic Torso Phantom with custom dimensions 45×33 cm that simulates anatomical structures and physiological parameters relevant to nuclear lung scans [116]. The phantom included partial (superiorly truncated) lung cavities which were filled with Styrofoam beads to emulate the low density of air-filled lung tissue. 779 MBq of Technetium-99m (^{99m}Tc)-

pertechnetate were diluted into a batch of approximately 500 mL of tap water, which was then used to fill the space between the Styrofoam beads in the lung cavities. All remaining phantom cavities (thorax and liver) were filled with tap water to emulate soft-tissue attenuation.

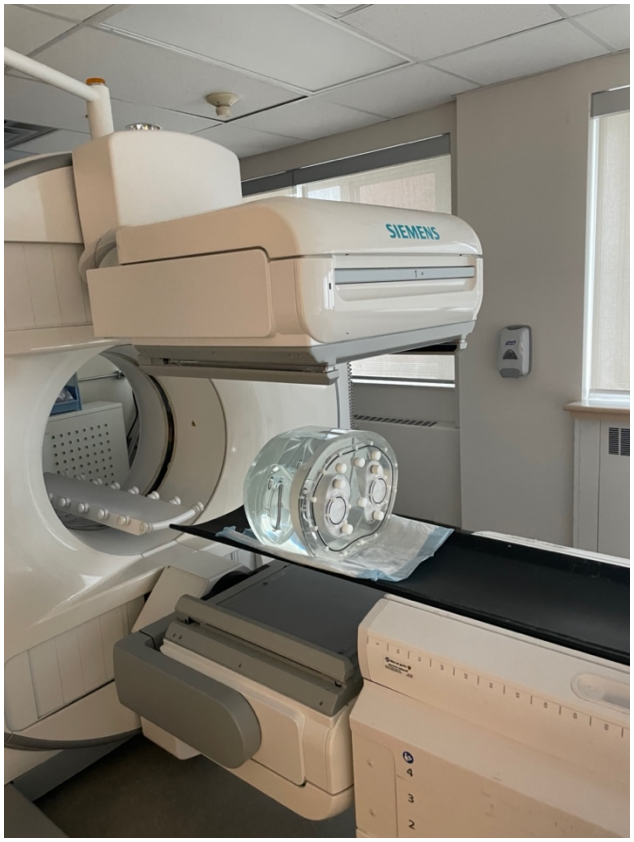

Figure 4.6- Imaging of the Data Spectrum Anthropomorphic Torso Phantom inside the Siemens Intevo Bold SPECT/CT. The phantom, tailored with dimensions of 45×33 cm, represents larger patient sizes and anatomical structures pertinent to nuclear lung scans.

4.3.2 Image Acquisition

Phantom images were acquired with a zoom factor of 1.45 at different spatial grids of 256×256, 128×128, and 64×64 with pixel spacings of 1.64 mm^2, 3.29 mm^2, and 6.59 mm^2, respectively. To easily generate planar images at various count levels, we performed dynamic acquisitions comprised of 100 frames of 1 second duration each, resulting in about 11 kcnts/image for a total of 1.1 Mcnts over the whole dynamic acquisition. Dead time was <3% to minimize the impact of count pile up. Each of the aforementioned acquisitions was performed for the typical six views of a V/Q scan: anterior (ANT), posterior (POST), left anterior oblique (LAO), right posterior oblique (RPO), left posterior oblique (LPO), and right anterior oblique (RAO). All acquisitions were performed in quick succession (within 1 hour) to minimize radioactive decay between image sets.

Data for the study were acquired with a dual head Siemens Intevo Bold SPECT/CT using low energy high resolution collimators. The energy window was set to 140 keV ± 7.5%, corresponding to the photon peak energy of ^{99m}Tc.

4.4 Results

4.4.1 Visual Inspection of Upsampled and Downsampled Images

Figure 4.7 illustrates the effect of the each upsampling and downsampling methods on image similarity with respect to real and synthetic scintigraphic images. As can be seen, upsampling with naïve linear interpolation maintains the original increased contrast of high-to-low count areas of the lower resolution image, which is less pronounced on the higher resolution image. The application of a Poisson resampling correction visually seems to restore the natural contrast of the image and re-introduce the typical Poisson noise of the real and synthetic images of the target higher resolution grid.

Downsampling with either linear or a sliding window summation seems to yield images with similar contrast of high-to-low count areas. However, the noise of downsampled with the sliding window appears more similar to both real and synthetic images of the target lower resolution grid.

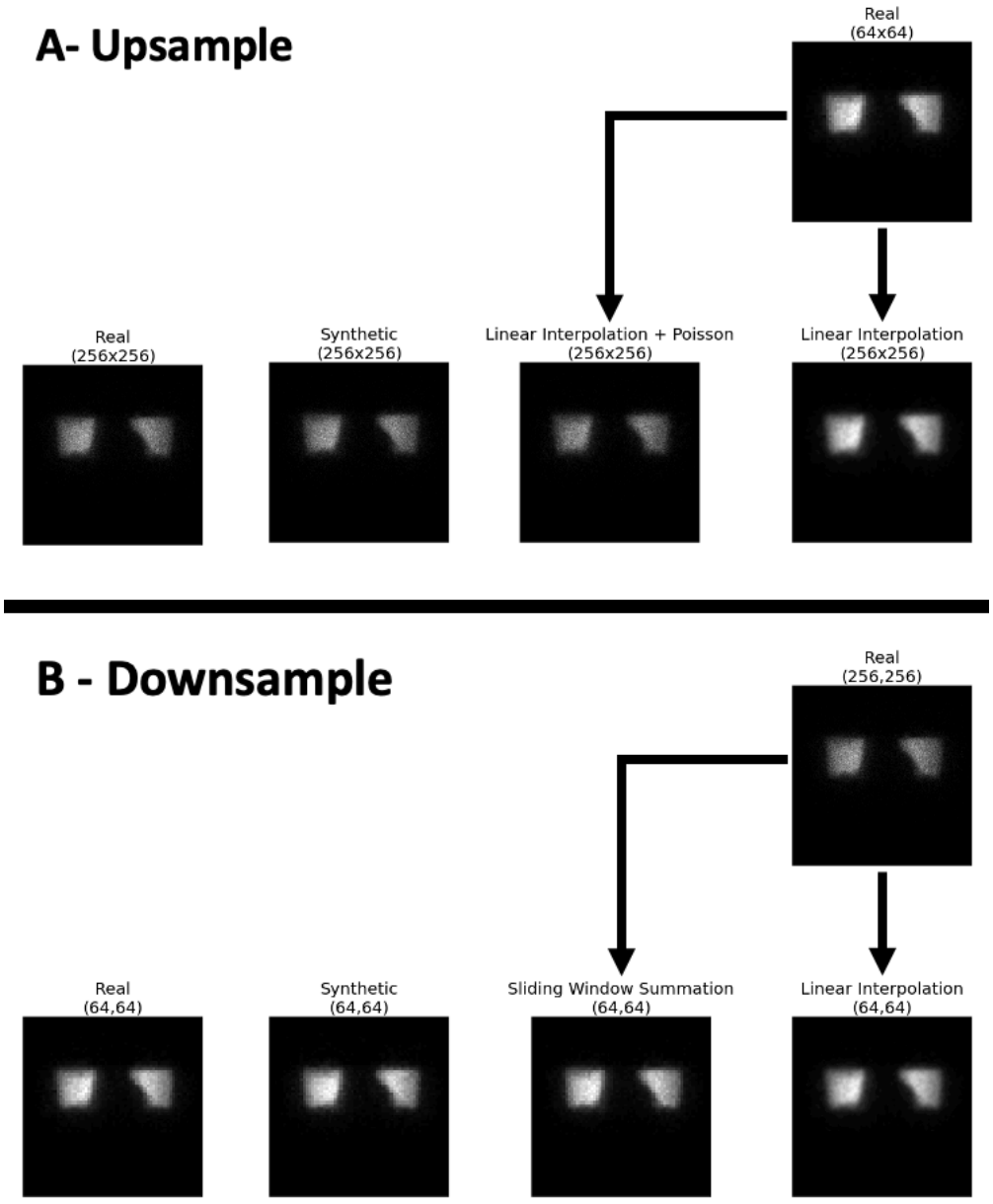

Figure 4.7- An illustrative example of the effect of the various upsampling (A) and downsampling (B) methods on image similarity with respect to real and synthetic scintigraphic images at approx.. 105 kcnts. (A) From left to right: a Real (256×256) reference image reconstructed from the dynamic acquisition, a Synthetic image obtained from Poisson resampling of another 550 kcnts image at the target spatial resolution (256×256), and a Real (64×64) 550 kcnts image that was upsampled with linear interpolation and Poisson resampling correction, and the same Real (64×64) image upsampled with only linear interpolation. (B) From left to right: a Real (64×64) 550 kcnts reference image reconstructed from the dynamic acquisition, a Synthetic image obtained from Poisson resampling of another 550 kcnts image at the target spatial resolution (256×256), a Real (256×256) 550 kcnts image downsampled with sliding window summation, and the same Real (256×256) downsampled with only linear interpolation. All images were rendered in grayscale, with their intensity levels scaled according to their respective minimum and maximum values.

4.4.2 Agreement Between Similarity Curves for Real and Synthetic Images

The reference curves for all similarity metrics from real and synthetic images overlapped nearly perfectly across all projections, count levels, and spatial resolutions, confirming that the synthesized (count reduced) images accurately modeled the actual count reduced images (Figure 4.8). A general finding also was that image similarity increased (i.e., higher SSIM and lower MSE) as the spatial resolution of the scintigraphic images decreased.

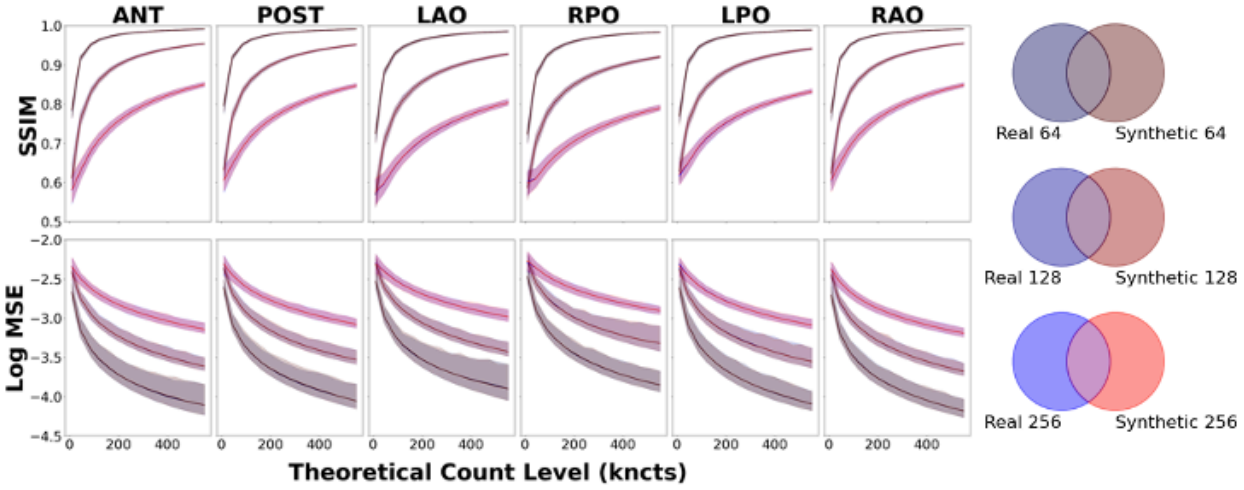

Figure 4.8- Evaluating similarity metrics SSIM and Log MSE using a dynamic phantom acquisition on six common 2D planar projections: ANT, anterior, LAO, left anterior oblique, LPO, left posterior oblique, POST, posterior, RAO, right anterior oblique, and RPO, right posterior oblique on spatial grids of 64, 128, and 256 pixels. Real curves were generated by summing pairs of scintigraphic images up to highest possible count value of 550 kcnts. Synthetic curves were created by synthesizing low-count versions of 550 kcnts using Poisson resampling correction.

4.4.3 Effect of Resizing Methods on Image Similarity

Upsampling with naïve linear interpolation yielded image similarity curves that deviated significantly from the reference curves for real and synthetic data regardless of the source or target spatial grid, count level, and projection, as shown in (Figure 4.9). In particular, when compared with scintigraphic images natively acquired on the target spatial grid, upsampled images with naïve linear interpolation produced higher SSIM and higher MSE than the reference curves for real and synthetic data. The deviations were more marked (i.e., less overlap of the confidence intervals) when upsampling from either 64×64 or 128×128 to the highest grid of 256×256, whereas there was more overlap of the similarity curves when upsampling from 64×64 to 128×128 (Figure 4.9, Figure 4.11, Figure 4.12), However, following the Poisson resampling correction, the similarity curves realigned with the reference curves both with respect to the mean and confidence intervals for each target spatial grid, count level, and projection.

With regards to downsampling, the most striking result was that when resizing by a factor of 2 (i.e., from 256×256 to 128×128 or from 128×128 to 64×64), linear interpolation and sliding

window summation methods yielded similar similarity curves, both of which overlaid nearly perfectly on the reference curves (Figure 4.13, Figure 4.14). However, when downsampling from 256×256 to 64×64, naïve linear interpolation yielded similarity curves that significantly deviated from the reference curves (Figure 4.10). In this case, when compared with scintigraphic images natively acquired on the target spatial grid (64×64), downsampled images with naïve linear interpolation resulted in lower values for SSIM as well as higher MSE with respect to the reference curves for real and synthetic data. Sliding window summation, on the other hand, produced a high level of agreement with real and synthetic data reference curves.

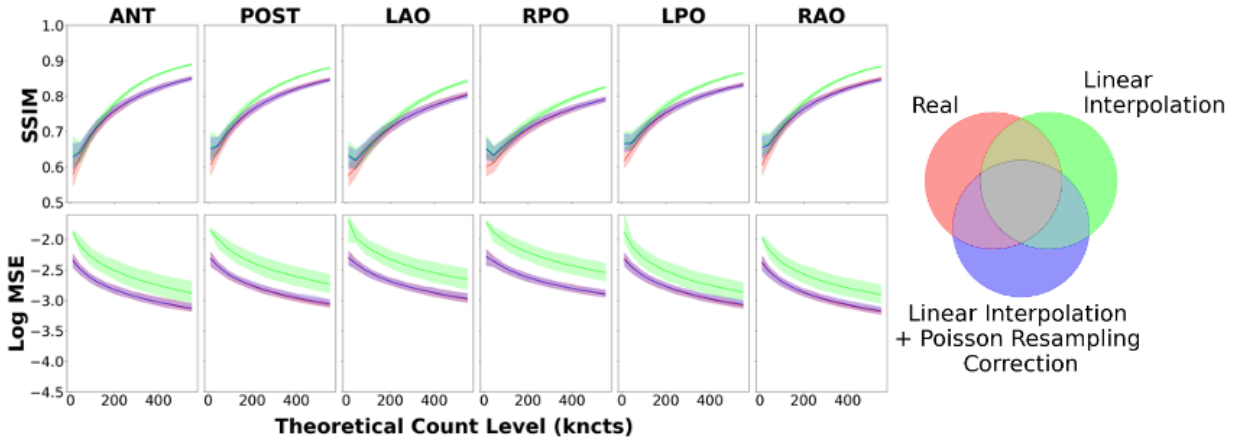

Figure 4.9- Image similarity metrics for 64×64 images upsampled to 256×256 with reference curves from the real data.

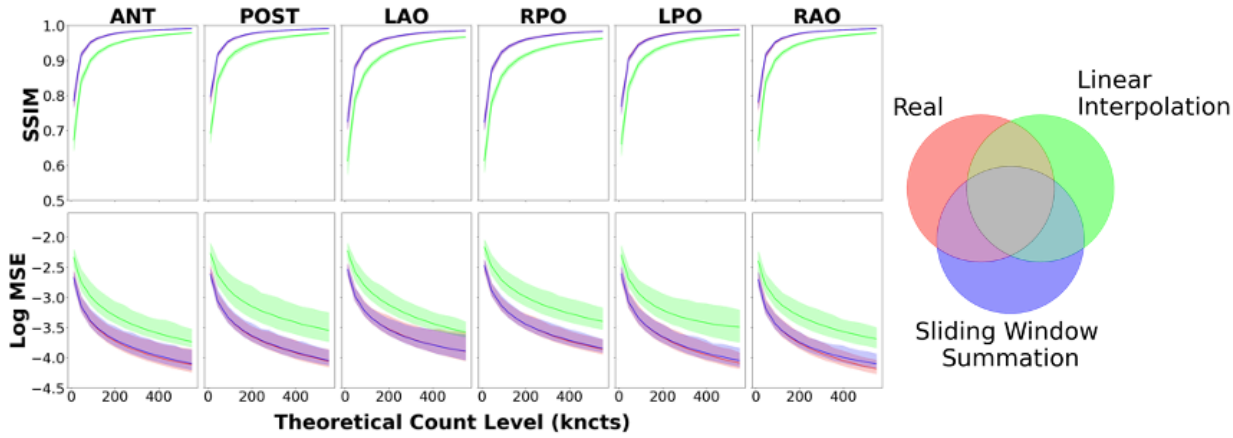

Figure 4.10- Image similarity metrics for 256×256 images downsampled to 64×64 with reference curve from the real data.

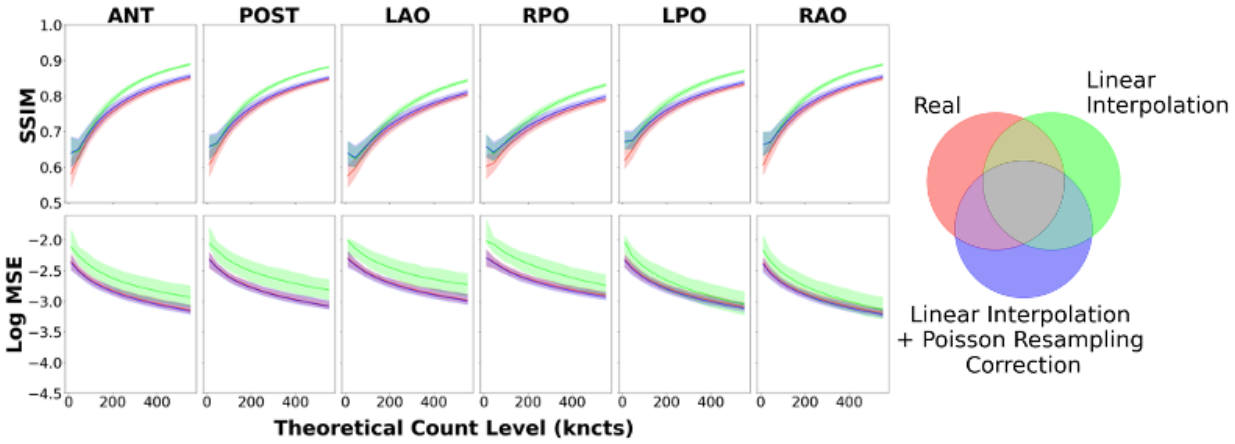

Figure 4.11- Image similarity metrics for 128×128 images upsampled to 256×256 with reference curve from the real data.

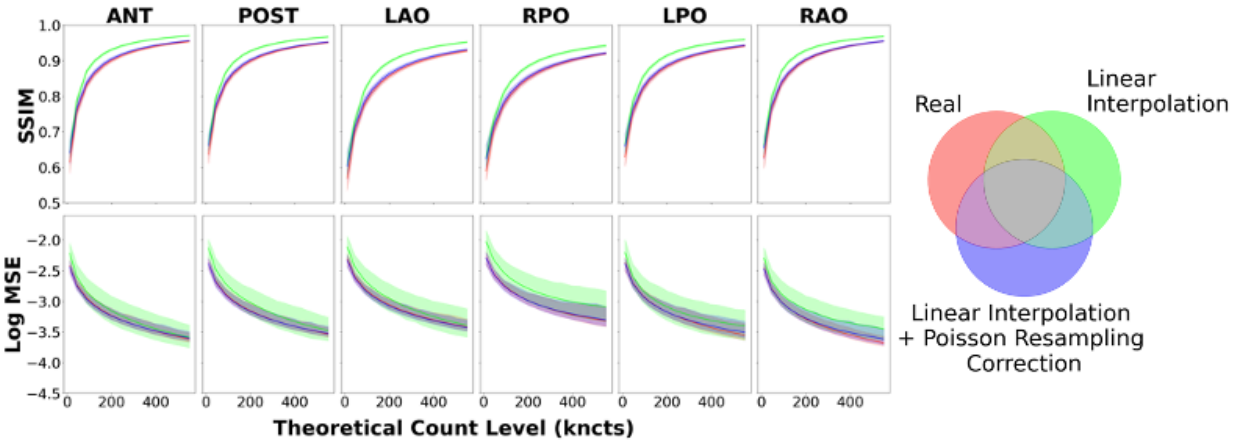

Figure 4.12- Image similarity metrics for 64×64 images upsampled to 128×128 with reference curve from the real data.

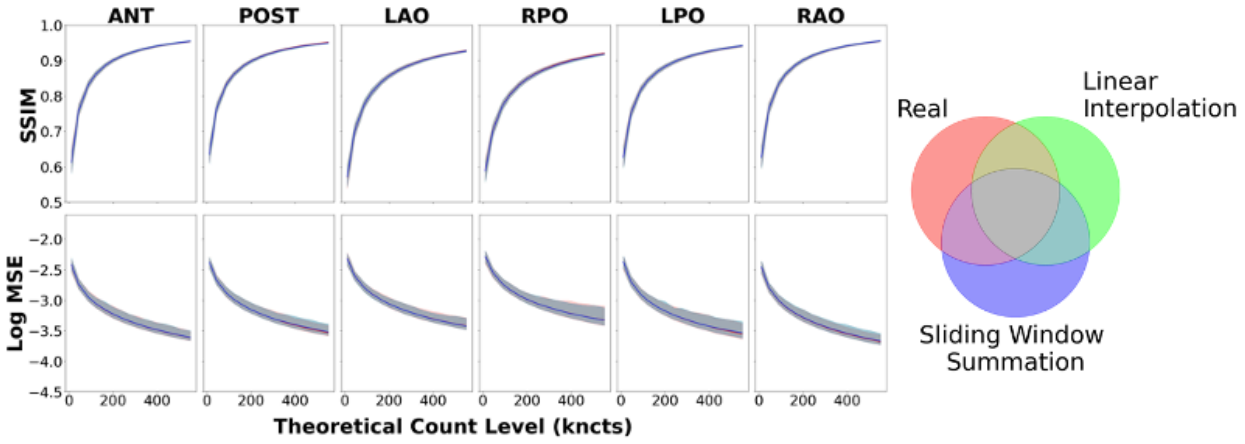

Figure 4.13- Image similarity metrics for 256×256 images downsampled to 128×128 with reference curve from the real data.

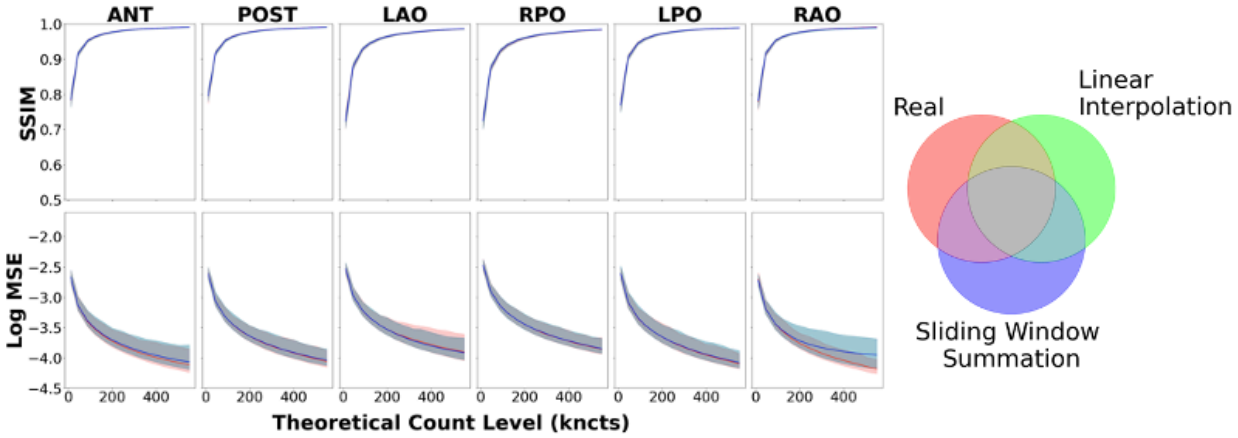

Figure 4.14- Image similarity metrics for 128×128 images downsampled to 64×64 with reference curve from the real data.

4.5 Discussion

In this study, we investigated methods of image resizing in the context of nuclear medical imaging in which the native image units represent counts of detected events and are governed by Poisson statistics [107]. This exploration is important to preserving the underlying noise and statistical properties of scintigraphic images [110], particularly when they are the focal point of analysis. By examining upsampling and downsampling techniques used for resizing and their impact on preserving image noise properties, our goal was to guide the research and development community on proper methodology. While potentially applicable to many image processing applications, we believe these lessons are particularly applicable to nuclear medical imaging where image pre-processing frequently includes image resizing.

4.5.1 Upsampling

In the case of image upsampling, we demonstrated that naïve application of linear interpolation violates inherent noise characteristics of scintigraphic images. Specifically, as images undergo greater degrees of upsampling (i.e., transitioning from a 64×64 to a 256×256 (Figure 4.9) grid as opposed to 128×128 to 256×256 (Figure 4.11)), the difference in similarity becomes more evident. These results can be partly explained by the mechanism by which linear interpolation operates [104]. To the first order, linear interpolation by a factor of 2 averages pixel intensities in a small vicinity, close to a neighborhood of 2×2 pixels. However, as the scaling factor increases (64×64 to a 256×256) (Figure 4.9), linear interpolation no longer includes in its average those pixels that are further than the immediate vicinity from the center of the interpolated pixel.

Furthermore, it is evident that solely relying on interpolation does not provide an accurate representation of the target resolution image, especially in relation to its noise characteristics. Hence, the recommended method for image upsampling is as follows: first, resize the image with linear Interpolation to the new sampling grid; second, apply a Poisson resampling correction by

resampling the linearly interpolated image, rounded to the nearest integer, with a binomial distribution where the interpolated integer pixel value constitutes the number of trials and the probability of success in our case is given the by the ratio between the old and new pixel spacing.

Perhaps surprisingly, interpolation-based upsampling resulted in higher SSIM values (with target images) than comparing two images acquired at the target sampling grid (Figure 4.9, Figure 4.11, Figure 4.12). This may be explained by the interpolation being performed from larger pixels, with more counts and less relative noise being reused to derive many small pixels in the upsampled image. Consequently, the local variance component of the SSIM equation in the upsampled image is reduced (as confirmed in Figure 4.7A, bottom-right) leading to inflated SSIM measures. However, counterintuitively, in this scenario a higher SSIM does not correspond to a more realistic upsampled image, simply because it deviates too much from the SSIM between two images acquired at the target resolution, which goes against our postulate of a successful resizing.

4.5.2 Downsampling

For downsampling images, we explored two techniques: linear interpolation [104] and sliding window summation [117]. Both methods yielded similar similarity metrics (and in agreement with reference values) when downsampling by a factor of 2 (i.e., from 256×256 to 128×128 (Figure 4.13) or from 128×128 to 64×64 (Figure 4.14)). However, when downsampling images by a larger factor (i.e., from 256×256 to 64×64 (Figure 4.10)) linear interpolation deviated from the reference similarity trends. The observed outcomes can be attributed, in part, to the principles governing linear interpolation when downsampling by a factor of 2. In this scenario, the position of the pixels in the downsampled image will fall perfectly in the center of a 2x2 window in the original image, resulting in the average of all four values in the window. By introducing a global scale correction to the linearly interpolated downsampled image, we recover the result of using a 2x2 sliding window summation. But with an increasing scaling factor (as demonstrated in Figure 4.1, downsampling), linear interpolation begins to exclude pixels that aren't in the immediate proximity of the center of the interpolated pixel. In contrast, sliding window summation accounts for all pixels in the window for all sizes. Furthermore, since summing original image pixel values mimics how a gamma camera would have aggregated photon counts within a larger pixel size, the resulting image necessarily has preserved total image counts and the correct noise characteristics. Hence, the recommended procedure for downsampling nuclear scintigraphic images is to apply sliding window summation rather than resampling with interpolation.

4.5.3 Implications on Nuclear Image Processing

One domain where resizing of nuclear medicine images can have a profound impact is in AI development. Indeed, there has been a strong emphasis on leveraging precise and representative data for AI model training and evaluation. The pitfalls of relying on inadequately simulated or non-representative data have been underscored in current literature [118]. When machine learning models are trained on datasets that do not encompass the complexities of real-world scenarios, there is an inherent risk of these models yielding untrustworthy or inaccurate results [119][120].

This disjunction between training data and real-world samples can severely impede a model's proficiency in image interpretation, directly influencing clinical decisions and patient outcomes. Moreover, models fed with improperly resized scintigraphic images that are non-representative may demonstrate excellent performance during validation phases using similar datasets, showcasing high accuracy and precision. However, their performance might be compromised in real-world applications where similar resizing is not applied [121].

Image resizing can also play an important role in multimodal co-registration where a source image is often resampled to the target image's reference frame and thus pixel/voxel size. Therefore, in cases where planar scintigraphy images are the source image, care should be taken to adjust the counts in the resulting image to account for the change in pixel/voxel spacing. This could be relevant, for instance, in registering scintigraphy bone scan data onto x-ray images for anatomical localization of metabolic abnormalities to harmonize inter-patient observations [122]. This is especially crucial if photon count statistics (mean, max, standard deviation, etc.) were to be manually extracted from these registered images and used as features in a downstream machine learning or radiomics task [99][123] .

4.5.4 Limitations

Despite our unequivocal results on the proper methods for resizing nuclear medicine images, our study is not without its limitations. First, we investigated convenient resizing factors of 2 and 4 on the most commonly used sampling grids in nuclear medicine scintigraphic applications (i.e., 256×256, 128×128 and 64×64). Therefore, we are not able to provide informed recommendations on resizing beyond factors of 4 or by non-integer factors. While upsampling by non-integer factors can still be accomplished with the recommended Poisson resampling correction, the correct procedure for downsampling by non-integer factors was not demonstrated in our study. Second, our study utilized phantom images to simulate a lung perfusion scan; however, because our methodology was grounded in fundamental principles of Poisson counting statistics and scintigraphy image acquisition, this should carry over to any real-world clinical applications in nuclear medicine scintigraphy. Third, other interpolation techniques exist, but for the scope of this study, we focused primarily on linear interpolation due to its prevalent use in the field. Fourth, we did not directly assess the potential effects of resizing on 3D SPECT reconstructions, as there are many post-processing steps in reconstructing tomographic volumes from sinograms (such as spatial smoothing), that may invalidate our assumptions and methodology. In the case of PET imaging, since voxel units are Bq/mL, which is a measure of activity density, they do not directly encode event count, further complicating estimation of voxel of count statistics. Finally, we did not investigate how to adapt our recommendations to non-linear registrations where local parts of the image are shrunk and expanded independently. While in theory one could locally apply Poisson resampling corrections in enlarged areas or sum counts in contracted areas, this technical development was outside the scope of this study yet merits further attention.

4.6 Conclusion

Image resizing is a common process in medical imaging that many neglect to reflect on its finer nuances. We make the case that in the context of nuclear image scintigraphy, one must take care to adopt methods that preserve total image counts and maintain realistic image noise properties. This is crucial during the preprocessing step of neural network training where images are harmonized to a common grid size. We provide a recipe for simple upsampling and downsampling of scintigraphic images to enable the scientific community to properly perform image rescaling operations in practice.

Chapter 5: Count Enhancement of Scintigraphic Images

Scintigraphic image quality is directly related to the number of photons captured by the camera which is dependent on the following factors: camera efficiency, amount of radioactivity in the field of view, and image acquisition time. Increasing any of these factors will result in increased count statistics and hence reduced image noise. However, any such increase comes at a cost. Increasing camera sensitivity requires more imaging hardware, or collimators that favour sensitivity over spatial resolution. Increasing radioactivity is at the peril of patient exposure to ionizing radiation and cost. Finally, increasing image acquisition time is at the expense of patient comfort, risk of patient motion, and reduced clinical throughput. In this chapter we explore the prospect of increasing the *effective* sensitivity of the imaging system by post-processing the scintigraphic images to enhance their quality.

Traditional image processing approaches to reduce image noise utilized low-pass filters, at the expense of image spatial resolution and signal to background contrast. While more advanced approaches have been proposed, very few have achieved clinical utility, with “Pixon” image enhancement being perhaps the rare exception [124]. Pixon achieves a minor improvement in image enhancement equivalent to ~20% boost in effective sensitivity improvement, by subjective visual assessment at the TOH clinic.

Recently, the use of machine-learning image enhancement technologies has been proposed in the domains of PET with practical applications entering clinical routine [125]. Relatively little similar efforts have been committed to planar scintigraphy. Thus, in this chapter we explore the prospect of ML-based image count enhancement in scintigraphic images and speculate on its utility to shorten image acquisition times and to generate pseudo-planar images.

5.1 Methods

We designed an image enhancement algorithm that converts simulated low count planar scintigraphic images to diagnostic quality planar scintigraphic images (Generator in Figure 5.1). A low count image is akin to a low dose image since it is acquired for a shorter period. Inputs to the generator were simulated low count images that were generated from diagnostic quality images with 10% of the counts using Poisson resampling as described in [110].The corresponding, original diagnostic planar image was treated as a gold standard image. The Generator output represents a predicted full count image, which for training and validation was compared against the ground-truth images.

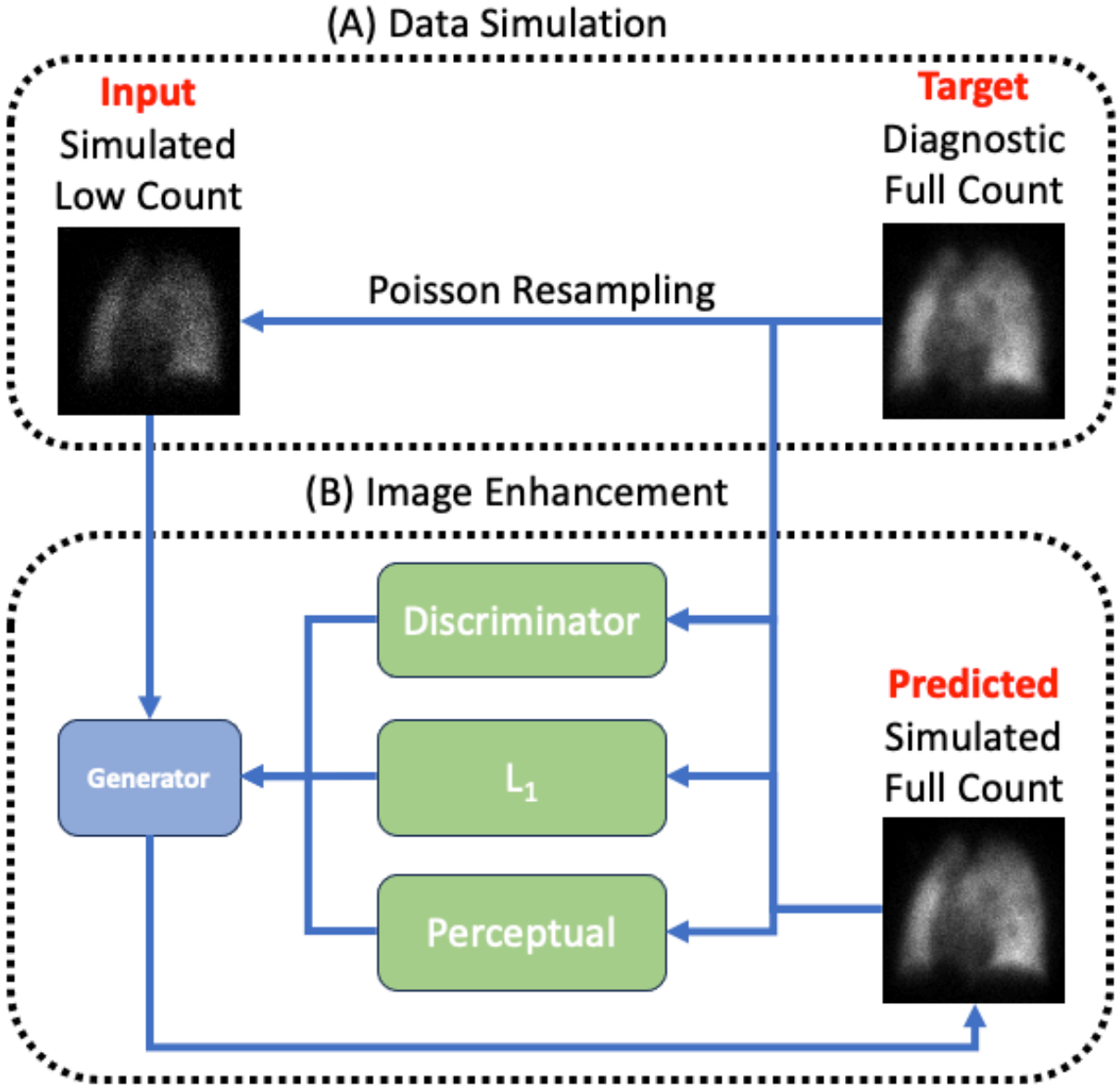

Figure 5.1- Proposed architecture of count enhancement algorithm that simulates low count scintigraphic images from full count perfusion planar with 10% counts using Poisson resampling. Images are enhanced (B) using a pix2pix architecture with a U-Net generator. The difference between predicted and target images were minimized by the following loss functions: Discriminator (GAN), L_1, and Percetpual.

5.1.1 Data Collection

For this study we used clinical images to construct an anonymized image database (approved by the Ottawa Health Science Network Research Ethics Board protocol #20220303). We included all lung V/Q studies performed at The Ottawa Hospital between June 2017 and January 2023 who underwent both perfusion planar and SPECT image acquisitions. Only the planar perfusion images were used in this study. Image acquisition consisted of 600k counts per image using a ±7.5% photopeak window centered at 140 keV on a 256×256 sampling grid. Consequently, image

acquisition times varied from ~60 seconds to as much at 5 minutes. Our data extraction yielded data from 704 patients with perfusion planar scans totalling 3900 images in up to 6 projection angles (anterior, posterior, right anterior oblique, left anterior oblique, right posterior oblique and left posterior oblique). Few studies only had anterior and posterior projections.

5.1.2 Data Preparation

All images were standardized in terms of subject position and image intensity. The center of mass of the image was determined and a corresponding image shift was performed to center the image signal (lungs) in the image space. Image intensity was normalized to [0,1] based on the pixel of maximum intensity. This standardization was performed to enhance model training performance [126][127].

The dataset was split into train (80%), validation (10%) and test (10%) using the Scikit-Learn library [128]. The projections of every patient were kept in the same split to prevent data leakage between splits. To optimize the training process, the data was batched in sets of 32 and shuffled; the buffer size was equated to the total data length. Additionally, data was prefetched to ensure efficient GPU utilization.

Corresponding low-count images were generated for each planar projection by simulating a 10% count image (i.e., a 90% count loss) using the method in [110] which was confirmed and detailed in Chapter 4:Resizing of Scintigraphic Images. I verified that these low-count simulated images maintained their structural integrity and no new defects were added or removed. This resulted in image pairs consisting of the original full count planar image (target) and the low-count simulation which are the candidates for image enhancement (input). Since the target full count planar images are acquired on a 256×256 grid at TOH, the input was also simulated on a 256×256 grid.

5.1.3 Model Implementation

The core architecture leverages the image-to-image translation model known as pix2pix, as detailed in section 2.4.2.2. The implementation closely aligns with the procedural framework outlined in the TensorFlow tutorial [107]. The pix2pix model employs two types of loss functions for training: generative adversarial loss (L_{GAN}) and L_1 loss. In this work, we wanted to see the effect of applying different loss functions on the generated image when using scintigraphic images as input.

The first loss function was the L_1 loss, a traditional approach in image processing. This loss function measures the absolute difference between predicted and target images, helping to maintain a similar global structure. L_1 loss has demonstrated effectiveness in various medical imaging modalities [129][93][130], enhancing the structural integrity of predicted images.

To complement L_1 loss, a GAN was added, which is increasingly popular in medical image translation tasks [131][132] and image enhancement tasks [93][133][129]. Following Isola et al.'s method [61], adversarial training was applied to transfer low-count images to corresponding full-count targets, enhancing the realism of predicted features. The inclusion of L_1 loss alongside adversarial training is critical, as adversarial loss alone cannot guarantee structural similarity with target full-count images.

Incorporating perceptual loss was another strategy to elevate image quality, to ensure perceptual resemblance to target full-count images. However, traditional pre-trained models like VGG [64], which are trained on ImageNet [134], fall short in capturing medically relevant features. To address this, a pre-trained ResNet50 model was utilized [56], trained on the RadImageNet database [135], comprised of 1.4 million medically relevant images. For content ($L_{content}$) and style (L_{style}) losses, we extracted layers conv5_block2_1_conv, and 'conv2_block3_3_conv', 'conv3_block4_3_conv', conv4_block6_3_conv', 'conv5_block3_3_conv' respectively. These specific layers were chosen for their role as bottlenecks in the ResNet50 model, based on findings in [136].

The integration of perceptual loss within the GAN architecture (2.4.2) is another approach that has shown promise in improving the quality of predicted images in PET [129] and MRI [133][137]. Therefore, we combined with L_1, GAN, and perceptual losses to optimize the image generation process.

The discriminator found in GANs often learns faster than the generator [138]. To mitigate this, a dual-phase training approach was adopted. For each training iteration of the discriminator, the generator was trained twice. This strategy was designed to maintain a balanced learning process, preventing either component from overpowering the other.

We explored four combinations of loss functions:

1- L_1
$L_{total} = L_{L1}$

2- L_1 + Perceptual
$L_{total} = L_{L1} + L_{content} + L_{style}$

3- L_1+ GAN
$L_{total} = \lambda_l L_{L1} + L_{GAN(G,D)}$ where $\lambda_l = 100$

4- L_1 + Perceptual + GAN
$L_{total} = \lambda_l L_{L1} + \lambda_c L_{content} + \lambda_s L_{style} + L_{GAN(G,D)}$ where $\lambda_l = \lambda_c = \lambda_s = 100$

It was crucial to ensure that images conformed to the input specifications of ResNet50 [56]. The native input shape for ResNet50 is (256,256,3), catering to RGB images. However, scintigraphic images are grayscale with a single channel. The single channel was copied and concatenated three times along the channel axis. Images were again normalized between [0,1]. For optimization, the Adam optimizer was utilized with a learning rate of 2.0×10^{-4} and a beta value of 5.0×10^{-1} for both the generator and discriminator.

5.1.4 Model Training

Training was conducted on NVIDIA GeForce RTX 3090 Ti, averaging 2.5 hours per loss function combination. During training, the impact of different loss functions on image generation was closely monitored. The objective was to evaluate how each loss function combination influenced the quality of generated images. Ten images were randomly selected from the validation set for

each of the four loss functions. While each model was trained for 300 epochs, the final model was arbitrary chosen based on visual assessment of generated images in the validation set – these are referred to as “selected epoch”.

5.1.5 Statistical Analysis

A statistical analysis was done to assess the quality of images generated using different loss function configurations. This analysis involved evaluating the images with established similarity metrics: SSIM, PSNR, and MSE.

For the statistical examination of the impact of varied loss configurations on image quality, a one-way repeated measures ANOVA (rmANOVA) was performed. This statistical method is apt for comparing the mean differences across groups where identical subjects or datasets have paired measurements in different conditions. This method determined whether changes in loss functions significantly influenced image quality as gauged by SSIM, PSNR, and MSE.

In this analysis, the within-subject factor were the different loss functions, each being applied to the same set of patient images. The dependent variables under scrutiny were the similarity metrics: SSIM, PSNR, and MSE. This examines if any notable variance existed in these metrics across the different loss function methods. The patient IDs variable served to label the repeated measurements belonging to the same patients and controlling for variability within each patient.

For the rmANOVA, the null hypothesis proposed that there is no significant difference in image quality metrics arising from training a model with the various loss functions. A significant result from the ANOVA test would suggest that at least one loss function yields statistically significant image quality metrics from the others. In the case of a significant rmANOVA, post-hoc pairwise comparisons t-tests were conducted to identify which losses resulted in significantly different image quality metrics. Descriptive statistics were also produced as the mean and standard deviation of differences between loss function pairs. P-values were corrected using the using Benjamini/Hochberg false discovery rate (FDR) correction [139]. This was only done during the post-hoc pairwise comparison with a significant cut-off set at $p<0.05$ after correction.

5.1.6 Clinical Evaluation

A clinical evaluation was done to assess the practical utility of the images predicted using different loss functions. This assessment was done to determine the real-world applicability of our AI-generated images in a clinical setting.

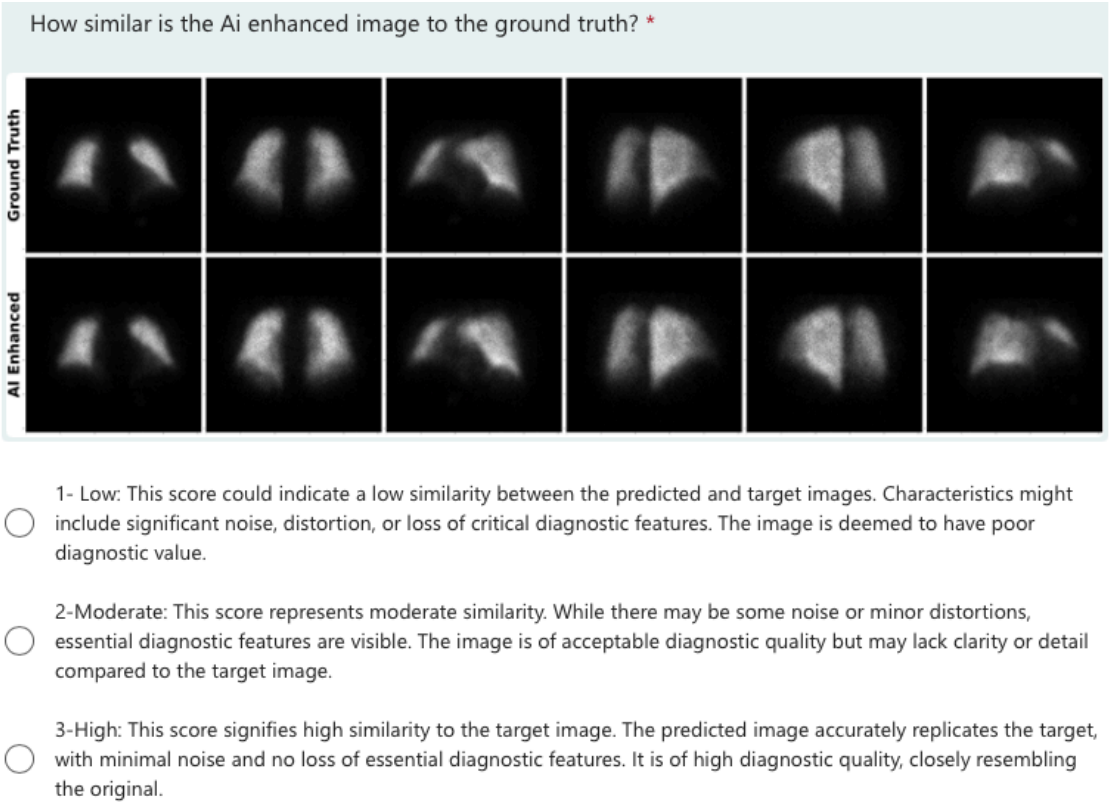

Figure 5.2- Clinical survey using L_1+Perc+GAN loss function for the AI enhanced images agaisnt diagnostically relavent full count perfusion planars (ground truth).

For an in-depth evaluation, I chose the best loss function based on the criteria below for a survey focusing on a side-by-side comparison of full count planars (ground truth) images against their predicted counterparts. All projections available for a patient were included as demonstrated in Figure 5.2. Out of 71 selected studies, 65 had all 6 projections and 6 only had the anterior and posterior projections. The survey participants were asked to rate the AI enhanced images against the corresponding ground truth based on the following:

> 1- Low: This score could indicate a low similarity between the predicted and target images. Characteristics might include significant noise, distortion, or loss of critical diagnostic features. The image is deemed to have poor diagnostic value.
>
> 2-Moderate: This score represents moderate similarity. While there may be some noise or minor distortions, essential diagnostic features are visible. The image is of acceptable diagnostic quality but may lack clarity or detail compared to the target image.
>
> 3-High: This score signifies high similarity to the target image. The predicted image accurately replicates the target, with minimal noise and no loss of essential diagnostic features. It is of high diagnostic quality, closely resembling the original.

Survey participants consisted of experienced researchers, senior medical students, and medical residents versed in V/Q scans. The readers were split into two groups: experienced and novice readers. The experienced group consisted of the two researchers and one certified nuclear medicine

physician well versed in reading V/Q scans and the novice group consisted of 8 people who are proficient at reading V/Q scans.

5.2 Results

5.2.1 Visual Analysis

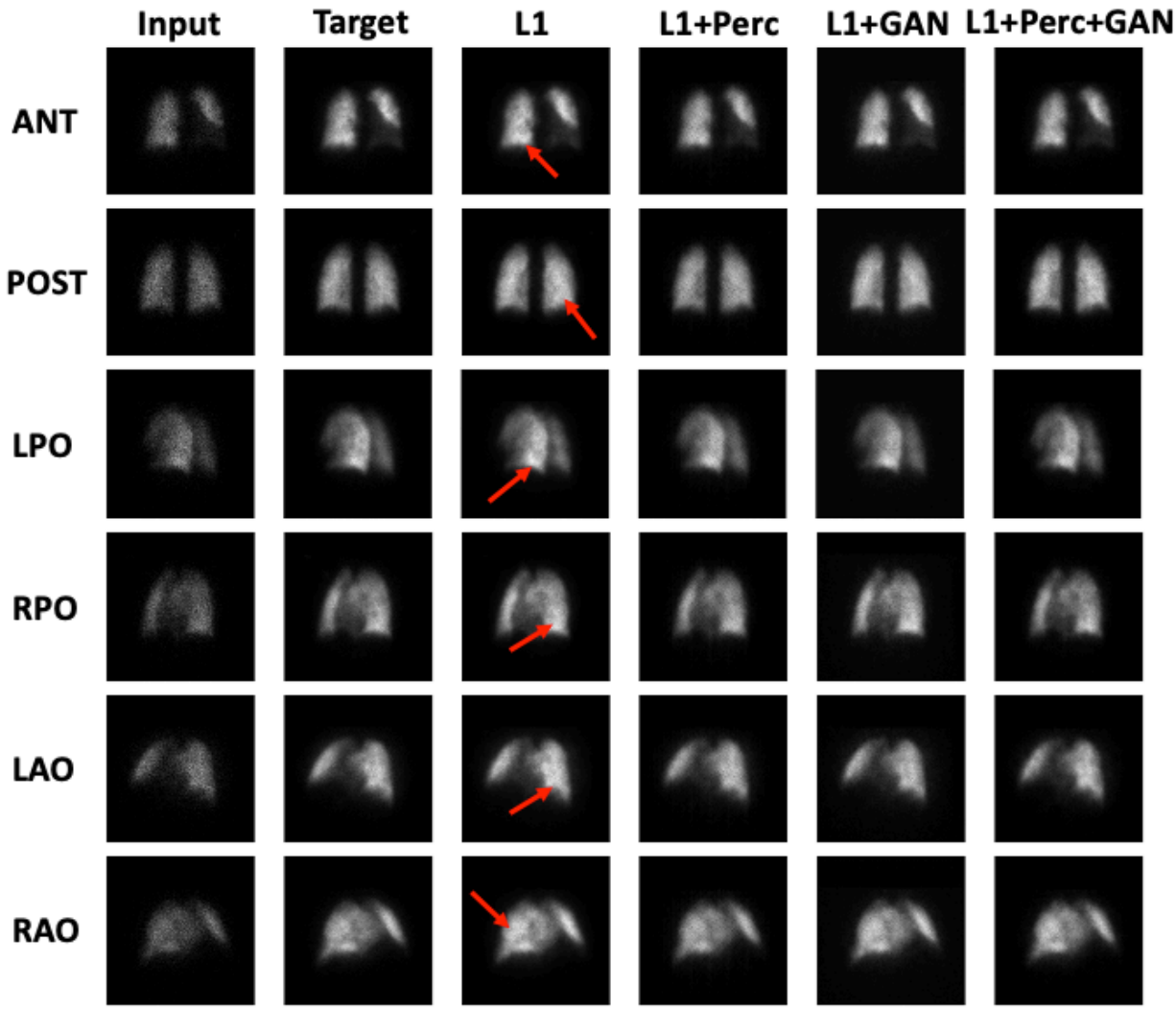

Figure 5.3- Qualitative results comparing the target to 4 different loss function combinations. These include L_1, L_1+Perc (perceptual), L_1+GAN, and L_1+Perc+GAN. This was done for all planar projections: anterior (ANT), posterior (POST), left poster oblique (LPO), right posterior oblique (RPO), left anterior oblique (LAO), and right anterior oblique (RAO).

5.2.1.1. L_1 loss

Figure 5.3 reveals that the L_1 looks similar to the ground truth, with no additional perfusion defects and preserved "salt and pepper" effect. However, discrepancies in brightness are observed in certain areas (arrows). Based on Table 5.1 and Figure 5.4, the model has overfit based on the

train/val ratio since the loss function validation curve is higher than the training curve and the corresponding curves diverge with increased training epochs.

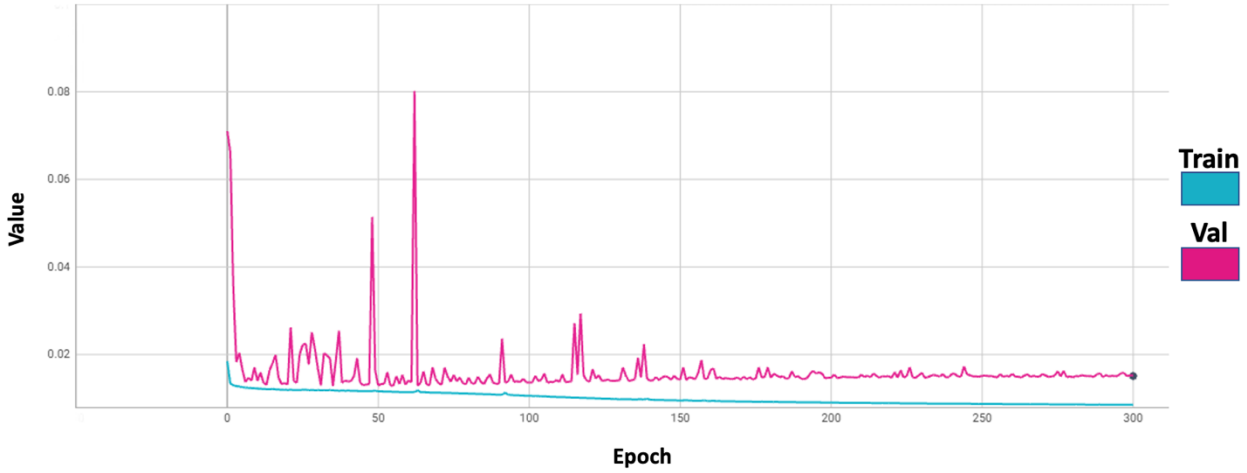

Figure 5.4- L_1 as the loss function for training and validation curves in a pix2pix model.

5.2.1.2. L_1+Perc loss

As demonstrated in Figure 5.3, a visually perceptible "checkerboard" pattern was observed in the lungs of L_1+Perc without introducing new perfusion defects. The training (blue) and validation (red) curves of the image similarity metrics (Figure 5.5, Table 5.2) align closely, suggesting effective learning and low overfitting. A marked improvement in similarity metrics is observed around epochs 30 to 60, with a stabilization phase from epoch 60 to 200, followed by another boost in similarity in epochs 200 to 250, with a final plateau. This trend is particularly noticeable in PSNR and SSIM, but also in MSE. Analysis of content and style losses (Figure 5.6, Table 5.1) shows a pronounced disparity in magnitudes, despite equal weighting. The content loss overfit to the training set, however the style loss did not. When combined with the L_1 loss (Figure 5.7), the perceptual loss also overfit to the training set.

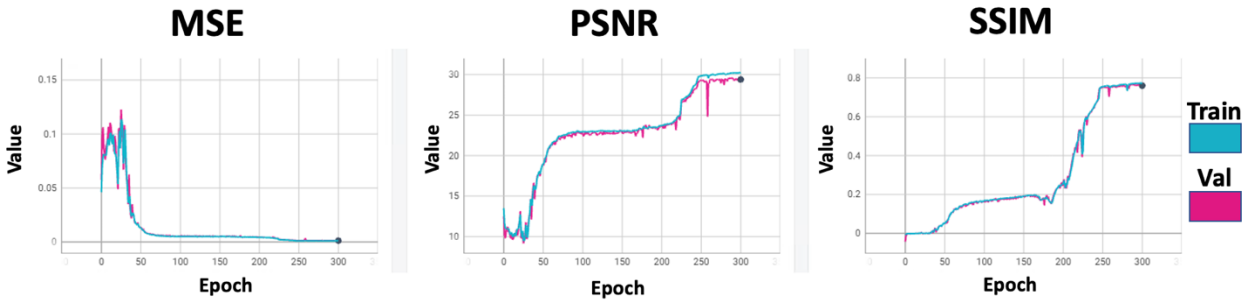

Figure 5.5- Image similarity metrics MSE, PSNR, and SSIM for training and validation curves with L_1 and perceptual losses in a pix2pix model.

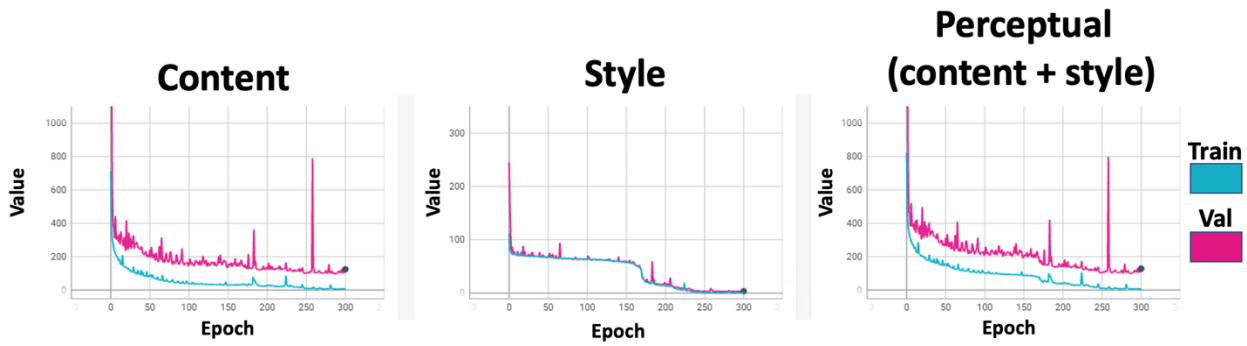

Figure 5.6- From left to right, content, style, and both combined for training and validation curves in a pix2pix model.

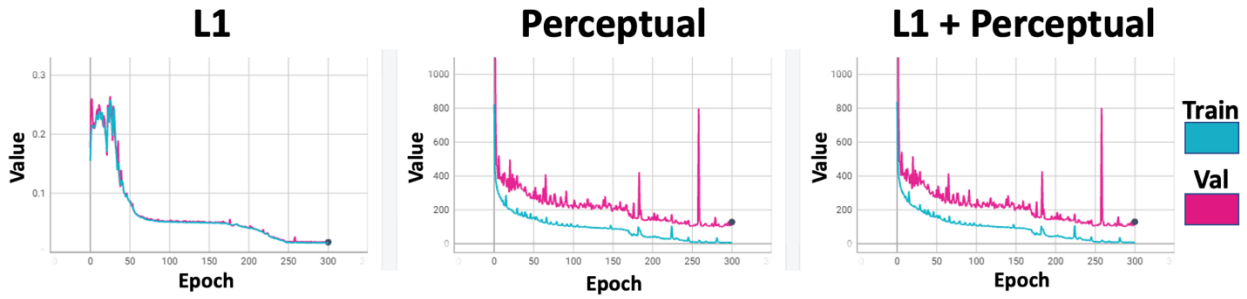

Figure 5.7- From left to right, L_1, perceptual and both are combined for training and validation curves in a pix2pix model.

5.2.1.3. L_1+GAN loss

L_1+GAN in Figure 5.3 did not introduce any visible artifacts. Examining the image similarity metrics (Figure 5.8, Table 5.2), the training curves remain stable, whereas validation curves exhibit large fluctuations between epochs, with SSIM showing fewer spikes after epoch 100. Figure 5.9 indicates a stable L_1 training curve with sporadic validation spikes. GAN training, expectedly, shows instability in both generator and discriminator curves (Figure 5.9), a characteristic of GAN training dynamics.

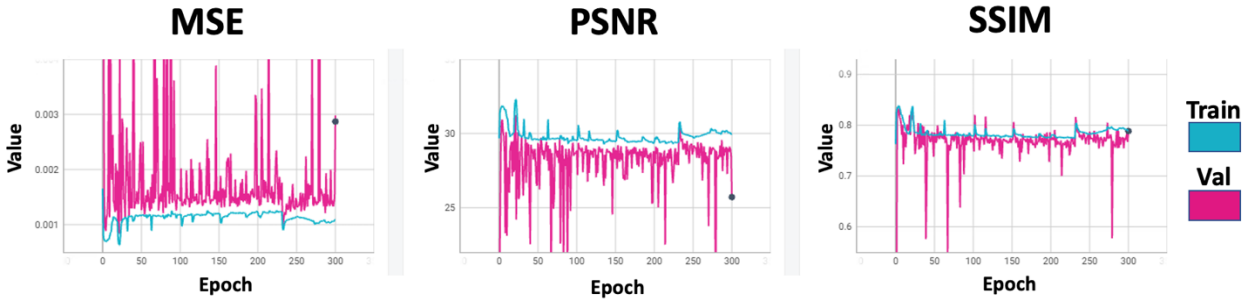

Figure 5.8- Image similarity metrics MSE, PSNR, and SSIM for training and validation curves with L_1 and GAN losses in a pix2pix model.

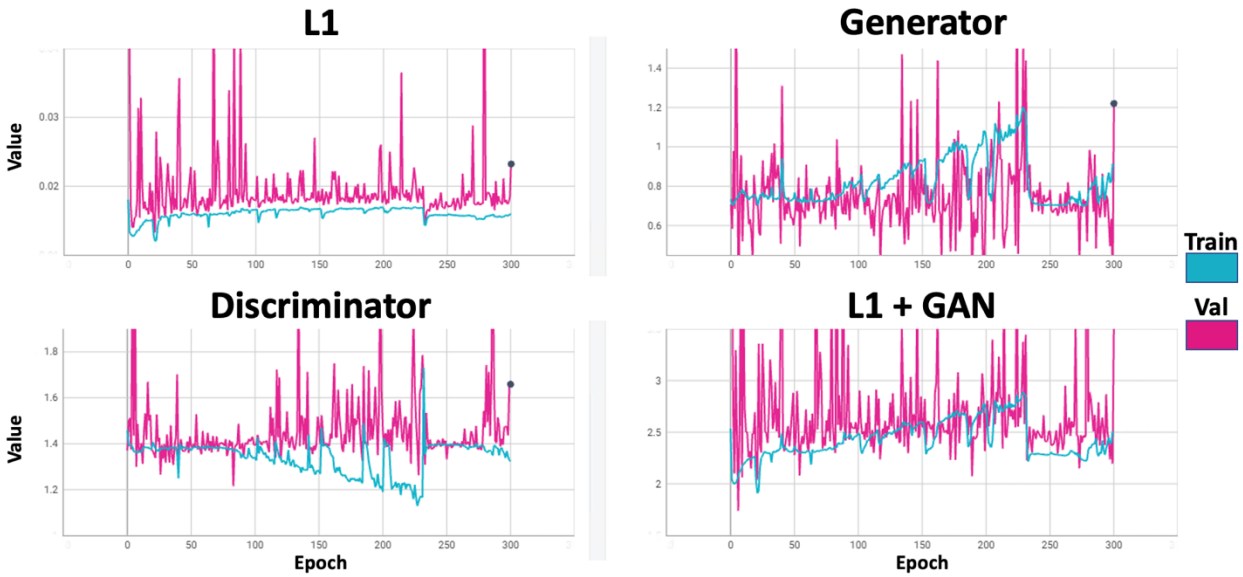

Figure 5.9- Combined L_1 and GAN for training and validation curves in a pix2pix model.

5.2.1.4. L_1+Perc+GAN loss

Visual assessment of Figure 5.3 suggests that images generated using this loss combination best agreed with the ground truth, with no additional defects or noise artifacts. Image similarity metrics (Figure 5.10, Table 5.2) indicate a consistent improvement during initial epochs, followed by stabilization and gradual improvement. The content, style and perceptual (content + style) loss (Figure 5.11) all demonstrate the model overfit. There are notable spikes in magnitude across the epochs. Figure 5.12 reveals overlapping L_1 training and validation curves, whereas perceptual loss and GAN training exhibit periodic fluctuations, with validation curves generally higher than their training counterparts.

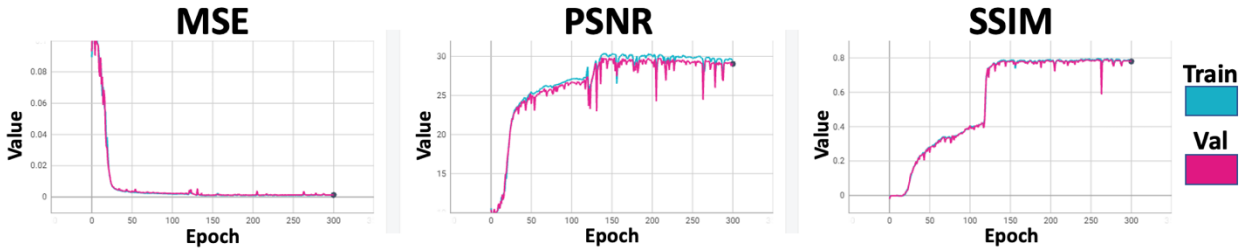

Figure 5.10- Image similarity metrics MSE, PSNR, and SSIM for training and validation curves with L_1, perceptual and GAN losses in a pix2pix model.

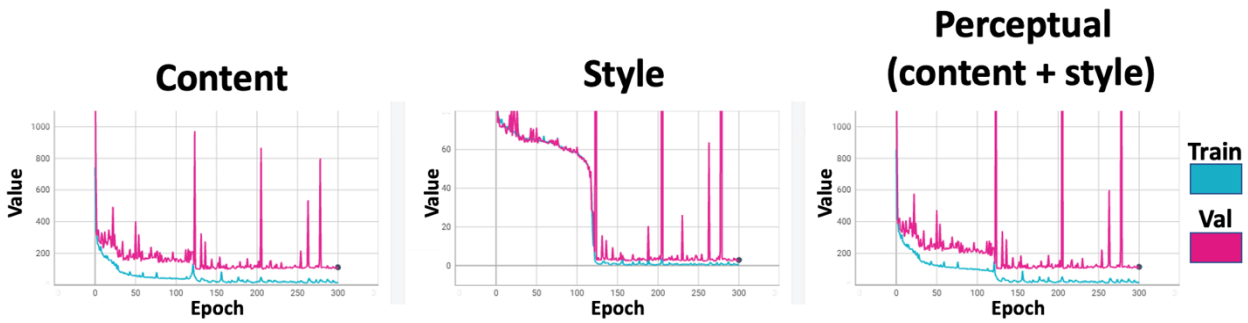

Figure 5.11- From left to right, style, content, and both combined for training and validation curves in a pix2pix model.

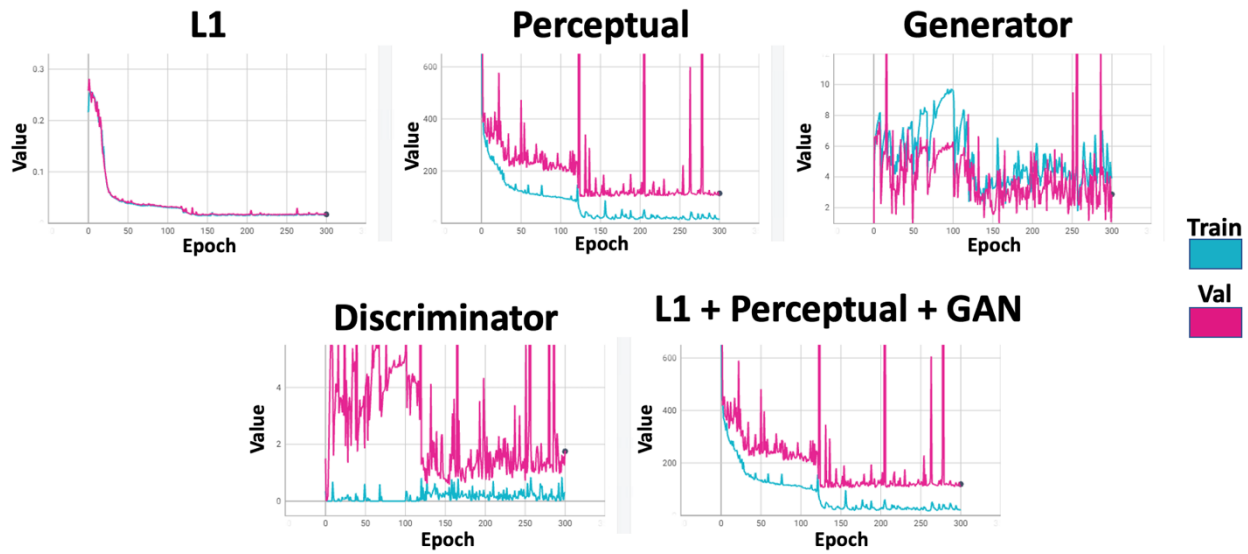

Figure 5.12- Combined L_1, perceptual, and GAN for training and validation curves in a pix2pix model.

5.2.2 Quantitative Analysis

Loss Functions		Selected Epoch	Train Loss	Val Loss	Train/Val Ratio
L_1	L_1	285	8.57×10^{-3}	1.49×10^{-2}	0.575
L_1+Perc	Content	300	6.61×10^{0}	1.18×10^{2}	0.056
	Style		1.87×10^{-1}	2.98×10^{0}	0.628
	Perc (Content + Style)		6.98×10^{0}	1.21×10^{2}	0.058
	L_1		1.56×10^{-2}	1.72×10^{-2}	0.907
	L_1+Perc		8.36×10^{0}	1.23×10^{2}	0.068
L_1+GAN	Generator	175	9.74×10^{-1}	1.04×10^{0}	0.937
	Discriminator		1.25×10^{0}	1.49×10^{0}	0.839

	L_1		$1.67x10^{-2}$	$1.89x10^{-2}$	0.884
	L_1+GAN		$2.65x10^{0}$	$2.88x10^{0}$	0.920
L_1+Perc+GAN	Content	300	$1.41x10^{1}$	$1.06x10^{2}$	0.133
	Style		$4.48x10^{-1}$	$2.55x10^{0}$	0.176
	Perc (Content + Style)		$1.46x10^{1}$	$1.08x10^{2}$	0.135
	Generator		$3.91x10^{0}$	$2.86x10^{0}$	1.37
	Discriminator		$3.01x10^{-1}$	$1.78x10^{0}$	0.169
	L_1		$1.64x10^{-2}$	$1.73x10^{-2}$	0.948
	L_1+Perc+GAN		$2.01x10^{1}$	$1.13x10^{2}$	0.178

Table 5.1- Training (train) and validation (val) losses for arbitrarily selected epoch based on visual assessment of ten images validation figures.

	SSIM		**PSNR**		**MSE**	
Loss Functions	**Train**	**Val**	**Train**	**Val**	**Train**	**Val**
L_1	**$8.88x10^{-1}$**	**$8.06x10^{-1}$**	**$3.50x10^{1}$**	**$2.99x10^{1}$**	**$3.28x10^{-4}$**	**$1.11x10^{-3}$**
L_1+Perc	$7.73x10^{-1}$	$7.63x10^{-1}$	$3.02x10^{1}$	$2.94x10^{1}$	$1.01x10^{-3}$	$1.27x10^{-3}$
L_1+GAN	$7.91x10^{-1}$	$7.91x10^{-1}$	$3.00x10^{1}$	$2.92x10^{1}$	$1.07x10^{-3}$	$1.33x10^{-3}$
L_1+Perc+GAN	$7.90x10^{-1}$	$7.84x10^{-1}$	$2.96x10^{1}$	$2.91x10^{1}$	$1.14x10^{-3}$	$1.34x10^{-3}$

Table 5.2- Image similarity metrics for different loss functions at selected epoch (same as in Table 5.1).

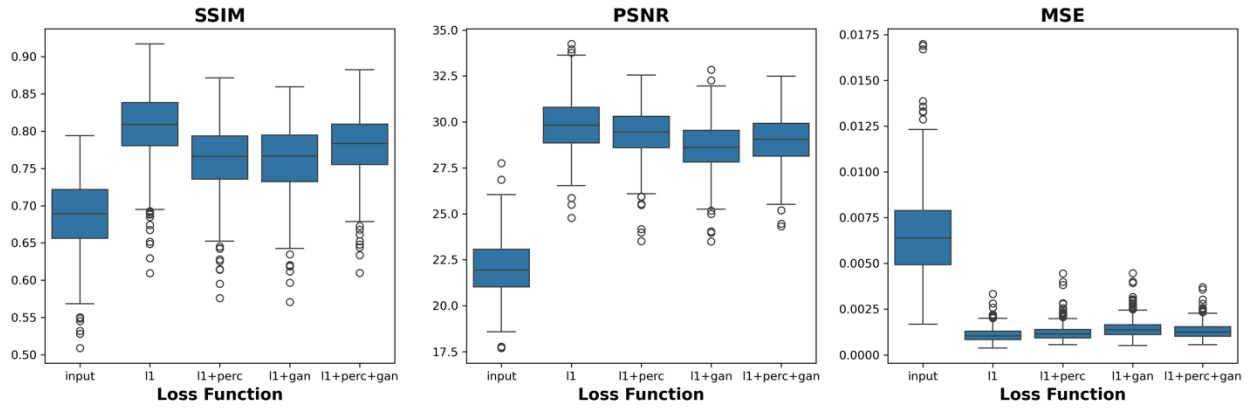

Figure 5.13- Boxplot comparing different loss functions (at selected epoch) and the input using image similarity metrics: (structural similarity index (SSIM), peak signal to noise ratio (PSNR) and mean squared error (MSE). Data was from the test set. N=402 images.

A boxplot analysis is shown in Figure 5.13 for the different loss combinations for each image similarity metrics (SSIM, PSNR, and MSE) on the test set. L_1 performed the best in all three metrics with the other methods having similar results.

5.2.2.1. ANOVA

ANOVA indicated significant performance change with the model using all three performance metrics ($p<10^{-5}$). A detailed pair-wise analysis for SSIM is shown in Table 5.3, again indicating that L_1 loss outperformed all other loss functions ($p<10^{-5}$). The same pattern followed for the remaining metrics but are not shown for brevity.

A	B	P-Value	Delta	Standard deviation
L_1	L_1+GAN	$1.72x10^{-65}$	$4.28x10^{-2}$	$5.26x10^{-3}$
L_1	L_1+Perc	$1.00x10^{-74}$	$4.39x10^{-2}$	$3.93x10^{-3}$
L_1	L_1+Perc+GAN	$1.72x10^{-46}$	$2.48x10^{-2}$	$5.85x10^{-3}$
L_1+GAN	L_1+Perc	$2.73x10^{-03}$	$1.11x10^{-3}$	$3.03x10^{-3}$
L_1+GAN	L_1+Perc+GAN	$5.07x10^{-43}$	$1.80x10^{-2}$	$4.83x10^{-3}$
L_1+Perc	L_1+Perc+GAN	$6.69x10^{-55}$	$1.91x10^{-2}$	$3.38x10^{-3}$

Table 5.3- ANOVA Post-hoc analysis for SSIM across different loss types.

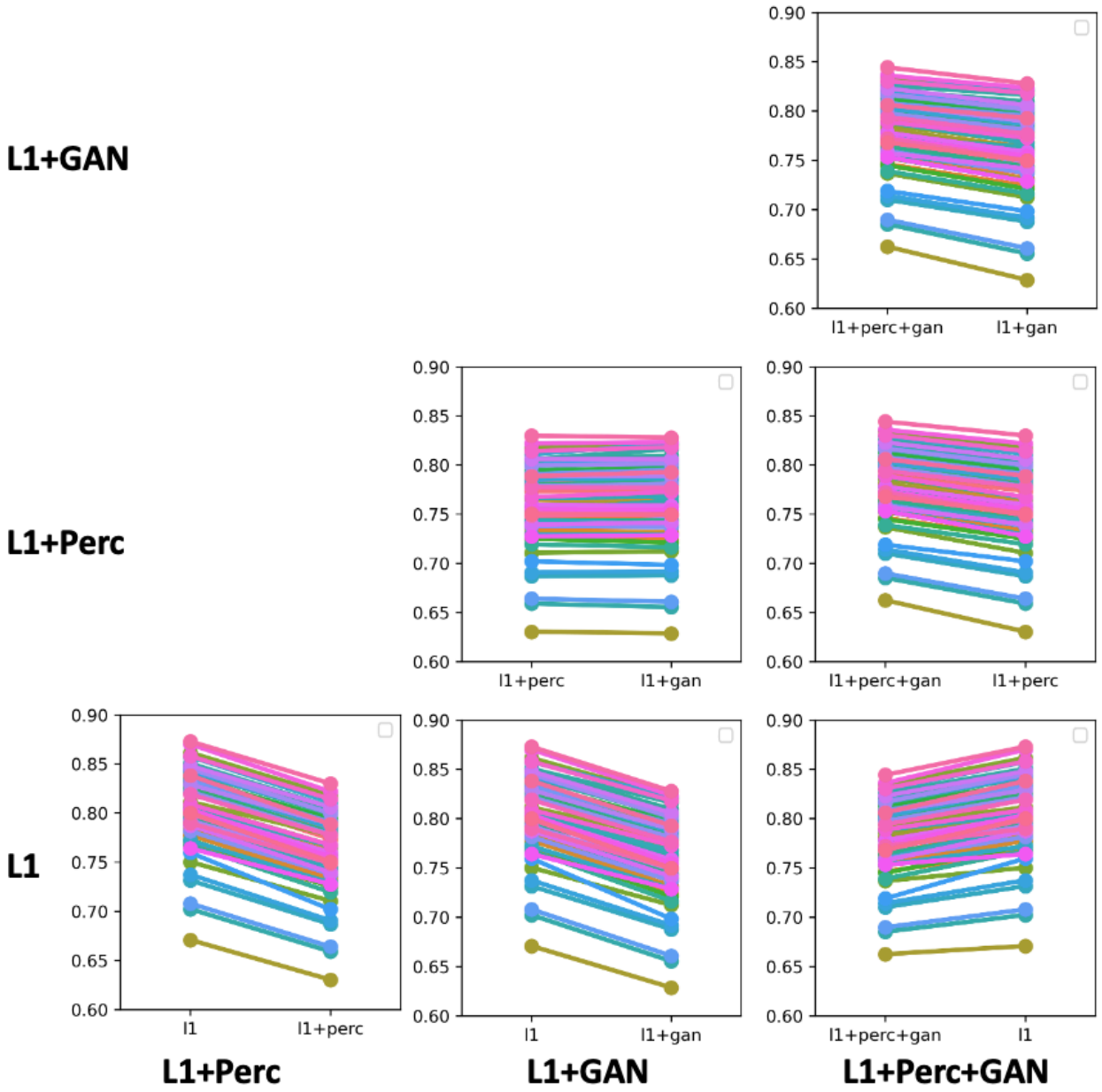

Figure 5.14- Pairwise comparison of SSIM between models as part of ANOVA post-hoc analysis (t-test) corresponding to Table 5.3 and Figure 5.13 (left). Images from individual patients are represented by a single line of a given color.

The p-value for SSIM was statistically significant ($p<10^{-5}$). Post hoc analysis, as detailed in Table 5.3 and illustrated in Figure 5.14, further support these findings, highlighting the superiority of L_1 over other loss functions. Similar patterns were observed in PSNR and MSE and were not included for brevity.

5.2.2.2. Clinical

While L_1 performed best in the quantitative analysis, by visual assessment, L_1+Perc+GAN loss function was chosen as the best to conduct the survey. Table 5.4 illustrates the average results in

each category for novice and experienced readers. For 71 patients in the test set (402 images), experienced and novice readers rated the predicted vs target images as low at a similar rate (4% for both). For the remaining scores, experienced viewers rated a smaller proportion as having high similarity to the ground truth (55% vs 67%), likely reflecting an attention to small details that comes with experience. The high similarity signifies high diagnostic quality similar to that of the ground truth.

	Low	**Moderate**	**High**	**Total**
Novice (n=8)	2.5 (4%)	21 (30%)	47.5 (67%)	71
Experienced (n=3)	3 (4%)	29.3 (41%)	38.7 (55%)	71
Combined (n=11)	2.64 (4%)	23.3 (33%)	45.1 (64%)	71

Table 5.4- Visual scoring for clinical quality by comparing predicted vs ground truth images by novice and experienced readers. Average number of responses across readers (% of total responses).

5.3 Discussion

In this work we implemented and tested four machine learning models for enhancing low count scintigraphic images of the lungs. We tested enhancement at a 10:1 count ratio, which in practice simulate a 10:1 shortening of the image acquisition time or 10:1 dose reduction. By successfully implementing and testing these models, our study aimed to demonstrate that deep learning can effectively generate high-quality images from low-count ones.

The simulated reduction in counts suggests a theoretical ability to reduce acquisition time or dose as a step towards greater patient comfort and safety without significant loss to diagnostic accuracy. Furthermore, it could be a cost-effective imaging procedure in clinical settings due to improved patient throughput. Another potential application is to produce pseudo-planar projections from raw tomographic sinograms. Typical tomographic projections in lung V/Q scans are on the order of 5 to 20 seconds, thus a 10:1 count enhancement should produce image quality of 50 to 200 second acquisitions – on the same order as typical planar projections.

It is perhaps interesting to note that the raw data had 600k counts and the low count images simulated 60k counts. Looking at Figure 4.8 (on a 256×256 sampling grid), corresponding SSIM should be approaching 0.8 for the ground truth and ~0.6 for the low count images. Our enhanced images (Figure 5.13) showed SSIM values in this range, with low count images ~0.7 and averages between 0.75 and 0.81 depending on the loss function.

5.3.1 Statistical Analysis

Overall, all the models evaluated in this work produced similar performance statistics (SSIM, PSNR and MSE) with the L_1 loss showing small, but statistically significant numerical superiority

in the test and validation sets. The statistical analysis using ANOVA allowed for the quantification of the impact of different loss configurations on image quality. The statistically significant p-values across SSIM, PSNR, and MSE (Table 5.3, Figure 5.14) confirm that the choice of loss function affects the quality of generated images.

It is perhaps, a foregone conclusion that the L_1 method had the lowest MSE, since this is precisely the definition of the L_1 loss function. Furthermore, SSIM and PSNR are inversely related to MSE, and it is therefore expected that methods with low MSE will tend to produce high PSNR and SSIM (as demonstrated in Figure 4.8). Consequently, these metrics alone are not sufficient to select the preferred image enhancement method.

5.3.2 Visual Analysis

Visual inspection of generated images was used in evaluating the efficacy of different loss functions applied for the image enhancement algorithm. The visual assessment, as depicted in Figure 5.3, revealed insights and limitations of each loss function.

For L_1, the visual similarity to the ground truth images was notable. This is expected since the L_1 focuses on the average magnitude between pixels. However, certain areas in the predicted image exhibited brighter pixels, suggesting minor deviations in pixel intensity values. Because the images were scaled to the brightest pixel (a common display approach) pixel intensities are dependent on the most intense single pixel in the image, which is prone to noise. Despite this, the maintenance of the "salt and pepper" effect in the images indicated the preservation of essential textural details, a key aspect in scintigraphic imaging. The trends in image similarity metrics (Table 5.2) and the L_1 (Table 5.1) demonstrate that the model overfit. The model may have learned the noise and detail characteristics but did not generalize well to the validation set. However, it did not introduce or remove defects in the lungs as a result of overfitting.

The L_1+Perc introduced a checkerboard pattern within the lung regions. Based on the literature [140][141][142], the checkerboard pattern arises from using a CNN network such as a U-Net. Within the U-Net [58] architecture, the upsampling layers using a deconvolution generate this pattern. As future work, one should look to Sugawara *et al* [140] to prevent the checkerboard pattern. Also, the checkerboard pattern was only present in L_1+Perc, however, the same U-Net generator was used for all for all loss functions. The perceptual loss, driven by content and style losses, seemed to emphasize content over style, as indicated by the overfitting of the content loss to the training set. This might explain the preservation of general lung structures at the cost of introducing subtle artifacts which is associated with the style loss.

In the case of L_1+GAN, the addition of a GAN loss provides adversarial feedback to the generator. While L_1 loss minimizes pixel-level differences, it falls short in rendering realistic images. In a GAN framework, the discriminator's role is to differentiate between real (target) and generated (predicted) images. The practice of training the generator twice for each discriminator update plays a role in this balance [138]. This approach allows the generator to refine its outputs more effectively against the dual criteria of L_1 loss and the discriminator. It enables the generator to overcome the inherent tendency of L_1 loss to produce blurred or averaged results by incorporating finer, realistic details encouraged by the discriminator. Furthermore, this training strategy mitigates the risk of the discriminator overpowering the generator too early in the training process. An overly effective discriminator can prematurely constrain the generator, limiting its ability to capture the

complex nuances of scintigraphic images. By training the generator more frequently, we allow it to learn and adapt, thus generating a better output. This frequent update of the generator also plays a crucial role in preventing mode collapse. It ensures that the generator does not settle on producing a narrow set of solutions that might satisfy the L_1 loss but lack diversity and realism. Instead, it encourages exploration and learning of a wider array of patterns and details, which is essential in medical imaging like scintigraphy where accuracy and detail are critical.

Lastly, the L_1+Perc+GAN combination (Figure 5.3) present the most visually appealing results. The images closely resemble the ground truth, with no apparent introduction of artifacts or noise. Based on the training and validation curves (Figure 5.12), the combined loss function resembles that of perceptual loss curve. This may indicate that the perceptual loss is driving the training process, and the GAN removes the checkerboard pattern found in perceptual loss. Based on Table 5.1, the L_1 demonstrates a good model was produced between training and validation curves (train/val ratio), the perceptual loss shows the model has overfit to the training set, the generator demonstrates typical GAN instability, the discriminator has overfit, and the overall function has overfit as well. Nevertheless, based on visual assessment and the absence of L_1 overfitting we selected this model as the superior choice and used it for future clinical evaluation.

5.3.3 Clinical Significance

In this study, the L1+Perc+GAN model was chosen based on visual assessment from the validation set. The subsequent survey, which was designed to gauge the effectiveness of AI-enhanced images, revealed interesting trends in perception among different levels of expertise. Experienced readers, those familiar with the nuances of lung scintigraphic images, were as likely to rate AI-enhanced images as having low similarity to full-count planars compared to novice readers. The overall response from most readers, regardless of their experience level, leaned heavily towards high similarity, indicating that critical diagnostic features were preserved in the AI-enhanced images.

The implementation of pseudo-planar images from simulated low count images has clinical implications. Simulated low count images were 10% of the counts in the diagnostic planar imaged, which is on the same order as SPECT raw data count level. Hence the count enhancement method together with the image resizing methodology of Chapter 4 hold promise for generating high quality planar image from raw SPECT projection data. Such a tool is valuable for clinicians who have become accustomed to planar lung scintigraphy, especially those who may be hesitant to transition fully to SPECT. At The Ottawa Hospital, where the shift from performing both SPECT and planar imaging to exclusive SPECT imaging has already occurred, the introduction of pseudo-planar images is timely and relevant. While not all physicians will require pseudo-planar images as a matter of routine, these may be invaluable in cases of suspected patient motion or other imaging artifacts that can jeopardize the diagnostic accuracy of the reconstructed SPECT.

As mentioned, the ability to reduce imaging times or the dosage of radiotracers through AI-enhanced images addresses key patient concerns, such as discomfort and radiation exposure. Alternatively, this technology may be useful to further enhance full-count image data above current practice.

5.3.4 Limitations

One significant limitation of this study is the exclusive focus on perfusion images. While perfusion scintigraphy is a critical component in diagnosing PE, ventilation imaging also plays an essential role. The exclusion of ventilation images means the study's findings may not fully represent the complexities involved in comprehensive lung scintigraphy, which typically includes both ventilation and perfusion. The method used to simulate low count data has been shown to faithfully reproduce low count equivalents, however, further research with diagnostically relevant data remains the gold standard.

Additionally, the study's reliance on a single hospital's data could limit the generalizability of the findings. Different hospitals may have variations in imaging equipment, protocols, and patient demographics, which could influence the performance of the AI models. Future research should aim to include a more diverse dataset, encompassing multiple institutions to enhance the robustness and applicability of the models.

Another limitation is the potential overfitting issue, particularly concerning the use of RadImageNet for perceptual loss. While RadImageNet's clinically relevant features offer an advantage, there is a risk that models trained on this dataset might not generalize well to other medical image datasets or real-world clinical scenarios.

Lastly, the study's focus on image quality metrics like SSIM, PSNR, and MSE while important, does not encompass any aspect of clinical usefulness. While we sought visual assessment by a panel of viewers, it is evident that experience plays an important role in evaluating the clinical accuracy of the enhanced images. Future studies should include a comprehensive assessment of diagnostic accuracy by having a panel of nuclear medicine physicians blindly assess AI enhanced images and state-of-the-art clinical images, to ensure they arrive at equivalent clinical interpretations.

5.4 Conclusion

In conclusion, this chapter demonstrates the development, implementation, and evaluation of a model capable of enhancing low-count images to match the diagnostic quality of high-count planar images. By visual assessment, the combined L_1, perceptual, and GAN loss model performed the best.

Chapter 6: Conclusion and Future Work

6.1 Summary of Findings

This thesis examined count enhancement of lung scintigraphic images using artificial intelligence. The research encompassed three pivotal projects, each contributing to the overarching goal of improving V/Q scan image acquisition. The first project, a systematic review, established the necessity for renewing research into the use of AI in V/Q scans, identifying multiple possibly fruitful avenues (Chapter 3). This exploration into existing literature underscored the potential of AI in enhancing the efficiency and accuracy of V/Q scans, setting the stage for subsequent empirical investigations. Recognising the need to resize SPECT to planar grid size and vice versa, the second project (Chapter 4) focused on establishing accurate image resizing method that crucially, preserved the inherent noise characteristics of V/Q scintigraphic images. This resizing technique was put into practice for preprocessing images for AI model training and validation in the final project. The final project (Chapter 5) involved developing an AI model to enhance low-count perfusion planar images to resemble a diagnostically relevant full count perfusion planar.

This model holds promise to generate pseudo-planars from SPECT projection data and to reduce the duration of V/Q scans, thus improving the patient experience in nuclear medicine procedures. However, this work falls short of fully demonstrating the readiness of this approach for clinical application.

6.2 Recommendations for Future Research

6.2.1 Clinical Evaluation

Future research should aim to diversify the dataset by adding data from multiple hospitals. This approach will introduce a broader range of images from different patients, thereby enhancing the model's robustness, generalizability, and reduce overfitting. Additionally, involving certified nuclear medicine clinicians in the evaluation process is crucial. Their assessments can help determine whether they can distinguish between real and AI-enhanced images and offer insights into the model's real-world efficacy.

6.2.2 Pseudo-Planars from SPECT Acquisitions

Subsequent studies should focus on refining the model to enhance image counts by varying degrees, not restricted to just 10% of the full-count perfusion planars. Utilizing both SPECT and planar data obtained for each patient, future models can effectively use SPECT data as low-count images, with corresponding full-count planar images serving as targets. These input-target pairs, derived from SPECT data captured on a 128×128 grid and planar images on a 256×256 grid, should be processed to ensure they are on a matching grid size. This can be achieved by upsampling and downsampling methods demonstrated in Chapter 4: Resizing of Scintigraphic Images.

To train and evaluate the use of image enhancement to generate pseudo-planar, a need exists to match the best angled projection (input) corresponding to the planar image (ground-truth). This can be achieved with a preprocessing step that first centers the center-of-mass (COM) in the images to ensure the lungs are centrally positioned in the image frame. Next, the SPECT projections can be co-registered with the full-count planar images to select the SPECT projection with the highest Pearson correlation as that with the most similar angle.

More sophisticated approaches to generating pseudo-planars from SPECT may leverage the flexibility of machine learning models to account for additional data, such as from multiple SPECT projections (from slightly different angles). In doing so they can benefit of more count statistics, while compensating for view angles. While this has been attempted using averaging of SPECT projections, the results are not ideal due to the inherent motion blur. Machine learning algorithms may be able to effectively compensate for this motion.

6.2.3 Improved Models

Future studies can also explore alternative loss functions that go beyond spatial domain comparisons. One direction is the development of a hybrid loss function that combines traditional pixel-wise methods (such as L_1 loss) with a frequency amplitude loss, derived from a 2D Fourier transform. This approach is expected to preserve high-frequency details and provide shift invariance. The hybrid loss function, integrating spatial accuracy through L_1 loss and texture preservation via frequency amplitude loss, is particularly advantageous for scintigraphic images. This loss function is sensitive to blurriness while being robust against mild noise, a balance that is vital in medical imaging where maintaining high-frequency details and textures is essential, especially in cases involving a transformation from low-count to full-count images where noise can be a noticeable factor.

6.3 Final Thoughts

This thesis represents a step in the ongoing journey to integrate artificial intelligence into the realm of nuclear medicine, specifically focusing on the enhancement of lung scintigraphic images. By focusing on the enhancement of low-count perfusion planar images, there is potential to further refine the model and integrate it into clinical settings. It can help augment the physician workflow, enhance diagnostic accuracy which in turn contributes to better patient outcomes.

References

[1] B. J. Amorim, M. Y. Rigolon, and C. D. Ramos, "Diagnosis of pulmonary embolism," *Nucl. Cardiol. Basic Adv. Concepts Clin. Pract.*, vol. 168, no. 2, pp. 723–739, 2021, doi: 10.1007/978-3-030-62195-7_31.

[2] P. S. Wells, R. Ihaddadene, A. Reilly, and M. A. Forgie, "Diagnosis of venous thromboembolism: 20 Years of progress," *Ann. Intern. Med.*, vol. 168, no. 2, pp. 131–140, 2018, doi: 10.7326/M17-0291.

[3] M. Di Nisio, A. Squizzato, A. W. S. Rutjes, H. R. Büller, A. H. Zwinderman, and P. M. M. Bossuyt, "Diagnostic accuracy of D-dimer test for exclusion of venous thromboembolism: A systematic review," *J. Thromb. Haemost.*, vol. 5, no. 2, 2007, doi: 10.1111/j.1538-7836.2007.02328.x.

[4] G. D. Tourassi, C. E. Floyd, and R. E. Coleman, "Improved noninvasive diagnosis of acute pulmonary embolism with optimally selected clinical and chest radiographic findings.," *Acad. Radiol.*, vol. 3, no. 12, pp. 1012–1018, 1996, doi: 10.1016/S1076-6332(96)80035-3.

[5] G. Zantonelli *et al.*, "Acute Pulmonary Embolism: Prognostic Role of Computed Tomography Pulmonary Angiography (CTPA)," *Tomography*, vol. 8, no. 1. 2022. doi: 10.3390/tomography8010042.

[6] P. Reinartz, J. E. Wildberger, W. Schaefer, B. Nowak, A. H. Mahnken, and U. Buell, "Tomographic imaging in the diagnosis of pulmonary embolism: A comparison between V/Q lung scintigraphy in SPECT technique and multislice spiral CT," *J. Nucl. Med.*, vol. 45, no. 9, pp. 1501–1508, 2004.

[7] B. Hasibuan, Oea Khairsyaf, and Fenty Anggrainy, "Pulmonary Embolism: A Narrative Literature Review," *Biosci. Med. J. Biomed. Transl. Res.*, vol. 6, no. 16, 2022, doi: 10.37275/bsm.v6i16.705.

[8] P. Poredos and M. K. Jezovnik, "Endothelial Dysfunction and Venous Thrombosis," *Angiology*, vol. 69, no. 7. 2018. doi: 10.1177/0003319717732238.

[9] G. G. King, B. Harris, and S. Mahadev, "V/Q SPECT: Utility for investigation of pulmonary physiology," *Semin. Nucl. Med.*, vol. 40, no. 6, pp. 467–473, 2010, doi: 10.1053/j.semnuclmed.2010.07.006.

[10] M. Ciurzyński *et al.*, "An expert opinion of the Polish Cardiac Society Working Group on Pulmonary Circulation on screening for chronic thromboembolic pulmonary hypertension patients after acute pulmonary embolism: Update," *Kardiol. Pol.*, vol. 80, no. 6, 2022, doi: 10.33963/KP.a2022.0141.

[11] Y. C. Lai, K. C. Potoka, H. C. Champion, A. L. Mora, and M. T. Gladwin, "Pulmonary arterial hypertension: The clinical syndrome," *Circ. Res.*, vol. 115, no. 1, 2014, doi: 10.1161/CIRCRESAHA.115.301146.

[12] P. S. Marshall, K. S. Mathews, and M. D. Siegel, "Diagnosis and management of life-threatening pulmonary embolism," *Journal of Intensive Care Medicine*, vol. 26, no. 5. 2011. doi: 10.1177/0885066610392658.

[13] J. Bělohlávek, V. Dytrych, and A. Linhart, "Pulmonary embolism, part I: Epidemiology,

risk factors and risk stratification, pathophysiology, clinical presentation, diagnosis and nonthrombotic pulmonary embolism," *Exp. Clin. Cardiol.*, vol. 18, no. 2, 2013.

[14] M. Idiz, C. Konuralp, and M. Ateş, "Underdiagnosis of pulmonary embolism: A recurrent nightmare for surgeons," *Eastern Journal of Medicine*, vol. 8, no. 1. 2003.

[15] M. Miniati *et al.*, "Survival and restoration of pulmonary perfusion in a long-term follow-up of patients after acute pulmonary embolism," *Medicine (Baltimore).*, vol. 85, no. 5, 2006, doi: 10.1097/01.md.0000236952.87590.c8.

[16] C. E. Mahan *et al.*, "Venous thromboembolism: Annualised United States models for total, hospital-acquired and preventable costs utilising long-term attack rates," *Thromb. Haemost.*, vol. 108, no. 2, 2012, doi: 10.1160/TH12-03-0162.

[17] D. Jiménez *et al.*, "Simplification of the pulmonary embolism severity index for prognostication in patients with acute symptomatic pulmonary embolism," *Arch. Intern. Med.*, vol. 170, no. 15, 2010, doi: 10.1001/archinternmed.2010.199.

[18] F. B. Payolla, A. C. Massabni, and C. Orvig, "Radiopharmaceuticals for diagnosis in nuclear medicine: A short review," *Eclet. Quim.*, vol. 44, no. 3, 2019, doi: 10.26850/1678-4618eqj.v44.3.2019.p11-19.

[19] D. Papagiannopoulou, "Technetium-99m radiochemistry for pharmaceutical applications," *Journal of Labelled Compounds and Radiopharmaceuticals*, vol. 60, no. 11. 2017. doi: 10.1002/jlcr.3531.

[20] M. C. Adams, T. G. Turkington, J. M. Wilson, and T. Z. Wong, "A systematic review of the factors affecting accuracy of SUV measurements," *American Journal of Roentgenology*, vol. 195, no. 2. 2010. doi: 10.2214/AJR.10.4923.

[21] F. Yang, K. Yang, and C. Yang, "Development and Application of Gamma Camera in the Field of Nuclear Medicine," *Int. J. Sci.*, vol. 4, no. 07, 2018, doi: 10.18483/ijsci.1758.

[22] K. Kim, M. H. Lee, and Y. Lee, "Investigation of a blind-deconvolution framework after noise reduction using a gamma camera in nuclear medicine imaging," *Nucl. Eng. Technol.*, vol. 52, no. 11, 2020, doi: 10.1016/j.net.2020.04.032.

[23] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," 1987. doi: 10.1145/37401.37422.

[24] S. Jan *et al.*, "GATE: A simulation toolkit for PET and SPECT," *Phys. Med. Biol.*, vol. 49, no. 19, 2004, doi: 10.1088/0031-9155/49/19/007.

[25] X. Pan, E. Y. Sidky, and M. Vannier, "Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction?," *Inverse Problems*, vol. 25, no. 12. 2009. doi: 10.1088/0266-5611/25/12/123009.

[26] H. O. Anger, "Scintillation camera," *Rev. Sci. Instrum.*, vol. 29, no. 1, 1958, doi: 10.1063/1.1715998.

[27] K. Agrawal, F. Marafi, G. Gnanasegaran, H. Van Der Wall, and I. Fogelman, "Pitfalls and limitations of radionuclide planar and hybrid bone imaging," *Seminars in Nuclear Medicine*, vol. 45, no. 5. 2015. doi: 10.1053/j.semnuclmed.2015.02.002.

[28] D. Van Nostrand, "The benefits and risks of I-131 therapy in patients with well-

differentiated thyroid cancer," *Thyroid*, vol. 19, no. 12. 2009. doi: 10.1089/thy.2009.1611.

[29] G. E. Kim, J. H. Park, J. S. Kim, K. S. Won, and H. W. Kim, "Comparison of Tc-99m DMSA Renal Planar Scan and SPECT for Detection of Cortical Defects in Infants with Suspected Acute Pyelonephritis," *Indian J. Pediatr.*, vol. 86, no. 9, 2019, doi: 10.1007/s12098-019-02969-6.

[30] H. Zaidi, "Comparative methods for quantifying thyroid volume using planar imaging and SPECT," *J. Nucl. Med.*, vol. 37, no. 8, 1996.

[31] J. Dickson, J. Ross, and S. Vöö, "Quantitative SPECT: the time is now," *EJNMMI Phys.*, vol. 6, no. 1, 2019, doi: 10.1186/s40658-019-0241-3.

[32] O. Israel *et al.*, "Two decades of SPECT/CT – the coming of age of a technology: An updated review of literature evidence," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 46, no. 10. 2019. doi: 10.1007/s00259-019-04404-6.

[33] L. T. Chang, "A method for attenuation correction in radionuclide computed tomography," *IEEE Trans. Nucl. Sci.*, vol. 25, no. 1, 1978, doi: 10.1109/TNS.1978.4329385.

[34] P. P. Bruyant, "Analytic and iterative reconstruction algorithms in SPECT," *J. Nucl. Med.*, vol. 43, no. 10, 2002.

[35] K. S. Krane and W. G. Lynch, " Introductory Nuclear Physics ," *Phys. Today*, vol. 42, no. 1, 1989, doi: 10.1063/1.2810884.

[36] G. B. Saha, *Fundamentals of nuclear pharmacy*. 2010. doi: 10.1007/978-1-4419-5860-0.

[37] E. L. C. and P. R. Bevington, "Data Reduction and Error Analysis for the Physical Sciences.," *J. Am. Stat. Assoc.*, vol. 67, no. 337, 1972, doi: 10.2307/2284752.

[38] S. Cherry, J. Sorenson, and M. Phelps, *Physics in Nuclear Medicine*. 2012. doi: 10.1016/C2009-0-51635-2.

[39] Y. Berker and Y. Li, "Attenuation correction in emission tomography using the emission data - A review," *Medical Physics*, vol. 43, no. 2. 2016. doi: 10.1118/1.4938264.

[40] P. Y. Le Roux *et al.*, "Lung scintigraphy for pulmonary embolism diagnosis in COVID-19 patients: a multicenter study," *J. Nucl. Med.*, vol. 63, no. 7, 2022, doi: 10.2967/jnumed.121.262955.

[41] G. M. Currie and D. L. Bailey, "A Technical Overview of Technegas as a Lung Ventilation Agent," *J. Nucl. Med. Technol.*, vol. 49, no. 4, 2021, doi: 10.2967/jnmt.121.262887.

[42] P. Investigators, "Value of the Ventilation/Perfusion Scan in Acute Pulmonary Embolism: Results of the Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED)," *JAMA J. Am. Med. Assoc.*, vol. 263, no. 20, 1990, doi: 10.1001/jama.1990.03440200057023.

[43] M. Bajc, J. B. Neilly, M. Miniati, C. Schuemichen, M. Meignan, and B. Jonson, "EANM guidelines for ventilation/perfusion scintigraphy : PPPPart 1. Pulmonary imaging with ventilation/perfusion single photon emission tomography," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 36, no. 8. 2009. doi: 10.1007/s00259-009-1170-5.

[44] L. S. Zuckier, "Safe Pulmonary Scintigraphy in the Era of COVID-19," *Seminars in Nuclear Medicine*, vol. 52, no. 1. 2022. doi: 10.1053/j.semnuclmed.2021.06.021.

[45] P.-Y. Le Roux *et al.*, "Diagnosis of pulmonary embolism: planar images generated from V/Q SPECT are not a reliable substitute for traditional planar V/Q scan," *Nucl. Med. Commun.*, vol. 33, no. 7, 2012, [Online]. Available: https://journals.lww.com/nuclearmedicinecomm/fulltext/2012/07000/diagnosis_of_pulmonary_embolism__planar_images.3.aspx

[46] D. L. Bailey, G. P. Schembri, B. E. Harris, E. A. Bailey, R. A. Cooper, and P. J. Roach, "Generation of planar images from lung ventilation/perfusion SPECT," *Ann. Nucl. Med.*, vol. 22, no. 5, pp. 437–445, 2008, doi: 10.1007/s12149-007-0120-1.

[47] P. Y. Le Roux *et al.*, "Pulmonary scintigraphy for the diagnosis of acute pulmonary embolism: A survey of current practices in Australia, Canada, and France," *J. Nucl. Med.*, vol. 56, no. 8, pp. 1212–1217, 2015, doi: 10.2967/jnumed.115.157743.

[48] Y. LeCun, G. Hinton, and Y. Bengio, "Deep learning (2015), Y. LeCun, Y. Bengio and G. Hinton," *Nature*, vol. 521, pp. 436–444, 2015.

[49] Y. LeCun *et al.*, "Backpropagation Applied to Handwritten Zip Code Recognition," *Neural Comput.*, vol. 1, no. 4, pp. 541–551, 1989, doi: 10.1162/neco.1989.1.4.541.

[50] G. Zhang, C. Wang, B. Xu, and R. Grosse, "Three mechanisms of weight decay regularization," 2019.

[51] S. Ruder, "An Overview Optimization Gradients," *arXiv Prepr. arXiv1609.04747*, 2017.

[52] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," 2015.

[53] V. Nair and G. E. Hinton, "Rectified linear units improve Restricted Boltzmann machines," 2010.

[54] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, vol. 2.

[55] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, vol. 07-12-June-2015. doi: 10.1109/CVPR.2015.7298594.

[56] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-December. doi: 10.1109/CVPR.2016.90.

[57] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, 2017, doi: 10.1109/TIP.2017.2662206.

[58] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9351. doi: 10.1007/978-3-319-24574-4_28.

[59] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, 2014, vol. 3, no. January. doi: 10.1007/978-3-658-40442-0_9.

[60] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets Mehdi,"

arXiv1411.1784v1 [cs.LG] 6 Nov 2014 Cond., 2018.

[61] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 5967–5976, 2017, doi: 10.1109/CVPR.2017.632.

[62] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, vol. 2017-October. doi: 10.1109/ICCV.2017.244.

[63] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual Losses for Style Transfer and SR," *Eccv*, pp. 1–5, 2016, [Online]. Available: https://cs.stanford.edu/people/jcjohns/papers/eccv16/JohnsonECCV16.pdf%0Ahttp://arxiv.org/abs/1603.08155

[64] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[65] L. A. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2015, vol. 2015-January.

[66] L. A. Gatys, A. S. Ecker, and M. Bethge, "A Neural Algorithm of Artistic Style arXiv : 1508 . 06576v1 [cs . CV] 26 Aug 2015," *arXiv*, 2015.

[67] 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, 2004, doi: 10.1109/TIP.2003.819861.

[68] H. Liu and Z. Wang, "Perceptual quality assessment of medical images," *Encycl. Biomed. Eng.*, vol. 1–3, pp. 588–596, 2019, doi: 10.1016/B978-0-12-801238-3.64099-0.

[69] R. Gonzalez and R. Woods, *Digital Image Processing Global Edition*. Pearson Deutschland, 2017. [Online]. Available: https://elibrary.pearson.de/book/99.150005/9781292223070

[70] P. R. Derenoncourt *et al.*, "Ventilation-perfusion scan: A primer for practicing radiologists," *Radiographics*, vol. 41, no. 7, 2021, doi: 10.1148/RG.2021210060.

[71] P. J. Hagen, I. J. C. Hartmann, O. S. Hoekstra, M. P. M. Stokkel, G. J. J. Teule, and M. H. Prins, "How to use a gestalt interpretation for ventilation-perfusion lung scintigraphy," *J. Nucl. Med.*, vol. 43, no. 10, 2002.

[72] A. Vyas, S. Yu, and J. Paik, "Fundamentals of digital image processing," in *Signals and Communication Technology*, 2018. doi: 10.1007/978-981-10-7272-7_1.

[73] A. Bovik, *Handbook of Image and Video Processing, Second Edition*. 2005. doi: 10.1016/B978-0-12-119792-6.X5062-1.

[74] Y. Jiao and B. W. Huang, "A new adaptive threshold image-denoising method based on edge detection," in *Applied Mechanics and Materials*, 2014, vol. 678. doi: 10.4028/www.scientific.net/AMM.678.137.

[75] J. S. Lim, "Two-dimensional signal and image processing," *Englewood Cliffs, NJ, Prentice Hall*, vol. 710, 1990.

[76] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-D transform-domain collaborative filtering,” *IEEE Trans. Image Process.*, vol. 16, no. 8, 2007, doi: 10.1109/TIP.2007.901238.

[77] J. Canny, “A Computational Approach to Edge Detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, 1986, doi: 10.1109/TPAMI.1986.4767851.

[78] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, “Bilateral filtering: Theory and applications,” *Found. Trends Comput. Graph. Vis.*, vol. 4, no. 1, 2009, doi: 10.1561/0600000020.

[79] A. Buades, B. Coll, and J. M. Morel, “A non-local algorithm for image denoising,” in *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 2005, vol. II. doi: 10.1109/CVPR.2005.38.

[80] F. Mettler and M. Guiberteau, *Essentials of nuclear medicine imaging*. 2012. doi: 10.1016/C2009-0-61619-6.

[81] X. Fu, L. Bi, A. Kumar, M. Fulham, and J. Kim, “Multimodal Spatial Attention Module for Targeting Multimodal PET-CT Lung Tumor Segmentation,” *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 9, 2021, doi: 10.1109/JBHI.2021.3059453.

[82] R. Hu and H. Liu, “TransEM: Residual Swin-Transformer Based Regularized PET Image Reconstruction,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2022, vol. 13434 LNCS. doi: 10.1007/978-3-031-16440-8_18.

[83] X. Lu, S. Wu, Z. Xiao, and X. Huang, “An enhanced multiscale generation and depth-perceptual loss-based super-resolution network for prostate ultrasound images,” *Meas. Sci. Technol.*, vol. 34, no. 2, 2023, doi: 10.1088/1361-6501/ac99f1.

[84] L. Xiang *et al.*, “Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI,” *Neurocomputing*, vol. 267, 2017, doi: 10.1016/j.neucom.2017.06.048.

[85] F. Liu, H. Jang, R. Kijowski, T. Bradshaw, and A. B. McMillan, “Deep learning MR imaging-based attenuation correction for PET/MR imaging,” *Radiology*, vol. 286, no. 2, 2018, doi: 10.1148/radiol.2017170700.

[86] K. Armanious *et al.*, “MedGAN: Medical image translation using GANs,” *Comput. Med. Imaging Graph.*, vol. 79, 2020, doi: 10.1016/j.compmedimag.2019.101684.

[87] K. Zhao *et al.*, “Study of low-dose PET image recovery using supervised learning with CycleGAN,” *PLoS One*, vol. 15, no. 9 september, 2020, doi: 10.1371/journal.pone.0238455.

[88] Y. Lei *et al.*, “Whole-body PET estimation from low count statistics using cycle-consistent generative adversarial networks,” *Phys. Med. Biol.*, vol. 64, no. 21, 2019, doi: 10.1088/1361-6560/ab4891.

[89] N. Aghakhan Olia *et al.*, “Deep learning–based denoising of low-dose SPECT myocardial perfusion images: quantitative assessment and clinical performance,” *Eur. J. Nucl. Med. Mol. Imaging*, vol. 49, no. 5, 2022, doi: 10.1007/s00259-021-05614-7.

[90] H. Arabi, A. AkhavanAllaf, A. Sanaat, I. Shiri, and H. Zaidi, "The promise of artificial intelligence and deep learning in PET and SPECT imaging," *Phys. Medica*, vol. 83, 2021, doi: 10.1016/j.ejmp.2021.03.008.

[91] C. Lin, Y. C. Chang, H. Y. Chiu, C. H. Cheng, and H. M. Huang, "Reducing scan time of paediatric 99mTc-DMSA SPECT via deep learning," *Clin. Radiol.*, vol. 76, no. 4, 2021, doi: 10.1016/j.crad.2020.11.114.

[92] T. Ryden, M. Van Essen, I. Marin, J. Svensson, and P. Bernhardt, "Deep-learning generation of synthetic intermediate projections improves 177Lu SPECT images reconstructed with sparsely acquired projections," *J. Nucl. Med.*, vol. 62, no. 4, 2021, doi: 10.2967/jnumed.120.245548.

[93] B. Pan *et al.*, "Ultra high speed SPECT bone imaging enabled by a deep learning enhancement method: a proof of concept," *EJNMMI Phys.*, vol. 9, no. 1, 2022, doi: 10.1186/s40658-022-00472-0.

[94] A. Jabbarpour *et al.*, "The Past, Present, and Future Role of Artificial Intelligence in Ventilation/Perfusion Scintigraphy: A Systematic Review," *Seminars in Nuclear Medicine*. 2023. doi: 10.1053/j.semnuclmed.2023.03.002.

[95] H. N. Wagner, V. Lopez-Majano, J. K. Langan, and R. C. Joshi, "Radioactive xenon in the differential diagnosis of pulmonary embolism.," *Radiology*, vol. 91, no. 6, 1968, doi: 10.1148/91.6.1168.

[96] H. N. Wagner, D. C. Sabiston, M. Iio, J. G. Mcafee, J. K. Meyer, and J. K. Langan, "Regional Pulmonary Blood Flow in Man by Radioisotope Scanning," *JAMA J. Am. Med. Assoc.*, vol. 187, no. 8, 1964, doi: 10.1001/jama.1964.03060210051012.

[97] D. L. Bailey and P. J. Roach, "A Brief History of Lung Ventilation and Perfusion Imaging Over the 50-Year Tenure of the Editors of Seminars in Nuclear Medicine," *Seminars in Nuclear Medicine*, vol. 50, no. 1. 2020. doi: 10.1053/j.semnuclmed.2019.07.004.

[98] G. D. Tourassi, C. E. Floyd, and R. E. Coleman, "Acute pulmonary embolism: Cost-effectiveness analysis of the effect of artificial neural networks on patient care," *Radiology*, vol. 206, no. 1, pp. 81–88, 1998, doi: 10.1148/radiology.206.1.9423655.

[99] J. A. Scott, E. L. Palmer, and A. J. Fischman, "How well can radiologists using neural network software diagnose pulmonary embolism?," *Am. J. Roentgenol.*, vol. 175, no. 2, 2000, doi: 10.2214/ajr.175.2.1750399.

[100] P. B. Bonnefoy *et al.*, "Ventilation/perfusion (V/Q) scanning in contemporary patients with pulmonary embolism: utilization rates and predictors of use in a multinational study," *J. Thromb. Thrombolysis*, vol. 53, no. 4, 2022, doi: 10.1007/s11239-021-02579-0.

[101] R. C. Wang *et al.*, "Trends in Imaging for Suspected Pulmonary Embolism across US Health Care Systems, 2004 to 2016," *JAMA Netw. Open*, vol. 3, no. 11, 2020, doi: 10.1001/jamanetworkopen.2020.26930.

[102] D. Moher *et al.*, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *PLoS Medicine*, vol. 6, no. 7. 2009. doi: 10.1371/journal.pmed.1000097.

[103] W.-C. Siu and K.-W. Hung, "Review of image interpolation and super-resolution," in

Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference, 2012, pp. 1–10.

[104] T. M. Lehmann, C. Gönner, and K. Spitzer, "Survey: Interpolation methods in medical image processing," *IEEE Trans. Med. Imaging*, vol. 18, no. 11, 1999, doi: 10.1109/42.816070.

[105] A. E. Ilesanmi and T. O. Ilesanmi, "Methods for image denoising using convolutional neural network: a review," *Complex Intell. Syst.*, vol. 7, no. 5, 2021, doi: 10.1007/s40747-021-00428-4.

[106] P. J. Roach, D. L. Bailey, and B. E. Harris, "Enhancing Lung Scintigraphy With Single-Photon Emission Computed Tomography," *Seminars in Nuclear Medicine*, vol. 38, no. 6. 2008. doi: 10.1053/j.semnuclmed.2008.06.002.

[107] P. Hannequin and J. Mas, "Statistical and heuristic image noise extraction (SHINE): A new method for processing Poisson noise in scintigraphic images," *Physics in Medicine and Biology*, vol. 47, no. 24. 2002. doi: 10.1088/0031-9155/47/24/302.

[108] Y. Zhang, D. Zhao, J. Zhang, R. Xiong, and W. Gao, "Interpolation-dependent image downsampling," *IEEE Trans. Image Process.*, vol. 20, no. 11, 2011, doi: 10.1109/TIP.2011.2158226.

[109] P. Gravel, G. Beaudoin, and J. A. De Guise, "A method for modeling noise in medical images," *IEEE Trans. Med. Imaging*, vol. 23, no. 10, 2004, doi: 10.1109/TMI.2004.832656.

[110] D. White and R. S. Lawson, "A Poisson resampling method for simulating reduced counts in nuclear medicine images," *Phys. Med. Biol.*, vol. 60, no. 9, pp. N167–N176, May 2015, doi: 10.1088/0031-9155/60/9/N167.

[111] R. Petrocelli, A. Vija, E. Hawman, and A. Yahil, "The role of signal-to-noise ratio in preserving diagnostic performance and reader confidence level in count-reduced planar nuclear studies undergoing Pixon image processing," *J. Nucl. Med.*, vol. 47, no. suppl 1, pp. 371P LP-371P, May 2006, [Online]. Available: http://jnm.snmjournals.org/content/47/suppl_1/371P.2.abstract

[112] C. C. Kuni, B. H. Hasegawa, and W. R. Hendee, "Noise reduction in nuclear medicine images," *J. Nucl. Med.*, vol. 24, no. 6, 1983.

[113] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," *Nat. Rev. Cancer*, vol. 18, no. 8, pp. 500–510, 2018, doi: 10.1038/s41568-018-0016-5.

[114] S. A. Abdelaziz Ismael, A. Mohammed, and H. Hefny, "An enhanced deep learning approach for brain cancer MRI images classification using residual networks," *Artif. Intell. Med.*, vol. 102, no. December 2019, p. 101779, 2020, doi: 10.1016/j.artmed.2019.101779.

[115] D. Gourdeau, S. Duchesne, and L. Archambault, "On the proper use of structural similarity for the robust evaluation of medical image synthesis models," *Med. Phys.*, vol. 49, no. 4, pp. 2462–2474, 2022, doi: 10.1002/mp.15514.

[116] M. F. Smith and J. Jenkins, "PET Myocardial Perfusion Imaging of Extremely Obese Patients: A Study with an Ultra-Large Anthropomorphic Cardiac-Torso Phantom," 2019. doi: 10.1109/NSS/MIC42101.2019.9059682.

[117] R. Nirthika, S. Manivannan, A. Ramanan, and R. Wang, "Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study," *Neural Computing and Applications*, vol. 34, no. 7. 2022. doi: 10.1007/s00521-022-06953-8.

[118] B. Nestor *et al.*, "Rethinking clinical prediction: Why machine learning must consider year of care and feature aggregation," no. NeurIPS 2018, pp. 1–7, 2018, [Online]. Available: http://arxiv.org/abs/1811.12583

[119] J. K. Winkler *et al.*, "Association Between Surgical Skin Markings in Dermoscopic Images and Diagnostic Performance of a Deep Learning Convolutional Neural Network for Melanoma Recognition," *JAMA Dermatology*, vol. 155, no. 10, pp. 1135–1141, Oct. 2019, doi: 10.1001/jamadermatol.2019.1735.

[120] M. A. Badgeley *et al.*, "Deep learning predicts hip fracture using confounding patient and healthcare variables," *npj Digit. Med.*, vol. 2, no. 1, 2019, doi: 10.1038/s41746-019-0105-1.

[121] S. E. Davis, R. A. Greevy, C. Fonnesbeck, T. A. Lasko, C. G. Walsh, and M. E. Matheny, "A nonparametric updating method to correct clinical prediction model drift," *J. Am. Med. Informatics Assoc.*, vol. 26, no. 12, pp. 1448–1457, 2019, doi: 10.1093/jamia/ocz127.

[122] D. J. Hawkes *et al.*, "Registration and display of the combined bone scan and radiograph in the diagnosis and management of wrist injuries," *Eur. J. Nucl. Med.*, vol. 18, no. 9, 1991, doi: 10.1007/BF00173434.

[123] M. Sadik *et al.*, "Computer-assisted interpretation of planar whole-body bone scans," *J. Nucl. Med.*, vol. 49, no. 12, pp. 1958–1965, 2008, doi: 10.2967/jnumed.108.055061.

[124] A. Yahil, A. H. Vija, and E. G. Hawman, "Enhanced feature extraction in planar nuclear medicine using pixon® minimum-complexity image processing," *IEEE Nucl. Sci. Symp. Conf. Rec.*, vol. 4, pp. 2037–2041, 2006, doi: 10.1109/NSSMIC.2006.354314.

[125] A. Mehranian *et al.*, "Deep learning–based time-of-flight (ToF) image enhancement of non-ToF PET scans," *Eur. J. Nucl. Med. Mol. Imaging*, vol. 49, no. 11, pp. 3740–3749, 2022, doi: 10.1007/s00259-022-05824-7.

[126] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient BackProp," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7700, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 9–48.

[127] A. Goodfellow, Ian J.Bengio, Yoshua Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org

[128] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, 2011.

[129] J. Ouyang, K. T. Chen, E. Gong, J. Pauly, and G. Zaharchuk, "Ultra-low-dose PET reconstruction using generative adversarial network with feature matching and task-specific perceptual loss," *Med. Phys.*, vol. 46, no. 8, pp. 3555–3564, 2019, doi: 10.1002/mp.13626.

[130] V. Ghodrati *et al.*, "MR image reconstruction using deep learning: Evaluation of network structure and loss functions," *Quant. Imaging Med. Surg.*, vol. 9, no. 9, pp. 1516–1527, 2019, doi: 10.21037/qims.2019.08.10.

[131] A. Ben-Cohen *et al.*, "Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection," *Eng. Appl. Artif. Intell.*, vol. 78, 2019, doi: 10.1016/j.engappai.2018.11.013.

[132] D. Nie *et al.*, "Medical Image Synthesis with Deep Convolutional Adversarial Networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 12, 2018, doi: 10.1109/TBME.2018.2814538.

[133] S. Pasumarthi, J. I. Tamir, S. Christensen, G. Zaharchuk, T. Zhang, and E. Gong, "A generic deep learning model for reduced gadolinium dose in contrast-enhanced brain MRI," *Magn. Reson. Med.*, vol. 86, no. 3, pp. 1687–1700, 2021, doi: 10.1002/mrm.28808.

[134] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.

[135] X. Mei *et al.*, "RadImageNet: An Open Radiologic Deep Learning Research Dataset for Effective Transfer Learning," *Radiol. Artif. Intell.*, vol. 4, no. 5, 2022, doi: 10.1148/ryai.210315.

[136] J. Song, H. Yi, W. Xu, X. Li, B. Li, and Y. Liu, "Dual Perceptual Loss for Single Image Super-Resolution Using ESRGAN," *Heliyon*, Jan. 2022, doi: 10.1016/j.heliyon.2023.e15134.

[137] M. Seitzer *et al.*, "Adversarial and perceptual refinement for compressed sensing MRI reconstruction," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11070 LNCS, pp. 232–240, 2018, doi: 10.1007/978-3-030-00928-1_27.

[138] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," 2016.

[139] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *J. R. Stat. Soc. Ser. B*, vol. 57, no. 1, pp. 289–300, 1995, doi: https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

[140] Y. Sugawara, S. Shiota, and H. Kiya, "Super-Resolution Using Convolutional Neural Networks Without Any Checkerboard Artifacts," *Proc. - Int. Conf. Image Process. ICIP*, pp. 66–70, 2018, doi: 10.1109/ICIP.2018.8451141.

[141] Z. Fu, Y. Zheng, T. Ma, H. Ye, J. Yang, and L. He, "Edge-aware deep image deblurring," *Neurocomputing*, vol. 502, pp. 37–47, 2022, doi: 10.1016/j.neucom.2022.06.051.

[142] M. Deng, A. Goy, S. Li, K. Arthur, and G. Barbastathis, "Probing shallower: perceptual loss trained Phase Extraction Neural Network (PLT-PhENN) for artifact-free reconstruction at low photon budget," *Opt. Express*, vol. 28, no. 2, pp. 2511–2535, Jan. 2020, doi: 10.1364/OE.381301.