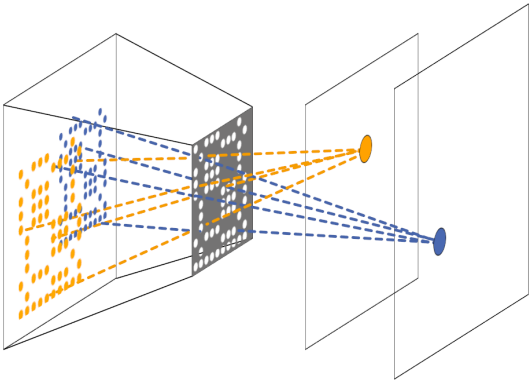
29
KARLSRUHE TRANSACTIONS
ON BIOMEDICAL ENGINEERING

TOBIAS MEIßNER

Development and Evaluation of Coded Aperture Reconstruction Methods for Intraoperative Gamma Cameras



Scientific Publishing

Tobias Meißner

Development and Evaluation of Coded Aperture Reconstruction Methods for Intraoperative Gamma Cameras

Vol. 29
Karlsruhe Transactions on Biomedical Engineering

Editor:
Karlsruhe Institute of Technology (KIT)
Institute of Biomedical Engineering

Eine Übersicht aller bisher in dieser Schriftenreihe erschienenen Bände finden Sie am Ende des Buchs.

Development and Evaluation of Coded Aperture Reconstruction Methods for Intraoperative Gamma Cameras

by
Tobias Meißner

KIT Scientific Publishing

Karlsruher Institut für Technologie
Institut für Biomedizinische Technik

Development and Evaluation of Coded Aperture
Reconstruction Methods for Intraoperative Gamma Cameras

Zur Erlangung des akademischen Grades eines Doktors der Ingenieurwissenschaften (Dr.-Ing.) von der KIT-Fakultät für Elektrotechnik und Informationstechnik des Karlsruher Instituts für Technologie (KIT) genehmigte Dissertation

von Tobias Meißner, M.Sc.

Tag der mündlichen Prüfung: 11. Oktober 2024
Referent: Prof. Dr. rer. nat. Werner Nahm
Korreferent: Prof. Dr. rer. nat. Jürgen Hesser

Impressum

KIT Scientific Publishing

Karlsruher Institut für Technologie (KIT)
KIT Scientific Publishing
Straße am Forum 2
D-76131 Karlsruhe
E-Mail: info@ksp.kit.edu

www.ksp.kit.edu
https://www.bibliothek.kit.edu/ksp.php

Print on Demand 2024 – Gedruckt auf FSC-zertifiziertem Papier

ISSN 1864-5933
ISBN 978-3-7315-1394-0
DOI 10.5445/KSP/1000176008

Abstract

The overall aim of this thesis is to improve the outcome of Radioguided Surgeries (RGSs) by developing and evaluating reconstruction methods to exploit the advantages of Coded Aperture Imaging (CAI) for Intraoperative Gamma Cameras (IGCs).

Motivation: Female breast cancer has surpassed lung cancer as the most commonly diagnosed cancer. The biopsy and examination of lymph nodes that receive lymphatic drainage from the primary tumor are guided by IGCs. CAI has been proposed as an alternative to parallel-hole or pinhole collimators to produce an image. It offers a better trade-off between sensitivity and spatial resolution, but requires image reconstruction.

Methods: First, planar reconstruction was investigated where the source's depth is assumed to be known. A Convolutional Encoder-Decoder (CED) was developed and trained on synthetic source distributions and a low-fidelity simulation. It was quantitatively compared to reconstruction methods from the literature, such as MURA Decoding and a Maximum Likelihood Expectation Maximization (MLEM) algorithm. The computing time and the Contrast-to-Noise Ratio (CNR) served as key metrics throughout this thesis. Moreover, the ability of super-resolution was investigated by reconstructing bilinear interpolations of simulated low-resolution detector images. In the second part of this thesis, the assumption of a known source-to-mask distance was loosened and the axial resolution quantified using the Full Width at Half Maximum (FWHM) of the axial CNR profile of a point-like source. Lastly, an Iterative Source Localization (ISL) algorithm incorporating an Exponentially Modified Gaussian (EMG) fitting was developed to localize sources in all three dimensions.

Results: While MURA Decoding quickly provides robust reconstructions with good quality, MLEM takes around 170 times longer with a 1.2 times

higher CNR. The CED performed the best with an on average 2.7 times higher CNR and a runtime comparable to MURA Decoding, despite the low-fidelity simulation of the training data. The simulation study indicated that super-resolution is feasible. Regarding 3D reconstruction, it was found that, for MURA Decoding, the axial resolution degraded from 5.3 mm FWHM at 12 mm mask-to-source distance to 42.2 mm at 100 mm. The ISL combined with an EMG fit achieved a mean localization error of 0.8 mm on the simulated and 2.6 mm on the experimental data in the imaging range of 20−100 mm.

Conclusion: The MLEM algorithm yields a higher CNR and better axial resolution, but is not suitable for RGS in its current form, due to its computational complexity. MURA Decoding provides robust reconstructions. Its fast computation enables 3D reconstruction which allows the localization of point-like sources with an accuracy comparable to that of stereoscopic cameras.

Zusammenfassung

Das Ziel dieser Arbeit ist es, den Ausgang von Sonden-geführten Eingriffen (SGE) zu verbessern, indem Rekonstruktionsmethoden für die Bildgebung mit einer kodierten Aperture (KA) entwickelt und evaluiert werden. Damit sollen die Vorteile von KA für intraoperative Gamma-Kameras (IGK) nutzbar gemacht werden.

Motivation: Brustkrebs bei Frauen hat Lungenkrebs als die am weltweit häufigste diagnostizierte Krebsart überholt. Die Biopsie und Untersuchung von Lymphknoten, die einen Lymphabfluss vom Primärtumor erhalten, wird mit Hilfe von IGKs durchgeführt. Als Alternative zu Parallelkollimatoren oder Lochblenden, wurde zur Bilderzeugung KA entwickelt. Sie bieten einen besseren Kompromiss zwischen Sensitivität und räumlicher Auflösung, erfordern allerdings eine Bildrekonstruktion.

Methoden: Zunächst wurde die planare Rekonstruktion untersucht, bei der die Tiefe der Quelle als bekannt angenommen wird. Ein faltungsbasiertes Encoder-Decoder-Netzwerk (FED) wurde entworfen und auf synthetischen Quellverteilungen und einer einfachen Simulation trainiert. Danach wurde es mit Methoden aus der Literatur, u.a. MURA Decoding und einem Maximum-Likelihood-Schätzer (MLS), quantitativ verglichen. Die Laufzeit und das Kontrast-zu-Rausch-Verhältnis (KRV) dienten hierbei als Hauptmetriken. Des Weiteren wurde die Möglichkeit der Super-Resolution untersucht, bei der niedrig aufgelöste Bilder vor der Rekonstruktion mittels bilinearer Interpolation skaliert wurden. Im zweiten Teil dieser Arbeit wurde die Annahme einer bekannten Quelltiefe fallen gelassen und die axiale Auflösung als Volle Breiter bei Halber Höhe (VBHH) des axialen KRV-Profils einer punktähnlichen Quelle bestimmt. Zuletzt wurde die Genauigkeit analysiert, mit welcher eine Quelle in allen drei Raumrichtungen lokalisiert werden kann. Ein Iterativer

Quell-Lokalisierungsalgorithmus (IQL) einschließlich einem Fitting mit einer symmetrischen Gauß-Verteilung oder einer exponentiell modifizierten Gauß-Verteilung (EMG) wurde dazu entwickelt.

Ergebnisse: Während MURA Decoding robust und schnell eine gute Rekonstruktionsqualität liefert, brauchte der MLS für ein 1.2 mal höheres KRV ca. 170 mal länger. Das FED erzielt die besten Ergebnisse mit einem durchschnittlich 2.7 mal höheren KRV, trotz seiner simplen Trainingssimulation. Die Ergebnisse der Simulationsstudie deuten daraufhin, dass Super-Resolution realisierbar ist. Im Bezug auf 3D-Bildgebung wurde herausgefunden, dass für MURA Decoding die axiale Auflösung von 5.3 mm VBHH bei 12 mm Quelldistanz zu 42.2 mm VBHH bei 100 mm abnimmt. Der IQL erreicht in der Kombination mit dem EMG-Fitting einen durchschnittlichen Lokalisierungsfehler von 0.8 mm auf simulierten und von 2.6 mm auf experimentellen Daten für Quelldistanzen im Bereich von $20-100$ mm.

Schlussfolgerungen: Der MLS bieten generell ein höheres KRV und eine bessere axiale Auflösung, ist aber aufgrund seines Rechenaufwands in seiner derzeitigen Form für SGE nicht geeignet. MURA Decoding bietet eine robuste Rekonstruktion. Die schnelle Berechnungszeit erlaubt die Generierung von 3D-Rekonstruktionen. Mit diesen ist eine Lokalisierungsgenauigkeit von punktähnlichen Quellen erreichbar, die mit Stereo-Kameras vergleichbar ist.

Acknowledgments

Jürgen and Werner, I am very grateful that you gave me the opportunity to pursue an academical career. I really value our discussions, your beneficial advise, and the stable and reliable working conditions you managed to provide. I would have never finished this thesis without the support of both research groups from the IBT and the MIISM I had the chance to be a part of. Thank you, Alex, Alice, Andrei, Daniel, Carmen, Claudia, Cristian, Jonathan, Joshua, Katharina, Lei , Lorena, Lu, Marlen, Marcus, Matthias, Miriam, Moritz, Patricia (I am still sorry about the flower bouquet), Peter, René, Sara, Silvia, and Simon. A special thanks goes out to our secretaries Astrid and Christiane, who were always dedicated to simplify my work life. Another special thanks goes to Sven Köhler, who kept me going with his discussions when I was doubting the entire undertaking.

Additionally, I would like to say thank you to my collaborators Vladislav Rozhkov, Prof. Russo, and Laura Antonia Cerbone, who took the effort and time to cooperate, discuss, argue, and still wrote papers with me together and, incidentally, showed me the best pizza spots in Napoli (sono “Napoli 1820” e “Elettroforno Giulia”). Getting the opportunity to work with Paolo and Laura, advanced my work significantly and I really enjoyed my stay in the fascinating city of Napoli.

None of this would have been possible without the work of my students: Pauline Seubert, Jannik Petersen, Dominik Zehender, Saverio Pietrantonio, Elisa Stegmeier. Through their brains and hands, important parts of this thesis were realized. I am grateful for my friends who put up with my complaining from time to time, but always kept on listening and asking. Flo, Mathis, Dominik, Lukas, Yannick, Lena, Dominik, Sven, Corinna, Isa, Dariusch, Billy, Gloria, ...

I hope all those not mentioned know how grateful I am for their friendship and companionship.

Finally, and most importantly, I would like to express my gratitude to my family. My parents, my mum, my dad, Dietmar, Susanne, gave me the opportunity to attend university and encouraged my decision to do a PhD. Many thanks to my brothers and sister, Philip, René, Alexander, and Marayke for your invaluable emotional support.

“ It isn’t only art that is incompatible with happiness, it’s also science. Science is dangerous, we have to keep it most carefully chained and muzzled. ”

– MUSTAPHA MOND, 2540 AD
Resident World Controller of a soulless, horribly inhumane, and deeply uneducated society in Aldous Huxley’s “Brave New World”

Contents

Abbreviations

3D-MLEM	3D Maximum Likelihood Expectation Maximization
ALND	Axillary Lymph Node Dissection
ASIC	Application-Specific Integrated Circuit
CAI	Coded Aperture Imaging
CED	Convolutional Encoder-Decoder
CMOS	Complementary Metal-Oxide Semiconductor
CNN	Convolutional Neural Network
CNR	Contrast-to-Noise Ratio
EMG	Exponentially Modified Gaussian
FOV	Field of View
FWHM	Full Width at Half Maximum
GT	Ground Truth
IGC	Intraoperative Gamma Camera
ISL	Iterative Source Localization
LCR	Largest Connected Region
MC	Monte Carlo
ML	Machine Learning
MLEM	Maximum Likelihood Expectation Maximization
MRI	Magnetic Resonance Imaging
MURA	Modified Uniformly Redundant Arrays
NTHT	No-Two-Holes-Touching
PET	Positron Emission Tomography
PMT	Photomultiplier Tube
PSF	Point Spread Function
RGS	Radioguided Surgery
ROI	Region of Interest

SLN	Sentinel Lymph Node
SLNB	Sentinel Lymph Node Biopsy
SPECT	Single Photon Emission Computed Tomography
SSIM	Structural Similarity Index Measure
THT	Two-Holes-Touching
URA	Uniformly Redundant Array

Chapter 1

Introduction

1.1 Motivation

Female breast cancer has surpassed lung cancer as the most commonly diagnosed cancer, with an estimated 2.3 million new cases worldwide in 2020 [1]. In Germany alone, the Robert Koch institute estimates that 74,500 women and 600 men are diagnosed with breast cancer per year [2]. Over the decades, axillary staging of breast cancer has evolved, becoming less invasive and more conservative. It has moved from dissecting the entire axillary lymph system to the less invasive Sentinel Lymph Node Biopsy (SLNB) [3], which is nowadays the gold standard for early-stage patients with clinically negative axilla [4]. The objective of SLNB is to locate, dissect and to pathologically examine the Sentinel Lymph Nodes (SLNs), i.e. lymph nodes in the lymphatic drainage of the primary tumor, which are the first nodes to receive lymph-borne metastatic cells [3, 5]. In the last decades, surgical procedures like SLNB are driving the development of Intraoperative Gamma Camera (IGC) to perform Radioguided Surgery (RGS) [6–8]. IGCs are gamma cameras that are used to visualize the distribution of biological structures during the surgery that were marked with radioactive tracers, emitting gamma photons from inside the body. In hospitals world-wide, IGCs guide the surgeons in SLNB as well as in surgical resection for localized tumors [5]. In order to neither stall nor interrupt the surgical workflow, IGCs have to be compact and lightweight to remain maneuverable. Additionally, they are required to provide fast acquisitions of images that are

both high in contrast and in spatial resolution [7, 8]. The collimator represents the component of IGCs that has the largest impact on the imaging performance. Compared to conventional forms of collimation, Coded Aperture Imaging (CAI) offers a superior trade-off between acquisition time and spatial resolution [7]. However, it requires image reconstruction that adds a layer of complexity to the imaging procedure. In the field of computer vision, Machine Learning (ML) approaches have gained considerable popularity, especially in image regression tasks [9, 10] and, thus, present a promising candidate for CAI reconstruction. Besides an accelerated imaging procedure and more detailed images, CAI is able to provide additional information about the source-to-camera distance. Overall, the development and evaluation of reconstruction methods, both for planar and 3D imaging, with regards to an intraoperative application has enormous potential to provide depth information and, hence, to improve the outcome of RGS [7, 8].

1.2 State of the Art

Stationary gamma cameras, such as Single Photon Emission Computed Tomography (SPECT) systems, are an established tool in nuclear medicine to provide comprehensible visualizations of the internal distribution of radiotracer materials. The goal is to measure the distribution of radiotracers that are specifically designed to accumulate in organs or tissue of interest in order to detect abnormalities which are characteristic of a disease [11–13]. Over the last few decades, intraoperative imaging has increasingly gained importance, and surgical procedures have become less invasive, leading to the development of RGS, where a radiation detection probe is employed during the surgery and within the surgical suite [8, 14]. In most clinics, RGS relies on a combination of preoperative imaging with intraoperative guidance with a handheld gamma probe. Such gamma probes provide instantaneous feedback on the local radioactivity concentration by displaying the count rate and offering variable-pitch acoustic feedback, which is essential for a rapid and precise detection in RGS [12]. However, several limitations were identified [14]: First, acoustic gamma probes do not provide image documentation for the medical record. Second, the outcome is highly operator dependent, and requires a lot of experience about the correct positioning. Third, SLNs, that are either deeply located or have a low uptake

of radiotracer, are hard to locate and may be missed due to the low emission and high attenuation inside the body. IGCs are small, compact, and lightweight imaging probes that offer solutions to the short-comings of acoustic gamma probes. They offer two key advantages: Instead of discrete point measurements, IGCs allow for a broader overview of the incision site making surveying larger areas less tedious and difficult for the surgery team. Additionally, the provided images from IGCs are considered more intuitive guidance than numerical or acoustic feedback [8].

Commercially available IGCs exist. Among them are, for example, the CrystalCam [15, 16], distributed by Crystal Photonics GmbH (Berlin, Germany), and the Sentinella 102 [17] by Oncovision (Valencia, Spain). In the research community, more diverse forms of IGCs are being investigated: the ultra-portable wireless PGC [18], λ-Eye, which is specifically designed for SLNB [19], and stereoscopic gamma cameras for 2D imaging with additional depth estimation [20], to name a few. The performance of IGCs are dominated by their collimator design and most of the current devices use either pinhole or parallel-hole collimators [7]. However, collimation always represents a trade-off between harvesting as many photons as possible (sensitivity) and high-resolution imaging (spatial resolution). For a pinhole collimator, this translates into the following: the smaller the opening, the sharper the image, but at the same time, a small opening also means that a large proportion of the incoming photon flux is blocked, resulting in a longer exposure time [11]. To improve the compromise between fast image acquisition and high-resolution images the concept of coded apertures was introduced independently by Dicke [21] and Ables [22] in 1968. Initially proposed for X-ray astronomy, a mask with multiple pinholes between the radiating object and detector encodes the directional information of incoming gamma rays. An increase in photon detection of nine to 20-fold is possible [23, 24], and a superior spatial resolution of below 1 mm was reported [25]. Another advantage of CAI is that, unlike parallel-hole collimation, magnification can be achieved, allowing larger areas to be surveyed, making it particularly useful for the source localization. However, image reconstruction is required to obtain an interpretable image of the scene. The reconstruction method is a major driver of the final image quality and several reconstruction methods have been suggested [23, 24, 26–29]. However, a thorough and quantitative analysis of reconstruction methods with regard to both the computational burden and the image quality is still missing. A few ML approaches have been

reported [29, 30], but not for high-resolution detectors and an evaluation on data from an experimental IGC is still pending. Furthermore, CAI offers the possibility of 3D imaging: In contrast to pinhole or parallel-hole collimation, an in-focus plane must be selected. In x-ray astronomy or other far-field use-cases an image plane infinitely far away can be chosen, but this does not apply to RGS. Interestingly, the depth dependency of the in-focus plane can be exploited to estimate the source-to-camera distance, particularly in the near-field where the image projection is more sensitive to the depth than in the far-field. More common technologies for estimating the source-to-camera distance are stereo gamma cameras [20, 31, 32], external tracking [33], and additional optical stereo [34] or depth cameras [35].

In summary, coded aperture collimators combine the high geometric sensitivity of parallel-hole collimators with the excellent resolution and magnification properties of pinhole collimators. Despite the advantages of CAI, it remains the disfavored choice of collimation in IGCs. Image reconstruction is required and adds a layer of complexity, but offers the possibility of 3D imaging. A recent review article on IGCs has identified great potential for CAI in general and its capabilities of 3D imaging in particular, where the reconstruction methods play a central role. These promising prospects, however, are currently impeded by the considerable burden of conducting experiments with radioactive material, coupled with the lack of publicly available datasets [7].

1.3 Research Question and Hypotheses

The goal of this thesis is to develop and evaluate appropriate reconstruction methods in order to exploit CAI for RGS. Despite the increased sensitivity and spatial resolution, there is little research carried out in employing CAI in the surgical suite [7]. Even though, several planar reconstruction methods have been proposed [23, 24, 26–29], they have not been quantitatively compared on real-world data to analyze their strengths and weaknesses with regards to RGS. Additionally, the high spatial resolution of CAI is achieved by an aperture with very small pinholes, which requires such a high number of detector pixels that can only be achieved with less efficient semiconductor-based detectors. Thus, it is desirable to combine a low-resolution but sensitive detector with a high-resolution coded aperture mask, which is referred to as super-resolution.

Furthermore, the research community lacks publicly available datasets with a variety of source distributions, which heavily impedes investigating CAI in general and reconstruction methods specifically. These problems lead to the following research question, thesis, and two accompanying hypotheses that will guide the reader through this work:

Research Question I

How do planar reconstruction methods for CAI suited for RGS perform on real-world images?

Thesis I

A ML approach is able to achieve better reconstruction results in less time than known planar reconstruction methods.

Hypothesis Ia

A Convolutional Encoder-Decoder (CED) network trained on a synthetic dataset that is generated with a low-fidelity simulation achieves on average a higher Contrast-to-Noise Ratio (CNR) with less computation time than known planar reconstruction methods.

Hypothesis Ib

Even with a detector that cannot resolve the coded aperture mask's small pinholes, a CED network still outperforms other methods and thus, achieves super-resolution.

In order to investigate the proposed hypotheses and to answer the research questions of this thesis, the following milestones must be addressed:

(M_1) A literature research must be conducted to find the planar reconstruction methods for CAI that are suitable for RGS and its requirements, which are a fast reconstruction time and a computational complexity that fits into the surgery room. They need to be implemented in a common framework for a fair comparison.

(M_2) In ML, a plenitude of architectures for image-to-image regression tasks exist. Therefore, a suited ML architecture must be implemented, trained, and compared. This involves developing a fast simulation of coded

aperture images to generate a dataset for training and an appropriate training procedure. Additionally, a domain of training data must be selected.

(M_3) A sensible and established metric is required to compare image quality. Additionally, this metric should be deterministic and not require manual labelling of either source or background as this would compromise reproducibility.

(M_4) Real-life images captured with experimental gamma cameras suffer from noise and artefacts that are difficult to predict or simulate. Semiconductor-based detectors, for example, tend to suffer from cracks in the substrate, defective pixels and additionally environmental gamma or alpha particles that can heavily disturb image acquisition and reconstruction. Thus, an appropriate form of preprocessing must be found.

The application described above is linked to the strong assumption that the distance between the camera and the source is known when reconstructing an image. Reconstruction becomes more complex, when this assumption is loosened and at the same time opens up the challenge of 3D imaging. Yet, the capabilities of CAI to produce 3D reconstructions from a single acquisition have not been fully explored. Planar reconstruction methods require selecting an in-focus plane, yet, e.g. the depth of a SLN in the tissue is not known beforehand. Thus, the second research question emerges and the following thesis and hypotheses can be derived:

Research Question II

How can point-like gamma sources be localized with high precision and high accuracy in all three spatial dimensions?

Thesis II

By applying planar reconstruction methods at succeeding in-focus planes in the near-field, a 3D reconstruction can be generated, allowing to estimate the 3D position of point-like sources.

Hypothesis II

Single point-like gamma sources can be localized with an accuracy of approximately 5 mm in all three dimensions in the range of 20 mm to 100 mm from a stack of planar reconstructions, and is, thus, comparable to a stereoscopic camera.

When investigating the potential of 3D imaging, it is necessary to address the following milestones:

(M_5) From the planar reconstruction methods investigated, a suitable method must be chosen to analyze both the precision and accuracy for localizing point-like gamma sources. The existing 3D Maximum Likelihood Expectation Maximization (3D-MLEM) algorithm must be generalized to our setup in order to serve as a reference method.

(M_6) The axial resolution, as a measurement for the precision, must be determined for different distances between the camera and the source in a typical range for RGS.

(M_7) A 3D-localization method for point-like sources must be developed, where from a single image captured with a coded aperture camera all three spatial coordinates of the source are estimated. Comparing the estimated position with the true source position yields the localization error and, thus, the accuracy.

(M_8) A dataset of real-life images captured with an experimental IGC is required. A dataset of detector images where a point-like source is positioned at different distances from the camera covering the entire Field of View (FOV) is needed to evaluate the axial resolution, as well as the localization accuracy.

In summary, the research presented in this thesis, aims to leverage the coded aperture technology for IGC by developing and evaluating reconstruction methods. It focuses on two cases: First, when the source is at a well-known distance. Second, when the source is close to the camera, which is desirable for a high photon yield, 3D imaging is possible and the question of the axial resolution and localization accuracy arises. These finding would pose the basic to further investigate the 3D imaging capabilities of CAI.

1.4 Structure of This Thesis

This thesis consists of five parts: First, in Part I the fundamentals about gamma cameras in nuclear medicine (Sec. 2), the basic principles of CAI (Sec. 3) and the basics of ML (Sec. 4) are presented. Part II describes the experimental datasets as well as the synthetic datasets acquired and used in the course of this thesis. Part III investigates planar CAI guided by research hypotheses Ia and Ib. It presents planar reconstruction methods that are suitable for RGS before showing the development and training of the ML approach. The selected reconstruction methods are then quantitatively compared regarding their computational costs and reconstruction quality. Afterwards, the concept of super-resolution is introduced and analyzed. In Part IV, the possibility of 3D imaging with a coded aperture gamma is investigated, as stated by research hypothesis II: the axial resolution of an experimental IGC is assessed, followed by analyzing the accuracy in localizing point-like gamma sources in all three spatial dimensions. Finally, Part V gives an outlook on further research, summarizes the main results of this work, and eventually completes this thesis by drawing its main conclusions.

PART I

FUNDAMENTALS

Chapter 2

Intraoperative Gamma Cameras in Nuclear Medicine

Intraoperative Gamma Cameras (IGCs) in nuclear medicine are specialized imaging devices used during Radioguided Surgery (RGS) aiding surgeons to locate marked tissue. This chapter provides an overview on how an image is generated and their application in nuclear medicine. In order to assess the imaging capabilities of IGCs, the main performance parameters of IGCs are describes in Sec. 2.1, while Sec. 2.2 and 2.3 give a detail overview over the state-of-the art technologies used. Finally, Sec. 2.4 provides an overview of Sentinel Lymph Node Biopsy (SLNB), the most popular procedure guided by IGCs.

2.1 Performance Parameters

IGCs combine two components to produce an image: a position-sensitive detector to measure the position and energy of an impinging gamma photon and a collimator between the detector and source to obtain its directional information. Both components heavily influence the camera's imaging capabilities. The following main performance characteristics of IGC were identified [7, 8]:

Sensitivity The sensitivity describes the efficiency in converting impinging radiation into a usable signal and is an important performance parameter due to the low activity in RGS. This parameter is mainly influenced by the detector choice (see Sec. 2.2). One can distinguish between *geometrical* sensitivity and *intrinsic* sensitivity. The geometrical sensitivity is the fraction of emitted photons which interact with the detector and is directly proportional to the sensitive detector area and inversely proportional to the square of the detector-to-source distance. Intrinsic sensitivity, on the other hand, describes the fraction of radiation that impinges in the sensitive area and is efficiently detected, i.e. converted to a readable electrical signal and attributes to the final detector image. In general, higher sensitivities have been reported for devices with scintillator-based detectors compared to semiconductor-based detectors [7].

Spatial resolution The spatial resolution expresses the capability of the device to resolve two sources as individual objects. Even though the spatial resolution is defined as a measure of the smallest distance between two objects that can be resolved, multiple ways of measuring it are in use [8]. For example, measuring the Full Width at Half Maximum (FWHM) of a line spread function, by taking the derivative of an edge response function, or combining measurements with mathematical approximation of the geometrical resolution of the collimator are used to quantify the spatial resolution [7]. The most common way, however, is to measure the FWHM of a point-like source, where its profile is considered to be a Gaussian curve with a mean γ and a variance δ^2. With the correspondence of Eq. 2.1 the FWHM is calculated from a fitted Gaussian curve [36].

$$\mathrm{FWHM} = 2\sqrt{2\ln 2}\delta \approx 2.35\delta. \tag{2.1}$$

That indicates that two point sources with a distance of FWHM have a relative intensity difference between both peaks of 0.5. The collimator is the biggest determinant factor for a good spatial resolution, but it implies a reduction of geometrical sensitivity. Consequently, the choice of collimation is always a trade-off between high spatial resolution and low sensitivity and vice versa.

Energy resolution A high energy resolution enables the device to discriminate between the different energies of detected photons. This allows to separate

gamma photons emitted by the target structure from scattered photons or background activity [37]. Usually, an energy window if $\pm 10\,\%$ around the target energy is applied to avoid large image degradation from scattered or background radiation. The energy resolution is especially important to reject image noise caused by Compton-scattered photons (see Sec. 3.5).

More performance parameters exists, such as contrast, count rate capability, shielding and the device's geometrical design, but are mainly influenced by the parameters above:

Contrast This parameter is used to describe the device's ability to separate high activity enclosed by low activity background. It is directly related to the device's sensitivity, spatial resolution and energy resolution, more specifically by the selected energy window.

Count rate capability It defines the maximum number of counts per seconds that the detector can detect before the material reaches its saturation. The amount of activity can only be derived correctly in the quasi-linear conversion domain of the detector. Both the detection crystal, and the readout electronic can limit the count rate capability [6]. The count rate capability of scintillator-based detectors is usually more limited than that of semiconductor-based detectors.

Shielding The better the shielding, the fewer photons from outside the target area reach the detector and degrade the image quality. While side shielding helps to reduce noise and thus improves contrast and spatial resolution, it can reduce the device's Field of View (FOV).

Geometrical design The distance between the collimator and the area under examination affects all of the parameters above. While being close to the source maximizes the sensitivity, it reduces the spatial resolution. On the contrary, with a larger collimator-to-detector distance, the sensitivity declines, because it also elongates the source-to-detector distance, and at the same time the spatial resolution improves, because the projection of a point-source becomes more of a parallel projection. The size of sensitive area of the detector acts analogously [8].

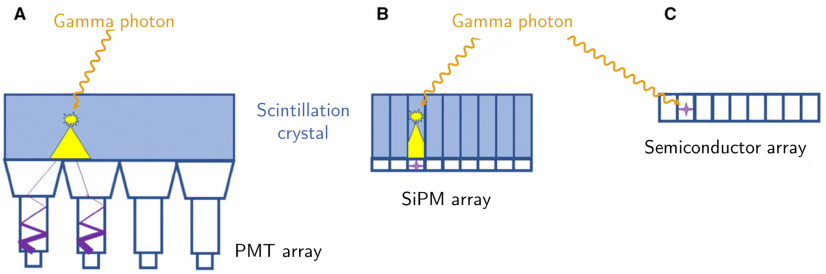


Figure 2.1: There are three basic detector technologies. Scintillator-based with a continuous crystal (A), with an pixelated array of single isolated crystals (B) and semiconductor-based detectors (C). Incoming gamma photons are converted inside the scintillator crystals and are either amplified by photomultiplier tubes (PMT) (A) or directly converted to an electrical signal by an array of silicon photomultipliers (SiPM). In semiconductor-based detectors the conversion takes place directly in the substrate and allows for a larger pixelation but at the costs of a lower conversion rate (C). Figure modified from [38].

2.2 Detector Technology

Three main types of detector technologies for gamma imaging can be distinguished: scintillator-based detectors with continuous or pixelated scintillation crystal and semiconductor-based detectors [8, 39]. An overview is presented in Fig. 2.1.

Continuous scintillator detectors The design of Single Photon Emission Computed Tomography (SPECT) cameras has remained unchanged for decades, consisting of a single large scintillation crystal made of sodium iodide doped with thallium ions (abbreviated by NaI:Tl) with an array of large Photomultiplier Tubes (PMTs) and a parallel-hole collimator [38]. This design is so renowned that it is commonly referred to as *Anger camera*, named after its inventor Hal Oscar Anger in 1957 [12]. The detection unit consists of three components: A scintillator, a photodetector and readout electronics. A scintillator is a slab of material that absorbs the energy of incoming gamma photons and re-emits it as light in response. In order to detect these light photons, they are amplified by PMTs. Behind the scintillator is an array of PMTs and readout electronics to compute photon energy and the location of the photon interaction within the

crystal. The spread of the emitted light photons increases with the scintillator's thickness, degrading the precision of the localization of the gamma ray interaction, resulting in a poorer intrinsic resolution of the device. However, a thick crystal improves the detector's sensitivity, because the probability that a gamma photon passing through the material will interact increases. Since the invention of the Anger camera, the scintillator materials have developed towards materials with higher stopping powers and improved light yields. Nowadays the most popular choices of scintillator material for continuous detectors are sodium-doped cesium iodide CsI:Na and cerium-doped lanthanum bromide ($LaBr_3$:Ce) [7].

The readout electronics have also undergone substantial improvement, with a progress especially in semiconductor-based readout technology. While many devices currently available or under investigation still employ photodetectors based on vacuum technology like PMTs or multi-channel PMTs, semiconductor-based photodetectors such as photodiodes (PINs) or photon multiplies made of silicon (SiPMs) are on the rise [7]. The latter has the advantage of remaining unaffected by magnetic fields allowing the usage together in an Magnetic Resonance Imaging (MRI) device, higher mechanical durability, and an overall more compact design [7, 8]. All in all, the general advantage of using a continuous scintillator is its higher sensitivity, at the costs of a poorer spatial resolution and a heavy and bulky design compared to semiconductor-based detectors.

Pixelated scintillator detector An emerging alternative to a single continuous scintillator crystal is packing a large number of small rectangular scintillation crystals (for example $1.9 \times 1.9 \times 5\,\mathrm{mm}^3$ cubes [19]) that are coated with a reflective material into a larger array to form a pixelated scintillator. Thereby, the re-emitited light is much more focused and can be converted into an electrical signal more precisely than in a continuous design. Furthermore, pixelated scintillators are capable of higher counting rate due to the light pulses emerging in isolated pixels, so only single pixels can reach saturation without effecting others [38]. Since each pixel can be produced separately, materials are chosen that are normally harder to grow uniformly and defect-free but achieve a higher light yield like thallium-doped cesium iodide (CsI:Tl) and cerium-doped gadolinium aluminium gallium garnet (GAGG:Ce) [7]. Even though pixelated detectors can be used with PMTs, the semiconductor-based SiPMs have become

the new photodetector of choice [38]. Concluding, pixelated scintillators are increasing in popularity, making up over half of all analyzed IGCs in a recent review paper [7], due to their improved spatial resolution, less complex readout electronics, and in total, a smaller size.

Semiconductor-based detectors Semiconductor-based detectors for gamma radiation are fairly new [40–42]. The incoming gamma photons are directly absorbed by the semiconductor and generate electron-hole pairs, which are converted in an electrical signal by a bias voltage of up to $-500\,\mathrm{V}$ without the detour of producing light photons [42]. Most current devices utilize a hybrid pixel architecture, where a continuous crystal sensor is bump-bonded to a Complementary Metal-Oxide Semiconductor (CMOS) ASIC readout chip with pixelated anode contacts [7, 8, 43]. The most common crystals are silicon (Si), cadmium-zinc-telluride (CdZnTe, or CZT) or cadmium telluride (CdTe), where the latter has a sensitivity around five times higher compared to Si but is also prone to more crystal defects [44]. Even though, the sensitivity is still below that of scintillator-based detectors, the direct conversion allows for an improved energy resolution. The energy resolution of semiconductor-based detectors detectors have even become so precise, that spectroscopic imaging is achievable: An example of this is the CrystalCam device, which has 4,095 energy channels per pixel, enabling a pixel-specific energy spectrum to be generated from any capture [16].

Similarly to the scintillator devices, a trade-off between sensitivity and energy resolution must be made when deciding for the thickness of the crystal. Furthermore, the thicker the crystal, the higher are the chances of crystal defects and impurities.

Another decision that affects the trade-off between spatial and energy resolution is the choice of pixelation: While the detector in the CrystalCam has an excellent energy resolution, its spatial resolution with its 16×16 pixels is rather poor. On the other hand, a high pixelation of 128×128 or even 256×256 leads to *charge sharing*, where a single gamma interaction will be registered across multiple neighboring pixels [7, 45]. To compensate, a specific software solution was developed: With a multitude of short-exposure frames and a cluster centroid localization method the actual event position is determined. With this aid, a superior spatial resolution of close to 1 mm was achieved [46]. Besides the aforementioned cracks and crystalline defects, another disadvantage is the small

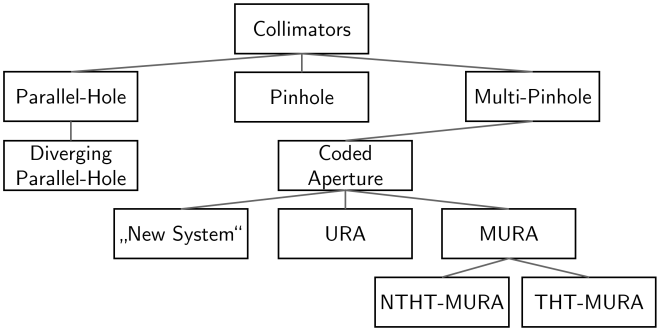


Figure 2.2: An overview of the most common collimation techniques. The majority of IGCs utilize pinhole or parallel-hole collimators where parallel-hole is more popular. Coded aperture collimators can be regarded as an extension to the multi-pinhole technique with several existing mask patterns. The Modified Uniformly Redundant Arrays (MURA) pattern, and especially its No-Two-Holes-Touching (NTHT) version is the most popular pattern.

drift of the detection efficiency with irradiation time [40, 44, 46]. According to a recent review article, semiconductor-based IGC detectors are the rarest deployed detector technology despite achieving the current best spatial and energy resolution [7].

2.3 Collimation Technology

The word *collimator* has its etymological root in the Latin word *collimare*, a misreading of *collineare*, meaning "to direct in a straight line"[1]. The collimator plays the dominant role in the performance of modern IGCs [7]. First, general information about collimators in IGC will be provided. Afterwards, the two most commonly used collimation techniques, together with their mathematical description of the two main performance parameters (spatial resolution and sensitivity) will be presented.

[1] "Collimate". Merriam-Webster.com Dictionary, https://www.merriam-webster.com/dictionary/collimate. Accessed 4 Mar. 2024.

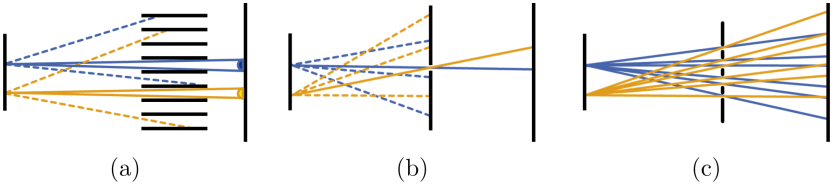


Figure 2.3: The three most common types of collimation: (a) Parallel-hole collimators only allow photons with a close to perpendicular direction to pass. (b) A pinhole collimator blocks photons that do not pass through the single opening. They provide a higher resolution with the costs of a low sensitivity compared to parallel-hole collimation. (c) A coded aperture mask consists of multiple small pinholes, allowing a better trade-off between sensitivity and spatial resolution. However, image reconstruction is required to obtain an interpretable image.

The general purpose of collimators is to restrict the influx of photons in such a way, that together with the information from the position-sensitive detector, the directional information of the influx can be extracted. Just a small fraction of the emitted photons pass through the aperture (typically $< 10^{-2}$), which makes it the main driver of the camera's sensitivity. Here lies the fundamental target conflict of collimation: enlarging the apertures, i.e. the holes, increases the sensitivity but will results in a degradation of the resolution. This phenomenon is referred to as the *sensitivity-resolution trade-off* [12, 47]. An overview of pinhole, parallel-hole and coded aperture collimation is presented in Fig. 2.2.

Refractive or reflective camera systems are practically not feasible in gamma imaging, because gamma photons are highly penetrative. Mirroring systems were and are in use for space telescopes but come with immense manufacturing costs and expensive quality control [48]. Thus, nearly all IGCs are equipped with pinhole or parallel-hole collimators to acquire the spatial information of incident gamma photons [6, 11, 49]. Application specific forms of collimators exist, like diverging parallel-hole collimators for stationary heart perfusion imaging exists, but they only play a niche role because their advantages are small [7, 38]. Collimators are made of high density materials such as lead (Pb), tungsten (W), gold (Ag) or platinum (Pt), to block as many photons as possible except for the photons that are directed through the aperture.

2.3.1 Parallel-Hole Collimation

Parallel-hole collimators, sometimes also called polycapillary optics [23, 50], consist of a thick mask with elongated parallel openings in a honeycomb pattern that resembles a pack of straws (see Fig. 2.3). In contrast to pinhole collimators, parallel-hole collimator block photons not based on where they hit the collimator, but based on their direction. This structure allows only photons to pass that are close to perpendicular to the detector area. The openings are usually of circular, square or hexagonal shape [7, 8]. The ratio of the hole size and the collimator's length is the primary factor for the camera's resolution and sensitivity. The septal thickness, the amount of material between the single straws, must be chosen according to the photon energy of the targeted radiotracer, since a thin septum does not prevent photons from crossing from hole to hole. Parallel-hole collimators are classified into categories like Low-Energy High Resolution (LEHR), Low-Energy High Sensitivity (LEHS), or compromises such as Low-Energy General Purpose (LEGP) or Medium-Energy General Purpose (MEGP) [51]. Parallel-hole collimators are the most popular choice of collimators for IGC, because their sensitivity is around 10 to 100 times higher than for pinhole collimators [7, 8] Because parallel-hole collimators produce a parallel projection on the detector, the FOV is approximately constant for all source-to-collimator distances, which makes it one of the most fundamental differences between the pinhole collimator and the parallel-hole collimator [39].

The main performance parameters for gamma cameras with parallel-hole collimators can be estimated with the following equations. The geometric sensitivity of parallel-hole collimators is given by

$$g_{\text{parallel}} = \frac{1}{4\pi l_{\text{eff}}^2} \frac{A_{\text{aperture}}^2}{A_{\text{unit}}}, \tag{2.2}$$

with the area of a single aperture A_{aperture}, and the area of a single unit cell A_{unit}, combining the areas of the opening and the septal thickness. The effective aperture length l_{eff} is an approximation of the collimator length l corrected for the imperfect attenuation and is calculated as follows: $l_{\text{eff}} = l - 2\mu^{-1}$.

The resolution is estimated by

$$R_{\text{parallel}} = \frac{d\,(l_{\text{eff}} + z)}{l_{\text{eff}}}, \tag{2.3}$$

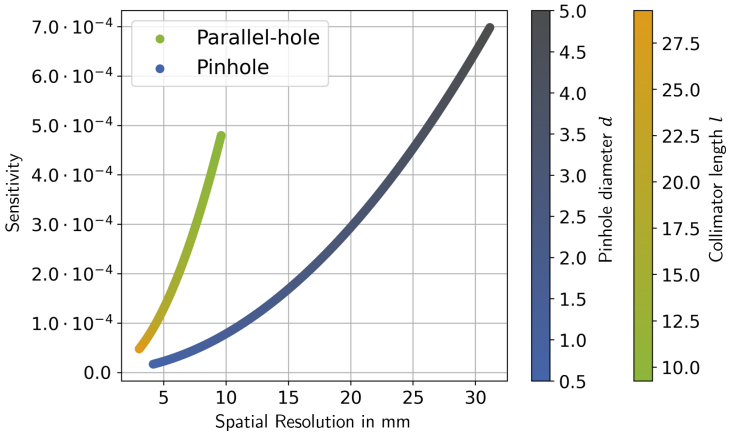


Figure 2.4: A comparison between pinhole and parallel-hole collimation. The optimal IGC has a high sensitivity and good spatial resolution, i.e. small values. This graphs shows the trade-off between the two major performance parameters spatial resolution (x-axis) and sensitivity (y-axis). Pinhole collimator is with varying diameter d, and parallel-hole collimator with varying length l. Further specifications can be found in [16] (parallel-hole) and [46] (pinhole).

where z represents the distance between the source and source-facing side of the collimator. Reported values of recent IGC with parallel-hole collimators of the geometric sensitivity g_{parallel} ranged from $1.0 \cdot 10^{-4}$ to $2.5 \cdot 10^{-3}$ [7]. The geometric resolution R_{parallel} of the same two devices from above was 3.6 mm and 11.7 mm at a distance of 50 mm, demonstrating the inverse proportionality between resolution and sensitivity.

Generally, parallel-hole collimators offer a higher sensitivity than pinhole collimators at the costs of a worse spatial resolution making them the most used collimator type nowadays [7]. The trade-off between those two performance parameters is depicted in Fig. 2.4, where a pinhole collimator with varying diameter d, and a parallel-hole collimator with varying length l and their effect on the geometrical sensitivity and spatial resolution are shown.

2.3.2 Pinhole Collimation

A pinhole collimator consists of a mask that contains only a single small opening in the center, that projects the scene on the detector. Thus, a reflected image of the scene is captured. In order to increase sensitivity for off-center sources the edge of the pinhole opening are tapered to form so called *knife-edges* with an opening angle α. Knife-edges enable a higher sensitivity at the costs of more photon penetration through the edges and, thus, generally cause more scattering noise. The FOV of pinhole collimators is approximated as a cone that is only determined by the mask-to-detector, the source-to-mask distance, and the detector size.

The main performance parameters – sensitivity and spatial resolution – for gamma cameras with pinhole collimators can be estimated with the commonly used formulas based on its geometry and the collimator material [7, 52]. The on-axis geometric sensitivity $g_{pinhole}$ under consideration of penetration through the knife-edges is given by

$$g_{pinhole} = \frac{d^2}{16z^2} + \frac{\tan^2\frac{\alpha}{2}}{8z^2\mu^2} \cdot \left(1 + \mu d \cot\frac{\alpha}{2}\right), \tag{2.4}$$

where d denotes the aperture diameter, z the distance between source and pinhole center, α the acceptance angle of the knife-edge aperture, and μ the linear attenuation coefficient [7]. The geometric resolution of a pinhole collimator can be approximated with

$$R_{pinhole} = \frac{d_{eff}(b+z)}{b}, \tag{2.5}$$

where b represents the distance between pinhole center and detector and d_{eff} the resolution-effective diameter, that factors the photon penetration of the knife-edges into the pinhole diameter and is given by:

$$d_{res} = d - \frac{\ln\frac{1}{2}\tan\frac{\alpha}{2}}{\mu}. \tag{2.6}$$

The calculated parameters of existing IGC, both in commercial and in development stage were analyzed and compared in a recent review study from Farnworth et al [7]. Common sensitivities $g_{pinhole}$ range from $1.10 \cdot 10^{-5}$ to $1.20 \cdot 10^{-4}$ with corresponding resolutions R_{pinhole} of 3.66 mm and 6.27 mm, both for source-to-mask distance of 50 mm [7].

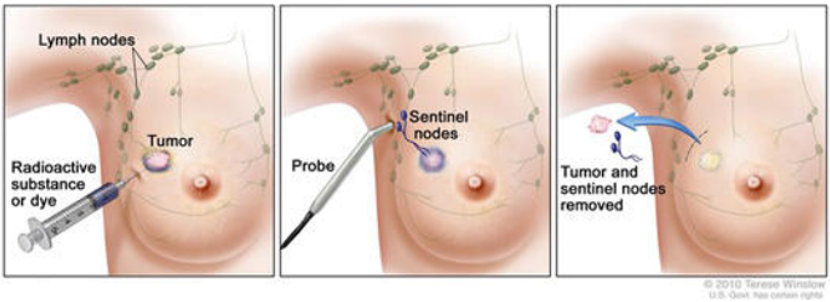


Figure 2.5: SLNB has the goal to find the first few lymph nodes that receive lymphatic drainage from the primary tumor. Therefore, a radiotracer is injected into the tumor, and a gamma probe or IGC is used to find the Sentinel Lymph Nodes (SLNs). Analyzing the nodal state is crucial in staging breast cancer. Figure with permission from [53]. ©(2024) Terese Winslow LLC, U.S. Govt. has certain rights.

2.4 Medical Application

SPECT, together with Positron Emission Tomography (PET), are the most common imaging techniques in nuclear medicine and have a wide variety of applications ranging from brain mapping over diagnoses and monitoring of heart diseases to cancer detection [13]. While SPECT can only provide pre-operative images, RGS utilizes a mobile gamma probe to detect the injected radiotracers inside the patient's body with the goal to precisely locate and possibly remove abnormal or cancerous tissue [12]. While in most clinics, procedure like SLNB relies on preoperative imaging in combination with an intraoperative acoustic gamma probe guiding the surgeons during the procedure, in the last decade IGC underwent considerable advancements both in technology as well as in applications. The most common procedure for the application of IGC is SLNB, which is explained in detail in the following section. Furthermore, the most commonly administered radiotracers are described.

2.4.1 Sentinel Lymph Node Biopsy

SLNB is nowadays the most common application of RGS including patients suffering from breast cancer, melanoma, hand and neck tumors, parathyroid surgery, urogenital malignancies, and bone tumors [14]. While breast cancer has become the most prevalent form of cancer [3], it is also the most common application of SLNB. The procedure of SLNB in the case of female breast cancer is depicted in Fig. 2.5.

The Sentinel Lymph Node (SLN) is any lymph node that receives lymphatic drainage from a primary tumor and is, hence, the first node to receive lymph-borne metastatic cells [19]. In breast cancer, the SLNs are part of the axillary lymph system. Thus, an accurate assessment of the SLN involvement is an essential component in staging breast cancer, as metastases in the SLNs are the most important predictor of overall recurrence and survival [3]. If and how many SLNs have become metastatic is essential for defining prognosis and highly affects treatment planning [5], i.e. the extent of the mastectomy, reconstruction methods, radiation therapy and if neoadjuvant chemotherapy is recommended or not [3]. Dissection of the entire axillary lymph system, known as Axillary Lymph Node Dissection (ALND), has been the standard approach [5], but can cause side effects like uncomfortable postoperative drains, pain, lymphedema [3], shoulder stiffness, arm swelling, and an overall increased length of hospitalization [54]. Over the last two decades, axillary staging in early breast cancer has evolved, becoming less invasive and more conservative, from ALND to SLNB and is nowadays gold standard in staging in early-stage patients with clinically negative axilla [3, 4]. Even though ALND is more effective by reducing the risk of recurrence of axillary lymph node metastases by 1-3 %, the risk of developing lymphedema is 14 % with ALND compared to only 5-7 % after SLNB [54]. This means that deciding on a treatment plan based only on the nodal status of the SLN is less invasive and may spare many women the serious side effects of ALND. SLNs are detected by injecting a radiopharmaceutical into the primary tumor. Sometimes additionally a dye (blue dye or indocyanine green) is administered as well, varying from clinician to clinician [54]. After several hours, the radiotracer accumulates in the SLNs and emits gamma photons that can then be detected by a gamma probe. A study comparing preoperative and intraoperative imaging with an IGC found out that

Table 2.1: Most used radiotracers in nuclear medicine, their half-life, main photon emission and the application their are used for. Reproduced from [56, 57].

Radiotracer	Half-life	Main photon emission [keV]	Application
^{99m}Tc	6 h	140.5 keV	SLNB, myocardial perfusion, thyroid imaging, bone metabolism
^{123}I	13.3 h	27-35 keV, 159 keV	Cerebral blood flow, myocardial imaging
^{111}In	2.8 days	171 keV, 245 keV	Tumor or extravascular imaging
^{131}I	8.04 days	364 keV	Obliteration of thyroid cancer,
^{125}I	59 days	27-32 keV	Brachytherapy, brain imaging
^{241}Am	432.6 years	59.5 keV	Calibration and educational purpose

the most frequently detected number of SLNs a patient was two SLNs with 15 out of 30 patients (50 %). One SLN was detected 7 patients (23.3 %), 3 SLNs were detected in 4 patients (13.3 %), and 5 SLNs were detected in one patient (3.3 %) [55]. Even though these frequencies are based on a relatively small number of participants ($n = 30$), it suggests that the possibility of more than a single SLN must be taken into account.

2.4.2 Radiotracer

Numerous radiotracers have been developed for RGS, varying from radioactive isotopes themselves, small molecules, peptides, antibodies and colloids. The choice of radiotracer depends on several factors, such as its half-life, the radiation type, and the emitted photon energy [14]. While the half-life should be long enough to allow accumulation and subsequently imaging of the targeted tissue, it must not be too long such that radiation dose to the patient and medical staff is minimal. The energy of the emitted gamma photons should be high

enough to avoid scattering and attenuation within the body and thus an internal smearing of the source distribution. If the energy is too high, though, problems described above in detecting the gamma photons both because of collimation and detection occur [57]. A list of common radiotracers and their application is presented in Tab. 2.1.

The most commonly used radioatracer in nuclear medicine is ^{99m}Tc. Over 75 % of all nuclear studies use this radiopharmaceutical [58]. ^{99m}Tc is a metastable nuclear isomer of ^{99}Tc which itself is an isotope of technetium [12]. Before administered to the patient, it is combined to form a colloid with a ligand that is chosen according to the affinity for the targeted organ. Here, the particle size plays an essential role, as particles larger than 500 nm have very limited drainage, while particles smaller than 12 nm rapidly migrate through the lymphatic system. Typically, SLNB is carried out with colloid sizes between 20 nm and 100 nm. ^{99m}Tc is the "workhorse isotope"[58] for gamma imaging because of its near-ideal imaging properties: More than 98 % of the gamma decays result in the emission of photons with an energy of 140.5 keV which is close to the wavelength used in conventional X-ray diagnostics, and the short half-life of approximately 6 h makes it relatively safe to work with [49]. Additionally, ^{99m}Tc has a widespread commercially availability and the generation of ^{99m}Tc is also quite convenient for hospitals and larger medical centers [58]. Depending on the time between injection and imaging procedure, the administered overall activity ranges from 150 to 250 MBq for 25 h before surgery and between 10 to 50 MBq for same-day surgeries [47, 59]. The second most used radiotracers are derived isotopes of iodide: ^{123}I, ^{123}I, and ^{131}I with main emission energies of 159, 20-35 keV, and 364 keV and increasing half-life times of 13.3 h, 59 days, and 8.2 days [60]. A special role has ^{241}Am. It is not used in medical applications, however, its long half-life of more than 400 years and rather low energy of ca. 60 keV make it a relatively safe and suitable candidate for equipment calibration in laboratories.

Chapter 3

Coded Aperture Imaging

This chapter starts by introducing the basic principles of Coded Aperture Imaging (CAI) with Sec. 3.1, explains different mathematical models to describe the imaging process in Sec. 3.3 and presents the most common mask patterns in Sec. 3.2. Finally, the Field of View (FOV) is explained in Sec. 3.4 and the occurring artefacts in near-field applications in Sec. 3.5.

3.1 Basic Principle

CAI was independently proposed by Ables [22] and Dicke [21] for x-ray and gamma-ray astronomy and, thus, only intended for a far-field application. Its concept can be regarded as an extension to a simple pinhole camera: A mask between the object of investigation and the detector consists of a radiopaque material with transparent pinholes that encode the directional information of the incoming gamma flux, as Fig. 3.1 shows. For point sources, this corresponds to casting a shadow of the mask pattern on the detector. General source distributions can be considered as a superposition of point sources: As each transparent element of the mask acts as a pinhole, it generates a projection of the source distributions on the detector. Thus, the resulting image, also known as *shadowgram*, consists of a multitude of overlapping images, making it incomprehensible for the human observer. Image reconstruction is therefore necessary to obtain an interpretable result [61].

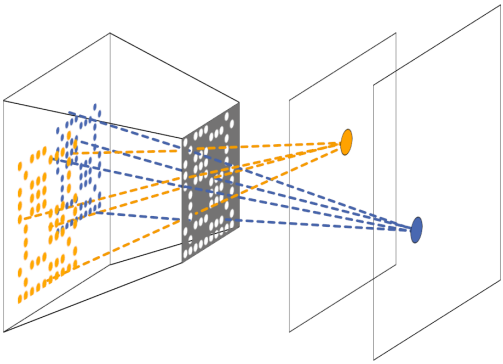


Figure 3.1: The basic principle of CAI: the lateral position of a point source is encoded by the shift of the mask's shadow, while the source-to-mask distance determines the size of the shadow.

Coded apertures can also be considered as an evolution of so-called multi-pinhole collimators (see. Fig. 2.2). While they leverage a similar concept, multi-pinhole collimators generally have much fewer pinholes (usually between 5 and 15) and are specifically designed to prevent overlap on the detector [62–64]. Coded aperture masks can have a total of up to 1,924 pinholes [25]. This large number of pinholes increases the camera's sensitivity, because it lets more photons shine on the detector. A factor of increase between 9 and 11 compared to a parallel-hole collimator based on the exact same design have been reported [23, 24]. In relation to a pinhole collimator, a 20-fold increase was shown while maintaining a comparable spatial resolution [24]. Another advantage of CAI is that opposed to parallel-collimation magnification can be achieved allowing to adjust the area to observe. Among proposed Intraoperative Gamma Camera (IGC), a recent review article found that a device equipped with a coded aperture collimator obtained the best measured spatial resolution of just below 1 mm [7]. It can therefore be stated that CAI combines the high sensitivity of parallel-hole with the high spatial resolution of pinhole collimators.

The following geometric properties that are illustrated in Fig. 3.2, emerge from CAI: The lateral position of a point source is encoded by the shift of the mask's shadow, while the source-to-mask distance is related to the size of the shadow. The lateral position of the cast shadow depends on the lateral

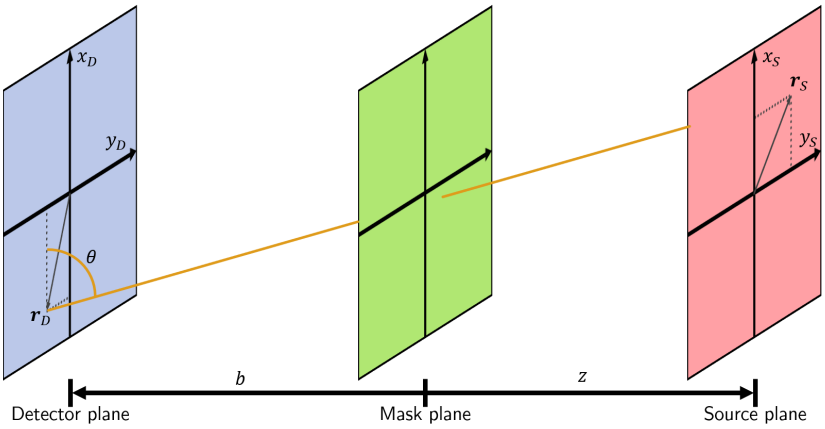


Figure 3.2: The geometry of coded aperture imaging. The detector and the mask have the distance b. The source plane is z distant from the mask. The position of a point in the source plane is denoted by $\mathbf{r}_S$ and the position where the ray hits the detector plane passing through the mask plane by $\mathbf{r}_D$. The resulting incident angle between the gamma ray and the detector plane is denoted by θ. Figure modified from [24].

position in the in-focus plane $[x_S, y_S]$, the detector-to-mask distance b and the mask-to-source distance z:

$$\begin{aligned} x_D &= -\frac{b}{z} x_S \\ y_D &= -\frac{b}{z} y_S \end{aligned} \tag{3.1}$$

The shadow size (D_s) is the size of the mask D_m projected on the detector. Note that others refer with z to the detector-to-source distance [24, 26], but since the detector is encased within the camera body and thus not directly visible by the user, we chose this more intuitive definition. Derived from the theorem of intersecting lines, the magnification factor M and the shadow size can be derived from the ratio of the actual mask size and the projected mask size:

$$\begin{aligned} D_s &= M \cdot D_m \qquad \text{with} \\ M &= \frac{z+b}{z} = 1 + \frac{b}{z}. \end{aligned} \tag{3.2}$$

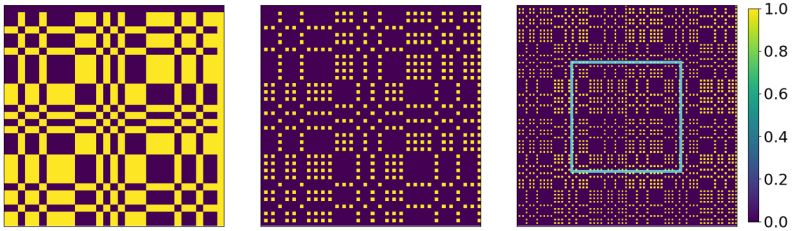


Figure 3.3: From left to right: MURA basic pattern h'_{single} of rank 31, the NTHT version of it with added rows and columns of 0 and the final mask design in a 2×2 arrangement, where the central pattern h'_{central} is indicated by the cyan-colored box.

From Eqs. 3.1 and 3.2 we can find the 3D position of a point-source by estimating the shadows center point, its size and the camera parameters D_{m}, D_{d}, and b.

Outside the field of Radioguided Surgery (RGS), coded apertures are investigated for lensless photo cameras[65, 66], depth estimation from a single photography[67, 68], or in compressed spectral sensing in x-ray imaging [69]. Additionally, there are applications for which special mask types were designed, e.g. a temporal coded aperture for improving x-ray tomography quality [70] or a large curved coded aperture collimator that fits inside a van designed for monitoring large areas like parking lots in homeland security [71]. For further information about the history of CAI, the reader is referred to references [48, 72] and more specifically in the medical context to references [37, 73].

3.2 Mask Pattern

The position of the pinholes in the mask heavily influence the imaging properties. The position can be described by a binary encoding matrix E, which is 0 everywhere except for positions of a pinhole, where its entry is 1. In the very beginning of CAI, pinholes were randomly distributed on the mask, however, problems occurred due to systematic noise emerging from the Fourier transformation during the reconstruction [74]. This problem was solved by the invention of the first structural coded aperture pattern in 1978 called Uniformly Redundant Array (URA) [75]. URAs offer a functionality to generate both E and a match-

ing decoding matrix G, so that their correlation with addition of some noise signal approximates a delta function, which in turn theoretically guarantees an optimal reconstruction [74]. It has been experimentally shown that URAs offer significant improvements in comparison with the previously used pattern of randomly distributed pinholes [74]. However, since the URA pattern is based on pseudo-noise arrays it has the major disadvantage that only rectangular patterns are construable where vertical and horizontal sides dimensions, p and q, must satisfy the condition $p-q=2$, limiting the choice of patterns immensely [75]. More than 10 years later, the Modified Uniformly Redundant Arrays (MURA) pattern was proposed [76] and recognized as an improved version of URA patterns. They enable rectangular, square and hexagonal patterns, where the latter is a configuration preferred in space telescopes [48]. The underlying mathematical concept changed from pseudo-random to the principle of quadratic residues. With MURA patterns available, researchers were now able to built encoding arrays in any side length L, called the rank, that satisfies the condition $L=4m+1$, where m is a natural number greater than 0 and L a prime number, to establish locations of transparent and opaque elements in the coding array E [76]. Equations 3.3 and 3.4 describe the MURA encoding pattern:

$$h'[i,j]=\begin{cases}1 & \text{if } i+j=0,\\ 1 & \text{if } j=0, i\neq 0,\\ 1 & \text{if } C[i]C[j]=1,\\ 0 & \text{otherwise}\end{cases} \qquad \text{with} \tag{3.3}$$

$$C[k]=\begin{cases}1 & \text{if } k \text{ is a quadratic residue modulo } L,\\ -1 & \text{otherwise}\end{cases} \tag{3.4}$$

and the index values i and j referring to rows and columns starting at 0. Here and for the remainder of this work, functions with square brackets, for example $f[x,y]$, will refer to discretized functions, while round brackets, $f(x,y)$, refer to continuous functions. The most commonly used MURA ranks are rank 7 [31, 35, 77], rank 13 [77, 78], rank 19 [31, 79] and rank 31 [25, 36, 80], even though strictly speaking the last one is not a valid rank. While the raw MURA pattern has a ratio of approximately 50 % between opaque and transparent elements, what most research groups use is the No-Two-Holes-Touching (NTHT) version.

It ensures a self-supporting mask, where material between two neighboring pinholes prevents to obtain a mask consisting of multiple unconnected parts. Thus, to generate a NTHT pattern, a column of zeros is inserted between every two columns and a line of zeros is inserted between every two lines of the original pattern [23], which reduces the overall open area of the mask to around 12.5 % [81].

Usually, the NTHT-MURA pattern is repeated to form a 2×2 arrangement of the basic pattern omitting the last row and column. In such cases, h' will refer to the entire mask pattern, h'_{single} to the single MURA pattern, and h'_{central} to the central quarter of h'. Due to the pattern repetition, h'_{central} corresponds to h' shifted both in x and y direction by half the number of rows and columns. An example of all three patterns is depicted in Fig. 3.3. The 2×2 arrangement serves two purposes: First, it simplifies the computation, because for reconstruction only h'_{central} in combination with a cheaper circular convolution can be used (see Sec. 3.3.2). Second, to provide the same FOV with a single basic pattern, a four times larger detector would be required [72]. In this work, we only consider NTHT mask designs where the pinholes have the same diameter as half the center-to-center distance between two pinholes. Mask with larger pinholes can be found [82], but are outside the scope of this work.

After the invention of the MURA patterns and their promising results, the transfer of CAI to the surgery room was investigated in the 1980s. However, interest waned because of the added complexity of reconstruction and the presence of artifacts whenever sources moved closer to the camera [73]. Two developments revived interest in CAI for nuclear medicine in the beginning of the 2000s: first, computing power had significantly improved, and second, the invention of dual image acquisition, also referred to as composite image technique, 90°mask rotation algorithm, or mask/anti-mask imaging [26, 83]. This technique was found to practically eliminate near-field artefacts. The anti-mask is the inverse mask pattern with pinholes, where the mask is opaque and vice versa. Certain MURA patterns inhibit the anti-mask by rotation of 90°. This makes dual acquisition easy to implement, as no additional mask needs to be manufactured or installed [73]. This approach, though, is less practical in an intraoperative setting.

In summary, a single acquisition with a 2×2 arrangement of the NTHT version of a basic MURA pattern is nowadays the most common mask pattern

due to its good balance between design complexity, a high throughput and good theoretical reconstruction capability.

3.3 Mathematical Models

In contrast to parallel or pinhole collimators, a coded aperture adds another layer of complexity to the key performance parameters (sensitivity and spatial resolution) and are generally less well understood [7]. Different mathematical models of CAI exist, where the most common one is the Convolutional Model because of its simplicity. Additionally, there are the Integral Model and Monte Carlo (MC) simulation, which are interesting for both simulation as well as developing and evaluating reconstruction methods.

3.3.1 Integral Model

The Integral Model assumes three co-planar planes that are centrally aligned as Fig 3.2 shows: the detector plane, the mask plane and the source plane. The detector-to-mask distance is denoted as b and the source-to-mask distance as z. In these three planes, $f'(\mathbf{r}_S)$ describes the source distribution and $h'(\mathbf{r})$ represents the binary encoding pattern. Then, the detector image denoted as $p(\mathbf{r}_D)$ is determined by

$$p(\mathbf{r}_D) = \iint_{\mathbf{r}_S} f'(\mathbf{r}_S) h'\left(\frac{z}{z+b}\mathbf{r}_D + \frac{b}{z+b}\mathbf{r}_S\right) \cos^3(\theta)\, C(\theta)\, \mathrm{d}^2\mathbf{r}_S \tag{3.5}$$

with the incident angle between the gamma ray and the detector plane θ as

$$\theta = \arctan\left(\frac{||\mathbf{r}_S - \mathbf{r}_D||_2}{z+b}\right), \tag{3.6}$$

where $||\bullet||_2$ denotes the Euclidean norm. The incident angle characterizes the trajectories from source to the detector and is the origin of the difference between the near-field and the far-field case [26]. The term $\cos^3(\theta)$ represents the projection of a spherical photon flux on a planar detector, which causes the outer areas of the detector to be less illuminated than the center. The

collimation factor $C(\theta)$ describes the blocking of photons by the inner sidewall of the mask's pinholes. Only approximations under certain assumptions of $C(\theta)$ exists and are described below in Sec. 3.5. This mathematical model of CAI as a continuous integral is used to describe near-field effects and to develop compensation methods [24, 26]. However, almost all researchers work with the simpler and discretized Convolutional Model.

3.3.2 Convolutional Model

The Convolutional Model can directly be derived from Eq. 3.5. When assuming that θ remains small and constant and the mask to be infinitely thin and thus no collimation occurs, i.e. $C(\theta) \equiv 1$ and that the photon flux is homogeneous, i.e. $\cos^3(\theta) \equiv 1$, Eq. 3.5 becomes

$$p(\mathbf{r}_D) = \iint_{\mathbf{r}_S} f'(\mathbf{r}_S)\, h'\left(\frac{z}{z+b}\mathbf{r}_D + \frac{b}{z+b}\mathbf{r}_S\right) \mathrm{d}^2\mathbf{r}_S, \tag{3.7}$$

Further, with

$$\xi = -\frac{b}{z}\mathbf{r}_S, \tag{3.8}$$

$$f(\mathbf{r}) = f'\left(-\frac{z}{b}\mathbf{r}\right), \text{ and} \tag{3.9}$$

$$h(\mathbf{r}) = h'\left(\frac{z}{z+b}\mathbf{r}\right), \tag{3.10}$$

Eq. 3.7 becomes

$$p(\mathbf{r}_D) = \iint_{\xi} f(\xi)\, h(r_D - \xi)\, \mathrm{d}^2\xi \tag{3.11}$$

$$= f(\mathbf{r}_D) * h(\mathbf{r}_D), \tag{3.12}$$

where "$*$" denotes the 2D linear convolution operator. Here, $f(\mathbf{r}_D)$ is the scaled and reflected form of the source image $f'(\mathbf{r})$, and $h(\mathbf{r}_D)$ is the scaled version of the mask pattern $h'(\mathbf{r})$, which will be denotes as Point Spread Function (PSF). The scaling factor for $f(\mathbf{r}_D)$ is $\frac{z}{b}$ and corresponds to the magnification of a pinhole camera with the same distances and the minus sign indicates a reflection.

The scaling factor for the mask pattern h is the ratio of the mask size to the size of its projection on the detector [26]. Its inverse is the magnification factor M from Eq. 3.2 depending only on b and z.
Equation 3.11 corresponds to the continuous version of the Convolutional Model from Eq. 3.13. When converting the Convolutional Model in its discrete form, two aspects are of importance:

1. The discretization of f and h must be the same as the pixel size of the detector, but the dimension of the PSF can be larger or smaller, depending on the geometry and the distances between the planes.
2. The pixel size of the source image must correspond to the size of the FOV divided by the number of pixels of the detector. For the discretized source image $f[x,y]$ we will automatically assume a correct discretization.

Thus, as a linear approximation of CAI, with the 2D convolution of the discretized source distribution $f[x,y]$, that covers the pinhole FOV and the scaled version of the PSF $h[x,y]$ of the mask, we yield the detector image $p[x,y]$:

$$p[x,y] = (f * h)[x,y] = f[x,y] * h[x,y] = \sum_y \sum_x f[x,y]\, h[i-x, j-y]. \qquad (3.13)$$

For the sake of simplicity the coordinates $[x,y]$ are omitted where not required. Intuitively, the Convolutional Model can be motivated as well: Because each pinhole of the masks projects an image of the object onto the detector, the entire detector image becomes a superposition of all projections shifted by the position of each pinhole encoded by the encoding matrix.

With finite signals, which is the case for CAI, two types of convolutions must be distinguished: The *linear* convolution, which will be denoted by the "$*$"-symbol, and the *circular* convolution denoted by the "$\circledast$"-symbol. The difference is the underlying assumption about the signals beyond the image borders. While the linear convolution assumes that all entries outside the image are 0, in the circular convolution the signals are assumed to be repeating themselves periodically infinitely many times. When two signals are convolved via a multiplication in the Fourier domain and no explicit zero-padding is carried out, the circular convolution is performed. Both convolutions are used in simulation and reconstruction of CAI and serve different purposes: Generally, a

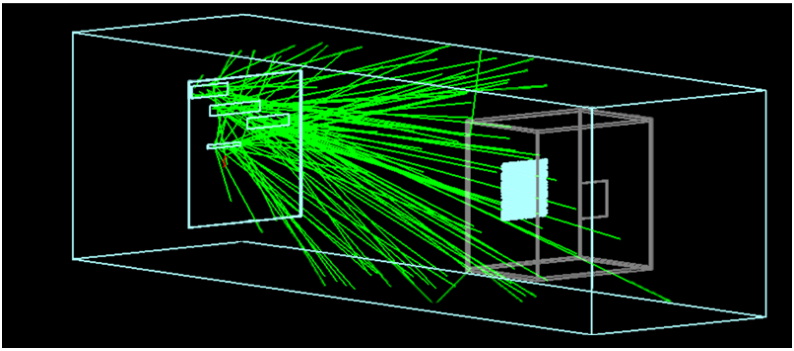


Figure 3.4: A visual representation of the MC simulation toolkit TOPAS MC. The dark grey box in the bottom right corner represents the IGC with the coded aperture mask in cyan and the detector in gray. The green lines represent the traces of the gamma photons emitted from the three horizontal rectangles that act as sources.

linear convolution of the PSF $h[x,y]$ (which might be bigger than the detector image itself) with the source image $f[x,y]$ gives the detector image. Instead, a faster circular convolution can be carried out, when a 2×2 arrangement of a single mask pattern is utilized. Then, and additionally, when the PSF is not smaller than a quarter of the detector image, instead of the entire scaled mask pattern only h_{central} can be used to calculate the detector image.

3.3.3 Monte Carlo Simulation

The fundamental idea behind Monte Carlo (MC) simulations is the tracing of single gamma photons, based on Monte Carlo methods [84]. This means, the passage of single photons through matter are simulated including the major physical interactions like Compton scattering and photon absorption, based on the photon energy and physical properties of the environment. MC simulations can be considered the gold standard in CAI simulation, since all the major physics aspects of the imaging system are considered, including all near-field effects mentioned below [74, 85, 86]. For CAI, MC simulation consist of three components. A source, the coded aperture mask and a detector. The source can be either flat or three-dimensional and build up from simple geometrical shapes like rectangles, or spheres, but most software libraries allow the user to include

volumetric sources in the form of “stl” or “ply” files as well. Geant4 [87] is the most popular software package. Figure 3.4 shows a visualization of TOPAS MC [88], a wrapper library around Geant4, where three rectangular sources are imaged by a coded aperture camera. The detector can either be modelled with the exact scintillator material and the energy deposition of impinging gamma photons, or with a simpler approach using a so-called phase space [43, 85, 89]. The phase space is an imaginary plane that collects all particles that pass through it and stores the position, photon energy, the incident angle, particle ID etc.. These files are then post-processed, e.g. discarding hits from particles with an energy outside the energy thresholds or smeared by the intrinsic resolution from the photon cascade in scintillator-based detectors.

MC simulation are widely used to simulate IGCs for predicting and evaluating the properties of collimator, shielding and detector design [7, 90–93]. However, due to simulating millions of photons, a computation time of multiple minutes up to 7 h for the simulation of one detector image must be expected [89].

3.4 Field of View

In contrast to pinhole or parallel hole collimation, in CAI, the FOV depends to the reconstruction methods. When using a 2×2 arrangement of a basic MURA pattern, only the central detector portion is used that corresponds to the size of a projection of one pattern, which ultimately defines the FOV. When neglecting the mask thickness, the FOV is defined as

$$\mathrm{FOV}(z) = \frac{z+b}{b}\frac{D_\mathrm{m}}{2}, \tag{3.14}$$

where D_m represents the side length of the entire 2×2 mask. There is a limitation in depth direction as well: In order to reconstruct an image, at least one entire single pattern must be projected on the detector, from which the minimal source-to-mask distance can be deducted:

$$z_\mathrm{min} = b\frac{D_\mathrm{m}}{4D_\mathrm{d} - D_\mathrm{m}}, \tag{3.15}$$

where D_d denotes to the size length of the detector’s active area. The FOV of a pinhole collimator is defined as follows:

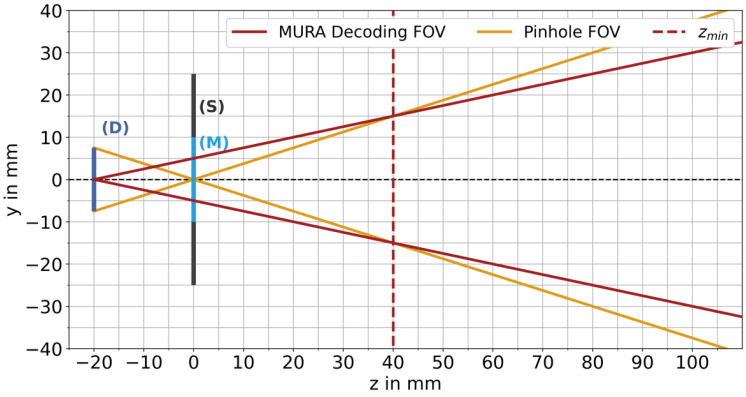


Figure 3.5: Different FOVs in CAI: A point source closer than z_{min} projects less than one MURA pattern on the detector, which inhibits image reconstruction. For this plot, the following geometrical parameters were used: $b = 20\,\text{mm}$, $h_{\text{d}} = 15\,\text{mm}$, $D_{\text{m}} = 20\,\text{mm}$ and a mask thickness of 1 mm. M, S, and D refer to the mask, shielding and detector.

$$\text{FOV}(z) = z\frac{D_{\text{d}}}{b} \tag{3.16}$$

However, because the FOV plays a minor role in this work, the reader is referred to [73] and its supplementary material, which can be found online for a more thorough description[1].

3.5 Near-Field Effects

As already mentioned in Sec. 3.3, in the case of near-field imaging both incident angle θ [81] as well as the mask thickness [24, 94] cause effects that impede image reconstruction. This section presents the most influential near-field effects, also called close-up effects [95], and their mathematical models with a focus on improving CAI simulations based on the Convolutional Model as is shown in Fig. 3.6.

[1] https://static-content.springer.com/esm/art%3A10.1007%2Fs11307-011-0494-2/MediaObjects/11307_2011_494_MOESM1_ESM.pdf

3.5.1 Planar Wave Effect

In near-field applications, the incoming photon flux can no longer be regarded as parallel rays, which means that the incident angle θ between impinging photons and the detector plane differs from 90°. Instead, a spherical wave hits the flat detector plane leading to steeper angles and thus a smaller flux density towards the edge of the detector compared to its center [26]. This effect is included in the integral form (Eq. 3.5) by the term $\cos^3(\theta)$, but remains unconsidered in the Convolutional Model (Eq. 3.13). The planar wave effect will increase the closer a source is to the camera. Accorsi et al. suggested to assume that all radiation stems from a single point source in the center, because that allows to make θ a function only of the detector coordinates and a correction term C_{planar} can be derived [26]:

$$\theta(\mathbf{r}_i = 0, \mathbf{r}_D) = \theta(\mathbf{r}_D) = \arctan\left(\frac{||\mathbf{r}_D||_2}{z+b}\right) \tag{3.17}$$

$$C_{\text{planar}}(\theta) = \cos^3(\theta). \tag{3.18}$$

Hence, the detector image can be corrected for the planar wave effect by dividing the simulated image by C_{planar}.

3.5.2 Collimation Effect

An ideal mask is infinitely thin and allows photons of all angles to pass through the pinholes. However, in reality the mask requires a certain thickness to prevent photons from reaching the detector unimpeded and to allow only photons that passed through a pinhole. The collimation effect arises from photons that are blocked by hitting the inner sidewall of the pinholes. Thus, the collimation effect is mainly influenced by the mask's thickness, which limits the pinhole's acceptance angle. The collimation effect increases together with the angle of incidence, e.g., when the source is off-center [96]. Analog to the planar wave effect (Sec. 3.5.1), Mu et al. [24] simplified the calculations by assuming a point source at the center of the plane in focus. Together with Eq. 3.6 the following equations approximate the collimation effect:

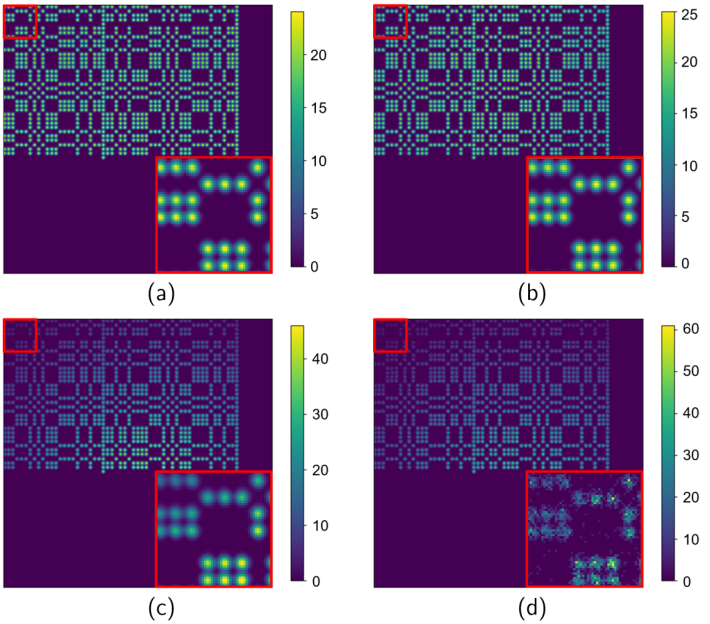


Figure 3.6: The visual influence of near-field effects and Poisson noise. (a) detector image of a point source in the bottom right corner of the FOV simulated with the Convolutional Model. (b): with 1 % transmission and the planar wave effect, (c): additionally with the collimation effect ($t = 0.11\,\text{mm}$), (d): the final simulation result after Poisson randomization, where a total of $500{,}000$ photons were set to hit the detector. Figure is taken from [96].

$$d = \min\left(t\tan\left(\theta\right), 2r\right) \tag{3.19}$$

$$\psi = \cos^{-1}\left(\frac{d}{2r}\right) \tag{3.20}$$

$$C_{\text{Coll}}\left(\theta\right) = \frac{2\psi r - d\sin\left(\psi\right)}{\pi r}, \tag{3.21}$$

with r denotes the pinholes' radius, t the mask's thickness, ψ the angle of displacement, and d the displacement of the projection of the aperture's front opening. For a more in-depth explanation the reader is referred to [24].

3.5.3 Other Effects

Additionally to the near-field effects described above, transmission, Poisson noise, scattering, and septal penetration are further contributors that are not limited to CAI, but generally deprive the image quality in gamma imaging.

Transmission High-energy photons like gamma photons always have a certain probability to pass through matter, and this probability T is proportional to the negative exponential function of the material's mass attenuation coefficient μ at the photon energy of interest, its density ρ, and the mask's thickness t:

$$T = e^{-\mu\rho t}. \tag{3.22}$$

For example, in order to block 99 % of all photons with 140 keV with a mask made of Tungsten ($\rho = 19.3\,\mathrm{g/cm^3}$, $\mu = 1.76\,\mathrm{cm^2/g}$ at 140 keV) a thickness of around 1.36 mm is necessary. This effect is independent from the object-to-detector distance. For a more in depth discussion of mask transmission the reader is referred to [47]. The detector image that includes transmission (p_T) can be obtained by weighting and adding to the detector image from both the Integral Model (Eq. 3.5) as well as from the Convolutional Model (Eq. 3.13) $p(\mathbf{r}_D)$.

$$p_T(\mathbf{r}_D) = (1-T)p(\mathbf{r}_D) + T \iint_{\mathbf{r}_S} f(\mathbf{r}_S)\,\mathrm{d}^2\mathbf{r}_S. \tag{3.23}$$

Poisson Noise Both the Integral and the Convolutional Model give deterministic results of the detector image. However, because ionizing radiation is harmful to the patient and medical staff, the guiding principle in RGS is to reduce the exposure to as low as reasonable achievable (the ALARA principle [77]). Thus, the photon flux is generally low and the number of detected photons per pixel must be regarded as a *Poisson process*, leading to Poisson noise which is sometimes also called *shot noise*. A Poisson process is a random process following a discrete probability distribution that expresses the probability for the number of detected photons in a pixel, if theses events can be considered to occur with a known constant mean rate λ and independent from the time of the last event [97]. Under a Poisson distribution with the expectation

of λ detected photons per pixel, the probability that k photons hit the pixel (here denoted by the random variable X) is given by

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}. \tag{3.24}$$

Only a discrete number of photons can be detected, and thus, k is always a non-negative natural number $\mathbb{N}_0$. For small λ, the probability distribution is heavily skewed towards 0, but becomes more and more symmetrical and converges toward a Normal distribution for $\lambda \gtrsim 30$ [97].

Scattering Gamma photons interact with the mask or the shielding, causing scattering, when they are not fully absorbed. Compton scattering plays the primary contributor in gamma imaging, where the collision of a gamma photon with an atom causes the reflection of a gamma photon with a lower energy, depending on the scattering angle [63]. When these lower energy photon hits are registered by the detector they cause additional image noise. Scattering noise can be reduced when the detector's energy resolution is high enough to separate scattered photons from the unscattered photons with a known photon energy [8].

Septal penetration Some gamma photons are not fully blocked by the inner sidewall of the pinholes, but pass through the matter, because it is thinner than the mask's general thickness. It has been suggested to empirically add a constant term κ to the collimation function $C_{\text{Coll}}(\theta)$ from Eq. 3.21 to account for septal penetration [24].

Chapter 4

Machine Learning

During the last decade, Machine Learning (ML) has gained a tremendous popularity in the computer vision community for tasks like image segmentation [98, 99], image denoising [100], or image reconstruction [9, 10, 101, 102]. In image reconstruction tasks, ML refers to the development and deployment of models that "learn" to map between image domains, generalize to unseen data and are, thus, able to perform tasks without any explicit instructions [103]. Such ML models are inspired by the neural networks of the nervous system and consist of hierarchical layers of mathematical operations with trainable parameters that transform an input into an output image [104]. This transformation is not derived from a mathematical description of the mapping. Instead, an approximate transformation from the input domain into the output domain is deduced by providing a large amount of training images and their corresponding Ground Truth (GT) data. For a more general insight in ML, there is excellent literature available [104, 105]. This chapter presents the typical training procedure (Sec. 4.1), gives an overview of the basic components of modern ML models (Sec. 4.2), and presents the Convolutional Encoder-Decoder architecture (Sec. 4.3).

4.1 Training Procedure

The goal of the training procedure is to obtain the optimal parameters of a given network architecture, that has generalized from the training data to unseen data

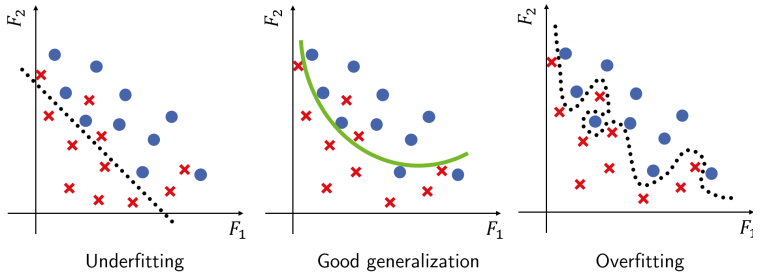


Figure 4.1: Overfitting, underfitting and the desired generalization in ML at the example of a simple 2D classification network.

and performs exactly the intended task. Because training datasets are usually so large that they do not fit into the memory in its entirety, the training procedure is performed iteratively on small portions of the training dataset, the so-called *mini-batches* or short *batches*. When all batches of a training set have been used once and the entire dataset has been used for updating the parameters, the network has been trained for one *epoch*. Usually, networks are trained for multiple dozens epochs [102]. Several other problem formulations, such as unsupervised or reinforcement learning, exist, but supervised learning is the most relevant training paradigm for this thesis: Supervised learning for image-to-image tasks, is a practice that can only be applied to problems, where the training set contains both input as well as the indented output images (often denoted as Ground Truth (GT)). One training step consists of three stages: First, a *batch*, usually consisting of 1 to 64 images, is propagated through the network to obtain its output, often referred to as *prediction*. Second, the prediction is compared to the GT and the difference is considered as the current error. Third, the error is propagated backwards through the network's layer and an optimizer adapts each layer's weights and biases according to their contribution to the error [105]. The following paragraphs go into detail of these essential steps:

Loss function The loss function is typically a mathematical metric[1], that quantifies the network's prediction error, i.e. how close a network's predictions

[1]Not all but most loss functions fulfill all mathematical properties of a metric, which are non-negativity, identity of indiscernibles, symmetry, and the triangle inequality [106]

are to the GT. By averaging the observed error over the entire dataset, a scalar value is obtained that represents the discrepancy between the network performance and the ideal outcome. It is assumed that if the loss function is minimal, the network has learned its intended task. In this manner, the loss function poses as a bridge between ML and classical optimization [105]. For image prediction the L_2 is commonly used, even though there are indications that training with the L_1 norm can deliver more "natural" looking images [107].

Backpropagation The backpropagation algorithm, which is based on the chain rule for differentiation, was invented in 1986 and poses an important milestone in ML history, because it allows training networks with more than one layer [104]. The algorithm propagates the network's error backwards through the layers of the network (hence the name) and distributes the "blame" of each trainable weight to the error according to its and the succeeding layers' derivatives. Thereby, the gradients of the loss function with respect to the weights are computed and combined by the chain rule for partial derivatives. By converting the network architecture into a computational graph and applying *automatic differentiation*, these partial derivatives can be computed efficiently, which was another main contributor of modern ML. To minimize the loss function the weights are adjusted by an optimizer.

Optimizer The optimizer determines how the weight updates are applied until the error converges to a minimum. There are simple approaches like the *stochastic gradient descent* (SGD), which is called stochastic, since updating happens iteratively on batches that only allow an approximation of the true gradient. The objective is to repeatedly take steps in the opposite direction to the approximated gradient of the loss function at the current position, as this is the direction of steepest descent leading to the minima. The step size is known as the *learning rate* in ML and is usually a small number in the range of 10^{-1} to 10^{-9}. More recently another optimizer has mostly replaced SGD, due to its strong empirical performance: the Adam optimizer [108]. It derives the optimal learning rate for each training parameter from estimates of first and second moments of the gradients, where the upper bound is fixed by the learning rate. Although Adam requires tracking not only each gradient but also each individual learning rate, making it computationally more expensive, its advantages often outweigh the costs.

Validation strategy ML engineers are typically more interested in how well a model performs on previously unseen data rather than achieving a good metric on the training dataset, since the GT for the training data is already known. Instead, the ability to perform well on previously unobserved input data is of major importance and is called *generalization* [104]. To quantify generalization, the loss function or other metrics like accuracy for classification or the Structural Similarity Index Measure (SSIM) for image regression are computed for both the training dataset as well as for a separate validation set. When a separate portion of the available data is used for validation it is called a *held-out validation strategy* [105]. Another strategy, especially used when the total amount of available data is limited, is *cross-validation*. In cross-validation, the whole dataset is split into N-fold subsets where in all N variations the network is trained on all sets but one and the overall validation metric is composed of the N subset validation values [109].

The natural antagonist of generalization is *overfitting*. Overfitting means that the network was unable to derive abstract features to guide its prediction. Instead, it has learned excessive and irrelevant details of the training set, and is thus unable to perform well on previously unseen data. In contrast, *underfitting* happens when the network fails to capture a sufficient level of structure in the data. In the course of the training, overfitting occurs as soon as the validation error increases while the training error still decreases. The concept of underfitting, a well generalization, and overfitting on the example of a classification network is presented in Fig. 4.1. For more information about unsupervised learning, automatic differentiation, the backpropagation algorithm and its historical development, the interest reader is referred to Ref. [104] and [105].

4.2 Basic Components

This section gives an overview of the basic components and layers that most neural network architectures are composed of.

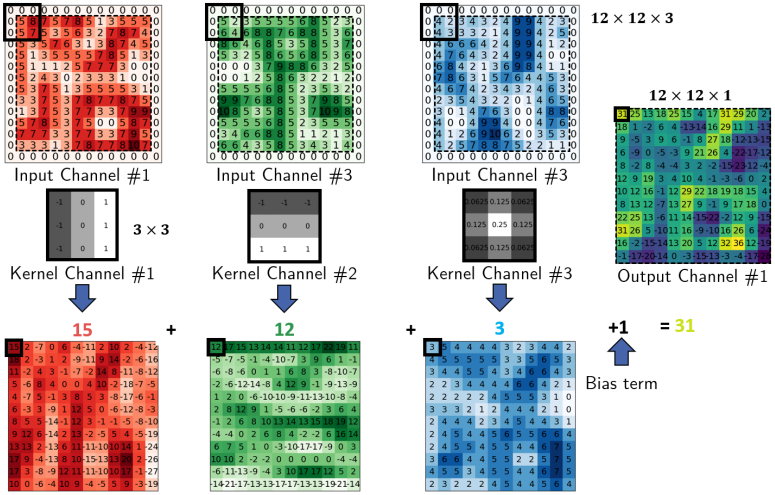


Figure 4.2: This graphic shows the convolution of an image with three channels ($H \times W \times 3$) with a set of 3×3 kernels. The convolutional layer produces an output image by shifting a kernel over the input image, multiplying element-wise and summing the intensities, and aggregating the results per input channel. The input image is zero-padded to obtain an equal size output image. A convolutional layer performs this procedure multiple times to produce a feature maps with C_{out} output channels. Figure modified from [110].

4.2.1 Convolution Layer

Convolution is an efficient way of applying the same linear transformation of a small local region, like edge detection or smoothing, to an entire input image. In ML terms, the first argument (function I in Eq. 4.1) to the convolution is often referred to as the input and the second argument (K) as the kernel. The output is usually referred to as a *feature map* or *activation map* (F). A convolutional layer performs multiple convolutions simultaneously, depending on the number of input channels and the desired output channels: From an input image I with the dimensions $H \times W \times C_{\text{in}}$ and a set of k trainable filter kernels, each of dimension $H_F \times W_F \times$, a feature map F is produced. For each input channel C_{in} and output channel C_{out}, there is a separate filter that slides across the input image, where for each position the kernel is multiplied element-wise with the overlapping image and summed up to generate the output pixel value (see Fig 4.2). Usually,

zero-padding is carried out prior to the convolutions in order to maintain image size. An additional trainable offset b_k, called *bias term*, is added to elevate the output values of an entire output channel which helps in exploiting the succeeding activation function. Each individual convolution can be described by:

$$F[i,j,k] = (I * K_k)[i,j] = \sum_{c=0}^{C_{\text{in}}-1} \sum_{m=0}^{H_F-1} \sum_{n=0}^{W_F-1} I[m,n,c]\, K_k[i-m, j-n, c] + b_k \quad (4.1)$$

Technically, since the filter is not flipped, a convolutional layer does not perform a convolution, but a correlation. However, this name is generally accepted [104]. Besides the number of filters, which determines C_{out}, most ML software packages allow the setting of further hyperparameters to adapt a convolutional layer to its intended purpose. First, the *kernel size*, which is typically rather small compared to the size of the input image and always uneven, most commonly 3×3 or 5×5 kernels [10, 98]. Second, the *stride* configures the step size at which the kernel shifts across the input and helps to reduce the spatial dimension of the output if a stride ≥ 2 is selected. Third, filter kernels can be *dilated*. That means, untrainable rows and columns of zeros are inserted between the elements to increase overlap between filter and input image. This increases the field of perception of a single filter but keeps a small number of trainable parameters. The convolutional layer allows to extract spatially relevant information with only relatively few trainable parameters, because an output pixel is only connected to a small portion of the input image or preceding feature map. From analysing the activations of a trained Convolutional Neural Network (CNN), it has been observed that the first few convolutional layers often act as smoothing and edge filters [111].

4.2.2 Activation Layer

Activation functions are usually a scalar-to-scalar function, which propagate the output of one layer to the next one. They are essential to neural networks, because they are the component that adds non-linearity to the networks capability [105]. Without activation functions, two consecutive convolutional layers could be shorten to one, because both perform linear operations. Figure 4.3

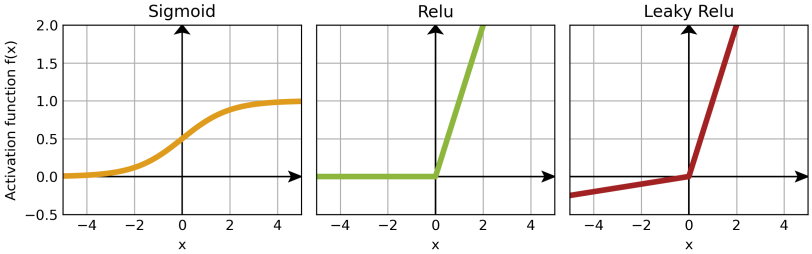


Figure 4.3: Activation functions add non-linearity to a ML network. x represents the input and $f(x)$ the output of the three presented activation functions *Sigmoid*, *Relu*, and *Leaky Relu*. Note the small negative slope of $p = 0.01$ in the negative half-plane with the Leaky Relu. When p is a trainable parameter it is called *Parametric Relu*.

shows the three most common activation functions which will be explained in the following paragraphs.

Sigmoid function It transforms variables from an infinite range into simple probability values between 0 and 1, in a smooth monotonically increasing matter. It has the advantage of having a smooth derivative that can be expressed with the activation itself, making it easy to compute and, in turn, allowing for fast training [104]. But because the gradient becomes very small towards the negative and positive limits, the simgoid activation function has lost popularity in contemporary image-to-image network architectures [105].

Rectified linear unit The rectified linear unit (often abbreviated as *Relu*) is a discontinuous activation function which returns zero until the input rises above a threshold, from where it has a linear relationship:

$$f(x) = \max(0, x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{otherwise} \end{cases} \tag{4.2}$$

Despite being discontinuous, Relu can be considered the state of the art activation function, as its fast and simple computation has been proven to work well in many different areas of ML [105].

Leaky Relu Leaky Relu activation function was developed to mitigate the difficulty during training called *dying Relu issue*. Some units might never be

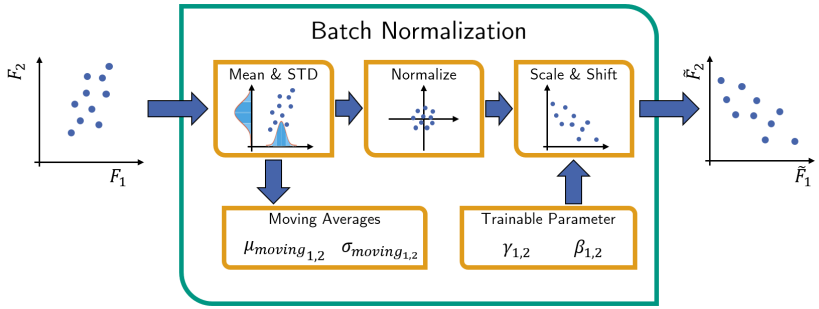


Figure 4.4: The Batch Normalization Layer explained on a batch of 16 samples with two features F_1 and F_2. First, the mean and standard deviation (STD) are calculated and combined with all previous batches to the moving averages for the mean $\mu_{moving_{1,2}}$ and standard deviation $\sigma_{moving_{1,2}}$. In a second step, the features are normalized and then scaled and shifted in a third step by the trainable parameters $\gamma_{1,2}$ and $\beta_{1,2}$. The result is a batch with an optimized location and extension of the new features $\tilde{F}_1$ and $\tilde{F}_2$.

activated during the entire training set, however, these units will never contribute to the network's output, posing a waste of computational power and effective network capacity. Instead of an all-zero output for negative input values, the negative half-plane of the Leaky Relu has a small negative slope (for example $p = 0.01$) and is defined as follows:

$$f(x) = \begin{cases} px & \text{if } x < 0 \\ x & \text{otherwise} \end{cases} \tag{4.3}$$

When p is a trainable parameter adjusted by the optimizer during training, this activation function is denoted *Parametric Relu* or simply as *Prelu* [104].

4.2.3 Batch Normalization Layer

Calculating the updates for all parameters happens according to the backpropagation algorithm from the output error backwards to the input layer. Applying the updates to the parameters of a single layer is based on the assumption that all other layers remain constant. However, since powerful networks consist of many layers in sequence, updating a previous layer can significantly change how a later layer influences the final output. Thus, the update might be sup-optimal. To

tackle this problem, Ioffe & Szegedy [112] proposed the Batch Normalization Layer in 2015, which was soon recognized as a revolutionary tool that stabilizes the training and allows faster training and deeper networks.

Batch Normalization is usually applied between each convolutional and activation layer. It shifts and scales the input data to a trainable mean with a trainable variance. It does this by multiplying the normalized values by a factor, γ, and adding it to the factor β, which is illustrated by Fig. 4.4. In image-to-image networks, each channel of the feature maps is normalized, shifted and scaled independently of each other. What differentiates this layer from many other layers, is the fact that it has sets of parameters: the trainable parameters γ and β and the non-trainable moving means of the average and variance μ_{moving} and σ_{moving}. They are updated in each training set based on the entire batch and they act as a proxy of the mean and variance of the entire training dataset [104]. Hence, it is capable of capturing more features about the dataset than other normalization techniques. Even though the authors initially explained the strong performance of Batch Normalization layers with reducing the internal covariate shift, more recent research attributes their improved performance to the smoothing of the objective function [113].

4.3 Convolutional Encoder Decoder Networks

The most popular architecture in image-to-image networks are CNNs, where the main components are convolutional layers. A CED is a subtype of CNNs, with the origin in the famous U-Net architecture proposed by Ronneberger et al. [98] in 2015. A CED uses convolutional, normalization, and activation layers, to successively transform an input image via a compressed low-dimensional representation into an output image. Figure 4.5 shows an exemplary CED that transforms an image of size $I \times I \times K$ over the latent space representation ($I/16 \times I/16 \times 1024$) into an output image of $I \times I \times K$. The input image is compressed into a *latent space representation* by strided convolutions or pooling layers [104], which represents the final feature map before the decoding stage. Thereafter, it is progressively up-sampled to generate the output image. The bottleneck convolutions effectively regularize the ill-posed inverse mapping

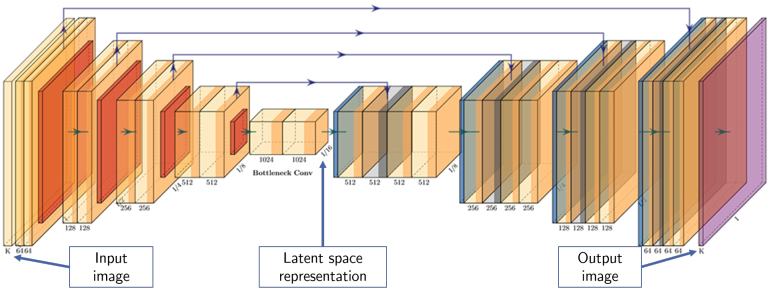


Figure 4.5: The U-Net architecture: an input image of size $I \times I \times K$ is successively spatially down-sampled until it has reached a 16^{th} of its original size. While doing so, the number of channels per feature map increases to $1{,}024$. The last feature map before the decoding is called *latent space representation*. The decoder part extends the image and simultaneously reduces its channels until the output dimensions are reached. The blue arrows represent a copy and concatenation of feature maps from encoder to the decoder part.

by acting as an implicit denoising algorithm [100]. Thus, CNNs and CEDs in particular, have been deployed successfully on direct reconstruction of Positron Emission Tomography (PET) scans from sonograms [10], Radon inversion [101], metal artifact reduction in computer tomography [114] or image deblurring [115], to mention a few examples.

PART II

DATASETS

Chapter 5

Experimental & Synthetic Datasets

Developing and evaluating reconstruction methods for Coded Aperture Imaging (CAI) requires realistic gamma camera images, as simulations do not provide the same level of realism as actual measurements. Therefore, in collaboration with other research groups three experimental datasets were captured and shared with the research community. The Intraoperative Gamma Cameras (IGCs) used, along with the details of the measurements are explained in Sec. 5.1, Sec. 5.2, and Sec. 5.3. For training a Machine Learning (ML) approach, however, a large quantity of training images is necessary. Two image domains serving as synthetic source images were investigated and are presented in Sec. 5.4, and 5.5.

5.1 Rozhkov Dataset

The purpose of the Rozhkov Dataset was to investigate different planar reconstruction methods at a medium source-to-mask distance. Since this dataset was used for a quantitative comparison of planar reconstruction methods, most of the following description is taken from Ref. [116]. The experimental IGC used to captured this dataset consists of a 1 mm thick tungsten mask and a semiconductor-based detector where a 2 mm thick CdTe crystal is coupled to a Timepix3© readout circuit. The anode is divided into 256×256

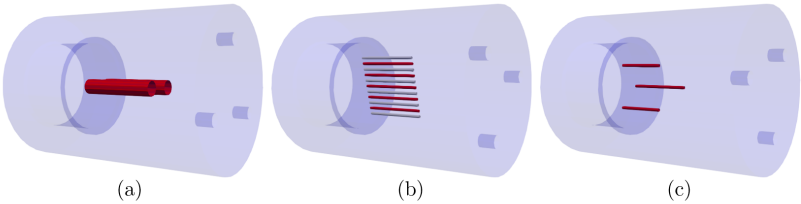


Figure 5.1: The three hot-rod phantoms of the Rozhkov dataset: (a) spatial resolution phantom (SRP), (b) linear resolution phantom (LRP), (c) contrast phantom (CP). Marked in red are the tubes that were filled with ^{99m}Tc. The notches on both sides were used to clamp and rotate the containers.

pixels with a pixel size of 0.055 mm. The detector has an active area of $D_d \times D_d = 14.08 \times 14.08\,\text{mm}^2$. The coded aperture pattern is a a 2×2 arrangement of a No-Two-Holes-Touching (NTHT) Modified Uniformly Redundant Arrays (MURA) pattern of rank 31 with pinholes of 0.34 mm in diameter, has a side length D_m of 22 mm and the mask was placed 42 mm in front of the detector, i.e. $b = 42$ mm. The mask is anti-symmetrical and can be mechanically rotated by 90°to form the anti-mask. The in-focus plane was set a mask-to-source distance of $z = 172$ mm, which, according to Eq. 3.16, results in a Field of View (FOV) of $57.75 \times 57.75\,\text{mm}^2$ for a pinhole collimator. A summary of the experimental IGC can be found in Table 5.1.

Three different phantoms were designed and manufactured: A spatial resolution phantom (SRP), a linear resolution phantom (LRP) and a contrast phantom (CP). All three phantoms have the basic shape of a cylinder with a height of 80 mm and a diameter of 50 mm, where tubes along the vertical axis were filled with ^{99m}Tc. The phantoms were designed using a CAD software and milled out of Lucite and a depiction of all three phantoms is shown with Fig. 5.1. The SRP has three tubes with a diameter of 1.1 mm, two of which are 15 mm and one 20 mm long. The tubes are arranged in parallel at three different positions. The LRP has eleven tubes in a straight line in radial direction. All tubes are 20 mm long and have a diameter of 1.1 mm. Only every second tube was filled with ^{99m}Tc. The CP consists of two larger tubes, each 25 mm long and 5 mm in diameter, centered around the long axis. All three phantoms were filled with ^{99m}Tc before beginning of the imaging process, where the total initial activity

Table 5.1: A summary of the experimental IGCs used to acquire the datasets used in this thesis.

Experimental IGC		**Rozhkov Dataset**	**Axial Resolution and Localization Dataset**
Detector	Material	CdTe	Si
	Thickness	2 mm	0.5 mm
	Side length D_d	14.08 mm	
	Pixelation	256×256	
Mask	Pattern	NTHT MURA rank 31 in 2×2 arrangement	
	Side length D_m	22 mm	9.92 mm
	Thickness	1 mm	0.11 mm
	Pinhole diameter	0.34 mm	0.08 mm
Source	Detector-to-mask distance b	42 mm	20mm
	Source-to-mask distance z	172 mm	$12 - 100$ mm
	Main gamma emission	140.5 keV	59.5 keV
	Type	^{99m}Tc-filled hot-rod phantoms	Spherical ^{241}Am source ($\varnothing$1 mm)

was 83 MBq, 50 MBq and 75 MBq for the SRP, LRP and CP. Each phantom was placed on a rotational table in front of the camera. The detector was exposed for 2 min and afterwards the phantom was rotated along its vertical axis by 3°. One complete rotation per hot-rod phantom resulted in 120 images per phantom. Each phantom was captured once with mask and once with anti-mask and thus a total amount of 720 images were acquired. Detected photons outside the energy window of 10 keV centered at the photon peak of 140 keV were discarded. The entire dataset is available upon request to the co-author of [116]: rozhkov@jinr.ru.

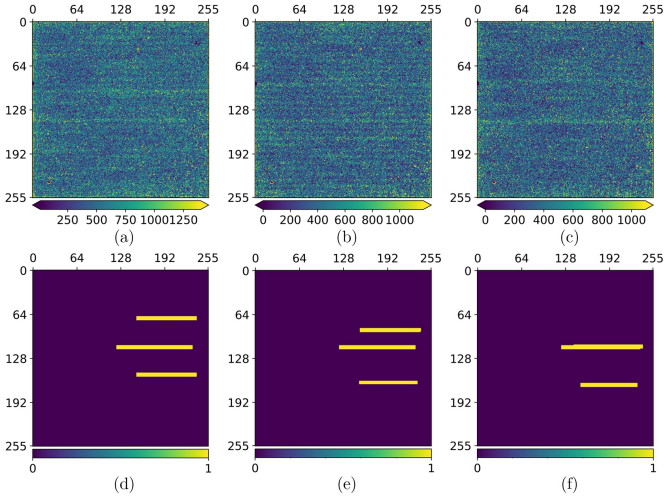


Figure 5.2: Exemplary images from the Rozhkov dataset: (a), (b), and (c) show the detector images of the spatial resolution phantom (SRP) at $0°$, $21°$, and $45°$ rotation. (d), (e), (f) show the corresponding Ground Truth (GT) images calculated based on their CAD models and the visualization software Paraview.

To compare the reconstructed images the perfect reconstruction (referred to as Ground Truth (GT) images) is essential. Therefore, GT images were obtained by loading the CAD models of the phantoms in the 3D visualization software tool Paraview (version 5.9.1) [117]. The containers and all un-filled tubes were set to transparent, in order to only show the ^{99m}Tc-filled tubes. The camera was then positioned at the corresponding distance and orientation, with focal length and resolution adjusted to match the imaging properties of our experimental gamma camera. The background and ^{99m}Tc filled tubes were set to black and white, respectively, and the camera resolution was set to 256×256. An automatic rotation around the long axis was initiated while automated screenshots were captured. Finally, the screenshots were binarized and shadows were cleaned by applying a simple thresholding operation with the threshold of 0.5 and all pixels belonging to the source was set to 1 while all other pixels were assigned to 0. A set of exemplary images from the SRP and their corresponding GT images are shown in Fig. 5.2.

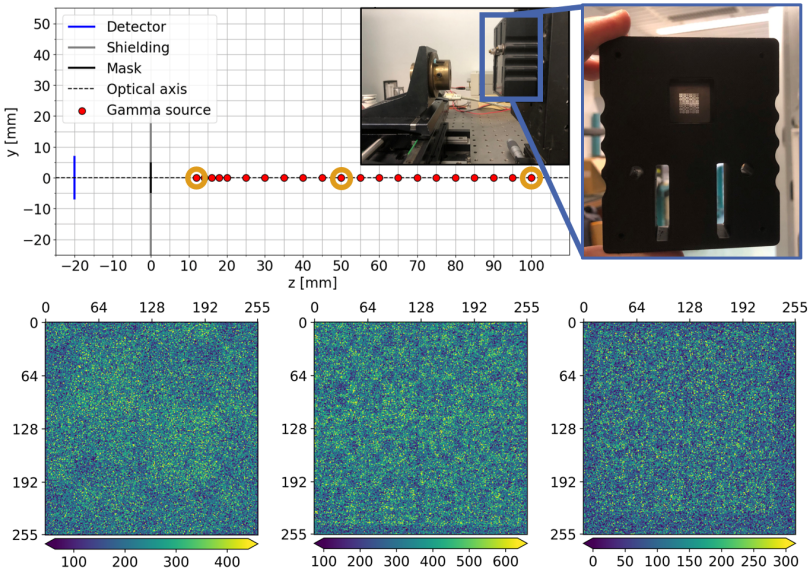


Figure 5.3: The measurement setup for the Axial Resolution Dataset. Top: The setup consisting of a Timepix3 detector, a MURA mask of rank 31 (see close-up photography on the top left) and a ^{241}Am source captured at 21 mask-to-source distances. Bottom: The captured detector images at a mask-to-source distance of 12 mm, 50 mm, and 100 mm (see orange circles above). Note, that for better visualization in each detector image pixels below and above the 1st and 99th percentile are presented in the lowest and highest corresponding color.

5.2 Axial Resolution Dataset

In order to assess the 3D imaging capabilities of CAI, a set of images with of a point-like source at several different positions in the FOV was required. This dataset, along with the one described below, was collected during the course of the publication [118], and most parts of this description are taken from there.

The experimental IGC we used for image acquisition was composed of a rank 31 NTHT MURA mask having 0.08 mm diameter holes in a 0.11 mm thick tungsten sheet and a semiconductor-based detector unit. The basic MURA pattern was duplicated in a 2×2 arrangement leading to a total mask size of $D_\mathrm{m} \times D_\mathrm{m} = 9.92 \times 9.92\,\mathrm{mm}^2$. Similar to the Rozhkov dataset, the detector

consisted also of a Timepix3© readout Application-Specific Integrated Circuit (ASIC) with a pixelation of 256×256. However, instead of a CdTe crystal, here it was bump-bonded to a 0.5 mm thick Silicon cystal with a sensitive area of $D_\mathrm{d} \times D_\mathrm{d} = 14.08 \times 14.08\,\mathrm{mm}^2$. This detector is known as the "MinipixEDU camera" designed and distributed by Advacam[1]. A 3D-printed case made of acrylonitrile butadiene styrene (ABS) keeps the detector and the collimator in a fixed distance and axially aligned and holds the mask in a detector-to-mask distance b of 20 mm.

For keeping a point-like source in a well defined distance to the camera, an automatic linear axis was used. It enabled us to move the source automatically and with high precision without interrupting the measurement procedure. An L-shaped holder was attached to the axis to which the ^{241}Am source was clamped. The source's nominal diameter is 1 mm but was previously measured to have a Full Width at Half Maximum (FWHM) of 0.65 mm [119] and emits mainly gamma photons of 59.5 keV. The source holder together with the gamma camera is depicted in the top right corner of Fig. 5.3. The coded aperture mask was originally designed for sources of 30 keV. Nevertheless, the source-mask combination used for this acquisition was the only one available to us. However, it has been shown that usage of this mask at higher energies (80 kV X-ray beam) may reduce the image contrast but does not generally impede the image reconstruction [45].

The images were acquired using a software tool provided by Advacam called "Pixet". Instead of collecting a predefined number of photons per image, we kept the acquisition time constant. We recorded 9,000 single frames with an acquisition time of 0.1 s in "Tracking mode". In this acquisition mode, the energy deposited by the interacting particles in the sensor is registered in each pixel, together with the time instant at which the interaction is revealed in the pixel. This information allows for the reconstruction of tracks released in the sensor from the impinging radiation, via a clustering algorithm based on time correspondence and spatial proximity of hits. This process, from which the name of the acquisition mode is derived, ultimately allows us to infer the type of radiation detected by the sensor based on its energy and the shape of its track. The time interval of 0.1 s per frame was chosen to avoid double counting of photons in a single pixel. Photon hits that resulted in a energy deposition of

[1] https://advacam.com/camera/minipix-edu

5 keV or less were discarded, and no further energy windowing was applied. Each acquisition lasted approximately 20 minutes, including 15 minutes of active acquisition and 5 minutes for the intermediate processing of the single frames. Thus, the pixel values of the final image represent the energy deposited in keV in each pixel integrated over the entire acquisition duration.

We captured images at 21 different source-to-mask distances, ranging from12 mm to 20 mm in 2 mm increments and from 20 mm to 100 mm in 5 mm increments. The measurement setup along with three exemplary images captured at $z = 12$ mm, $z = 50$ mm, and $z = 100$ mm can be seen in Fig. 5.3. Instead of GT images, the source size (1 mm diameter, 0.65 mm FWHM) as well as the source-to-mask distance z (ranging from 12 mm to 100 mm) serve as GT.

The experimental IGC used for this dataset, with varying detector-to-mask distances b, is known under the name "MediProbe2" [25, 46] and mainly developed by the research team surrounding Prof. Russo at the Università di Napoli Federico II in Naples, Italy, where we also took the measurements. MediProbe2 with its combination of a high-resolution detector and the coded aperture mask with very small pinholes enable a lateral resolution, was recently found to yield the best lateral resolution among current IGCs [7].

5.3 Localization Dataset

The purpose of acquiring the Localization Dataset was to evaluate the 3D-localization capabilities of a single gamma camera equipped with a coded aperture collimator. The same experimental IGC, source, and imaging procedure as for the Axial Resolution Dataset from Sec. 5.2 was used.

The key difference is that the source position was varied in two directions (z and y) while keeping it central along the x-axis. A total of 17 images were captured with lateral shifts y from the center at several mask-to-source distances z: 0 mm for $z = 20$ mm; 0 mm, 2 mm, 4 mm, 6 mm, and 8 mm for $z = 50$ mm, $z = 75$ mm, and at $z = 100$ mm additionally at $y = 14$ mm. The x-coordinate was kept constant at 0 mm. We chose these positions for two reasons. First, it was shown that off-center sources are reconstructed with a lower contrast than centered sources [46]. However, neither was a difference in the two lateral coordinates reported, nor can we think of a reason to assume an asymmetry. Thus, to spend our measurement time efficiently, we captured only data of a source moving

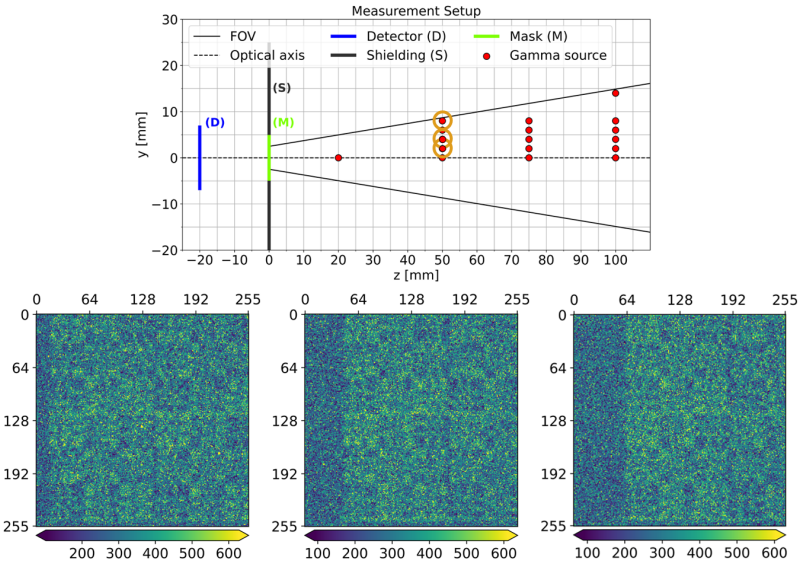


Figure 5.4: The measurement setup for the Axial Localization Dataset. Top: The experimental gamma camera setup and the source are the same as for the Axial Resolution dataset. Bottom: Three captured detector images at a mask-to-source distance of 50 mm and a lateral position of 2 mm, 4 mm, and 8 mm (see orange circles above).

along the positive y-coordinate. Second, practical limitations were imposed by the two automatic linear axes (Physik Instrumente (PI), Karlsruhe, Germany) available to us.

In order to have a better control over the exact source position, we reproduced the experimental setup in-silico using a Monte Carlo (MC) simulation, as described in Sec. 3.3.3. With the software package TOPAS (version 3.8.1) [88], the source was modeled as a homogeneously radiating sphere of ^{241}Am with a 1 mm diameter and emitting a total of 10^9 photons of 59.5 keV. The mask was modeled as a sheet of tungsten with the parameters described above. Because simulating the deposited energy in a semiconductor detector with a bias voltage is a complex process and an area of ongoing research [43], we decided for the following image generation process: Instead of the deposited energy in each pixel, we captured the photon hits. All photons that passed the front surface of the detector were scored in a phase space file. From this list, we generated

a detector image by computing the 2D histogram according to the detector specifications. By doing so, we ignored the charge-sharing effect between neighboring pixels and defective pixels, which are rare in thin silicon-based detectors compared to CdTe [44].

The setup is depicted in Fig. 5.4 together with three exemplary images from the dataset and a table with all positions can be found in the Tab. 5.2, including the total integrated energy per image and the number of photon hits for the MC simulation.

Table 5.2: We both simulated and captured detector images of a ^{241}Am radioactive source with a nominal diameter of 1 mm emitting mainly gamma photons of 59.5 keV at 17 positions given in millimeters within the near-field of our gamma camera. Additionally, the total energy captured by our experimental gamma camera as well as the number of photon counts from the MC simulation are presented.

Source position / mm			Experimental data	MC Simulation
x	y	z	Integrated Energy / keV	Photon counts
0	0	20	34,691,600	5,204,466
0	0	50	21,581,440	1,734,918
0	2	50	21,378,936	1,724,074
0	4	50	3,062,625	1,706,909
0	6	50	20,536,964	1,684,588
0	8	50	20,173,747	1,656,675
0	0	75	11,640,180	930,834
0	2	75	11,572,164	929,798
0	4	75	11,524,095	925,509
0	6	75	11,408,447	917,916
0	8	75	11,257,055	909,253
0	0	100	7,141,504	580,302
0	2	100	7,146,672	579,403
0	4	100	7,157,586	578,895
0	6	100	7,271,324	576,351
0	8	100	7,158,355	572,672
0	14	100	6,938,520	558,671

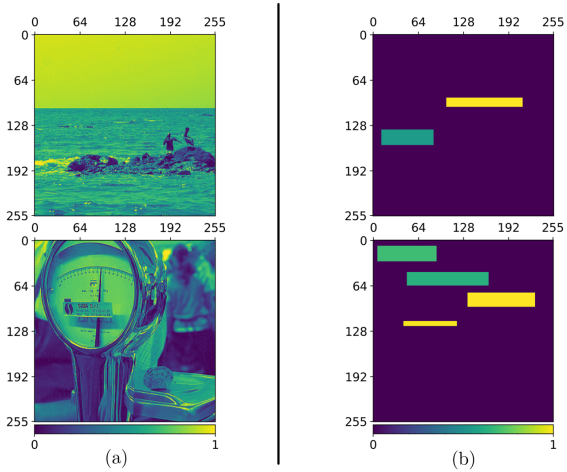


Figure 5.5: Two different sets of image data that serve as GT for training: (a) Imagenet, and (b) Lines. Note, that the bottom image of (b) is the same image that is shown in Fig. 3.4.

5.4 Imagenet Dataset

What distinguishes a ML approach from analytical reconstruction methods is the need for a large quantity of training images. A common way to solve this data problem is by using simulated training data [10, 50, 120]. Thus, two different image domains were used to generate detector images by leveraging the Convolutional Model from Sec. 3.3.2 to generate detector images: A dataset based on the Imagenet dataset (abbreviated by *IN*), and a dataset of images with horizontal lines (*Lines*).

The IN dataset is based on a large collection of natural photographs called *ImageNet* [121]. The first $50,000$ images from the ImageNet's validation set were chosen, all three color channels averaged to yield a gray-scale image and a random region of 256×256 pixels was cropped. The final image was finally normalized to the range of 0 to 1. Two exemplary images from the Imagenet Dataset are shown in Fig. 5.5

5.5 Lines Dataset

The Lines Dataset is fully synthetic and contains horizontal lines with varying thickness, length and intensity on dark background. These image were designed to represent a wide variety of source distributions in a single in-focus plane in front of our IGC with the goal to reconstruct the images from the Rozhkov dataset described above. For this training dataset, $50,000$ images with black background and different number of horizontal lines with varying thickness were created using the openCV library (version 3.4.2) [122]. The number of lines varied between 2 and 6, with a height and width between 4 and 24 pixels and between 60 and 122 pixels, respectively. The center points of the lines were randomly positioned across the image, with any portions of lines extending beyond the image borders being cropped. Each line was assigned a uniform intensity between 0 and 1, and the intensities of overlapping lines were added to ensure a non-binary intensity distribution. For both sets, the final images were normalized to the range of 0 to 1.

PART III

2D IMAGING

Chapter 6

Quantitative Comparison of Planar Reconstruction Methods

This chapter examines Hypothesis Ia by first, giving an overview of common planar reconstruction methods from the literature with Sec. 6.2.1. Afterwards, the architecture, training, and generation of training data for the Machine Learning (ML) approach are explained in Sec. 6.2.2. Section 6.3 shows the results of comparing the reconstruction methods in terms of reconstruction quality and runtime, which are discussed in Sec. 6.4. Finally, Sec. 6.5 concludes this study's main findings.

This chapter is taken in most parts from the related open-access publication licensed under CC-BY 4.0 in IOPscience "Journal of Instrumentation" [116]. It was complemented by a more detailed comparison of the influence of the preprocessing, and a more in-depth demonstration on how the Two-Holes-Touching (THT) pattern compared to the No-Two-Holes-Touching (NTHT) pattern affects the reconstruction results.

6.1 State of the Art

Coded Aperture Imaging (CAI) has been proposed as an alternative collimation technique for Radioguided Surgery (RGS), because it persuades a better

compromise between resolution and photon efficiency [21, 22]. However, it requires image reconstruction to obtain an interpretable image. When the depth dimension of an object of interest is small compared to the camera-object distance, one can reasonably assume that all gamma radiation emerges from a fixed plane parallel to the detector. Subsequently, CAI can be regarded as an image-to-image mapping and is referred to as *planar CAI*. Several reconstruction methods for planar CAI have been proposed but yet, no comprehensive quantitative comparison has been carried out to assess their advantages and disadvantages. A first study has been carried out by Kulow et al. [50], but it is only a qualitatively comparison based on only few captured images between reconstruction methods that are not widely used in CAI. ML approaches have been investigated by Zhang et al. [29, 120]. However, their architectures are not capable of processing high-resolution images, the evaluation is entirely based on simulated data and a comparison to state-of-the-art methods is yet to be presented. Additionally, the phenomenon of *domain shift* has not been quantitatively investigated with respect to CAI reconstruction. Domain shift refers to when a network trained on a source domain encounters a different data distribution when applied to new unseen data [104]. Usually, the larger the domain shift, the worse the network performs on the unseen data [9, 123]. The main contributions of this chapter to the state of the art can be summarized as follows:

1. A thorough and quantitative comparison of the most commonly used planar reconstruction methods with regards to runtime and reconstruction quality is presented.
2. A Convolutional Encoder-Decoder (CED) is proposed that, even though trained on simulated CAI images, is able to outperform state-of-the-art methods and proved worthy of further investigations.
3. We offer the research community access to three datasets of high-resolution coded aperture images of different hot-rod phantoms acquired with an experimental gamma camera.
4. This chapter describes, to the best of our knowledge, the first re-implementation and evaluation on experimental data of the convolution-based Maximum Likelihood Expectation Maximization algorithm proposed by [24].

6.2 Methods & Material

In the following, the chosen analytical reconstruction methods from literature and the development of a novel ML approach will be presented. Furthermore, the Contrast-to-Noise Ratio (CNR) as a metric for comparing the reconstruction quality is introduced before the preprocessing of the acquired detector images is described. The comparison is carried out based on the Rozhkov dataset described above in Sec. 5.1.

6.2.1 Analytical Reconstruction Methods

Several methods for planar image reconstruction have been proposed within the last few decades: MURA Decoding, also called inverse filtering, or cross-correlation analysis [25, 26, 31, 32, 73, 78, 83, 124–127], Wiener Filter [23, 27], standard Maximum Likelihood Expectation Maximization (MLEM) [28, 128], convolutional MLEM algorithm [24, 129], genetic algorithm [23], least-square optimization [23], and data-driven ML approaches [29, 50, 120]. A number of the proposed methods (standard MLEM, genetic algorithm, and least-square algorithms) require a system matrix which results in high computational costs and runtimes of multiple minutes up to 16 h were reported [23]. Therefore, these reconstruction methods were not considered in this study.

The runtimes for all methods were measured by averaging the elapsed time for the reconstruction of one image over $1,000$ runs on the CPU of a computer with a 6-kernel Intel Core i7 processor (2.6 GHz) and 16 GB of RAM. The ML library TensorFlow [130] enables the use of the GPU for the analytical reconstruction methods as well. Hence, the runtime with GPU support was also evaluated with a Nvidia GeForce RTX 2070 and 8 GB dedicated RAM. All reconstruction methods were implemented, processed, and analyzed with NumPy (1.24), SciPy (1.10.1), and TensorFlow (2.7.2) in Python (3.9.7).

6.2.1.1 MURA Decoding

The MURA patterns were developed with the idea to automatically yield an optimal decoding pattern without using the Fourier transformation [74, 131].

Under considerations of the convolutional model (see Sec. 3.3.2) and purely Gaussian additive white noise, the decoding pattern g' is optimal and the correlation of the decoding and the encoding pattern h' results in a δ-distribution. The decoding pattern g' is directly derived from the encoding pattern h':

$$g'[i,j] = \begin{cases} 1 & \text{if } i+j=0, \\ 1 & \text{if } h'[i,j]=1, i+j\neq 0, \\ -1 & \text{if } h'[i,j]=0, i+j\neq 0. \end{cases} \tag{6.1}$$

Practically this translates to changing all entries that are 0 to $-1$and all 1 stay 1 except for the central pixel if a 2×2 arranged mask is chosen [74]. Two different decoding patterns were constructed and evaluated on the Rozhkov dataset: A version according to the NTHT pattern where the intermitting rows and columns are considered part of the encoding pattern $h'[x,y]$ and thus turned negative as well, and a version according to the THT pattern, where $h'[x,y]$ does not have these rows. Both have been used as base pattern for reconstruction in literature [27, 127].

Generally, when using a 2×2 arrangement of the basic MURA pattern, only the central part of the detector image $p[x,y]$ is employed for reconstruction. This central part, $C(p[x,y],b,z)$ is obtained by cropping the portion of the detector with the central cropping operator $C(\cdot)$ onto which one basic MURA pattern is projected. For a more in-depth explanation the reader is referred to [73]. The size of this projection depends on the magnification factor M from by Eq. 3.2. As a result, the dimension of the reconstructed image depends on the detector-to-mask and source-to-mask distances b and z. Because of the geometrical properties of our camera setup, the central projection matched exactly the detector size and, thus, no cropping was required.

Instead of a circular convolution with the central mask patter, we opted for a linear convolution. The decoding pattern g' was scaled to twice the detector dimensions of 512×512 pixels by nearest neighbor interpolation and is denoted as $g[x,y]$ in analogy to the Convolutional Model from Sec. 3.3.2. Finally, the linear convolution of the given detector image $p[x,y]$ with the decoding pattern $g[x,y]$ yields the reconstructed image $\hat{f}[x,y]$:

$$\hat{f}[x,y] = p[x,y] * g[x,y] \tag{6.2}$$

Normally for MURA Decoding, the detector image is cropped to the central projection. However, because of the geometrical properties of this camera setup the entire detector was covered by the central projection. Thus, for MURA Decoding and the Wiener Filter, no cropping was required and the complete detector image could be used.

6.2.1.2 Wiener Filter

The Wiener Filter executes reconstruction in an optimal balance between inverse filtering and noise smoothing with regards to the mean squared error. As above, the underlying assumptions are the convolutional model and Gaussian additive white noise [131]. It can be regarded as a weighted cross correlation with the mask pattern and has the design parameter *signal-to-noise-ratio* (SNR), which is a scalar constant. For formulating the Wiener Filter $\mathbf{W}[u,v]$, the Point Spread Function (PSF) $h[x,y]$ must be transformed into the Fourier domain where it is denoted as $\mathbf{H}[u,v]$ via the discrete 2D Fourier transformation $\mathfrak{F}$ with the frequency coordinates u and v:

$$\mathbf{H}[u,v] = \mathfrak{F}\{h[x,y]\} \tag{6.3}$$

Hence, the Wiener Filter is defined as follows:

$$\mathbf{W}[u,v] = \frac{1}{\mathbf{H}[u,v]} \frac{||\mathbf{H}[u,v]||_2^2}{||\mathbf{H}[u,v]||_2^2 + \frac{1}{\mathrm{SNR}}} \tag{6.4}$$

As opposed to MURA Decoding, here, the reconstruction turns into a multiplication and thus a faster reconstruction can be expected. Therefore, the detector image $p[x,y]$ is transformed into the Fourier domain, element-wise multiplied with $\mathbf{W}[u,v]$ and afterwards transformed back into the image domain to yield the reconstruction result $\hat{f}[x,y]$:

$$\hat{f}[x,y] = \mathfrak{F}^{-1}\left\{\mathfrak{F}\left\{p[x,y]\right\} \cdot \mathbf{W}[u,v]\right\} \tag{6.5}$$

Analogously to MURA Decoding, two types of mask pattern were used as base pattern for the Wiener Filter: The THT and the NTHT version. However, since the convolution will be carried out in the Fourier domain, only the central 256×256 pixels of $h[x,y]$, corresponding to one basic MURA pattern, was required. Furthermore, no cropping took place and the entire detector image

was exploited. The Wiener Filter was calculated according to Eq. 6.4 with various SNR. A line search was carried out to find the optimal value for the design parameter SNR ranging from 10^{-1} to 10^{-9}. The SNR that obtained the highest median CNR for one of the hot-rod phantoms, the SRP, was chosen for all following analyses.

6.2.1.3 Maximum Likelihood Expectation Maximization

The MLEM algorithm is an iterative algorithm that estimates the source image with the highest likelihood for the captured detector image. The algorithm is derived from the assumption that the photon detection follows a Poisson process [106] that is described in Sec. 3.5. Originally, a transfer matrix A must be provided where the source and detector images are considered as vectors. The entries a_{ij} represent the fraction of photons emitted from source pixel j and detected at detector pixel i. Ergo, A grows rapidly for high-resolution imaging and would contain more than 4 billion entries for an image of 256×256 pixels resulting in a computational demand that makes this approach impractical for RGS [28, 128]. To overcome this major drawback, Mu & Hoang [24] proposed a convolutional-based version of the MLEM algorithm: Instead of using the system matrix A for forward- and back-projection, the PSF $h[x,y]$ is used:

$$\hat{f}^{(k+1)}[x,y] = \hat{f}^{(k)}[x,y] \cdot \left[\frac{p[x,y]}{\hat{f}^{(k)}[x,y] * h[x,y]} \times h[x,y] \right], \tag{6.6}$$

where "$\cdot$", "$*$" and "$\times$" denote element-wise multiplication, linear convolution and the correlation operator. The current iteration is indicated by the superscript letter k. The algorithm consists of four major steps: forward-projection, calculating the difference between the detector image and the forward-projection, back-projection of the error and updating the estimate. The convolution in the denominator represents the forward-projection step using current estimate $\hat{f}^{(k)}$. Dividing the detector image p by the current forward-projection yields the relative difference. The correlation of the relative difference with the PSF h represents the back-projection, which is multiplied element-wise with the current estimate $\hat{f}^{(k)}$ to obtain an updated estimation $\hat{f}^{(k+1)}$.

The algorithm according to Eq. 6.6 was implemented with an additional small ε of 10^{-7} added to the denominator to avoid zero divisions. Additionally, the correlation was carried out as convolutional with a flipped second argument

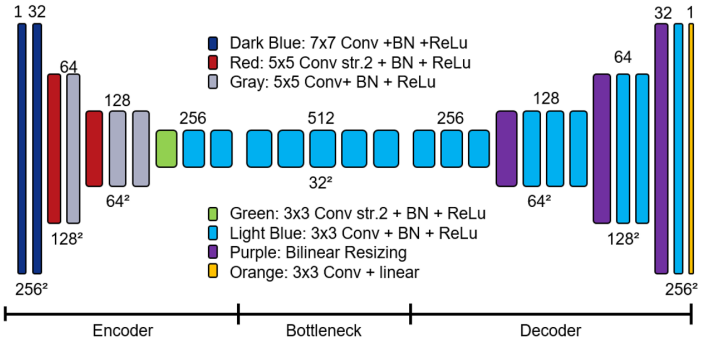


Figure 6.1: The architecture of the deployed CED network with the number of feature channels on top and the spatial resolution of the feature maps at the bottom. "Conv" indicates convolutional layers, "BN" Batch Normalization and "ReLu" an activation with the ReLu function.

to use the same implementations for both operations. The initial guess was set to a constant image of 0.5. Instead of applying an adaptive stopping criteria, we analyzed the results of one of the phantoms, the SRP, and chose the number of iterations for all further analyses where the median CNR was maximal.

6.2.2 Data-Driven Reconstruction Methods

Initial promising experiments have been carried out on the application of Convolutional Neural Networks (CNNs) to planar CAI reconstruction, but they were either validated on simulated low-resolution images [29, 120] or were only compared visually with other methods [50]. This section describes the development of a CED, a specific architecture of CNNs trained on two different training datasets, that performs high-resolution planar reconstruction. In the following, the above described reconstruction methods are referred to as *analytical* reconstruction methods, while trained CEDs are denoted as *data-driven* reconstruction methods.

6.2.2.1 Architecture and Training

Figure 6.1 presents the CED architecture used for this experiment. It is based on architectures from literature [98, 101, 132] and in particular on DeepPET [10].

A multitude of architectures for image-to-image tasks have been proposed. A well-known architecture called Automap [9] contains multiple layers of fully connected neurons which makes it infeasible for the detector size of our setup: For the reconstruction of an image with 256×256 pixels more than $4 \cdot 10^9$ trainable parameters would be required, exceeding the capabilities of available GPUs. The widely acclaimed U-Net architecture [98], which is another subtype of the CED architecture, has concatenating or additive skip connections directly transferring information from the encoder to the decoder part. However, as there is no similarity on a pixel level between input and output domain in CAI, it was decided to refrain from skip connections. Another architecture is the Mixed Scale Dense Convolutional Network (MSD-Net) [99], which relies on convolutional kernels with a large dilatation. The problem is that these kernel makes training highly inefficient and, hence, only small networks could be evaluated. In the appendix A.1, it is shown how other architectures compete against our proposed CED architecture. It is compared to the three MSD-Nets and two U-Nets that were trained in the same manner.

The main building block of the developed CED architecture is a convolution block. It consists of a convolutional layer with different kernel and strides, batch normalization and is concluded by a ReLU activation function. Details about the building blocks can be found above in Sec.4.2. To ensure a large receptive field, which is especially important in CAI, as information is spread across the entire input image, a large kernel size of 7×7 pixels and 5×5 pixels in the encoding part was chosen. In the encoder part, the spatial resolution is successively reduced to 32×32 pixels by convolutions with stride 2. Simultaneously, the number of channels increases from 1 to 512 in the bottleneck. The bottleneck is intended for the actual transformation between the domain of the detector image and the domain of the source distribution and thus the majority of trainable parameters were embedded here. The decoder takes the encoded representation from the latent space and progressively increases the spatial size while reducing the channels to reach a final output image with a single channel and 256×256 pixels. As described in [133], using transposed convolutions can lead to

checkerboard artifacts in the output image when performing image regression tasks. Therefore, bilinear resizing followed by unstrided convolutions was employed in this network's decoder. The kernel size remains 3×3 in the entire decoder section to produce high-resolution images. The final layer is a convolution followed by a linear activation function, because of its constant sensitivity. Overall, the CED network consists of $16,414,817$ trainable and $10,432$ non-trainable parameters.

The loss function used for training is the mean squared error of output and target image and was optimized by the Adam optimizer [108] with its standard parameters. A mini-batch size of 16 was chosen, as a compromise between short training time and a small generalization gap between training and validation data [104]. The network was trained on $35,000$ training images and validated during training on separate $10,000$ validation images for 20 epochs. If within three epochs the validation loss had not improved, the learning rate was automatically reduced by a factor of 0.4. An automatic stopping criteria was defined to efficiently use the development time of such large networks [104]: When the validation loss did decrease above a threshold of 10^{-6} for 4 epochs, training was automatically stopped. Training was carried out for both training sets, whereas the network trained on the Imagenet dataset served as pre-trained network for the Lines dataset. Afterwards, the CEDs with the best performance on the validation set were chosen and applied to the hot-rod phantom data.

The CED, as well as all other reconstruction methods, were implemented, trained and evaluated in TensorFlow. The network trained on the ImageNet dataset (see Sec. 5.4 is referred to as *CED-IN*, the network additionally fine-tuned on the Lines dataset (Sec. 5.5) as *CED-Lines*. In the following analysis both trained CEDs are considered as separate reconstruction methods.

6.2.2.2 Simulation of Training Data

In order to rapidly simulate the detector image from an arbitrary source images, the Convolutional Model described in Sec. 3.3.2 was chosen. Therefore, the mask pattern $h'[x,y]$ was scaled according to Eq. 3.8 with $b = 42\,\text{mm}$ and the source-to-mask distance $z = 172\,\text{mm}$ to obtain the camera's PSF $h[x,y]$. The linear convolution of the PSF and a source image resulted in an image of 498×498 pixels, that were then centrally cropped to 256×256 pixels yielding

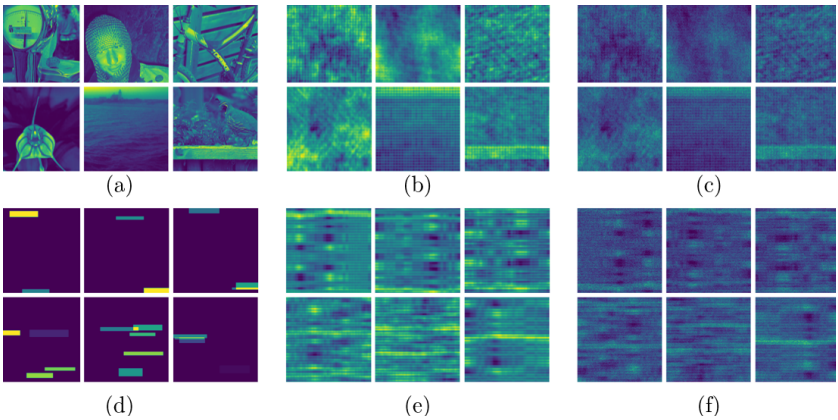


Figure 6.2: Steps in simulating the training data for the CEDs based on two different image datasets, IN (top) and Lines (bottom). (a) and (d) are the target images, (b) and (e) the simulated detector images, and (c) and (f) the detector images after Poisson randomization.

the simulated detector image. No additive noise or near-field artefacts from Sec. 3.5 were considered. However, different levels of Poisson noise were simulated. The sum of all pixel intensities, corresponding to the amount of the photon count, was scaled to reach a random number in the range of $1,000$ and $1,000,000$ photons. This scaled image represents the *expected photon count per pixel* and formed the basis for the Poisson randomization (see Eq.3.24). Finally, the simulated detector image was normalized to the range of 0 to 1. Exemplary images with their simulated forward-projection with and without Poisson randomization from the ImageNet dataset and for the horizontal line dataset can be seen in Fig. 6.2. The generation of training data was carried out on-line while training the CEDs.

6.2.3 Contrast-to-Noise Ratio

A widely used metric to quantify the reconstruction quality in nuclear imaging is the Contrast-to-Noise Ratio (CNR), because it takes into considerations both the source visibility and the degradation from noise [29, 120, 129, 134]. Thus, the CNR is considered to be more aligned with the human visual perception of image quality compared to the root mean squared error. Additionally, it is

well-known in the medical imaging field and applied across various imaging modalities, such as SPECT, MRI, and CT [134, 135]. The following definition of CNR is employed:

$$CNR = \frac{\bar{S} - \bar{B}}{\sigma_B}, \tag{6.7}$$

where $\bar{S}$ denotes the mean intensity of the signal S, $\bar{B}$ the mean intensity and σ_B the standard deviation of the background B. The binary ground truth images from Section 5.1 allowed the bisection of the reconstructed image into signal and background.

6.2.4 Preprocessing

As previously described in Sec. 2.2 semiconductor-based detectors are prone to defects in the crystaline structure of the active area of the sensor. The Rozhkov dataset is captured with a detector consisting of CdTe and small clusters of defective pixels are noticeable on first inspection of the captured images.

A negative impact on the reconstruction results must be assumed and, thus, a preprocessing step prior to reconstruction has been proposed [136]. The goal of this procedure is to mitigate degradation by clusters of defective pixels and replace them with more "reasonable" values. Thus, a two-step preprocessing is employed: First, the defective pixels were identified by averaging all 720 images taken with our experimental gamma camera. This average image was thresholded according to the 2^{nd} and 98^{th} percentile. Pixels outside this range were considered defective. Second, these defective pixels were consequently replaced by the median of their 3×3 neighborhood in each image of our dataset. The effect of preprocessing the detector images on the final reconstruction quality was analyzed based on the SRP.

6.3 Results

The results are presented in the following order: First, the preprocessing of the captured images is compared to the reconstruction quality based on the raw detector images. Afterwards, the reconstruction methods with the optimal

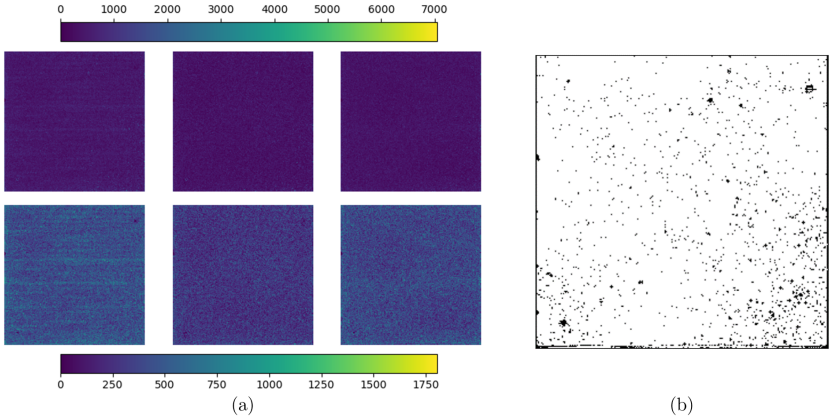


Figure 6.3: (a) shows the first detector images from each of the three captured hot-rod phantoms before (top) and after preprocessing. (b) shows the binary mask which mark the defective pixels in black according to the 2^{nd} and 98^{th} percentile of the average image. Figure from [116].

parameters are compared both regarding their reconstruction quality as well as the runtime.

6.3.1 Preprocessing

The first detector image for each of the three hot-rod phantoms is presented in Fig. 6.3a before and after preprocessing. Pixels with high intensities or clusters of black pixel corrupt the image and structure is barely visible as the top row shows. Replacing outliers by the median value of their neighborhood lead to a higher contrast and vertical stripes becoming visible. The binary mask generated by thresholding the average detector image (Fig. 6.3b) shows multiple clusters of erroneous pixels in addition to pixels along the edge. Despite the preprocessing, systematic pixel errors remain and are noticeable as black clusters in the top right corner of all detector images.

Table 6.1 presents the median CNR of each reconstruction method and phantom based on the un-preprocessed detector images (raw) and the difference to the preprocessed detector images. Regarding the SRP, where overall the best

Table 6.1: This table shows the median CNR and the interquartile range between the 25^{th} and 75^{th} percentiles in brackets of the SRP reconstructions based on the raw, the preprocessed detector images and their difference. The Wiener Filter is carried out with $SNR = 10^{-9}$ and the MLEM algorithm with 106 iterations. The values printed in **bold** indicate that the median CNR of the preprocessed is higher than of the raw images. The median CNRs for the preprocessed images for all phantoms are presented in Tab. 6.2.

Reconstruction method	Raw	Preprocessed	Difference
Wiener Filter	1.59 (0.45)	1.92 (0.63)	**+0.33**
MURA Decoding	1.90 (0.69)	1.92 (0.74)	**+0.02**
MLEM	2.57 (1.24)	2.31 (1.03)	−0.26
CED-IN	1.36 (0.32)	2.65 (1.05)	**+1.29**
CED-Lines	3.89 (1.70)	5.22 (2.82)	**+1.33**

reconstructions were obtained, it is noticeably that the Wiener Filter improves by 0.33 while the MLEM algorithm obtains a median CNR that is smaller by 0.26. The largest improvement were obtained by the data-driven methods: CED-IN and CED-Lines achieve improvements of 1.29 and 1.33.

6.3.2 Comparison of Reconstruction Methods

As Fig. 6.4 shows, the NTHT version for reconstruction causes a periodic noise causing the median CNR of MURA Decoding to drop from 1.92 to 0.35 on the SRP. An analogous behavior was observed for the analytical reconstruction methods as well. Thus, in the following, all presented results are based on the THT version.

For the Wiener Filter a SNR of 10^{-7} resulted in the highest median CNR for the SRP (see Fig. 6.5b) which will be considered for the remainder of this analysis. Instead of applying adaptive stopping criteria to the MLEM algorithm, the fixed number iterations was chosen based on where the reconstruction quality of the SRP as well. As Fig. 6.5b shows the median CNR rises from 1.89 after one iteration and declines after reaching the maximum median CNR of 2.31 after 106 iterations. Training the CED-IN took around 11:55 h and additional 3:45 h for the lines dataset. The automatic stopping aborted training after 19

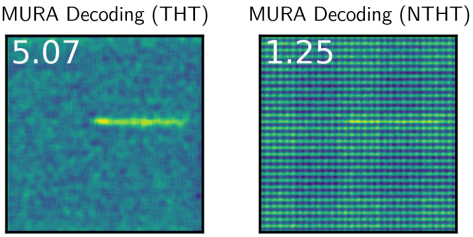


Figure 6.4: This graphic shows the reconstruction of one image from the SRP with MURA Decoding where the decoding pattern is based on the THT (left) and on the NTHT (right) version of the utilized MURA pattern. The CNRs of the reconstructions are printed in the top left corners. Note the heavy periodic noise when the NTHT version is used.

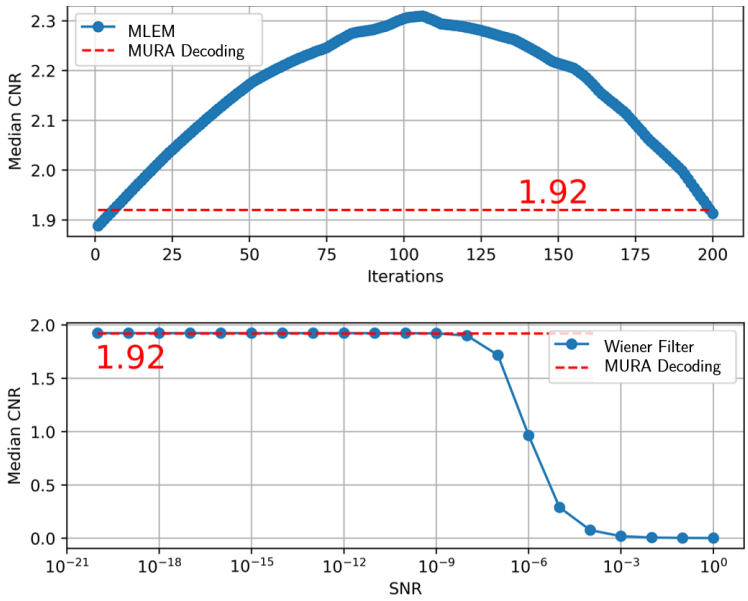


Figure 6.5: Top: The median CNR of the SRP dataset plotted against the number of iterations for the MLEM algorithm. Bottom: The median CNR of the SRP dataset plotted against the hyperparameter SNR of the Wiener Filter. Finally, 106 iterations and a SNR of 10^{-9} was chosen as optimal hyperparameters. The median CNR of the SRP dataset reconstructed with MURA Decoding of 1.92 (red dashed line) is plotted for reference.

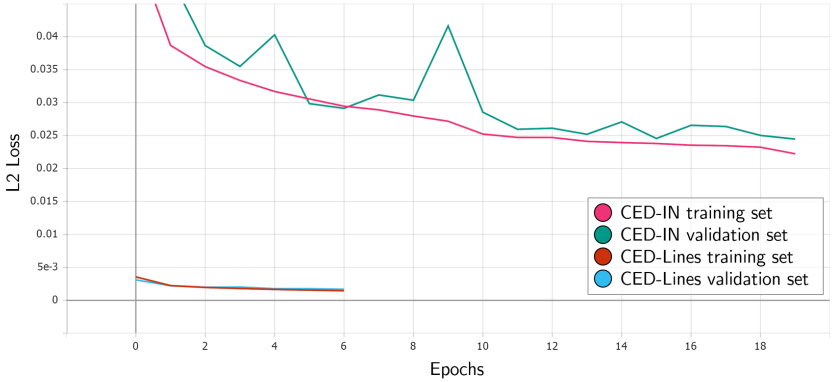


Figure 6.6: The L_2 loss of the CED training plotted against the number of training epochs. The top pair represent the loss of the CED-IN over the training set (green) and the validation set (pink). The bottom pair is the CED-Lines with training loss in red and the validation loss in blue.

epochs and 6 epochs for the CED-IN and CED-Lines, respectively, as Fig. 6.6 shows.

Table 6.2: The median CNR and the 25:75 interquartile range in brackets separated by the hot-rod phantoms and the average runtime of each algorithm in milliseconds. The reconstruction method with the highest CNR and the shortest runtime per phantom are printed in bold. Table modified from [116].

Reconstruction method	Median Quality in CNR			Average runtime in ms	
	SRP	LRP	CP	CPU	GPU
Wiener Filter	1.92 (0.63)	0.63 (0.69)	0.71 (0.36)	**67 ± 6**	**37 ± 44**
MURA Decoding	1.92 (0.74)	0.55 (0.62)	0.62 (0.33)	288 ± 8	53 ± 56
MLEM	2.31 (1.03)	0.39 (0.65)	0.50 (0.35)	48,962 ± 828	4,667 ± 292
CED-IN	2.65 (1.05)	**0.92 (1.46)**	1.23 (0.66)	285 ± 10	52 ± 37
CED-Lines	**5.22 (2.82)**	0.52 (2.45)	**1.42 (1.53)**	286 ± 9	52 ± 17

The distribution of the 120 CNR values for the five different reconstruction methods are presented as grouped boxplots in Fig. 6.7. The median CNR values for each reconstruction method and phantom are listed in Tab. 6.2. It is noticeable, that the SRP is reconstructed with a higher CNR than the other two phantoms (see also Fig. 6.8): While the median CNR for the SRP varies between 1.92 and 5.22, the median CNR for the CP ranges between 0.57 and 1.42 and is even smaller for the LRP with median CNRs below 1.0.

The Wiener Filter and MURA Decoding show similar reconstruction results, with the latter producing a smoother background and hence reaching an overall higher median CNR for all phantoms. However, among the analytical reconstruction methods, the MLEM algorithm produces the best reconstructions for the SRP and reaches a similar quality for the LRP and CP. Its reconstructions exhibit a darker and more uniform background, but also slightly weaker tubes.

Both CEDs reconstructed the phantoms with a higher median CNR than all analytical reconstruction methods for two of the three phantoms, namely the SRP and CP. While the CED-IN obtains a median CNR of 2.65 for the SRP, the CED-Lines achieves an even higher median of 5.22. The background generated by the data-driven reconstruction methods and especially for the CED-Lines is almost uniformly dark, except for a few cloudy artifacts. As Fig. 6.8 shows, the tubes of the SRP are clearly visible in all reconstructions, but are weaker for the LRP and CP, especially when the tubes are in an upright position. Overall, the second tube of the CP is missing from the reconstructions. In most reconstructions there is a thin bright stripe at the top of most images, regardless of the reconstruction method.

The average runtimes and their standard deviations are presented in Tab. 6.2: 106 iterations of MLEM executed on the CPU takes an average of 48,962 ms, while MURA Decoding takes 288 ms and the Wiener Filter even less with 67 ms. The runtime of both CEDs are almost identical with 285 ms and 286 ms on the CPU and 52 ms for both CEDs on the GPU. When deploying a GPU, the runtimes decrease, but are associated with a higher standard deviation.

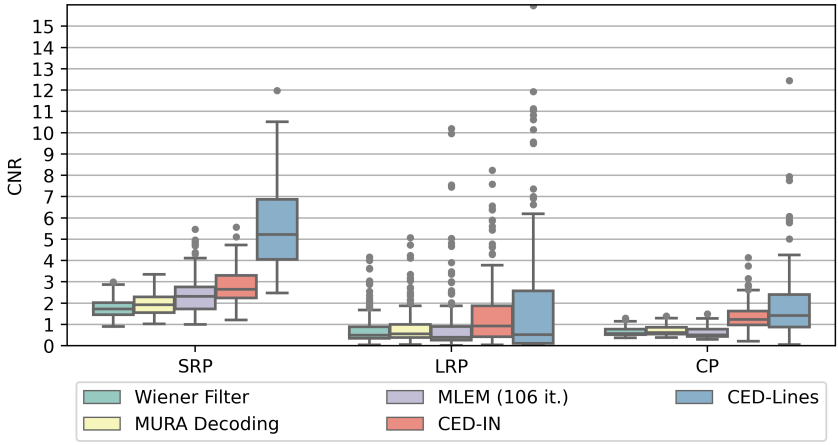


Figure 6.7: CNR distribution of the reconstructions of the preprocessed detector images separated by the three hot-rod phantoms (SRP, LRP and CP). Three data points above a CNR of 16 are not shown. The boxes indicate the 25:75 percentiles, lines are medians, whiskers stretch 1.5 times the interquartile range and data point outside this range are considered outliers. Figure from [116].

6.4 Discussion

In following sections, the reported results are interpreted starting with the Rozhkov dataset, the proposed preprocessing, before the planar reconstruction methods are compared.

6.4.1 Rozhkov Dataset and Preprocessing

Even after preprocessing the acquired images, noise dominates all detector images. Structures in the images are more distinctive for the SRP than for the CP and LRP. Despite the preprocessing, the acquired detector images still suffer from systematic pixel errors. The acquisition a flood-field and a dark-field image would allow a pixel-wise calibration process. However, defective pixels that always output 0 could not be corrected and remain a general disadvantage of semiconductor-based detectors.

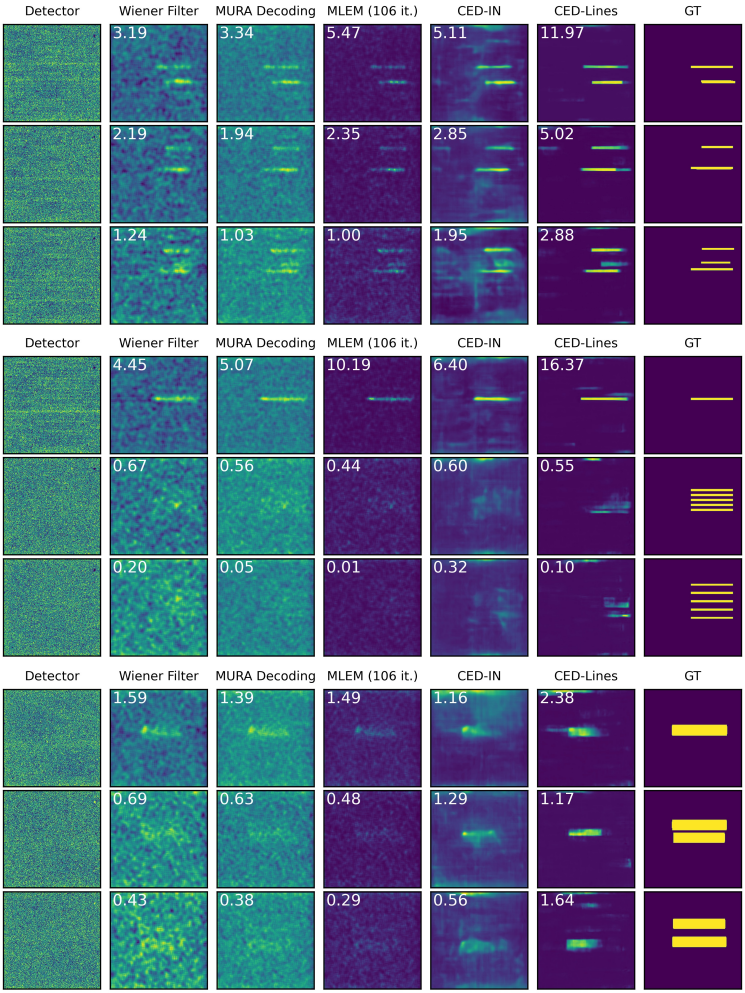


Figure 6.8: Exemplary reconstructions from the SRP (top), LRP (middle) and CP (bottom). For each hot-rod phantom from top to bottom row: The best, median and worst reconstructions from MURA Decoding and the corresponding reconstructions from all other methods are displayed. The outer columns represents the detector and ground truth (GT) image respectively. Note, that the histogram of the detector images were equalized for better visualization. Figure from [116].

Nonetheless, all reconstruction methods are equally affected by noise and defective pixels which assures a fair comparison. Overall, the reconstructions from the LRP and CP yield on average CNRs of around 1.0 and below, suggesting that not enough photons were detected, potentially due to a large amount of ^{99m}Tc that has already decayed before the acquisition took place. Thus, the SRP dataset where all three tubes are visible in most images and independently from the reconstruction method can be considered the most useful dataset in comparing. An air bubble trapped in one tube of the CP was found after the measurements and posed an additional challenge.

All reconstruction methods, not only the CEDs, have difficulties in reconstructing the phantoms when the tubes are in an upright position. In these cases, the radiating sources are spread over a larger portion of the Field of View (FOV). This leads to a more uniform distribution of the photons hitting the detector, which reduces the photon count per pixel. Hence, Poisson noise dominates these detector images. The almost blank reconstructions of the LRP show that the data-driven reconstruction methods cannot perform miracles and rely on a certain quality of the detector image.

Whether the proposed preprocessing improves the reconstruction quality depends on the reconstruction method. In the case of MURA Decoding, the choice of raw or preprocessed detector images did not affect the reconstruction outcome, rendering it the most robust reconstruction method. After the Wiener Filter, the CEDs benefited the most with improvements above 1.0. That makes it the most sensitive reconstruction methods, but also the best. They can only unfold their full potential when the utilized detector images look as similar as possible to the training data [137]. A further improvement can be expected when the detector technology advances and is able to provide crystals with fewer defects and a more homogeneous sensitivity.

6.4.2 Reconstruction Method

The most commonly used reconstruction method, MURA Decoding, proved to be good compromise between reconstruction quality and runtime, despite its negligence of Poisson noise. The Wiener Filter with an optimal hyperparameter (SNR) yields reconstructions of comparable quality but has the advantage of a more than four times faster runtime. This can be attributed to the convolutions

that are not carried out in the spatial domain but in the Fourier domain. This acceleration indicates that MURA Decoding should be carried out in the Fourier domain as well, since, compared to the Wiener Filter, no hyperparameter search is required.

The MLEM algorithm offers on average a 1.2 times better reconstruction quality on the SRP than MURA Decoding which comes with a higher computational cost: it takes more than 170 times longer for a single reconstruction compared to MURA Decoding. Each iteration consists of two convolutions and two element-wise operations, one division and one multiplication, where the reconstruction is updated based on the current forward projection. The potential of the MLEM algorithm is certainly not fully exploited yet. It is conceivable to accelerate the MLEM algorithm, for example, a warm-start based on MURA Decoding can be considered [132, 138], additional regularization [139], or an adaptive stopping criterion [140].

Overall, both CEDs reconstructed the phantoms better than all analytical reconstruction methods. Especially the SRP is reconstructed about 1.4 times better by the CED-IN compared to MURA Decoding and 2.7 times better by the CED-Lines. This is achieved, even though the simulated training data are based on the simplistic Convolutional Model (see Sec. 3.3.2) and did not account for any near-field effects. Neglecting near-field effects, transmission and scattering inherently limit the CED's reconstruction ability. It can be expected, that a more realistic generation of training data has the potential to improve CAI reconstruction. The runtime of both CEDs are identical with circa 286 ms and, thus, equal to MURA Decoding with 288 ms.

The fact that the data-driven reconstruction methods produce reconstructions with similar characteristic (e. g. the highlighted tip in the top image of the CP in Fig. 6.8) as all analytical reconstruction methods supports the hypothesis, that the CEDs learned an actual image-to-image mapping and did not just repeat known patterns. The data-driven CEDs, unlike analytical reconstruction methods, are not based on a formal description of the imaging system. Two problems arise from this: First, it has been shown, that CNNs can inherently possess instabilities [138, 141]: For CAI, instabilities mean that perturbations of the detector image can lead to nonsensical reconstructions, e.g. reconstructions with large uniform areas or negative pixel values. However, instabilities were only noticed during training and none of the final deployed networks showed instable behavior, but its possibility must be kept in mind, especially because

the application in RGS can have serious health consequences. Therefore, a more in depth analysis of the CED regarding its trustworthiness and the influence of training data is required. Second, the reconstruction results depend on the domain of the training data. The different reconstruction results between CED-IN and CED-Lines are solely caused by the different image domains used for training. This data dependency has already been described by Kulow et al. [50] and indicates, that ML approaches are not capable yet, to learn a domain-independent image-to-image mapping. The superior performance of the CED-Lines indicate, that the domain shift between the Lines dataset and the target domain (the hot-rod phantom data) is smaller compared to the Imagenet training set. This can be used as an advantage: The two training sets used in this paper represent different approaches to CAI reconstruction. First, using natural photographs as source distribution forms a general approach aiming at the question "What do we have in front of our camera?". Second, when the approximate amount, shape and size of radioactive sources is known, a training set can be tailored to that specific application. This approach contains a priori knowledge and seeks to answers the question "Where is the source located?". For Sentinel Lymph Node Biopsy (SLNB) the radiating sources can be considered of spherical shape with maximally a few centimeters in diameter [82, 142] and a set of training data could be generated for this specific task.

With respect to the state of art, the presented quantitative comparison provides researchers with an overview of the existing reconstruction methods for planar CAI. Additionally, research groups are encouraged to use the published Rozhkov dataset with its three hot-rod phantoms for evaluating and developing their own reconstruction methods. This study also gives concrete information about the advantages and disadvantages of the most common reconstruction methods: In summary, MURA Decoding balances quality and speed, the Wiener Filter is faster, MLEM provides superior quality but is slow, and the data-driven CEDs outperform analytical techniques but require further analysis regarding the domain shift and its reliability.

6.5 Conclusion & Outlook

In this chapter common reconstruction methods for planar CAI on data acquired by an experimental gamma camera and three hot-rod phantoms were compared. For the given set-up, MURA Decoding, the most commonly used CAI reconstruction method, provides robust reconstructions despite the assumption of a linear system model. The Wiener Filter delivers results of almost similar quality while requiring only a quarter of the computational time. The MLEM algorithm yields the best reconstruction quality among the analytical reconstruction methods, but comes with a 170 times longer runtime with the used hyperparameters. All in all, both developed CEDs outperformed the analytical reconstruction methods for two out of three phantoms, despite a simple and low-fidelity simulation of training data based on the Convolutional Model. The ML approach turned out to be especially successful, when a priori knowledge about the expected radiation sources in the form of a tailored training set is used. Even though, none of the CED reconstructions exhibited instabilities, its possibility must be kept in mind.

The main conclusions from this comparative study with regards to using CAI for Intraoperative Gamma Camera (IGC) can be summarized as follows: First, preprocessing the detector images to mitigate the impact of defective pixels might be beneficial and should be selected based on the reconstruction method used. Second, even though the actual imaging is carried out with a NTHT MURA pattern, it is beneficial to use its THT version for the reconstruction process. Third, the Wiener Filter is much faster than MURA Decoding and delivers the same quality. Therefore, it is advisable to transfer the convolution of MURA Decoding to the Fourier domain to obtain a fast and robust method without the need of adjusting hyperparameters. Fourth, the MLEM algorithm offers superior quality, but takes multiple seconds for what the Wiener Filter needs milliseconds. This might make the MLEM a more suitable reconstruction method for SPECT imaging where the requirements regarding the runtime is of minor priority. Fifth, CEDs offer both high quality reconstructions and a fast runtime. However, the unknown reliability and the dependency on the training domain of the training data must be further investigated.

Chapter 7

Simulation Study on Super-Resolution for Planar Coded Aperture Gamma Imaging

This chapter investigates Hypothesis Ib and aims at answering the question if and to what extend super-resolution imaging is possible in planar Coded Aperture Imaging (CAI).

This chapter has been accepted but not yet published at a conference proceedings to the "4EU+ International Workshop on Recent Advancements in Artificial Intelligence (4EU+IWAI)" where a preprint is publicly available under license CC-BY 4.0 on the arXiv [143].

7.1 State of the Art

The term super-resolution refers to the process of combining several "low resolution, noisy, slightly shifted observations" [144] to reconstruct an image of the underlying high resolution scene, as Fig. 7.1 illustrates. With regards to CAI, Kulow et al.[23] argues that because multiple projections from slightly different

Figure 7.1: The process of combining two or more low-resolution or noisy images to form an high-resolution image of the underlying scene is called super-resolution. Figure adapted from [143].

angles are captured by a coded mask, super-resolution is theoretically feasible. In practice, this means producing a high-resolution image reconstruction from a low-resolution detector image. Because the spatial resolution in CAI is mainly influenced by the mask's pinhole diameter [23], increasing the MURA rank and thus the amount of pinholes while reducing their diameter would increase the spatial resolution. So far, the pinhole diameter has been chosen such that the utilized detector can properly sample the resulting PSF [25, 82]. To the best of the author's knowledge, no research has been conducted about the combination of small pinholes and a low-resolution detector. Therefore, the investigated hypothesis of this chapter is as follows: Existing CAI reconstruction methods are capable of reconstructing point sources from an undersampling detector, and thus achieving super-resolution, at reasonable quality even though the detector is unable to resolve the higher spatial resolution of the aperture. This is due to the shifted but overlapping projections caused by the coded aperture. Planar reconstruction methods from the previous chapter (MURA Decoding, MLEM, and the CED-IN) are investigated and their results compared. With this chapter, the following contributions to the state of the art are made:

1. It is shown that in theory super-resolution for planar CAI is feasible, as long as the sampling theorem is fulfilled.
2. Smoothing the captured detector images can improve the CED's prediction.

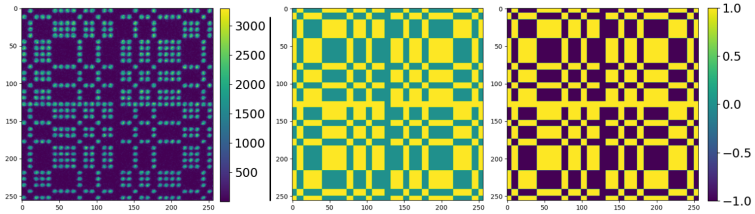


Figure 7.2: From left to right: The portion of the Point Spread Function (PSF) captured by the experimental gamma camera from the Rozhkov dataset, the Two-Holes-Touching (THT) encoding and the THT decoding pattern used for MURA Decoding. Figure from [143].

7.2 Methods & Material

First, the analyzed reconstruction methods are briefly summarized. Afterwards, an under-sampling detector of a test image is simulated with a state-of-the-art Monte Carlo (MC) simulation framework. Thereafter, the experimental data are shortly summarized, and the procedure of simulating a low-resolution detector image from a high-resolution image is explained. Finally, the concept of the critical super-resolution factor is introduced.

7.2.1 Reconstruction Methods

Three different methods for super-resolution reconstruction are analyzed and compared in this paper: MURA Decoding [75], a convolutional Maximum Likelihood Expectation Maximization algorithm (MLEM) [24] and a convolutional encoder-decoder network (CED) from previous work [116]. Instead of the real Point Spread Function (PSF) $h[x,y]$ with round pinhole projections, the Two-Holes-Touching (THT) version of the PSF without gaps between neighboring pinholes is used for reconstruction since it suppresses periodical noise (see Sec. 6.3). Both the THT-PSF and its corresponding decoding pattern, scaled according to Eq. 3.8, are of rectangular structure and a square of 8 bright pixels represents each projected pinhole. They also define the reconstructed resolution of 256×256 pixels. Figure 7.2 depicts the measured PSF (only the central portion that fits the detector) , the THT-PSF and the decoding pattern. Additionally, MURA Decoding with low-resolution THT-PSF was implemented, where

not the synthetic high-resolution THT-PSF was used, but the down-sampled THT-PSFs mimicking how the PSF would be sampled with a low-resolution detector. Since the detector images come in low-resolution, the reconstructed images were upsampled to 256×256 pixels by bilinear interpolation before the actual reconstruction step. The reconstruction quality was measured with the Contrast-to-Noise Ratio (CNR) in analogy to the previous experiment where details can be found in Sec. 6.2.3. For the MLEM algorithm, a fixed number of 25 iterations was heuristically found to be a good compromise between speed and quality.

7.2.2 Simulating the Test Image

For simulating a test image, the MC simulation toolkit TOPAS [88], a wrapper library of the popular Geant4 software package [87], is adopted. As described above in Sec. 3.3.3, MC simulations account for photon-mass interactions like scattering and mask penetration and are therefore considered the gold standard for simulating gamma cameras [36]. The geometrical components and dimensions were simulated according to the experimental gamma camera from the Rozhkov dataset where the exact specifications can be found in Tab. 5.1. The test image consists of three circular sources with diameters d_1, d_2 and d_3 of 1, 2 and 3 mm distributed within the Field of View (FOV) as Fig. 7.5 shows. 10^9 gamma photons with a photon energy of 140.5 keV (corresponding to the photon peak of ^{99m}Tc the most commonly used radiotracer in nuclear medicine [12]) were distributed to the three sources according to their area. Every photon penetrating the front plane of the detector was registered and stored in a so called *phase space file*.

In addition to the coded aperture a single pinhole collimator, as described in Sec. 2.3, with the same pinhole diameter was simulated to serve as a reference for the reconstructed images. The captured pinhole image was smoothed by Gaussian blurring with a σ of 2 pixels as is standard in gamma imaging [145, 146]. The Ground Truth (GT) image was generated from the geometrical model and remains binary: 1 for where a source is located and 0 everywhere else.

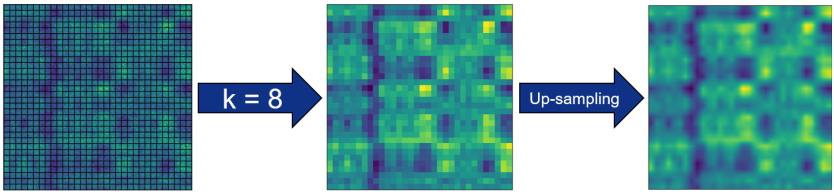


Figure 7.3: Pixels of the high-resolution detector image from the TOPAS simulation are accumulated (here with $k = 8$ into 32×32 pixels) to form the low-resolution detector image. Afterwards this image is upsampled by bilinear interpolation to the high-resolution of 256×256 pixels.

7.2.3 Experimental Data

Images from an experimental gamma camera were used to validate the effect of super-resolution on real-world experimental data. From the Rozhkov dataset (see Sec. 5.1) the 120 images captured of the Spatial Resolution Phantom (SRP) were used. The same preprocessing as described in Sec. 6.2.4 was applied to the detector images.

7.2.4 Simulating Low-Resolution Detector Images

To analyze the effect of different pixel sizes, low-resolution images of different resolutions were generated. The captured photons from the MC simulation and from the experimental gamma camera were binned into images of different low resolution. The actual detector served as reference with a resolution of 256×256 pixels, which corresponds to the resolution of the final reconstructed image. Therefore, the super-resolution factor k is introduced. This factor represents how many pixels are pooled to a single low-resolution pixel, i.e. $k \times k$ high-resolution pixels are reconstructed from a single low-resolution pixel. Note that the absolute detector size with the side length D_D of 14.1 mm remains the same. Only the size of each pixel s changes proportional to k with $s = 14.1 \cdot k/256$. Effectively, this down-sampling in combination with the following upsampling in the reconstruction step can be regarded as a low-pass filtering of the detector images. The process of generating low-resolution images is shown in Fig. 7.3 for $k = 8$.

7.2.5 Critical Super-Resolution Factors

The Nyquist-Shannon sampling theorem states that the sampling frequency of a pixelated representation must be larger than twice the maximum frequency of the image. Thus, when the smallest occurring structure is sampled by two pixels or less, an image is not represented unambiguously anymore, which leads to aliasing and hence signal degradation [131]. Since the aforementioned analytical reconstruction methods, MURA Decoding and MLEM, consist of one or more convolutions of two discretized signals, a reconstruction without aliasing artifacts is only possible when both images were sampled by enough pixels. Thus, critical super-resolution factors $\tilde{k}$ were determined both for the coded aperture test image and the THT-PSF $h[x,y]$. The smallest point source of the test image is 1 mm wide and therefore much larger than the pinhole diameter: $d_1 \gg d$. Hence, the smallest structure on the detector caused by the small point source can be approximated by $t = d_1 \cdot M = 1.244\,\text{mm}$ with the magnification factor M from Eq. 3.2. For $h[x,y]$ the smallest structure t is 8 pixels wide, i.e. $t = 8 \cdot s$. All in all, for the given gamma camera with its magnification factor of $M = (1 + b/z) \cong 1.244$, the smallest depicted structure t and the single pixel side length of $s = 0.0551\,\text{mm}$ $\tilde{k}$ is defined as

$$\tilde{k} = \left\lfloor \frac{1}{2} \cdot \frac{t}{s} \right\rfloor , \tag{7.1}$$

where $\lfloor \cdot \rfloor$ denotes rounding off to the nearest smallest integer value.

7.3 Results

The results of this study are presented as following: First, the critical super-resolution factors are given. Second, the reconstructions of the rest image and third, the reconstructions of the captured images are presented.

7.3.1 Critical Super-Resolution Factors

The following critical super-resolution factors $\tilde{k}$ were obtained from the Nyquist-Shannon sampling theorem: The THT-PSF $h[x,y]$ must be sampled by at least

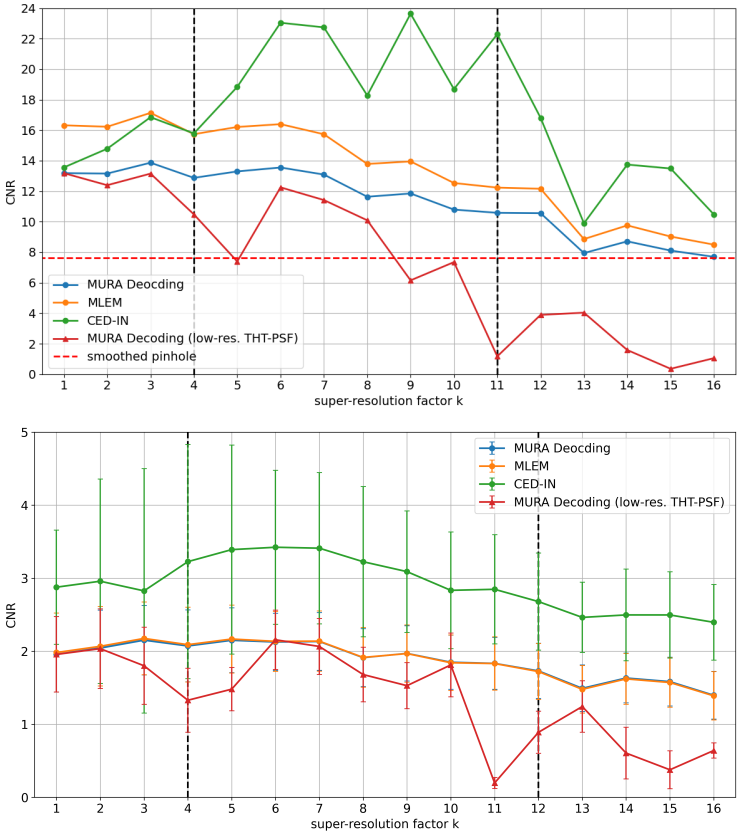


Figure 7.4: The CNR plotted against the super-resolution factor k. Top: Test image. The black dotted vertical lines mark the critical super-resolution factors $\tilde{k}_{\text{THT-PSF}} = 4$ and $\tilde{k}_1 = 11$. The red dotted line represents the CNR of the smoothed image captured by a pinhole collimator and serves as reference. Bottom: Measured phantom data. Error bars represent the standard deviation at each data point. The vertical dotted lines denote the critical super-resolution factors $\tilde{k}_{\text{THT-PSF}} = 4$ and $\tilde{k}_{\text{measurement}} = 12$.

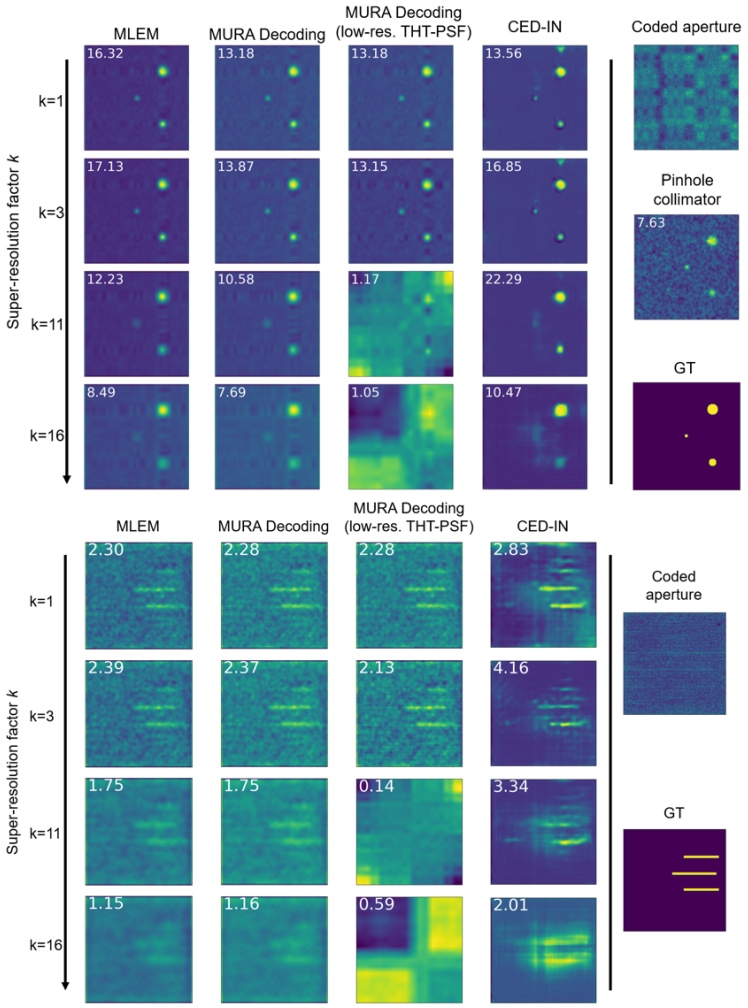


Figure 7.5: On the right-hand side the coded aperture detector and the GT images are shown The CNR are printed in the top right corner of each reconstruction. Top: Exemplary reconstructions of the test image generated by MC simulation at different super-resolution factors k. For reference, the coded aperture simulation in 256×256 pixels, the smoothed pinhole collimator simulation in the same resolution and the GT is shown on the right-hand side. Bottom: super-resolution evaluated on the SRP data captured with our experimental gamma camera.

64×64 pixels leading to $\tilde{k}_{\text{THT-PSF}} = 4$. This means that the synthetic high-resolution THT-PSF in this paper is 4-times oversampled. The test image with the 1 mm point source results in a higher critical super-resolution factor of $\tilde{k}_1 = \lfloor 11.29 \rfloor = 11$. The tubes of the phantom captured by the experimental gamma-camera have a diameter of 1.1 mm, that are magnified to approximately 1.37 mm and thus 24.83 pixels. This results in a critical super-resolution factor for the measured data of $\tilde{k}_{\text{measurement}} = \lfloor 12.42 \rfloor = 12$.

7.3.2 Results on the Test Image

Figure 7.4 shows the CNR of the four reconstruction methods over the super-resolution factor k. The red dotted line at 7.63 denotes the CNR of the smoothed image captured with a pinhole collimator where no reconstruction was required. It is depicted central on the right-hand side together with the GT image and the coded aperture test image in Figure 7.5. The left-hand side shows exemplary reconstructions for $k = 1, 3, 6, 11$ and 16. It can be seen that the CNRs of the reconstruction methods with the synthetic high-resolution THT-PSF increase until $k = 3$ and decline afterwards. The CED-IN is an exception, where the CNR increases further until falling below its baseline at $k = 13$. For all reconstruction methods using the synthetic high-resolution THT-PSF the smallest point source starts to disappear for $k > 11$ and is hardly visible for $k = 16$.
MURA Decoding with the low-resolution THT-PSF does not exceed its baseline CNR and falls beneath the pinhole reference at $k = 5$ and again for all $k \geq 9$. The reconstructions for $k \geq 11$ fail entirely and resemble no similarity to the GT anymore.

7.3.3 Results on the Measured Images

Analogously to the test image, Fig. 7.4 shows the CNR for the presented reconstruction method at different super-resolution factors k. Because 120 reconstructions were analyzed, the marker represents the median CNR and additional error bars represent the standard deviation for each reconstruction method and super-resolution factor k. In general, lower CNRs compared to the test image can be observed. Similarly, all methods except for the MURA Decoding with low-resolution THT-PSF slightly rise for small k and then fall

after approximately $k = 7$. This behavior can also be seen in the exemplary images in Fig. 7.5. Visually, a higher background noise is present compared to the test image and the three line sources are prominent until they start to disappear for larger k. The maximum median CNR is reached by the CED-IN with 3.42 at super-resolution factor $k = 6$.

7.4 Discussion

At first, the results from the test image will be discussed, before the experimental data and general limitations of this experiment are examined.

7.4.1 Simulated Test Image

For the given setup the Nyquist-Shannon sampling theorem states, that for super-resolution factors above $k = 4$ the PSF is not sufficiently sampled anymore. If the THT-PSF is undersampled it loses its characteristic to properly function for CAI reconstruction. The simulation study of this paper shows this behavior where CNRs of MURA Decoding with low-resolution THT-PSF drop notably for $k \geq \tilde{k}_{\text{THT-PSF}}$ and the reconstruction of $k = 11$ show major artifacts rendering the three sources unrecognizably.

But, when upsampling the low-resolution detector image to a high-resolution of 256×256 pixels and using the synthetic high-resolution THT-PSF for reconstructions, the CNRs drop slower for both analytical reconstruction methods. For $k \leq 6$ the CNRs even stay approximately constant. Additionally, the reconstructed images are visually closer to the expected output. For $k > \tilde{k}_1$ the smallest point source cannot be reconstructed properly and starts to dissolve which is in accordance to the prediction by the sampling theorem. In the simulation study super-resolution with small k and upsampling by bilinear interpolation even had a positive impact on the CNR and the reconstructions look smoother.

7.4.2 Measured Phantom Images

Similar behavior was observed for the measured phantom data. For all reconstruction methods, the CNR does not decrease until approximately $k = 7$, even

though undersampling of the THT-PSF starts at $k = 4$. However, the gain in CNR for the measured phantom data is far less for small k compared to the simulated test image.

It is remarkable that the CED-IN, even though not trained on processing upsampled low-resolution images, performs better than all other reconstruction methods for super-resolution factors of $k \geq 4$. For the measured phantom data it is the best reconstruction method for all k. This indicates that the CED-IN generalized from the training domain of natural photographs to discrete sources on a dark background, even when the input images where smoothed by the upsampling process. Specifically the background is reconstructed more uniform. This implies that the CED-IN has learned to compress the input detector image into a robust representation of the image, suppressing the influence of noise and defective pixels, like in the top right corner of the coded aperture image in Fig. 7.5.

7.4.3 Limitations

The test image generated by MC simulation shows in theory that super-resolution in CAI is possible and the simulated low-resolution detector images based on phantom data captured by an experimental gamma-camera strengthen this hypothesis. However, the question remains as to how a real low-resolution detector would affect the reconstruction. In this study low-resolution detectors were simulated by accumulating photon hits of neighboring pixels, but defective pixels on a low-resolution detector will have a higher impact since it captures a larger fraction of the mask's shadow. Furthermore, the critical super-resolution factor also depends on the source-to-mask distance z and thus on the magnification factor M from Eq. 3.2. Hence, additional experiments for 3D imaging are necessary. Another aspect not investigated in this study are other types of gamma sources. Especially extended sources are known to cause problems in CAI for reasons that are still unknown [129, 147].

7.5 Conclusion

The conducted simulation study indicates that super-resolution reconstruction for planar CAI is feasible even when the detector is not capable of sampling the PSF sufficiently according to the Nyquist-Shannon sampling theorem. Instead, upsampling the captured low-resolution detector image by bilinear interpolation is combined with a synthetic high-resolution THT-PSF. This way, established reconstruction methods were able to reconstruct the simulated test image. As a byproduct of our investigation, we discovered that the smoothing of the detector image can enhance the reconstruction quality of the CED-IN. However, for large super-resolution factors, the smallest point source could not be reconstructed as predicted by the sampling theorem. Applying the same technique to simulated low-resolution detector images from data of a hot-rod phantom captured with an experimental gamma-camera strengthen these findings. For future research, further experiments with a more realistic undersampling detector including defective pixels are required.

PART IV

3D IMAGING

Chapter 8

Assessment of the Axial Resolution

This chapter and the following one address Hypothesis II, where the assumption of a known source-to-mask distance is dropped. In analogy to the spatial resolution in x and y-direction, in this chapter the axial resolution of our experimental Intraoperative Gamma Camera (IGC) is assessed.

The content of this chapter is taken from a research paper that has been published in the "European Journal of Nuclear Medicine and Molecular Imaging Physics" under license CC-BY 4.0 [118].

8.1 State of the Art

Planar image reconstruction requires to select an *in-focus plane*, i.e. a distance at which the source is assumed to be located. The choice is either based on the assumption that the source is infinitely far away, as is the case in far-field applications such as space astronomy, or on a priori knowledge. In principle, this represents an inherent limitation of Coded Aperture Imaging (CAI) since other collimation technologies do not require selecting an in-focus plane. However, this can also be used to extract 3D information and has already been investigated for point sources [25, 95]. As described in the Fundamentals (see Sec. 3.1),

the lateral position of a point source is encoded by the shift of the mask's shadow, while the source-to-mask distance is related to the size of the projected shadow. The size of the projected shadow depends on the magnification factor M from Eq. 3.2 that, in turn, depends on the detector-to-mask distance b and the source-to-mask distance z. Hence, in theory, the 3D-localization of a point source based on a single image is feasible.

Accurately estimating the 3D position of gamma sources can substantially improve the effectiveness of Radioguided Surgery (RGS). Procedures such as Sentinel Lymph Node Biopsy (SLNB) for breast cancer staging highlight the critical need for detailed examination of the axial resolution in CAI [6, 8, 20]. So far, two reconstruction algorithms have been proposed for 3D CAI reconstruction: MURA Decoding and 3D convolution-based Maximum Likelihood Expectation Maximization algorithm (3D-MLEM). The first method retrieves a 3D reconstruction by a sequence of images with in-focus planes at multiple distances. With this method Russo et al. [25] were able to show that a ring-shaped object is captured with an axial resolution of about 3 mm. However, no systematic analysis of the axial resolution at different source-to-mask distances has been carried out. The second method, the 3D-MLEM algorithm, extends the MLEM algorithm already used in Chapter 6 to reconstruct an entire 3D source distribution simultaneously. Mu et al. [24, 129] demonstrated that this algorithm is capable to distinguish two sources positioned in succession. However, the axial resolution is not reported and the 3D-MLEM algorithm is not yet suitable for a more general camera setup.

All in all, in contrast to the extensive investigations into the lateral resolution of CAI [26, 36, 73], the axial resolution has received much less attention. Only a few articles exist, covering only a limited range of source-to-mask distances by different methodologies. Thus, the experiments of this chapter aim at closing this gap and make the following contributions to the state of the art:

1. A systematic experiment and assessment of the axial resolution of an experimental IGC equipped with a coded aperture collimator is presented.
2. We propose a reproducible method for measuring the axial resolution by calculating the FWHM of the CNR profile along the z-axis of a point-like source.
3. The 3D-MLEM algorithm from [24] is generalized to a broader camera setup by extending it by a normalization factor and mask transmission.

4. This study compares two coded aperture reconstruction methods and demonstrates that 3D-MLEM can be considered as the slower but superior reconstruction method compared to standard MURA Decoding which is faster but less precise.

The entire acquired dataset of 21 images and its preprocessed versions are publicly available at `https://zenodo.org/doi/10.5281/zenodo.8315861`.

8.2 Methods & Material

In this study we use the Axial Resolution Dataset that is described in detail in Sec. 5.2. In the following section, the preprocessing of the detector images is described, followed by the proposed modifications to the 3D-MLEM algorithm. Finally, the assessment of the axial and the lateral resolution are described.

8.2.1 Preprocessing

The detector images that are used in this study are captured with a detector based on Si instead of CdTe like in the previous chapters. Because this Si detector is thinner than Rozhkov's CdTe detector (0.5 mm instead of 2 mm) and Si is generally less prone to clusters of defective pixels [44], a less radical preprocessing step is used. Per detector image, all pixels with values outside the range of the 1^{st} and 99^{th} percentile were considered to be outliers and were replaced by the median value of their 3×3 neighborhood. Additionally, Gaussian smoothing with a sigma of 1 pixel was applied. Figure 8.1b shows a comparison between a raw and a preprocessed detector image.

8.2.2 Generalizing the 3D MLEM Algorithm

The Maximum Likelihood Expectation Maximization (MLEM) algorithm is an iterative algorithm that estimates the source distribution with the highest likelihood assuming the measured detected photons follow a random Poisson process [148]. The original MLEM algorithm was adapted to CAI by replacing

the computationally excessive system matrix with a convolutional approach [24], that is explained in more detail in Sec. 6.2.1. Additionally to this planar reconstruction method, Mu et al. [24] extended it to reconstruct planes at several source-to-mask distances simultaneously, hence, generating an entire 3D source distribution. The entire gamma camera can solely be defined by its Point Spread Function (PSF) $h_z[x,y]$, where the subscript z denotes the mask-to-source distance. The algorithm for the $(\mathrm{k}+1)^{\mathrm{th}}$ iteration of reconstructing the source distribution $\hat{f}_z^{(\mathrm{k}+1)}[x,y]$ at a source-to-mask distance z is given by

$$\hat{f}_z^{(\mathrm{k}+1)} = \frac{\hat{f}_z^{(\mathrm{k})}}{\sum_z h_z} \cdot \left[\frac{p - \sum\limits_{\forall z' \neq z} \hat{f}_{z'}^{(\mathrm{k})} * h_{z'}}{\hat{f}_z^{(\mathrm{k})} * h_z} \times h_z \right] \tag{8.1}$$

where "$*$" represents the linear 2D convolution, "$\times$" the 2D correlation and "$\cdot$" the point-wise multiplication. The given detector image is denoted as $p[x,y]$ and for a better readability the lateral coordinates $[x,y]$ have been omitted.

However, two conditions are not taken into account by the 3D-MLEM algorithm: first, a mask that is smaller in dimension than the detector and second, gamma photons that transmit through the mask. As these two aspects were present in working with our experimental setup, we extended the 3D-MLEM algorithm by two modifications: a more general normalization term s_z that accounts for the size difference between the mask and the detector and a forward simulation function $\mathbf{F}$ which integrates mask transmission.

8.2.2.1 Normalization

The 3D-MLEM formula is derived from the general MLEM formula in its vectorized form with the system matrix $\mathbf{A}$. The entries a_{ij} of $\mathbf{A}$ denote the probability that photons from source j are detected in detector pixel i. In the literature the following normalization term, also referred to as sensitivity, can be found [148]:

$$s = \sum_{i=0}^{m} \mathbf{A}_{ij} \tag{8.2}$$

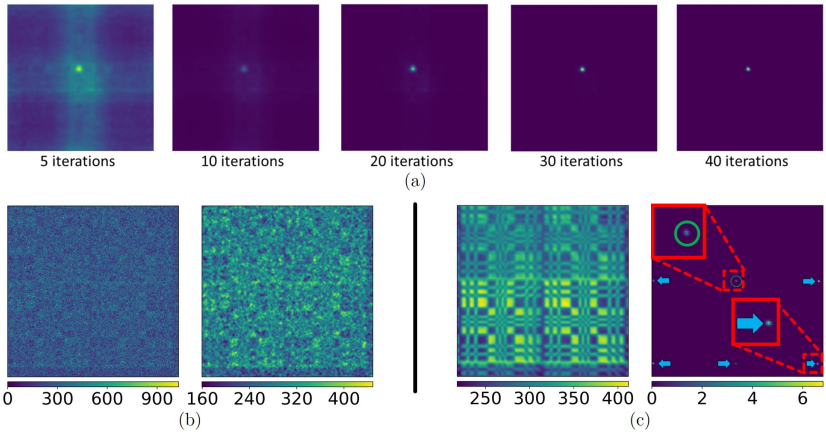


Figure 8.1: (a): Successive number of iterations for the 3D-MLEM algorithm applied to the source and reconstructed at 30 mm. (b): Raw (left) and preprocessed detector image (right). (c): The left image shows the forward projection of 3D-MLEM for the source at 40 mm and the right image shows the reconstruction at 40 mm. Due to the 2×2 arrangement of the basic MURA pattern, multiple ghost sources (blue arrows) along the image border emerge in addition to the true source (green circle) from the 3D-MLEM algorithm for sources that are more than 40 mm away. Figure from [118].

This summation over the matrix columns represent the summed likelihood that the photon from source j is detected by any pixel of the detector, i.e. by the entire detector. s is a vector and is generally not a multiple of an all-one vector.

To illustrate, when the coded aperture mask is smaller than the detector, photons emitted from a source positioned further away from the center have a greater probability of passing through the pinholes but missing the detector. This can be visualized by considering the shadow cast by such a source, where only parts of the shadow fall on the active area of the detector. The further off-center the source is, the higher the part of the shadow that is not hitting the detector area and the smaller the summed likelihood s becomes. The normalization factor s ensures that the inherent forward projection has approximately as many detected photons as the given detector image $p[x,y]$.

Transferred to the convolution-based 3D-MLEM algorithm the normalization factor s is an image that depends on the source-to-mask distance z: $s_z[x,y]$. It can be calculated in advance by a backward projection of an entirely illuminated detector to the in-focus plane. Thus, we obtain $s_z[x,y]$ by calculating

the cross-correlation between an all-one image ($\mathbf{1}$) and the PSF at the given distance z:

$$s_z[x,y] = h_z[x,y] \times \mathbf{1} \tag{8.3}$$

8.2.2.2 Forward Simulation with Transmission

Transmission noise emerges from photons that penetrate the mask and is here approximated as uniform background noise proportional to the transmission probability T of the coded aperture mask. With a mask thickness of 0.11 mm, a 59.5 keV source (see Sec. 5.2) and Eq. 3.22 the transmission probability T of our setup is approximately 46%, meaning that about half of the photons pass through the mask. Because this transmission is high, the forward projection $\mathbf{F}$ (2nd denominator of Eq. 8.1) of the 3D-MLEM algorithm must be adapted. The projected image becomes a weighted superposition of the projection of the reconstruction $\hat{f}_z^{(\mathrm{k})}$ and a uniform transmission noise image. The weight is selected according to the transmission rate T and the sum of emitted photons from the in-focus plane $\hat{f}_z^{(\mathrm{k})}$ in accordance with Eq. 3.23:

$$\mathbf{F}\left(\hat{f}_z^{(\mathrm{k})}, h_z\right) = (1-T)\left(\hat{f}_z^{(\mathrm{k})} * h_z\right) + T \sum_{x,y} \hat{f}_z^{(\mathrm{k})}. \tag{8.4}$$

As explained in Sec. 6.3, we do not use the NTHT-PSF but the Two-Holes-Touching (THT) pattern to avoid periodical background noise. In contrast to MURA Decoding, we decided to use the entire detector image for the reconstruction. This choice ensures that the output maintains a consistent size of 256×256 pixels. Because MLEM is based on the convolutional model, the Field of View (FOV) is equal to that of a single pinhole collimator. All in all, the proposed 3D-MLEM algorithm can be summarized as follows:

$$\hat{f}_z^{(\mathrm{k}+1)} = \frac{\hat{f}_z^{(\mathrm{k})}}{s_z} \cdot \left[\frac{p - \sum_{\forall z' \neq z} \mathbf{F}\left(\hat{f}_{z'}^{(\mathrm{k})}, h_{z'}\right)}{\mathbf{F}\left(\hat{f}_z^{(\mathrm{k})}, h_z\right)} \times h_z \right]. \tag{8.5}$$

From visual inspections we found 40 iterations to be a good compromise between noise amplification and reconstruction quality, which can be seen in Fig. 8.1a. The 3D-MLEM algorithm was implemented in Python (3.8) using

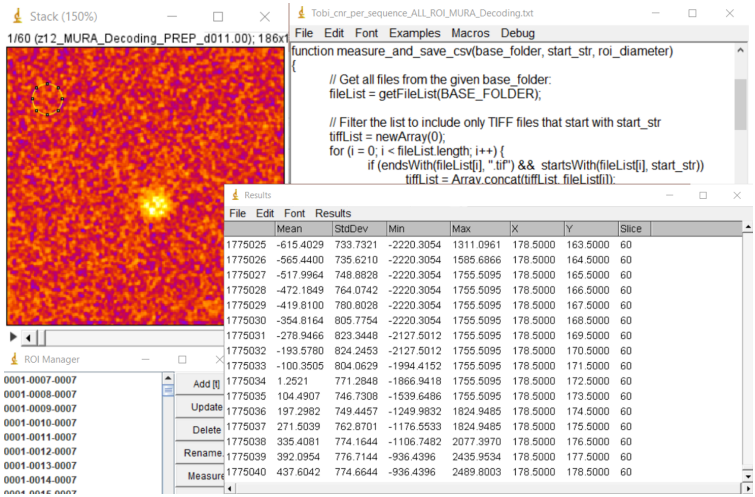


Figure 8.2: The semi-automatic CNR algorithm as a macro in imageJ. Clockwise from the top left to bottom left: the image stack that is analyzed, the macro that performs the selection of all possible ROIs, the results of analyzing all ROIs including the position (column X and Y), and the mean and standard deviation (Mean and StdDev), and the ROI manager that is controlled by the macro. The CNR profile is extracted with a Python script based on the tabular data, as explained in the text.

the NumPy (1.24) library and, similarly to MURA Decoding, all convolutions are performed in the Fourier domain.

8.2.3 Contrast-to-Noise Ratio

The profile of the Contrast-to-Noise Ratio (CNR) along the z-direction was used to determine the axial resolution as the Full Width at Half Maximum (FWHM) of this bell-shaped curve. The same definition of the CNR as in Sec. 6.2.3 was employed:

$$CNR = \frac{\bar{S} - \bar{B}}{\sigma_B}, \tag{8.6}$$

where $\bar{S}$ denotes the average pixel intensity of the circular Region of Interest (ROI) S that represents the signal, and $\bar{B}$ and σ_B the mean intensity and the

standard deviation of the background B which acts as a representative of the image noise.

For the following reasons the same approach to compute the CNR of the previous chapters could not be deployed: First, different from the previous chapters, we do not have complete certainty about the source position in lateral directions. During the acquisition (see Sec. 5.2) we tried to keep the source central, but a small deviation is present. Thus, we do not have a full Ground Truth (GT). Second, we do not require one metric per 3D reconstruction, but rather a quantification of how visible the source is in each slice of the 3D reconstruction along the z-axis. Third, processing each image individually and separating the signal region from the background region by hand is tedious, prone to fluctuations and, hence, not reproducible.

A semi-automatic algorithm was developed and implemented as an `imageJ` macro that samples ROIs over the entire image and, thus, eliminates the need for choosing regions for S and B manually. The only required input of the algorithm is the reconstructed slice and the ROI diameter in pixels. Depending on the image size and ROI diameter there were between approximately 14,000 and 31,000, and 54,000 and 65,000 ROIs per image. A screenshot of the procedure in imageJ is depicted in Fig. 8.2. For each ROI, its position in pixels, the mean intensity, and standard deviation were calculated and stored as a table in a text file. To derive the CNR from this table a Python script is used to select the regions S and B. The ROI with the highest mean intensity is selected as the signal S and $\bar{S}$ becomes the mean intensity. A constraint was introduced that restricted the signal ROI to be in the inner 50% of the image area to avoid measuring one of the ghost sources along the image border as visible in Fig. 8.1c. Afterwards, ROIs that overlapped with S were discarded from further processing. All remaining ROIs are considered as background B and $\bar{B}$ was computed as the average intensity of all background ROIs. The same was performed for the standard deviation $\bar{\sigma}_B$. To summarize, the axial resolution is determined based on the CNR profile along the z-direction by employing the combination of a semi-automatic algorithm in ImageJ and a Python script to ensure reproducibility and accuracy.

8.2.4 Assessing the Axial Resolution

The axial resolution expresses how well a point-like source can be localized along the z-axis, i.e. in depth direction. We use the profile of the CNR along the z-direction to determine the axial resolution as the FWHM of this Gaussian-like curve: this provides a more intuitive understanding of the spatial resolution and takes into account not just the source intensity - as in [73], where the pixel intensity of the source was used to compute the axial spatial resolution - but also how well the source is distinguishable from the background. Usually, a point source is used for measuring the spatial resolution. Consequently, the smallest source that was available for use was selected: a point-like ^{241}Am source with a nominal diameter of 1 mm and a previously measured FWHM of 0.65 mm [119].

The foundation for the assessment was the 3D reconstruction. First, we reconstructed each source image within a broad range from 5 to 100 mm in 5 mm steps to roughly locate the source. Second, images within a tighter z range containing the actual source position in 0.5 mm steps were reconstructed resulting in sets of images ranging from 60 to 240 images for MURA Decoding and 54 to 101 for 3D-MLEM. From here on, we will refer to a set of reconstructions of the same detector image at different depths as an *image stack*.

For MURA Decoding, the change of image sizes along the stack resulted in sharp jumps of the CNR which heavily impeded the following fitting procedure. Thus, the reconstructions from MURA Decoding were resized by bilinear interpolation to the image size of the reconstruction at the true source position. The impact of this resizing on the final axial resolution was found to be negligible for source positions where the fitting worked. The axial resolution for both the resized and the non-resized stacks for the source placed at 30 mm were 11.9 ± 0.5 mm and 12.5 ± 0.5 mm, respectively. Because they are compatible within the errors, we decided to continue our analysis on the resized stacks. No resizing was required for 3D-MLEM images, as the algorithm returns images of a fixed size.

The CNR profile was obtained in the following manner: the diameter of the ROI was chosen separately for each stack, i.e. each captured detector image, based on the true source size (see Sec. 5.2) and the true source distance. First, the FWHM source diameter of 0.65 mm was converted to pixels with

respect to the FOV at the true source distance which was then rounded to the nearest integer to obtain the ROI diameter. The FOV of MURA Decoding was calculated according to Eq. 3.14 and for 3D-MLEM the FOV of a pinhole camera is used [73].

The CNR of the image reconstructed at the z where the source was actually located was determined as described in Sec. 8.2.3. For all other reconstructions from the stack the position of S was kept constant, which assured that the resulting CNR profile corresponds to a measurement along the z-axis. Finally, a Gaussian curve with offset of the following form was fitted through the CNR profile:

$$\mathrm{CNR}(z) = \alpha + (\beta - \alpha) \exp\left(-\frac{(z-\gamma)^2}{2\,\delta^2}\right) \tag{8.7}$$

with the fitting parameters α, β, γ, and δ. The fitting procedure was carried out in Python (3.8) with the `curve_fit` function from SciPy (1.10). As initial guess, α and β were set to the minimal and to the difference between the maximal and minimal occurring value of the profile. The GT position and 1.0 were used to initialize γ and δ. The fitting parameters were bound to 0 and [500,000; 500,000; 170; 20]. Additionally, the analytical Jacobian matrix of Eq. 8.7 and a limit of 100,000 function evaluations. The axial resolution and its uncertainty introduced from the fitting procedure are reported as FWHM with correspondence between δ and the FWHM from Eq. 2.1. Additionally, the coefficient of determination (R^2) was calculated per fit.

8.2.5 Assessing the Lateral Resolution

In addition to the axial resolution we measured the lateral resolution in order to compare our results with values from the literature. This allows us to additionally report the ratio of axial to lateral resolution. We determined the lateral resolution for 30 mm, 50 mm, and 100 mm source-to-mask distance in accordance with the suggestion given in a recent review on IGCs [7].

To compute the lateral resolution, the in-focus image within the image stack used for the assessment of the axial resolution was selected. Then the source profile along the row with the highest pixel intensity was chosen for further processing. A Gaussian curve with offset, analog to Eq. 8.7, was fitted to this

profile and the FWHM value was obtained from the resulting standard deviation. As the FWHM value was given in pixels, it was converted to millimeters by using the FOV of the respective reconstruction method [118].

8.3 Results

First, the proposed modifications to the 3D-MLEM algorithm are evaluated before the general 3D reconstructions are compared. Afterwards, the axial and the lateral resolutions are presented and eventually the computation time of the reconstruction methods are addressed.

8.3.1 Proposed 3D-MLEM algorithm

A comparison between the original 3D-MLEM algorithm and our 3D-MLEM with the proposed modifications is shown in Fig. 8.3. Note, how in the initial 3D-MLEM reconstruction the source at 30 mm distance is barely visible. Additionally, the maximum intensity of the entire image stack is located in the reconstructed image at 45 mm. Similar artifacts are visible in all other image stacks of sources at different distances as well. The 3D reconstruction from our proposed 3D-MLEM algorithm shows a single bright spot at the distance where we indeed placed the source, despite the high transmission noise.

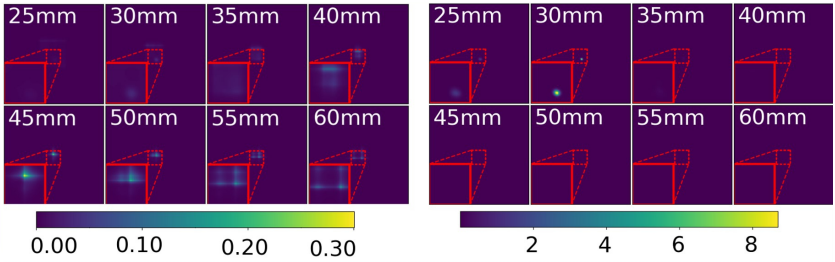


Figure 8.3: The original 3D-MLEM algorithm (left) from [24] in comparison to our proposed 3D-MLEM algorithm (right) applied to the source at 30 mm distance. The center, marked by the red dotted square, is magnified for better visibility. Figure from [118].

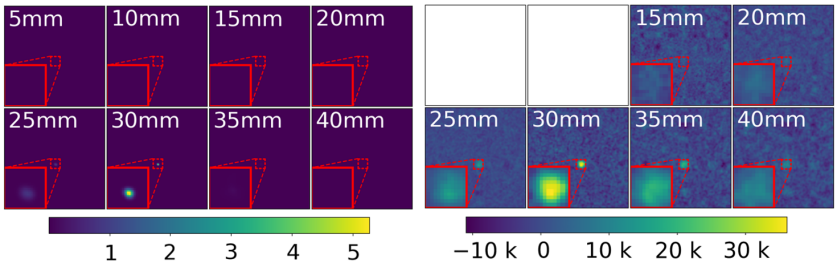


Figure 8.4: 3D-MLEM (left) and MURA Decoding (right) of the 30 mm source. The distance between the mask and the in-focus plane in mm is indicated in the top left corner. MURA Decoding is not capable of reconstructing planes that are closer than 11 mm. A magnification of the area around the source (red dotted square) is shown in the bottom left corner. Figure from [118].

8.3.2 3D Reconstructions

For a rough localization, all 21 detector images were reconstructed in the range of 5 to 100 mm in 5 mm steps. The first eight of the 20 images from the image stack of the source at 30 mm distance are shown in Fig. 8.4. The entire image stack for the sources at 50 mm and 100 mm can be found in Appendix A.2.

Both methods show bright spots in the center of the reconstructed image at the true distance, even though the pixel intensity differs largely between the methods, as Fig. 8.4 shows. In general, the background of the 3D-MLEM images appears more uniform, whereas MURA Decoding yields images with a higher background noise. For reconstructions at 50 mm with the source positioned at $z = 50$ mm and the image normalized to the range of 0 to 1, the obtained σ_B values from MURA Decoding and 3D-MLEM were 0.0281 and $8.7614 \cdot 10^{-}5$, respectively. From Fig. 8.4 alone a worse axial resolution of MURA Decoding can be observed, as the source is also visible in the reconstructions at 25 mm and 35 mm distance. Furthermore, 3D-MLEM is capable of reconstructing at distances that are closer than 15 mm.

Table 8.1: The FWHM axial resolutions are displayed separately for the two reconstruction methods (MURA Decoding and 3D-MLEM) and for raw and preprocessed detector images. The standard deviation values are obtained through the fitting algorithm. Table from [118].

Source distance [mm]	**MURA Decoding** raw [mm]	preprocessed [mm]	**3D-MLEM** raw [mm]	preprocessed [mm]
12	5.3 ± 0.6	5.3 ± 0.6	2.17 ± 0.03	1.75 ± 0.03
14	4.7 ± 0.1	4.7 ± 0.1	2.20 ± 0.02	1.80 ± 0.06
16	6.0 ± 0.2	5.8 ± 0.2	2.32 ± 0.04	1.85 ± 0.08
18	6.9 ± 0.2	6.2 ± 0.2	2.34 ± 0.02	2.02 ± 0.03
20	7.5 ± 0.1	7.3 ± 0.1	2.60 ± 0.09	2.26 ± 0.02
25	11.0 ± 0.4	9.8 ± 0.4	2.37 ± 0.02	2.54 ± 0.04
30	12.1 ± 0.5	11.9 ± 0.5	2.74 ± 0.02	2.76 ± 0.11
35	17.2 ± 0.9	15.1 ± 0.8	2.45 ± 0.10	2.01 ± 0.14
40	16.9 ± 0.9	18.5 ± 1.1	$\dagger 4.32 \pm 0.12$	$\dagger 3.51 \pm 0.15$
45	14.2 ± 0.8	18.4 ± 1.3	$\dagger 5.48 \pm 0.06$	$\dagger 4.69 \pm 0.13$
50	15.6 ± 1.0	17.5 ± 1.0	$\dagger 5.53 \pm 0.09$	$\dagger 5.97 \pm 0.09$
55	18.3 ± 1.1	18.8 ± 1.1	$\dagger 5.52 \pm 0.11$	$\dagger 4.73 \pm 0.08$
60	19.4 ± 0.9	19.9 ± 0.9	$\dagger 6.71 \pm 0.11$	$\dagger 5.24 \pm 0.10$
65	23.9 ± 0.9	28.0 ± 1.2	$\dagger 7.04 \pm 0.12$	$\dagger 6.67 \pm 0.10$
70	22.2 ± 0.7	23.8 ± 0.9	$\dagger 8.19 \pm 0.17$	$\dagger 7.37 \pm 0.11$
75	26.6 ± 0.7	27.8 ± 0.7	$\dagger 9.79 \pm 0.17$	$\dagger 9.10 \pm 0.17$
80	32.6 ± 0.5	35.4 ± 0.5	$\dagger 10.40 \pm 0.18$	$\dagger 11.64 \pm 0.27$
85	32.7 ± 0.2	35.9 ± 0.4	$\dagger 11.11 \pm 0.21$	$\dagger 12.34 \pm 0.24$
90	37.0 ± 0.4	37.8 ± 0.5	$\dagger 11.20 \pm 0.24$	$\dagger 10.37 \pm 0.24$
95	35.2 ± 0.8	38.3 ± 0.8	$\dagger 12.28 \pm 0.29$	$\dagger 14.81 \pm 0.41$
100	34.8 ± 0.4	42.2 ± 0.9	$\dagger 14.84 \pm 0.66$	$\dagger 13.48 \pm 0.86$

† Additional ghost sources appear in the resulting image stack.

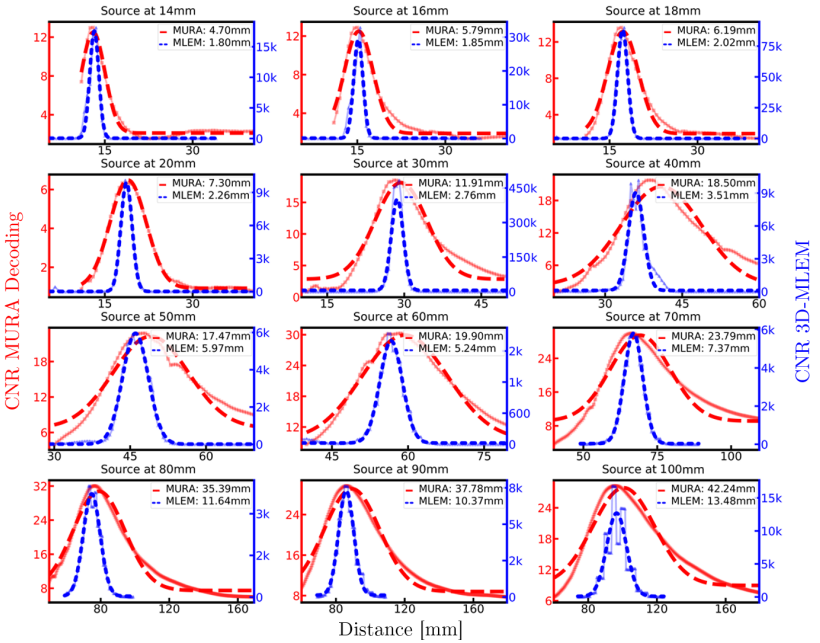


Figure 8.5: The CNR profiles over the distance used for reconstruction for a selection of source positions: the semi-transparent red line of squares and the blue line of triangles show the CNR profiles of MURA Decoding and 3D-MLEM reconstruction. The Gaussian curves with offset, represented by the bold red dotted line (MURA Decoding) and the blue dashed line (3D-MLEM), were fitted to the CNR profiles. These curves serve as the basis for determining the axial resolution, and the corresponding FWHM values are displayed in the top right corner of each graph. Figure from [118].

8.3.3 Axial Resolution

Table 8.1 shows the resulting FWHM axial resolution for each of the 21 acquired detector images. The CNR profiles and Gaussian fits of a few image examples, from which the FWHM was derived are presented in Fig. 8.5. A clear difference between MURA Decoding and 3D-MLEM reconstruction is evident: First, the axial resolutions obtained by 3D-MLEM are better (smaller in value), as their profiles are narrower than those obtained from MURA Decoding. Additionally,

the CNR values, i.e. the height of the Gaussian curves, are greater by a factor between approximately 60 and 30,000.

The axial resolutions of the raw detector images are comparable to the those obtained with preprocessing. The average ratio of the FWHM from preprocessed to raw detector images are 1.05 ± 0.10 (MURA Decoding) and 0.93 ± 0.12 (3D-MLEM). Nevertheless, the ultimate objective of CAI reconstruction is to generate clear images that are interpretable also by human observers. As illustrated in Fig. A.3 from the Appendix A, reconstructions based on preprocessed images exhibit less background noise and, thus, the subsequent sections will focus on the axial resolution of the preprocessed images. The axial resolutions are 11.9 ± 0.5 mm and 2.76 ± 0.11 mm (source at 30 mm, magnification $M = 1.67$), 17.5 ± 1.0 mm and 5.97 ± 0.09 mm (source at 50 mm, $M = 1.40$), and 42.2 ± 0.9 mm and 13.48 ± 0.86 mm (source at 100 mm, $M = 1.20$) for MURA Decoding and 3D-MLEM respectively. The average standard deviations introduced by the fitting procedure are 3.9 % (MURA Decoding) and 2.7 % (3D-MLEM). For sources at greater distances than 40 mm, 3D-MLEM reconstructions contain up to eight *ghost sources* in a regular pattern surrounding the true central position, as is shown in Fig. 8.1c. For a fair comparison with other imaging systems, the axial resolution is presented in Fig. 8.6 as a function of the magnification factor M, alongside the corresponding source-to-mask distances.

8.3.4 Lateral Resolution

The lateral resolutions based on the preprocessed detector images measured are 0.74 mm (MURA Decoding) and 0.27 mm (3D-MLEM) at a source distance of 30 mm, 0.80 mm and 0.29 mm at 50 mm and 1.04 mm and 0.4 mm at 100 mm. The resulting ratios between axial and lateral resolutions are 16 : 1 and 10 : 1, 22 : 1 and 21 : 1, and 41 : 1 and 34 : 1 for MURA Decoding and 3D-MLEM, respectively.

8.3.5 Computation Time

To compare the computational performance of the two reconstruction methods with their respective implementations, the average runtime for one image in an image stack is presented here. The total runtime for image stacks containing

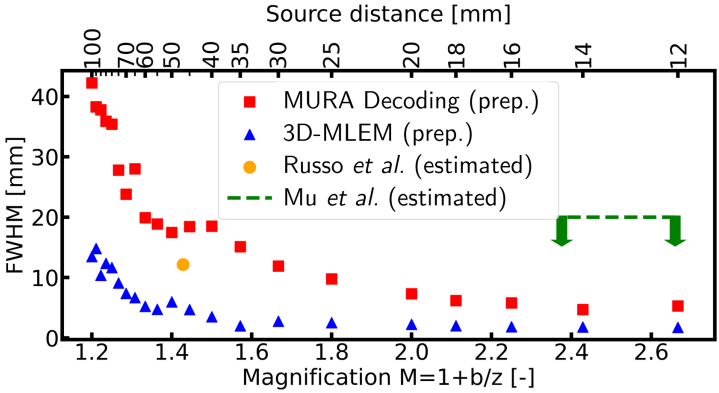


Figure 8.6: The axial resolution for both presented reconstruction methods plotted against the dimensionless magnification factor M. The orange circle and green dashed line represent reference values for the axial resolution estimated from literature (Russo *et al.* [25], and Mu *et al.* [24]). Note that the source-to-mask distance at the top axis only applies to this study. Figure modified from [118].

between 54 and 101 reconstructions with the 3D-MLEM algorithm for 40 iterations on a computer with a 6-kernel Intel Core i7-9750H processor (2.6 GHz) and 16 GB of RAM ranged from 348 s to 579 s. Relative to the number of images per stack the average runtime is (5.68 ± 1.40) s per image for 40 iterations.

The runtime of MURA Decoding was measured using a laptop equipped with a 12$^{\text{th}}$ generation Intel Core i7-12700H processor (2.3 GHz) and 32 GB of RAM. The total reconstruction time for image stacks containing between 60 and 240 reconstructions ranged from 31.2 ms to 547 ms with a mean of (1.3 ± 0.5) ms per image. As a result, MURA Decoding is approximately 4,400 times faster than 3D-MLEM with the considered implementations and the chosen number of iterations.

8.4 Discussion

In the following sections, the results of this study are discussed: It starts with a comparison of the two reconstruction methods and afterwards the axial resolu-

tion is compared to the sparse values from other research groups. Finally, the limitations and potentials of this study are discussed.

8.4.1 Reconstruction Methods

Two improvements were made to the initial convolutional-based 3D-MLEM: first, a general sensitivity term was added that takes into account that photons emitted from the edge of the FOV are less likely to hit the detector. Second, the forward projection was extended to account for transmission noise. The absence of artifacts and the lower background noise (see Fig. 8.3) justify our proposed modifications.

In contrast to MURA Decoding, the 3D-MLEM algorithm provides a complete 3D distribution, as the contribution of all slices are taken into account simultaneously. This, however, entails more convolutions: for each slice in each iteration the forward simulation, the backward simulation and another forward simulation with the updated slice are calculated. Additionally, the images which are processed – the entire detector image and the PSFs – are generally larger than for MURA Decoding. This makes it computationally expensive and therefore slow, but more accurate. Although the runtime comparison was not performed on the same computer and improvements in the implementation are conceivable, the order of magnitude by which 3D-MLEM is slower than MURA Decoding is sufficient to conclude that it is too slow for intraoperative use.

The ghost sources that appear in the 3D-MLEM reconstruction for sources more than 40 mm away (see image on the right of Fig. 8.1c are likely caused by the self-similarity of the mask pattern due to its 2×2 arrangement. A potential explanation is that the cast shadow is smaller than the detector. The surrounding margins (see the bottom of the detector images in Fig. 8.1b that are not illuminated by the mask pattern, are relatively small and contain noise, which means that the algorithm has only a limited area to penalize a set of ghost sources outside the FOV that are partially contributing to the correct pattern in the center. It is important to note that these ghost sources do not completely prohibit a reconstruction, but they add an element of ambiguity. This issue could be solved by modifying the algorithm to use only the central part of the detector image, similar to MURA Decoding. This, however, would come at the cost of a narrower FOV.

The presented findings of this work allow to answer the research question explicitly mentioned in [73]: "What role iterative reconstruction algorithms [...] will play in improving Z resolution": dividing the axial resolution obtained from the iterative 3D-MLEM algorithm by the axial resolution obtained from MURA-Decoding for each source position, gives an average factor of 0.3 ± 0.1. Hence, on average the axial resolution of 3D-MLEM is one third of the values from MURA Decoding, meaning with the first method we achieve a three times better axial resolution.

The high resolution in both lateral and axial directions makes 3D-MLEM interesting for large gamma cameras in Single Photon Emission Computed Tomography (SPECT) systems where runtime is of minor importance. Nevertheless, for an intraoperative application where the computation time substantially influences its practicability, the long runtime makes the 3D-MLEM algorithm unsuitable.

8.4.2 Comparison to the Literature

Setting the axial resolution in relation to the magnification factor M from Eq. (3.2), has the advantage of eliminating the dependency on the detector-to-mask distance b and hence allowing a comparison to other cameras from literature.

Only few papers exist with which we can compare our results [24, 25, 73]. In a previous experiment [25] with the same coded aperture collimator but a slightly different detector-to-mask distance b, and the same MURA Decoding as in this work the authors evaluated the zone of best-focus of a ring-shaped object. The zone of best-focus was defined as the zone where the image contrast is maximum and constant within about 1 % and was approximately 3 mm, while the lateral resolution was measured to be 0.6 mm at a source-to-mask distance of about 50 mm [25]. Their contrast profile is not equivalent to the CNR profile used in this work for assessing the axial resolution, but serves as a rough benchmark. When estimating the FWHM from the given graph, we obtain an FWHM of approximately 12 mm (denoted by the orange dot in Fig. 8.6). This axial resolution is slightly better than our values reported above, but still provides good plausibility. Comparing the relative resolution as the ratio of axial to lateral resolution, a ratio of $12\,\text{mm}/0.6\,\text{mm} = 20$ at 50 mm distance is

obtained. Our assessments of a ratio of 22 and 21 for MURA Decoding and 3D-MLEM at 50 mm distance are in good accordance with that.

A rough classification for the reported 3D-MLEM results can be deducted from [24]. In their described experiment, the authors placed sources shaped like an "H" and a ">"-symbol at 164 mm and 244 mm distances from the mask and reconstructed the scenery with 3D-MLEM. Since in their figure the two sources appear separately in their corresponding planes, the axial resolution must be smaller than half the distance between them, approximately 20 mm (green annotation in Fig. 8.6). Our values at comparable M are lower than that and thus not contradictory.

8.4.3 Limitations

As explained in Sec. 2.1, measuring the spatial resolution for pinhole and parallel-hole collimators as the FWHM of a point source is a common procedure. The underlying assumption is the superposition principle which allows to apply the spatial resolution to extended or multiple sources. However, the superposition assumption is problematic in CAI, since extended sources are known to be reconstructed in lower quality than point sources [24, 73]. That means further investigations regarding extended sources are required.

Furthermore, we do not have the GT of the entire 3D position of the sources from our dataset, only the source-to-mask distance. Thus, an entire 3D localization error as in [32] could not be presented. We also did not test different values for detector-to-mask distance b.

In an intraoperative experiment with pigs, a reduction of the dynamic range was observed when a very bright source was present. It was called "concentration effect" (also referred to as "shine-through" [14]) and makes weaker sources less visible [73]. The 3D-MLEM algorithm is supposedly immune to this effect, but this was not investigated yet. However, in previous experiments with the same semiconductor-based detector as used in this study it was observed that a ^{241}Am gamma source with an activity of 1 μCi is still visible when a 1 mCi ^{241}Am source is placed nearby and used as background noise [25]. We claim that this is possible thanks to the extended counting linearity range, the practical immunity to read-out noise and the pixel-wise functioning of the photon counting detector, compared to scintillator-based detectors.

Two more aspects, that were outside the scope of this study, are the impact of a lateral shift of the source and the mask thickness. First, an image degradation has been reported when the source is off-center [46]. This could imply a small decrease in axial resolution, but a quantification of this effect has yet to be performed. Second, the thin mask used in this experiment caused almost no collimation, but a high transmission rate. The question of how transmission affects the lateral and axial resolution remains open.

8.4.4 Potentials

The findings of this study indicate the following potentials. First, the proposed modified 3D-MLEM algorithm enables other researchers to use this algorithm for their gamma cameras, with masks tailored to their specific purpose, e.g. for SPECT systems. Additionally, integrating the transmission noise into the reconstruction step might allow to investigate masks that have a transmission rate of less than a typical rate of 98 % [36]. Second, this unique and publicly available dataset could be used to analyze how two point-like gamma sources in a row can be best distinguished, similar to [129]. Third, a key advantage of 3D-MLEM is that in addition to the reconstruction, it provides the forward projection of the estimated reconstruction (see, for example the left-hand side of Fig. 8.1c). This projection could be employed to assess the degree of correspondence between the estimated and the acquired detector image, offering insights into the influence of defective pixels or similar imaging issues.

Fourth, the improvements to the 3D-MLEM algorithms opens up the possibilities for further optimization in its design, potentially result in an enhanced axial resolution or a substantial increase in processing speed. Further research might be aimed to accelerate the algorithm or reduce the number of necessary iterations. Only then, it would be feasible for intraoperative usage.

Finally, the investigation of the localization error was beyond the scope of this study. With the availability of concrete figures for the axial resolution, however, further analyses are underway with the objective to evaluate the full 3D-localization capabilities for point-like sources. In order to further analyze the potential intraoperative use of an IGC with coded aperture collimator, it is essential to conduct experiments that more closely resemble real-world use cases. These experiments require a new mask, since the most commonly

administered radionuclide in nuclear medicine is ^{99m}Tc [12] with a main gamma emission at 140.5 keV. Consequently, a thicker mask is required to reduce transmission. A new coded aperture collimator with 0.25 mm holes on a 1 mm thick tungsten sheet is currently under development by the University of Naples and the INFN. In addition, the reported issues of CAI with extended sources require investigations of the influence on the axial resolution and, hence, on 3D-localization.

8.5 Conclusion

In this chapter, the axial resolution of an experimental IGC equipped with a coded aperture collimator was systematically assessed. The FWHM was derived from 3D reconstructions closely around the true source distance followed by fitting a Gaussian curve with offset to the extracted CNR profile. The CNR profile along the z-direction of the obtained image stack was calculated by a semi-automatic algorithm to assure reproducibility. In addition to the most commonly used reconstruction method (MURA Decoding) a 3D-MLEM algorithm was adapted to address the issue of transmission noise and a more general camera setup.

The results of this study complement our understanding of the spatial resolution in CAI for point-like sources to all three dimensions. It was shown that the CNR profile of a point-like source represents a monomodal curve with a width depending on the reconstruction method. Analog to the lateral resolution, this width, measured as FWHM, can be regarded the axial resolution. Between the two different reconstruction methods compared, a large difference in the obtained axial resolution was observed: MURA Decoding was found to be fast and yielding a mediocre axial resolution, while the 3D-MLEM algorithm reconstructs with a superior axial resolution, both in lateral and axial directions, but is much slower than MURA Decoding. In the context of 3D localization, the axial resolution can be interpreted as the precision, which indicates the number of in-focus planes that the detector image must be reconstructed to achieve an accurate depth estimation. Despite the inferior precision, it is hypothesized that MURA Decoding is capable of estimating the depth of single point-like sources with high accuracy in all three dimensions in a reasonable amount of time due

to its fast runtime. Meanwhile, 3D-MLEM in its current form, is not suitable for this task in an intraoperative setting, given its extensive computational demands. In addition to a higher resolution and sensitivity, estimating the depth of a source has been identified as a desirable feature of IGC [7]. This capability further distinguishes coded aperture collimators from other types, including parallel-hole and pinhole collimators.

Chapter 9

3D-Localization of Point-Like Gamma Sources

As demonstrated in the preceding chapter, point-like sources imaged with a coded aperture camera exhibit a monomodal Contrast-to-Noise Ratio (CNR) profile in the axial direction. This chapter aims at answering the question on how well this curve can be exploited to localize point-like gamma sources in all three spatial dimensions.

This chapter originates from a publication to the IOPscience journal "Physics in Medicine & Biology" which is publicly available under CC-BY 4.0 license [149].

9.1 State of the Art

Over the last few years, gamma probes have become an important tool in Radioguided Surgery (RGS) for a variety of cancerous diseases [6, 7, 150]. In contrast to counting probes which provide only an acoustic feedback and a count rate reading, Intraoperative Gamma Camera (IGC) allow for a precise detection of radioactively-marked structures and, generally, give a broader overview of the incision site [8].

A common use-case for IGCs is Sentinel Lymph Node Biopsy (SLNB) [150]. An accurate assessment of the axillary lymph node involvement is essential in staging breast cancer. Metastases in the axillary lymph nodes are the most important predictor of overall recurrence and survival [3, 7]. Thus, localizing gamma sources in all three spatial dimensions with a mobile gamma camera could provide valuable information to the surgeon [20] and is a first step toward providing valuable depth information in SLNB.

As a potential solution for 3D-localization of gamma sources, stereo camera systems are currently under investigation [20, 31, 32, 35]: While Bugby et al. [20] report mean localization errors of 1.23 mm on simulated, and 3.54 mm on experimental data, Kaissas et al. [31, 32] obtaining errors between 0.28 mm and 7.8 mm depending on the source size and distance. Nevertheless, the disadvantage of using stereo cameras is that they require two costly gamma cameras and two precisely placed coded aperture masks. An alternative approach is to combine a single mobile gamma camera with an external tracking that is capable of merging images from multiple viewpoints into a single 3D map [33]. They report an error between 2.9 mm and 7.4 mm depending on the user's experience. However, this approach still requires additional hardware and suffers from an increased acquisition time.

3D imaging by using a coded aperture is a research field where the capabilities have not been fully explored yet [25, 95]. By reconstructing the captured detector image at several successive planes, also referred to as *in-focus planes*, a 3D reconstruction of the scenery can be computed. In the previous chapter, it was demonstrated that the most popular reconstruction algorithm MURA Decoding is able to achieve an axial resolution that is approximately between 15 : 1 and 40 : 1 relative to the lateral resolution [118]. This chapter's goal is to answer the following research question: How accurate can a single high-resolution gamma camera with a coded aperture collimator localize a single point-like source in the near-field setting? 3D-localization means identifying the coordinates $[\hat{x}, \hat{y}, \hat{z}]$ in a camera-based coordinate system of a single point-like source where the source is assumed to be a isotropically radiating spherical gamma source from the 3D reconstruction $\hat{f}(x, y, z)$. Two different methods for 3D-localization are investigated: one is based on the Center Of Mass (COM) and the other is the Iterative Source Localization (ISL) method which relies on calculating the CNR profile in axial direction. For the depth estimation, two different fitting functions are compared: a Gaussian curve and an Exponentially

Modified Gaussian (EMG) distribution. After evaluating the performance of the two methods on a simulated dataset, the most accurate method is finally tested on the experimental dataset we acquired with our experimental IGC. The contributions of this study to the state of the art are as follows:

1. We show that the CNR profiles of point-like sources are best fitted by an EMG distribution.
2. We propose an iterative source localization method based on a deterministically calculated CNR for 3D-localization of point-like sources.
3. We demonstrate a localization accuracy of less than 3 mm for point-like sources at a distance of 20–100 mm, which is comparable to the accuracy of more complex technologies like stereo cameras.

To promote transparency and reproducibility, both the acquired datasets and the localization methods of this research are publicly available on GitHub at `https://zenodo.org/records/11449544`.

9.2 Methods

This section begins with the presentation of an enhanced algorithm for calculating the CNR. Afterwards, the methodology for generating 3D reconstruction is briefly explained, before two different localization methods are introduced. Finally, this section will conclude with an outline of the conducted sensitivity analysis.

The dataset used for investigating the 3D-localization error was acquired with the same experimental IGC from the previous chapter and is presented in Sec. 5.3. However, the localization methods are initially analyzed on a *in silico* reproduction of the experimental dataset from a Monte Carlo (MC) simulation, where more precise control of the actual source position was possible.

9.2.1 Convolutional Contrast-to-Noise Ratio

To determine the z-component of a point-like source, we use the profile of the CNR along the z-direction in analogy to Chapter 8. However, the previously used imageJ macro described in Sec. 8.2.3 was slow, took up a large amount

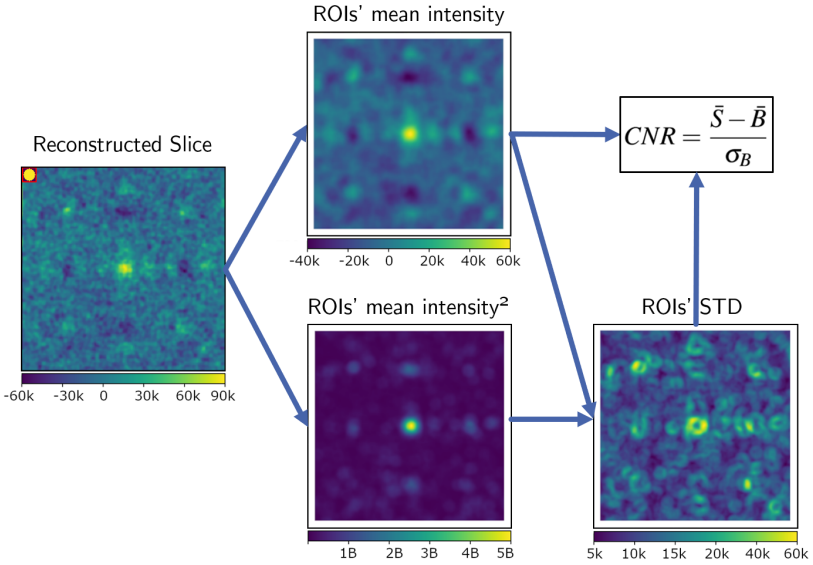


Figure 9.1: The CNR is determined by convolution of the reconstructed slice and a spherical kernel normalized to the sum of 1 which is depicted in the red rectangle in the top left corner in original size. The resulting image represents the average intensity of a Region of Interest (ROI) centered at each pixel position. With the identity $Std(X) = \sqrt{E(X^2) - E(X)^2}$, the standard deviation (STD) inside all possible Region of Interest (ROI) is calculated. Afterwards, the Region of Interest (ROI) with the highest average intensity becomes the signal $\bar{S}$ and the background's parameter $\bar{B}$ and σ_B is taken by averaging over all Region of Interests (ROIs) that do not overlap with the Region of Interest (ROI) of $\bar{S}$. The CNR of the exemplary image with a Region of Interest (ROI) size of 16 pixels in diameter is 4.71.

of disk space and required switching between computer programs manually. Therefore, a faster variant of this algorithm was implemented in Python and is presented here: Its foundation is the same definition of the CNR as Eq. 8.6 in Sec. 6.2.3.

In order to increase reproducibility we developed a semi-automatic algorithm that determines the CNR for a single image solely based on the source diameter. Its scheme is depicted in Fig. 9.1. The algorithm considers all possible Region of Interests (ROIs), which are defined as regions of a given size centred at each possible position on the pixelated grid. This is carried out by two convolutions of the reconstructed slice with a kernel that represents the circular

ROI. The kernel is a quadratic image twice the size of the ROI's radius in pixels and is 0 everywhere except for the central circular region where the intensities are set to 1. The kernel is normalized to the sum of 1, and thus a convolution generates an image in which each pixel represents the mean intensity in the ROI centered at this pixel. An unpadded convolution is used, so the kernel does not extend beyond the image boundaries. Yet, in order to keep the pixel coordinates consistent rows and columns of NaNs ("not a number") are appended. The second convolution is carried out to the reconstructed image with pixel-wise squared intensities. The resulting image is then used to calculate the standard deviation (STD) of each ROI by the following equation:

$$STD(X) = \sqrt{E(X^2) - E(X)^2} \tag{9.1}$$

with the random variable X and the expectation operator E. Combining the results from both convolutions with Eq. 9.1, we yield the standard deviation of all ROIs.

The two regions S and B that are required for the CNR were selected as follows: first, the ROI with the highest mean intensity was chosen as S. Then, all ROIs that overlap with S are removed from further processing. Second, the mean and the standard deviation of the remaining ROIs are considered as background, and by averaging we obtain B and σ_B, respectively. Finally, the CNR is calculated according to Eq. 8.6. In conclusion, the presented method enables a fast and memory-efficient calculation of the CNR directly in Python, which is solely depending on the ROI size and is independent of manual interaction, making it deterministic and reproducible.

9.2.2 3D Reconstruction

Among the two reconstruction methods for 3D reconstruction, we learned from the previous chapters that only MURA Decoding is capable of producing 3D reconstructions in an acceptable amount of time. To localize the point-like gamma source, the detector image was reconstructed at several successive in-focus planes and thereby a 3D reconstruction of the scene was obtained. We explored a depth range of 11 mm to 130 mm with steps of $\Delta z = 0.5$ mm, resulting in a stack of 239 images. Since the size of the reconstructed images depends on z, each image of the stack is resized to the maximally occurring image size

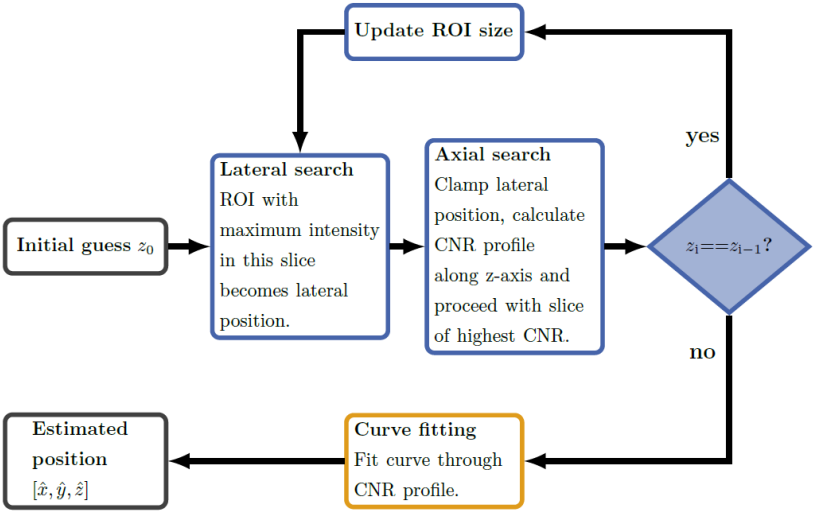
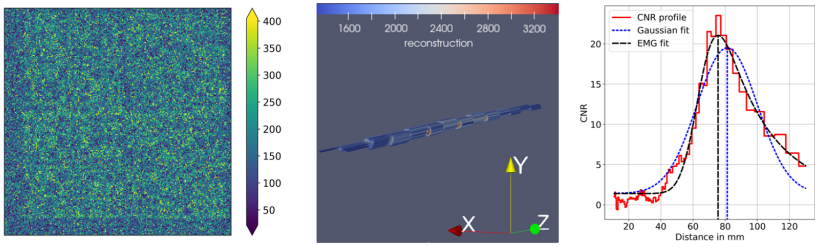


Figure 9.2: A flowchart representing the ISL algorithm with its alternating lateral and axial search and the final curve fitting to estimate the 3D position of the source. Figure from [149].

by bilinear interpolation, which is 254×254 pixels for $z = 11\,\text{mm}$. Resizing allows for easy handling of the 3D reconstruction, as it can be processed as a 3D matrix, where the z-coordinate corresponds to the extension of the source in z-direction. This approach ensures that the position of the source remains consistent with respect to the Field of View (FOV)'s center, which is beneficial during the localization process.

9.2.3 Localization with the Iterative Source Localization method (ISL)

In order to obtain the source position from a 3D reconstruction, the ISL was developed. Figure 9.2 shows a flowchart of the ISL algorithm. It iteratively alternates between a lateral and an axial search, followed by fitting a curve through the obtained CNR profile. The algorithm requires as input the 3D reconstruction, the FOV for each slice, and an initial guess denoted as z_0 for



(a) (b) (c)

Figure 9.3: (a) The detector image of the source located at $[0,2,75]$ mm captured with our experimental gamma camera. The color coding was limited to intensities between the 1st and the 99th percentile to visualize the projected mask pattern. (b) The 3D reconstruction of the detector image thresholded at the 99th percentile to visualize the reconstructed source. Note how the extension in z-direction is much larger than in the lateral directions. (c) The CNR profile with both the Gaussian fit and the EMG fit, resulting in a z-estimation of 73.7 mm ($R^2 = 0.97$) and 81.2 mm ($R^2 = 0.92$), respectively. Figure from [149].

the source-to-mask distance. The output of the ISL algorithm is the source's estimated 3D position $[\hat{x},\hat{y},\hat{z}]$ in millimeters.

In the first step, the lateral search, we begin with the image slice that is closest to z_0 and with the search of the lateral position, i.e. $[\hat{x},\hat{y}]$. Similar to the convolution described in Sec. 9.2.1, the mean intensities of all possible ROIs are calculated and the position of the highest value is assigned to the current lateral source position. The second step is the axial search, where the lateral position is fixed and a CNR profile along the z-direction is calculated as described above. Subsequently, the algorithm determines whether a slice with a higher CNR value exists than the current slice at z_i. If so, the algorithm initiates another lateral search, followed by an axial search. The ROI size is updated in each iteration based on the given source size and the FOV of the current slice.

The third and final step of the algorithms is entered as soon as z_i does not change within one iteration: the curve fitting. For more robust and accurate depth estimation, a curve fit is applied to the last CNR profile obtained. Two different fitting functions were compared: once a Gaussian curve with offset from Eq. 8.7, which will be denoted as Gaussian or Gauss, and a scaled EMG distribution with offset [151]. An EMG distribution emerges from the sum of two independent random variables, where one is normal distributed with mean γ and variance δ^2 and the other one exponentially distributed with a

rate of λ. This function was suggested as a potentially good candidate for the intensity distribution along the axial dimension in Chapter 4.3 of Ref. [152]. The resulting function has one additional fitting parameter more compared to the Gaussian function and can eventually be described as

$$\mathrm{CNR}(z) = \alpha + (\beta - \alpha)\,\tfrac{\lambda}{2}\exp\left(\tfrac{\lambda}{2}\left(2\gamma + \lambda\delta^2 - 2z\right)\right)\cdot \mathrm{erfc}\left(\tfrac{\gamma + \lambda\delta^2 - z}{\sqrt{2}\delta}\right),$$

where "erfc" represents the complimentary error function that is defined as

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty \exp\left(-t^2\right)\mathrm{d}t. \tag{9.2}$$

Equations for estimating the peak position (also referred to as *mode*) of an EMG function exist [151], but for the sake of simplicity, we utilize the monomodality, and sampled the fitted function with a step size of 0.01 mm and selected the z-value where the sampled function was maximal. The following initial guesses are directly derived from the CNR profile for the fitting procedure: $(\alpha, \beta, \gamma, \delta, \lambda) = \left(\min(\mathrm{CNR}(z)); \max(\mathrm{CNR}(z)); \mathrm{argmax}_z(\mathrm{CNR}(z)); 1; 1\right)$. The parameter λ is dropped for the Gaussian fitting. For the remainder of this chapter, we will refer to the ISL method with a final EMG fit as ISL-EMG and to the one with a Gaussian fit as ISL-Gaussian. An exemplary CNR profile with both fits is presented in Fig. 9.3c.

Overall, after a few iterations, the ISL method generated the estimated source position $[\hat{x}, \hat{y}, \hat{z}]$, where the x and y-component are determined by the brightest ROI position and the z-component by calculating the mode of the fitted function.

9.2.4 Localization with the Center of Mass (COM)

A more intuitive approach to finding the center of a 3D distribution is via the Center of Mass (COM). The advantage is that neither fitting nor a user input is required, which makes this method independent from any hyperparameters. The COM method relies on the fundamental assumption that the source position is the COM of the Largest Connected Region (LCR) in the given 3D reconstruction. Therefore, we built an algorithms that extracts the LCR and calculates the COM: First, the 3D reconstruction $\hat{f}(x, y, z)$ is thresholded by the 99.9^{th} percentile,

a processing step similar to the one proposed by [153]. By thresholding the 3D reconstruction, we assume that all voxels with a lower intensity than the 99.9^{th} percentile are background noise and therefore not the actual reconstructed source.

This led to multiple remaining connected regions, i.e. voxel clusters: a large region where the actual source is located and multiple smaller clusters closer to the camera that appeared roughly between 11 mm and 15 mm away from the mask. Hence, in a second step, the region with the largest number of connected voxels was extracted and is denoted as LCR: $\hat{f}_{\text{LCR}}(x,y,z)$. An exemplary LCR of the simulated source at $[0,2,75]$ mm is depicted in Fig. 9.3b. Finally, the COM $[\hat{x},\hat{y},\hat{z}]$ with the intensity of all voxels within this cluster is calculated via the 0^{th} (M_{000}) and 1^{st} moments (M_{100}, M_{010}, M_{001}) of a 3D distribution, which corresponds to the intensity-weighted mean of $\hat{f}_{\text{LCR}}$:

$$M_{000} = \sum_x \sum_y \sum_z \hat{f}_{\text{LCR}}(x,y,z) \tag{9.3}$$

$$M_{100} = \sum_x \sum_y \sum_z x\, \hat{f}_{\text{LCR}}(x,y,z) \tag{9.4}$$

$$M_{010} = \sum_x \sum_y \sum_z y\, \hat{f}_{\text{LCR}}(x,y,z) \tag{9.5}$$

$$M_{001} = \sum_x \sum_y \sum_z z\, \hat{f}_{\text{LCR}}(x,y,z) \tag{9.6}$$

$$[\hat{x},\hat{y},\hat{z}] = \tfrac{1}{M_{000}} [M_{100}, M_{010}, M_{001}] \tag{9.7}$$

In summary, the COM method estimates the source positions by performing automatic thresholding of the 3D reconstruction, selecting the LCR and calculating its COM. Both localization methods were implemented in Python (3.8.18) with NumPy (1.24.4), Tensorflow (2.10.1) for the CNR calculation, SciPy (1.10.1) for the fitting, the VTK library (9.3.0) for finding the LCR and Pandas (2.0.3) for the final analysis. All processing was carried out on a computer with a 6-kernel Intel Core i7-9750H processor (2.6 GHz), 16 GB of RAM and a NVIDIA GeForce RTX 2070 with 8 GB vRAM.

9.2.5 Sensitivity Analysis

While the COM method does not require any user input, the ISL method requires an initial guess of the source position by the user. Hence, the impact of the initial guess on the resulting z-coordinate was analyzed by running the ISL method with varying z_0. To mimic a user scrolling through the slices and selecting the slice where they assume the source, the initial guess z_0 was varied in three categories within a uniform distribution around the true value with $\pm 5\,\text{mm}$, with $\pm 10\,\text{mm}$ and $\pm 15\,\text{mm}$. With a Δz of 0.5 mm, the three categories correspond to ± 10, ± 20, and ± 30 slices. Additionally, we analyzed the localization error when we automatically selected the slice with the highest voxel intensity as z_0. The ISL method was applied five times to the simulated dataset with randomly varying initial, once with the highest voxel intensity, and once with the true z_0.

9.3 Results

First, the localization error on the simulation dataset are presented with a focus on comparing the localization methods. Then the results of the sensitivity analysis are shown before the final analysis on the experimental dataset are presented.

9.3.1 Simulation Results

The mean localization error and the standard deviation averaged over all 17 estimations for both localization methods are presented in Table 9.1. The mean localization errors of the detector images from the MC simulation are $(1.65 \pm 1.05)\,\text{mm}$, $(3.13 \pm 1.15)\,\text{mm}$ and $(0.77 \pm 0.62)\,\text{mm}$ for COM, ISL-Gaussian and ISL-EMG, respectively.

When comparing the localization methods, it becomes obvious that the ISL-EMG method yields overall statistically better results than the COM method. Using the EMG fit results in a mean coefficient of determination (R^2) of 0.97 ± 0.02 as opposed to the Gaussian with 0.93 ± 0.04. The localization error is larger with EMG $(0.77 \pm 0.62)\,\text{mm}$ compared to the Gaussian $(3.13 \pm 1.15)\,\text{mm}$ which makes ISL-EMG on average more accurate than ISL-Gaussian by a factor of

Table 9.1: This table shows the localization errors of the different methods. In addition to the mean error with its standard deviation (STD) over the 17 3D reconstructions, also the median error and the relative error of the z-component with respect to the true source distance in percent are presented. Table adapted from [149].

Data	Method	Mean	Localization Error in mm		Relative z-error in %	
		R^2	Mean±STD	Median	Mean±STD	Median
Sim.	COM	–	1.65 ± 1.05	1.34	2.32 ± 1.49	1.95
	ISL-Gauss	0.93	3.13 ± 1.15	3.13	4.38 ± 1.39	3.90
	ISL-EMG	0.97	0.77 ± 0.62	0.58	0.88 ± 0.68	0.60
Exp.	ISL-EMG	0.97	2.64 ± 0.71	2.59	3.06 ± 1.50	2.98

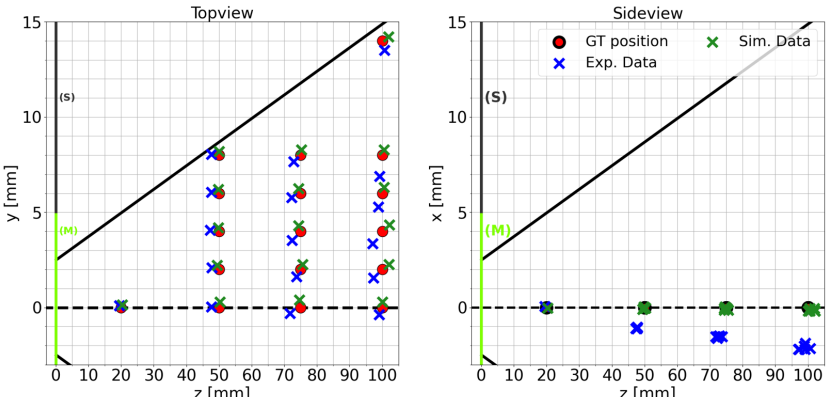


Figure 9.4: The true source positions (red circles) and the estimates from the ISL-EMG method applied to the simulated (green crosses) and experimental data (blue crosses) in a topview (left) and sideview (right). Note the different range in z-direction. Figure from [149].

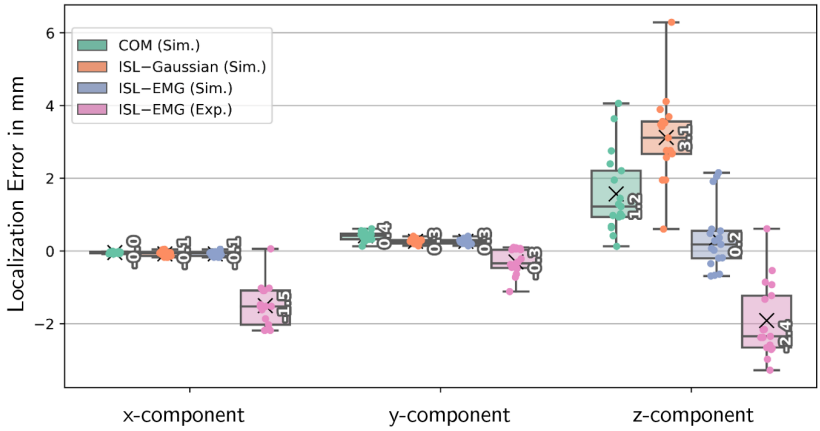


Figure 9.5: The localization error in mm broken down in x, y, and z-component by localization method (COM, ISL-Gaussian and ISL-EMG) and dataset (simulated and experimental). The boxes indicate the 25:75 percentile range, whiskers are maximum and minimum values, and lines are the median error which are also printed vertically in white. The crosses represent the mean values. Figure from [149].

4.06. Moreover, for the ISL-EMG method, the x, y and z-component contribute on average 3.2 %, 33.6 %, and 63.2 % to the localization error. The runtimes on the computer specified above with initialized GPU averaged over the 17 source positions were (1.88 ± 0.43) s, (1.48 ± 0.25) s, and (11.76 ± 0.46) s for the ISL-Gaussian, ISL-EMG and the COM method, respectively.

9.3.2 Sensitivity Analysis

The sensitivity analysis regarding the initial guess z_0 reveals that the mean error remains on average unaffected at (0.76 ± 0.66) mm with random variations of ± 10 slices, but increases to (0.88 ± 1.18) mm and (0.95 ± 1.65) mm for larger variations of ± 20 and ± 30 slices around the true z-value. When z_0 is selected according to the highest occurring intensity, the localization error increases to (49.85 ± 32.28) mm. Note, that the mean error is calculated based on different numbers of localization errors: While the initial guess with the true slice and

with the maximal voxel were calculated once on the dataset, the random guess with increasing variations were carried out five times on the simulated dataset.

9.3.3 Experimental Results

We applied the ISL-EMG method with the Ground Truth (GT) as an initial guess and the correct source Full Width at Half Maximum (FWHM) of 0.65 mm on the experimental data and obtained a mean localization error of (2.64 ± 0.71) mm. Especially the errors in x and z-direction are not centered around 0 but at approximately -1.5 mm and -2.4 mm as can be seen in Fig. 9.5. Overall, the localization error is about 3.4 times worse than that obtained from the MC simulation using the same method.

9.4 Discussion

The results presented above will be discussed by first examining the localization methods, before comparing the accuracy achieved with other localization technologies. Finally, the identified limitations of this study are addressed.

9.4.1 Localization Method

The MC simulation results reveal that the true mask-to-source distance corresponds fairly well to the peak of the CNR profile, which has already been noticed when investigating the axial resolution (see Sec. 8.3). However, as the CNR profile of a single source is skewed towards the camera, the mean value and the mode, i.e. the peak position, do not coincide as they would for symmetric functions. This explains why the COM method overestimates the z-coordinate and hence, has a localization error that is approximately 2.1 times higher than the ISL-EMG method. This demonstrates the necessity of considering the positive skewness of the intensity distribution in the z-direction.

The sensitivity analysis showed that the mean localization error achieved with variations of up to 30 reconstructed images is still below 1 mm, although with increasing standard deviation. This increase can be attributed to the presence of small clusters of high-intensity voxels at close distance to the camera

that are misidentified as the source. In addition, these clusters are the reason why it is not advisable to choose the slice with the highest pixel intensity for the initial guess as the large mean error of almost 50 mm demonstrates. In most cases, these cluster had the highest occurring intensities in the entire 3D reconstruction. This adds another type of systematic artifact caused by MURA Decoding besides the known cross-shaped artifact [83], near-field effects [24, 26], and ghost source effect [154]: the axial ghost sources. In conclusion, the ISL method can be considered robust with regard to the initial guess, provided that the user is able to identify the source position with an acceptable degree of precision.

The localization errors obtained from the experimental data exhibit a systematic error in the x and z-component, despite the overall low error seen in the simulation data. We postulate that these systematic errors emerged from inaccuracies related to the measurement setup. First, the estimated position in x-direction decreases linearly with respect to the mask-to-source distance (see Fig. 9.4), while it was expected to be zero. This indicates that the source was placed slightly off-center, and additionally, the camera was tilted around the y-axis, i.e. pointed upwards. Second, the systematic underestimation of the source-to-mask distance is probably caused by an inaccurate detector-to-mask distance b. Even though the camera case was 3D printed, printing tolerances and assembly of the case with the Timepix sensor may have led to deviations from the target distance of 20 mm. There is an additional uncertainty at which depth in the silicon sensor an impinging photon deposits its energy. When minimizing the error with respect to the rotation angle β and the detector-to-mask distance b from the estimated source positions we obtain an angle and a distance of approximately 1.2° and 20.6 mm.

Less surprising is the fact that circa $^2/_3$ of the error stems from the z-component. It is in accordance with a previous study, where it was established that with the same camera setup, MURA Decoding yields an axial resolution between 15 and 40 times worse than the lateral resolution and is degrading with increasing distances [118].

It is important to consider the runtimes of the localization methods in the context of their current implementations, which have not been optimized for runtime or computational efficiency. They rather represent a proof of concept.

9.4.2 Comparison to Other Localization Technologies

To compare our results with other localization technologies we focus on the z-component, which dominates the overall accuracy and set it in relation to the true source-to-mask distance. All distances in this paragraph were converted from source-to-detector to source-to-mask distances where applicable. Other technologies for localizing gamma sources are either already commercially available [33] or were recently proposed to the research community [20, 31, 32, 35]. The commercial freehandSPECT system was analyzed in regards to its localization accuracy of small gamma sources and mean errors between 2.9 mm and 7.4 mm depending on user experience were reported [33]. Though, explicit distances are not specified by the authors, from Fig. 2 we can derive an imaging range of 300-800 mm, which results in a relative error below 0.96 % and below 2.46 % for the respective user group.

Additionally, research groups are investigating stereo gamma cameras with coded aperture collimators [20, 31, 32, 35] or pinhole collimators [20]. Bugby et al. report a median z-error of and 1.23 mm (0.83 %) on simulated data and 3.54 mm (3.63%) on experimental data [20]. The research group Kaissas et al. report smaller z-errors of 0.28 mm (0.22 %) for a source placed at a source-to-mask distance of 130 mm (low rank and thus low-resolution mask) and 1.23 mm (0.94 %) (high-resolution mask) [31]. Additionally, the same research group analyzed extended gamma sources with a cylinder (24 mm in diameter and 9 mm in height) at increasing source-to-mask distances. Thereby, they show that the localization error deteriorates from 6.1 mm to 7.8 mm for source-to-mask distances of 140-200 mm [32]. Paradiso et al. [35] aim to use their gamma camera at much larger distances with source-to-mask distances between 360 mm and 4,000 mm. A ^{241}Am source at the distance of 1250 mm was estimated to be at 1200 mm and at 3000 mm to be at 2927 mm, resulting in localization errors of 50 mm (4.16 %) and 73 mm (2.43 %).

In this paper, with the ISL-EMG method we were able to achieve an error of $(0.88 \pm 0.68)\%$ and $(3.06 \pm 1.50)\%$ on the simulated and on the experimental data with median errors of 0.60 % and 2.98 %. That means for the range of 20–100 mm source-to-mask distance with a single gamma camera equipped with a coded aperture collimator we obtain a comparable localization accuracy without requiring additional hardware like external tracking or a second camera.

Stereo cameras which also use coded aperture collimators could benefit from our approach without any changes to the hardware. Their localization procedure, which is currently performed by triangulation, could potentially be made more robust by combining it with two separate single estimations using the here proposed methods.

9.4.3 Limitations

The experiment described in this chapter has several limitations that range from specific localization issues to more general challenges in Coded Aperture Imaging (CAI). First, the ISL algorithm in its current state is only be applicable in cases where no more than one source is present within the FOV.

Second, due to practical restrictions we were not able to evaluate the localization accuracy for sources beyond 100 mm. However, the advantage of our approach is that the camera can be positioned in close proximity to the source (up to around 11 mm for our setup), which allows for the capture of sources that may fall outside the FOV of stereoscopic cameras.

Third, the 0.5 mm thick silicon detector adopted has a low detection efficiency at 59.5 keV, which implied long acquisition times up to 15 min. However, alternative semiconductor-based detectors such as the Timepix3 employ a thicker cadmium-zinc-telluride (CdZnTe, or CZT) or cadmium telluride (CdTe) crystal, whose higher sensitivity could reduce the acquisition time to a tenth. Due to the known difficulties in simulating semiconductor-based detectors [43], the acquired dataset and the in silico reproduction did not measure the same quantity (see Tab. 5.2). To compare the photon count of both, the photon count of the acquired dataset must be estimated. When assuming that each photon deposited its full energy into a single pixel, i.e. no charge sharing occurs, the integrated energy per pixel can be divided by 59.5 keV. Based on this rough estimation, the detector images from our MC simulation contain on average 4.8 times more photons than the images captured with our experimental gamma camera. This difference in photon count could partially explain the difference in localization accuracy. It emphasizes that a more precise understanding of the relationship between the number of captured photons and the reconstruction quality and thus the localization accuracy is required, especially for a low-flux real-time application as RGS.

Fourth, the localization study with extended gamma sources of [32] indicate an increasing difficulty to localize sources with increasing size. A plausible reason can be found in the supplementary material from [73] (Supplementary Fig. 7) where it is shown that the reconstruction quality decreases exponentially with growing source size. Analyzing the influence of the source size was beyond the scope of this experiment, but the authors acknowledge that extended sources can represent a serious challenge in the development of a coded aperture camera for RGS. For example, in SLNB lymph nodes cannot be considered as point-like sources as their size varies from 5 to 20 mm [32, 73], and thus, the influence of the source size on the source localization remains to be investigated. Fifth, other 3D reconstruction methods exist, such as the 3D-MLEM described above (see Sec. 8.2.2) with a superior axial resolution. However its computational complexity renders it unsuitable for RGS.

Sixth, in this study, we have used a single camera setup to analyze the 3D-localization accuracy of a point-like source. However, we expect other factors to affect the accuracy, including the detector-to-mask distance which directly influences the magnification of the mask pattern. Furthermore, we assume the pinhole size, pinhole shape (round or square) and the MURA rank to affect the accuracy as well. These factors were not within the scope of this work, but their investigation could lead to a broader understanding of 3D-localization.

9.4.4 Potentials

The findings and limitations discussed in this study imply several potentials. First, a considerable drawback of the ISL method presented is the necessity for the user to possess knowledge of the source size and the initial guess z_0. It is conceivable that a user would rapidly navigate through the slices of the 3D reconstruction, locate the source, and draw a circle indicating both the rough source size and the initial guess z_0. Nevertheless, it would be more convenient if this process was automated as well. It is imaginable to combine both methods and use the COM to find an estimate for the initialization for the ISL method. This could potentially result in a fully automated source localization. Second, the ISL-EMG algorithms holds the potential for improving the capabilities to handle multiple sources within the FOV, for more complex scenarios with multiple radiation sources present. Third, a Machine Learning (ML) approach to

estimate the source position and source size directly from the detector image is conceivable. Fourth, this study can be considered a benchmark for investigating the following issues in order to leverage the full potential of CAI in RGS: The most used radiotracer in nuclear medicine is ^{99m}Tc which emits gamma photons of higher energy than the utilized ^{241}Am. Hence, a larger and thicker mask (1 mm thick with 0.25 mm pinholes) for ^{99m}Tc imaging is required and has already been designed. It is currently under testing by the team from the University of Naples and phantom studies will be conducted in the near future. Additionally, a more comprehensive understanding and a solution to the problem of imaging extended sources poses an important milestone in developing a fast and high-resolution gamma camera.

9.5 Conclusion

The goal of this experiment was to investigate how accurately a single gamma camera with a coded aperture collimator is able to localize in 3D single point-like sources in the near-field. Our proposed ISL algorithm iteratively searches for the source position based on a rough initial estimate. Mean localization errors below 1 mm based on data obtained by MC simulation and below 3 mm based on experimental data were achieved. It was demonstrated that when localizing point-like sources, it is crucial to consider the decreasing axial resolution which manifests in a positively-skewed CNR profile. Incorporating this behavior in form of an EMG fitting improves the localization accuracy by an average factor of 4 compared to using a standard Gaussian fitting. The occurrence of systematic localization errors in the experimental data emphasizes the importance of a thorough assembly and calibration process during image acquisition.

PART V

FINAL REMARKS

Chapter 10

Outlook

Several aspects for future research emerges from the described studies of this thesis. Specific suggestions toward the clinical usability of Coded Aperture Imaging (CAI) are elaborated in the following chapter.

10.1 ML Approach for 3D Reconstruction

In Chapter 6, the development of a Convolutional Encoder-Decoder (CED) was presented which is capable of reconstructing images with a higher quality in less computation time than all other planar reconstruction methods. However, successfully adapting this architecture for 3D reconstruction has proven challenging due to several difficulties: First, when predicting a full 3D reconstruction, the output dimension is much bigger than in the planar case. For example, instead of predicting a 256×256 pixel image, a $256 \times 256 \times 64$ data cube must be generated. In the reconstruction of point-like sources, the vast majority of voxels would be zero, making the training process extremely challenging, since the difference between the trivial solution, where all voxels are zero, and the correct solution is vanishingly small. Second, the convolutional layers in the decoder must be replaced by 3D convolutional layers which substantially increases the number of trainable parameters. Thus, depending on the available computing power, the batch size must be reduced which may influence the training results. Additionally, a substantial increase in training time must be expected.

In general, CEDs require a vast amount of training data. To address this issue, the simulation framework *ConvSim* has been developed. It is capable of simulating thousands of detector images of sources at various distances from the gamma camera. ConvSim has been compared to Monte Carlo (MC) simulations and shows high similarity to the detector images for small sources [96]. It also includes effects such as the collimation effect, the planar wave effect, and transmission for point sources. It is publicly available[1] and was specifically developed for training a CED for 3D reconstruction, making it ready for deployment.

It may also be worth to investigate the potential of reducing the computational load of CEDs by removing unnecessary or redundant weights from the architecture, a process called *pruning*. Furthermore, there has been a notable rise in popularity of novel architectures in the field of computer vision in the last few years, with transformer-based architectures as a prominent example [155, 156]. These architectures offer a functionality entirely different from CEDs, particularly in terms of their receptive field.

10.2 Reconstructing Extended Sources

Extended sources have reportedly been associated with lower reconstruction quality compared to point-like sources for MURA Decoding [25, 32, 73]. From conducting preliminary simulation experiments, we can confirm this behavior for the MLEM algorithm as illustrated in Fig. 10.1. Based on these yet unpublished studies we propose the following hypothesis as a starting point for future research: “In contrast to point sources, extended sources cause an increased average incident angle and thus, reconstruction methods for extended sources must consider the emerging near-field effects.” This hypothesis includes data-driven approaches, and first experiments with a CED and low-resolution images of extended source have already been reported [157]. According to this hypothesis, the CED from Chapter 6 trained on detector images with near-field effects generated by the ConvSim, should be capable of reconstructing extended sources.

[1] https://github.com/tomeiss/convsim

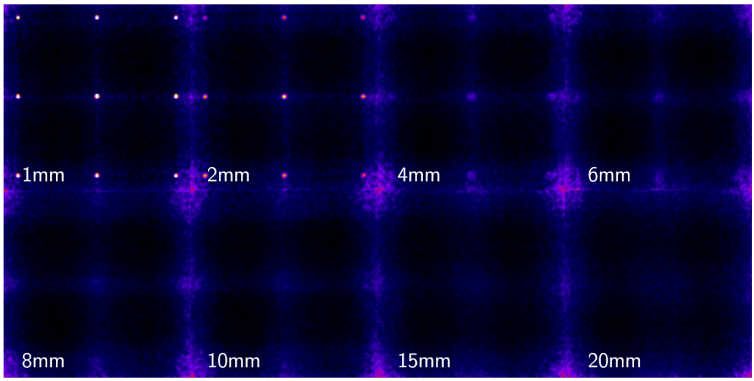


Figure 10.1: Reconstructions with the planar Maximum Likelihood Expectation Maximization (MLEM) algorithm of MC simulations of extended sources with increasing diameter (1 mm, 2 mm, 4 mm, 6 mm, 8 mm, 10 mm, 15 mm, and 20 mm) illustrate the known issue of CAI with extended sources. The simulated camera setup is the same as from Chapter 9 and a total of 10^9 photons were simulated. Ghost sources emerge at the image border as described above.

Another avenue of future research, which is closely related to this topic, is investigating how the number of captured photons influences reconstruction quality and, consequently, image quality. This issue was raised in Sec. 9.4.3, where it was estimated that the MC simulation contained on average 4.8 more photons than the experimentally captured images. Because the mask shadow projected by an extended source is more widespread than the one from a point source, the illuminated area is larger. Assuming an identical radioactivity implies that the same amount of photons impinge on a larger detector area, resulting in a lower photon count per pixel and an increased image degradation from Poisson noise.

In addition to planar imaging, the effect of extended sources on 3D localization capability is even less studied and largely unknown. The axial resolution reported above can only serve as a reference for small point-like sources and a spread of the CNR profile for extended sources must be expected. To ensure that robust depth estimations can effectively assist surgeons, this issue requires further investigation.

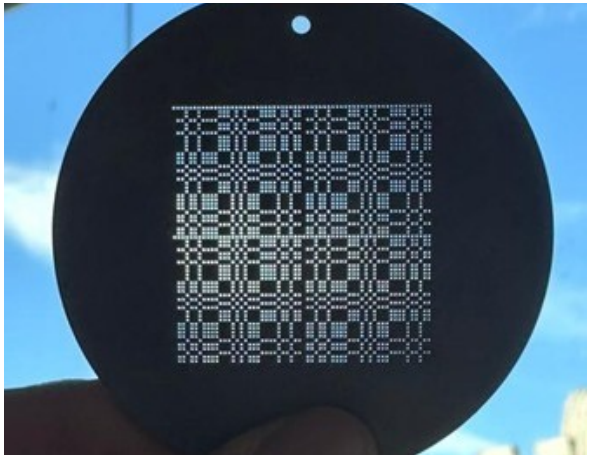


Figure 10.2: New MURA mask for ^{99m}Tc imaging with a larger and thicker mask that is 1 mm thick with 0.25 mm diameter pinholes developed and currently under testing by the team from Naples university.

10.3 99mTc Imaging

The most commonly used radiotracer in nuclear medicine is ^{99m}Tc with an estimated usage of 75% among all radiotracers. Its more energetic gamma photons (140.5 keV) massively penetrate thin masks, like the one that was used in the second part of this thesis. By incorporating this transmission into the 3D-MLEM algorithm, we were able to perform image reconstruction, but a transmission rate above 95 % from the combination of 0.11 mm tungsten and ^{99m}Tc would render CAI useless. Therefore, a thicker mask is required, which in turn, increases the impact of near-field effects as described in Sec. 3.5. A larger and thicker mask (1 mm thick with 0.25 mm pinholes, resulting transmission rate: 3.3 %) for ^{99m}Tc imaging has already been designed and is currently under investigation by the team led by Prof. Russo from the University of Naples. A photograph of this mask is shown in Fig. 10.2. In this work we were able to neglect the near-field effects, because the source was either far away (Part III) or the mask was thin compared to the pinhole size (Part IV). For ^{99m}Tc imaging, though, near-field effects must be addressed and reconstruction

methods adapted, potentially through the integration of existing compensation techniques [24, 26].

10.4 Semiconductor-Based Detectors

Since their first usage in mobile gamma cameras over two decades ago [158], semiconductor-based detectors have undergone a consistent advancements in terms of sensor area, material quality, and temperature resistance [7]. Despite these improvements, the non-ideal imaging fidelity remains a challenge. The causes are material impurities, flawed crystal growth that results in defective pixels and cracks, or nonuniform bump-bonding to the readout electronics, which have been encountered throughout this thesis. The CdTe detector used in the studies of Part III were more affected by these issues compared to the Si detector from Part IV.

An improvement in their manufacturing towards an overall more homogeneous light response is desirable. Alternatively, it would be beneficial to investigate image preprocessing methods to mitigate image degradation. Furthermore, the question arises as to how semiconductor-based detectors may be effectively incorporated into existing MC simulation frameworks. Nowadays, MC simulations play a major role in the development of Intraoperative Gamma Cameras (IGCs) [7], and therefore, a complete integration, considering photon hits, charge sharing, and deposited energy of CdTe or Si detectors is essential.

In conclusion, improving the imaging fidelity of semiconductor-based detectors, whether through hardware or software modifications, is a highly desirable objective. It would allow to fully harvest the superior resolution of high-pixelated semiconductor-based detectors and to improve gamma imaging in general, and CAI in particular.

Chapter 11

Conclusion

In the course of this thesis, both planar reconstruction methods and 3D reconstruction methods for Coded Aperture Imaging (CAI) were developed and evaluated with a focus on Radioguided Surgery (RGS). In this last chapter, conclusions will be drawn by reiterating the research hypothesis, and discussing the key findings. Three research hypotheses were proposed, where the first one stated:

Hypothesis Ia

A Convolutional Encoder-Decoder (CED) network trained on a synthetic dataset that is generated with a low-fidelity simulation achieves on average a higher Contrast-to-Noise Ratio (CNR) with less computation time than known planar reconstruction methods.

This hypothesis can be partially accepted. What contradicts the hypothesis is, that the Wiener Filter and generally a convolution carried out in the Fourier domain leads to a shorter computation time than the CED. Additionally, using a CED exhibits the following challenges: First, it was found that a CED trained on synthetic data is not as robust as for example MURA Decoding. Defective pixels and an inhomogeneous detector response limits the capabilities of a CED. The proposed preprocessing was effective, but in general, a semiconductor-based detector free of defects is highly desirable and topic of current research. Second,

an inherent lack of explainability and the possibility of instabilities must be kept in mind, when working with CEDs.

However, the CEDs yield on average a CNR 1.4 and 2.7 times better than the most commonly used MURA Decoding, depending on the training dataset. A training dataset tailored to the application-specific source distributions can be used to infuse a priori knowledge and further improve the reconstruction quality. Regarding the computation time, we showed that the CED with approximately 290 ms is around 170 times faster than the Maximum Likelihood Expectation Maximization (MLEM) algorithm, which represents the best analytical reconstruction method. It can be hypothesised that improvements are likely with a more accurate simulation of the training data, for example by integrating near-field effects, due to the fact that the CED is mainly determined by the training data. Therefore, CEDs are a promising candidate for overcoming the challenges presented by thicker masks and extended sources.

The second hypothesis was:

Hypothesis Ib

Even with a detector that cannot resolve the coded aperture mask's small pinholes, a CED network still outperforms other methods and thus, achieves super-resolution.

This hypothesis can be accepted with reservations. The simulation study of this thesis indicated that it is possible to achieve super-resolution with CAI. Even though, the Nyquist-Shannon sampling theorem must be fulfilled, combining a low-resolution but highly sensitive scintillator-based detector with a high-resolution coded aperture collimator. This hypothesis was investigated based on both a MC simulation and experimental data, but the fact that the low-resolution detector was simulated by aggregating the photon hits of neighboring pixels limits the validity of the hypothesis. Additionally, it was discovered that the upsampled detector image had a beneficial effect on the CED, due to its smoothing effect.

The third research hypothesis focused on 3D imaging with a coded aperture and was postulated as follows:

Hypothesis II

Single point-like gamma sources can be localized with an accuracy of approximately 5 mm in all three dimensions in the range of 20 mm to 100 mm from a stack of planar reconstructions, and is, thus, comparable to a stereoscopic camera.

The results of the presented studies confirm this hypothesis. From the assessment of the axial resolution, it was shown that it depends on the reconstruction method used for 3D reconstruction. While the proposed generalized 3D-MLEM algorithm yields on average a 0.3 times lower, and thus, better axial resolution, it takes approximately 4,400 times longer than generating a 3D reconstruction via sequential MURA Decoding. Generally it was shown, that the axial resolution degrades the further the source moves away from the camera. Despite its worse axial resolution, it was demonstrated, that with a sequence of planar reconstructions by MURA Decoding, the depth of a point-like source can be accurately estimated. The proposed ISL method in combination with an EMG fit, incorporating the decreasing axial resolution, achieved a 3D-localization error of about 0.8 mm on the simulated and 2.6 mm on the experimental data in the imaging range of 20 mm to 100 mm. It can be assumed that 3D-MLEM would achieve an even better localization accuracy, but its axial resolution requires a fine sampling in z-direction which renders it unsuited for RGS. All in all, our proposed experimental Intraoperative Gamma Camera (IGC) makes 3D source localization with a single gamma camera equipped with a coded aperture collimator as accurate as stereoscopic cameras with pinhole collimators. Providing valuable depth estimation to surgeons in a low-cost and computationally effective manner, constitutes an additional benefit of coded aperture collimators.

Overall, this thesis provides the research community with the first quantitative comparison of the most commonly used planar reconstruction methods. The reconstruction methods, along with the acquired datasets of our experimental IGC were made publicly available to ease the access for other research groups. Additionally, it has been shown that a Machine Learning (ML) approach is powerful and a highly promising candidate for both planar and 3D reconstruction, which is worth of further investigation. For the first time, two datasets were acquired and published where the three-dimensional Field of View (FOV) of

an IGC equipped with a coded aperture collimator was systematically sampled. Furthermore, it was demonstrated that a single gamma camera can achieve a localization accuracy for point-like sources that is comparable to stereo cameras, paving the way to provide valuable depth information to surgeons during RGS.

Appendix A

Appendix

A.1 Other Architectures for Planar Reconstruction

Other architectures used in the literature for image-to-image regression tasks have been evaluated as well. They were trained in the same manner as the CED-IN explained in Sec. 6.2.2. The MSD-Net architecture [99] was re-implemented in Tensorflow (Version 2.2) and, as suggested by [99], 100 layers with a channel width of 1 was selected. The hyperparameter, width and depth, determine how the dilated kernels are distributed along the network and vary in our case between 1 and 10. For further information, the reader is referred to the original publication. Three different mini-batch sizes were tested, since there are less reference values available in the literature: 8, 16, and 32 which are denoted by BS8, BS16, and BS32, respectively. Other than the original architecture in order to make a fair comparison, batch normalization layers (see Sec. 4.2) were inserted between each convolutional and activation layer. Additionally, the use of concatenating skip connections between the encoder and the decoder part were tested. The number of downsampling steps, the number of filters, the kernel size, etc. was kept the same compared to the CED-IN. Since the skip connections make the CED architecture closer to the original U-Net architecture from Ronneberger et al. [98], it will be referred to as U-Net. Two

mini-batch sizes of 8 and 16 were chosen, because they are typical sizes for such networks [104].

All networks were trained with the Adam optimizer for 20 epochs, a training and validation size of 35,000 and 10,000 images from the Imagenet database (see Sec. 5.4) and the learning rate schedule and automatic stopping as explained above. Finally, the networks were evaluated on the SRP dataset captured with our experimental gamma camera. The CNR distribution of the SRP phantom are depicted in Fig. A.1 together with our proposed CED-IN.

The CED-IN, which serves as a benchmark, achieves a median CNR of 2.65. The median CNRs of the MSD-Net BS8, MSD-Net BS16, MSD-Net BS32 are 1.83, 1.60, and 1.86. The U-Net BS8 and U-Net BS16 obtain similar values with 1.72 and 1.84. From the median CNR and a visual inspection of the boxplots, it becomes obvious that none of the tested networks beats the performance of the CED-IN.

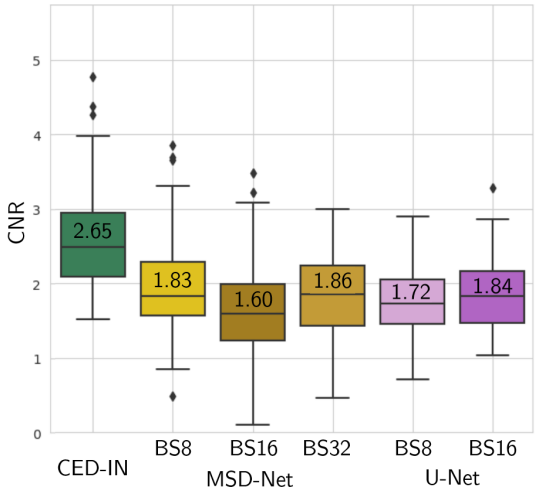


Figure A.1: The boxplots represent the CNRs of the SRP obtained from different network architectures. The vertical lines indicate the median value, the box the 25:75 quartiles, and the diamonds mark outliers. A total of 12 outliers are above 5.5 and not shown in this graph.

A.2 Additional 3D Reconstructions

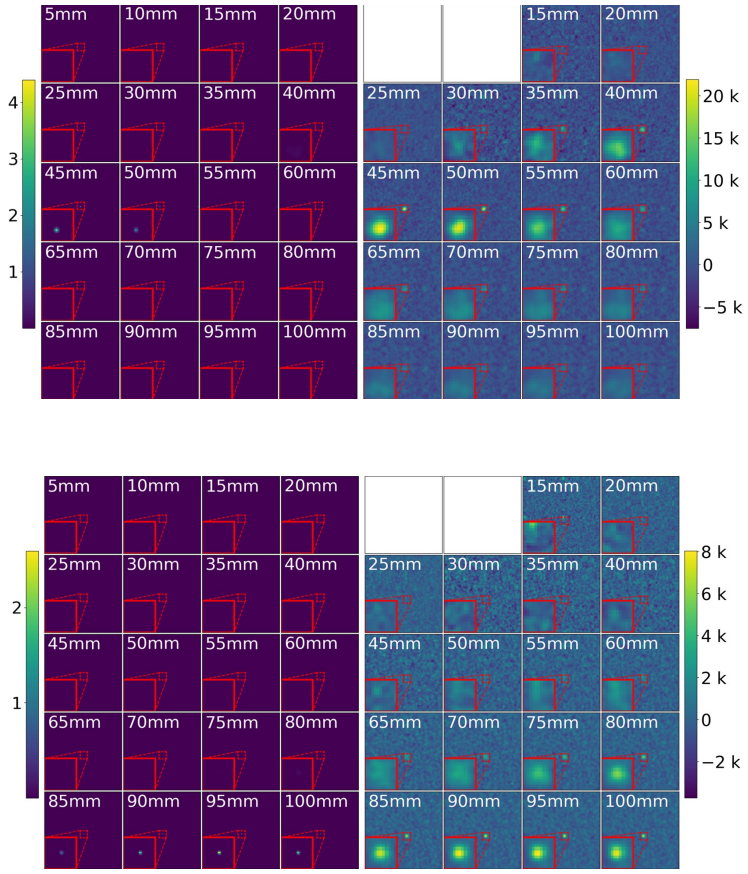


Figure A.2: The entire image stack of reconstructing the source at 20 equidistant image planes from 5 to 100 mm with MURA Decoding (right) and 3D-MLEM (left) for the source at 50 mm (top) and 100 mm (bottom). The area marked by the red square has been magnified for better visualization. Units are arbitrary. Figure from [118].

A.3 3D Reconstruction: Raw vs. Preprocessed Detector Images

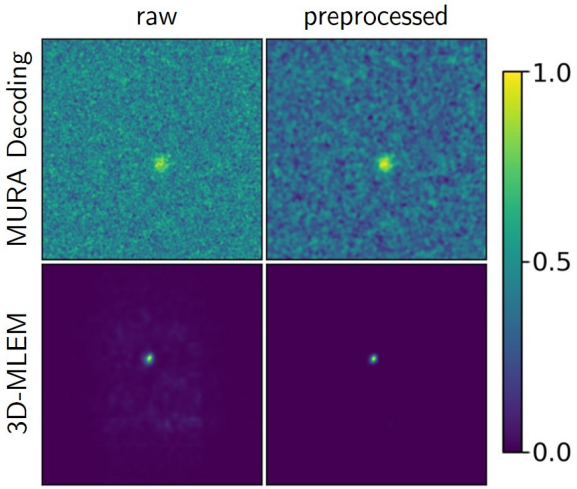


Figure A.3: All images show the reconstruction of the raw (left) and preprocessed (right) detector image of the source at source-to-mask distance of 20 mm at the 20 mm plane. The top row shows the reconstructions from MURA Decoding while the bottom row shows the 3D-MLEM results. Notice the higher background noise in the reconstructions from the raw detector image. For a fair comparison, images were normalized to 0 to 1. Figure modified from [118].

List of Figures

List of Tables

References

[1] H. Sung, J. Ferlay, R. L. Siegel, et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021, DOI: 10.3322/caac.21660.

[2] I.-A. Beckmann, *Die blauen Ratgeber - Brustkrebs. Antworten. Hilfen. Perspektiven.*, Deutsche Krebshilfe, Ed. Bonn: Deutsche Krebshilfe & Deutsche Krebsgesellschaft, 2013.

[3] J. M. Chang, J. W. T. Leung, L. Moy, et al., "Axillary Nodal Evaluation in Breast Cancer: State of the Art," *Radiology*, vol. 295, no. 3, pp. 500–515, jun 2020, DOI: 10.1148/radiol.2020192534.

[4] E. P. Mamounas, T. Kuehn, E. J. Rutgers, et al., "Current approach of the axilla in patients with early-stage breast cancer," *The Lancet*, vol. 6736, no. 17, 2017, DOI: 10.1016/S0140-6736(17)31451-4.

[5] A. S. Caudle, J. A. Cupp, and H. M. Kuerer, "Management of axillary disease," *Surgical Oncology Clinics of North America*, vol. 23, no. 3, pp. 473–486, 2014, DOI: 10.1016/j.soc.2014.03.007.

[6] S. Heller and P. Zanzonico, "Nuclear Probes and Intraoperative Gamma Cameras," *Seminars in Nuclear Medicine*, vol. 41, no. 3, pp. 166–181, may 2011, DOI: 10.1053/j.semnuclmed.2010.12.004.

[7] A. L. Farnworth and S. L. Bugby, "Intraoperative Gamma Cameras: A Review of Development in the Last Decade and Future Outlook," *Journal of Imaging*, vol. 9, no. 5, p. 102, may 2023, DOI: 10.3390/jimaging9050102.

[8] A. Gonzalez-Montoro, C. D. Vera-Donoso, G. Konstantinou, et al., "Nuclear-medicine probes: Where we are and where we are going," *Medical Physics*, vol. 49, no. 7, pp. 4372–4390, jul 2022, DOI: 10.1002/mp.15690.

[9] B. Zhu, J. Z. Liu, S. F. Cauley, et al., "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, mar 2018, DOI: 10.1038/nature25988.

[10] I. Häggström, C. R. Schmidtlein, G. Campanella, et al., "DeepPET: A deep encoder–decoder network for directly solving the PET image reconstruction inverse problem," *Medical Image Analysis*, vol. 54, pp. 253–262, 2019, DOI: 10.1016/j.media.2019.03.013.

[11] M. T. Madsen, "Recent advances in SPECT imaging," *Journal of Nuclear Medicine*, vol. 48, no. 4, pp. 661–673, 2007, DOI: 10.2967/jnumed.106.032680.

[12] T. E. Peterson and L. R. Furenlid, "SPECT detectors: The Anger Camera and beyond," *Physics in Medicine and Biology*, vol. 56, no. 17, 2011, DOI: 10.1088/0031-9155/56/17/R01.

[13] M. N. Wernick and J. N. Aarsvold, *Emission tomography: the fundamentals of PET and SPECT*, 1st ed., M. N. Wernick and J. N. Aarsvold, Eds. San Diego, CA: Elsevier, 2004.

[14] S. P. Povoski, "The History of Radioguided Surgery: Early Historical Milestones and the Development of Later Innovative Clinical Applications," in *Radioguided Surgery*, K. Herrmann, O. E. Nieweg, and S. P. Povoski, Eds. Cham: Springer International Publishing, jan 2016, pp. 3–12, DOI: 10.1007/978-3-319-26051-8_1.

[15] D. Roth, E. Larsson, J. Strand, et al., "Feasibility of 177Lu activity quantification using a small portable CZT-based gamma-camera," *EJNMMI Physics*, vol. 11, no. 1, 2024, DOI: 10.1186/s40658-023-00602-2.

[16] P. Knoll, S. Mirzaei, K. Schwenkenbecher, et al., "Performance evaluation of a solid-state detector based handheld gamma camera system," *Frontiers in Biomedical Technologies*, vol. 1, no. 1, pp. 61–67, 2014.

[17] F. Sánchez, J. M. Benlloch, B. Escat, et al., "Design and tests of a portable mini gamma camera," *Medical Physics*, vol. 31, no. 6, pp. 1384–1397, 2004, DOI: 10.1118/1.1755570.

[18] R. Massari, A. Ucci, C. Campisi, et al., "A novel fully integrated handheld gamma camera," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 832, pp. 271–278, 2016, DOI: 10.1016/j.nima.2016.06.124.

[19] M. Georgiou, G. Loudos, E. Fysikopoulos, et al., "λ-Eye: A high-sensitivity γ imaging probe for axillary sentinel lymph node mapping," *Nuclear Medicine Communications*, vol. 37, no. 10, pp. 1001–1009, 2016, DOI: 10.1097/MNM.0000000000000546.

[20] S. L. Bugby, J. E. Lees, W. K. McKnight, et al., "Stereoscopic portable hybrid gamma imaging for source depth estimation," *Physics in Medicine & Biology*, vol. 66, no. 4, p. 045031, feb 2021, DOI: 10.1088/1361-6560/abd955.

[21] R. H. Dicke, "Scatter-hole cameras for x-rays and gamma rays," *The astrophysical journal*, vol. 153, p. L101, 1968.

[22] J. G. Ables, "Fourier Transform Photography: A New Method for X-Ray Astronomy," *Publications of the Astronomical Society of Australia*, vol. 1, no. 4, pp. 172–173, dec 1968, DOI: 10.1017/s1323358000011292.

[23] A. Kulow, A. G. Buzanich, U. Reinholz, et al., "On the way to full-field X-ray fluorescence spectroscopy imaging with coded apertures," *Journal of Analytical Atomic Spectrometry*, vol. 35, no. 2, pp. 347–356, feb 2020, DOI: 10.1039/c9ja00232d.

[24] Z. Mu and Yi-Hwa Liu, "Aperture collimation correction and maximum-likelihood image reconstruction for near-field coded aperture imaging of single photon emission computerized tomography," *IEEE Transactions on Medical Imaging*, vol. 25, no. 6, pp. 701–711, jun 2006, DOI: 10.1109/TMI.2006.873298.

[25] P. Russo, F. Di Lillo, V. Corvino, et al., "CdTe compact gamma camera for coded aperture imaging in radioguided surgery," *Physica Medica*, vol. 69, no. June 2019, pp. 223–232, 2020, DOI: 10.1016/j.ejmp.2019.12.024.

[26] R. Accorsi and R. C. Lanza, "Near-field artifact reduction in planar coded aperture imaging," *Applied Optics*, vol. 40, no. 26, p. 4697, sep 2001, DOI: 10.1364/AO.40.004697.

[27] A. Haboub, A. A. Macdowell, S. Marchesini, et al., "Coded aperture imaging for fluorescent x-rays," *Review of Scientific Instruments*, vol. 85, no. 6, p. 63704, 2014, DOI: 10.1063/1.4882337.

[28] A. Martineau, J. M. Rocchisani, and J. L. Moretti, "Coded aperture optimization using Monte Carlo simulations," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 616, no. 1, pp. 75–80, 2010, DOI: 10.1016/j.nima.2010.02.261.

[29] R. Zhang, P. Gong, X. Tang, et al., "Reconstruction method for gamma-ray coded-aperture imaging based on convolutional neural network," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 934, pp. 41–51, aug 2019, DOI: 10.1016/j.nima.2019.04.055.

[30] Q. Zhang, J. Sun, and G. S. Mok, "Low dose SPECT image denoising using a generative adversarial network," *arXiv*, pp. 1–4, 2019, DOI: 10.48550/arXiv.1907.11944.

[31] I. Kaissas, C. Papadimitropoulos, K. Karafasoulis, et al., "3-D localization of gamma ray sources with coded apertures for medical applications," *Journal of Physics: Conference Series*, vol. 637, no. 1, 2015, DOI: 10.1088/1742-6596/637/1/012016.

[32] I. Kaissas, C. Papadimitropoulos, C. Potiriadis, et al., "Imaging of spatially extended hot spots with coded apertures for intra-operative nuclear medicine applications," *Journal of Instrumentation*, vol. 12, no. 1, pp. 1–9, 2017, DOI: 10.1088/1748-0221/12/01/C01059.

[33] B. Pouw, L. J. de Wit-van der Veen, M. P. Stokkel, et al., "Improved Accuracy and Reproducibility Using a Training Protocol for Freehand-SPECT 3D Mapping in Radio-Guided Surgery," *Clinical Nuclear Medicine*, vol. 40, no. 9, pp. e457–e460, sep 2015, DOI: 10.1097/RLU.0000000000000787.

[34] D. Hellingman, S. Vidal-Sicart, L. J. De Wit-Van Der Veen, et al., "A new portable hybrid camera for fused optical and scintigraphic imaging: First clinical experiences," *Clinical Nuclear Medicine*, vol. 41, no. 1, pp. e39–e43, jan 2016, DOI: 10.1097/RLU.0000000000000874.

[35] V. Paradiso, K. Amgarou, N. B. de Lanaute, et al., "3-D localization of radioactive hotspots via portable gamma cameras," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 910, no. August, pp. 194–203, 2018, DOI: 10.1016/j.nima.2018.09.081.

[36] V. Rozhkov, G. Chelkov, I. Hernández, et al., "Visualization of radiotracers for SPECT imaging using a Timepix detector with a coded aperture," *Journal of Instrumentation*, vol. 15, no. 06, pp. P06 028–P06 028, 2020, DOI: 10.1088/1748-0221/15/06/p06028.

[37] W. E. Smith, H. H. Barrett, and J. N. Aarsvold, "Coded-Aperture Imaging in Nuclear Medicine," NASA, Langley Research Center, Tech. Rep., 1989.

[38] S. Dorbala, K. Ananthasubramaniam, I. S. Armstrong, et al., "Single Photon Emission Computed Tomography (SPECT) Myocardial Perfusion Imaging Guidelines: Instrumentation, Acquisition, Processing, and Interpretation," *Journal of Nuclear Cardiology*, vol. 25, no. 5, pp. 1784–1846, 2018, DOI: 10.1007/s12350-018-1283-y.

[39] T. Nolte, N. Gross-Weege, and V. Schulz, *(hybrid) spect and pet technologies*. Cham: Springer International Publishing, 2020, pp. 111–133, DOI: 10.1007/978-3-030-42618-7_3.

[40] M. Chmeissani, C. Frojdh, O. Gal, et al., "First experimental tests with a CdTe photon counting pixel detector hybridized with a Medipix2 readout chip," *IEEE transactions on nuclear science*, vol. 51, no. 5, pp. 2379–2385, 2004, DOI: 10.1109/TNS.2004.832324.

[41] P. Russo, G. Mettivier, R. Pani, et al., "Imaging performance comparison between a LaBr3:Ce scintillator based and a CdTe semiconductor based photon counting compact gamma camera," *Medical Physics*, vol. 36, no. 4, pp. 1298–1317, 2009, DOI: 10.1118/1.3081412.

[42] D. Maneuski, V. Astromskas, E. Fröjdh, et al., "Imaging and spectroscopic performance studies of pixellated CdTe Timepix detector," *Journal of Instrumentation*, vol. 7, no. 1, 2012, DOI: 10.1088/1748-0221/7/01/C01038.

[43] R. Nabha, O. Van Hoey, C. Granja, et al., "A novel method to assess the incident angle and the LET of protons using a compact single-layer Timepix detector," *Radiation Physics and Chemistry*, vol. 199, no. May, p. 110349, oct 2022, DOI: 10.1016/j.radphyschem.2022.110349.

[44] M. Ruat and C. Ponchut, "Characterization of a Pixelated CdTe X-Ray Detector Using the Timepix Photon-Counting Readout Chip," *IEEE Transactions on Nuclear Science*, vol. 59, no. 5, pp. 2392–2401, oct 2012, DOI: 10.1109/TNS.2012.2210909.

[45] P. Russo and G. Mettivier, "Method for measuring the focal spot size of an x-ray tube using a coded aperture mask and a digital detector," *Medical Physics*, vol. 38, no. 4, pp. 2099–2115, 2011, DOI: 10.1118/1.3567503.

[46] P. Russo, A. S. Curion, G. Mettivier, et al., "Evaluation of a CdTe semiconductor based compact gamma camera for sentinel lymph node imaging," *Medical Physics*, vol. 38, no. 3, pp. 1547–1560, 2011, DOI: 10.1118/1.3555034.

[47] K. Van Audenhaege, R. Van Holen, S. Vandenberghe, et al., "Review of SPECT collimator selection, optimization, and fabrication for clinical and preclinical imaging," *Medical Physics*, vol. 42, no. 8, pp. 4796–4813, 2015, DOI: 10.1118/1.4927061.

[48] J. Braga, "Coded Aperture Imaging in High-energy Astrophysics," *Publications of the Astronomical Society of the Pacific*, vol. 132, no. 1007, p. 12001, 2020, DOI: 10.1088/1538-3873/ab450a.

[49] M. Tsuchimochi and K. Hayama, "Intraoperative gamma cameras for radioguided surgery: Technical characteristics, performance parameters, and clinical applications," *Physica Medica*, vol. 29, no. 2, pp. 126–138, 2013, DOI: 10.1016/j.ejmp.2012.05.002.

[50] A. Kulow, A. G. Buzanich, U. Reinholz, et al., "Comparison of three reconstruction methods based on deconvolution, iterative algorithm, and neural network for X-ray fluorescence spectroscopy with coded apertures," *Journal of Analytical Atomic Spectrometry*, 2020, DOI: 10.1039/d0ja00146e.

[51] P. Ritt, "Recent Developments in SPECT/CT," *Seminars in Nuclear Medicine*, vol. 52, no. 3, pp. 276–285, 2022, DOI: 10.1053/j.semnuclmed.2022.01.004.

[52] R. Moyer, "A low-energy multihole converging collimator compared with a pinhole collimator," *Journal of Nuclear Medicine*, vol. 15, no. 2, pp. 59–64, 1974.

[53] For the National Cancer Institute ©(2024) Terese Winslow LLC, U.S. Govt. has certain rights, "Sentinel lymph node biopsy," https://www.cancer.gov/about-cancer/diagnosis-staging/staging/sentinel-node-biopsy-fact-sheet, 2022, Accessed: 06.03.2024 at 16:30.

[54] A. S. Scheer, "Sentinel lymph node biopsy and axillary management for breast cancer," https://www.dynamed.com/evaluation/sentinel-lymph-node-biopsy-for-breast-cancer, 2024, Accessed: 14.04.2024 at 12:03.

[55] M. H. Ibraheem, M. Gamil, A. Tantawy, et al., "The Role of Intra-Operative Mobile Gamma Camera and Gamma Probe in Detection of Sentinel Lymph Node in Early Stage Breast

Cancer," *Journal of Cancer Science and Clinical Therapeutics*, vol. 03, no. 04, pp. 229–239, 2019, DOI: 10.26502/jcsct.5079037.

[56] S. R. Povoski, R. L. Neff, C. M. Mojzisik, et al., "A comprehensive overview of radioguided surgery using gamma detection probe technology," *World Journal of Surgical Oncology*, vol. 7, pp. 1–63, 2009, DOI: 10.1186/1477-7819-7-11.

[57] L. C. Johnson, "Development of a Small - Animal SPECT System with a High - Purity Germanium Detector," PhD thesis, Graduate School of Vanderbilt University, 2013.

[58] M. L. Waller and F. U. Chowdhury, "The basic science of nuclear medicine," *Orthopaedics and Trauma*, vol. 30, no. 3, pp. 201–222, 2016, DOI: https://doi.org/10.1016/j.mporth.2016.05.013.

[59] T. Kühn, A. Bembenek, H. Büchels, et al., "Sentinel-node-biopsie beim mammakarzinom. Interdisziplinär abgestimmter Konsensus der Deutschen Gesellschaft für Senologie für eine qualitätsgesicherte anwendung in der klinischen routine," *Pathologe*, vol. 25, no. 3, pp. 238–244, 2004, DOI: 10.1007/s00292-003-0661-6.

[60] S. R. Meikle, P. Kench, M. Kassiou, et al., "Small animal SPECT and its place in the matrix of molecular imaging technologies," *Physics in Medicine and Biology*, vol. 50, no. 22, pp. R45–R61, nov 2005, DOI: 10.1088/0031-9155/50/22/R01.

[61] Y. Li, G. Zhang, and Z. Liu, *Coded Aperture Imaging*. Singapore: Springer Nature Singapore, 2024, pp. 185–208, DOI: 10.1007/978-981-97-1455-1_7.

[62] J. Islamian, A. Azazrm, B. Mahmoudian, et al., "Advances in Pinhole and Multi-Pinhole Collimators For Single Photon Emission Computed Tomography Imaging," *World Journal of Nuclear Medicine*, vol. 14, no. 01, pp. 3–9, 2015, DOI: 10.4103/1450-1147.150505.

[63] N. U. Schramm, G. Ebel, U. Engeland, et al., "High-Resolution SPECT using Multi-Pinhole Collimation," *IEEE Nuclear Science Symposium and Medical Imaging Conference*, vol. 2, no. 3, pp. 774–777, 2002, DOI: 10.1109/nssmic.2002.1239437.

[64] B. H. Greenberg, R. Hart, E. H. Botvinick, et al., "Thallium-201 myocardial perfusion scintigraphy: Results of standard and multi-pinhole tomographic techniques," *American Journal of Cardiology*, vol. 43, no. 4, pp. 787–793, 1979, DOI: 10.1016/0002-9149(78)90896-2.

[65] J. Bacca, L. Galvis, and H. Arguello, "Coupled deep learning coded aperture design for compressive image classification," *Optics Express*, vol. 28, no. 6, p. 8528, mar 2020, DOI: 10.1364/oe.381479.

[66] M. S. Asif, A. Ayremlou, A. Sankaranarayanan, et al., "FlatCam: Thin, Lensless Cameras Using Coded Aperture and Computation," *IEEE Transactions on Computational Imaging*, vol. 3, no. 3, pp. 384–397, 2016, DOI: 10.1109/tci.2016.2593662.

[67] Y. Wu, V. Boominathan, H. Chen, et al., "PhaseCam3D - Learning Phase Masks for Passive Single View Depth Estimation," in *2019 IEEE International Conference on Computational Photography, ICCP 2019*, 2019, pp. 1–12, DOI: 10.1109/ICCPHOT.2019.8747330.

[68] A. Levin, R. Fergus, F. Durand, et al., "Image and depth from a conventional camera with a coded aperture," *ACM Transactions on Graphics*, vol. 26, no. 99, p. 70, 2007, DOI: 10.1145/1239451.1239521.

[69] K. Choi and D. Brady, "Coded aperture computed tomography," *Adaptive Coded Aperture Imaging, Non-Imaging, and Unconventional Imaging Sensor Systems*, vol. 7468, no. August 2009, p. 74680B, 2009, DOI: 10.1117/12.825277.

[70] D. Ching, S. Aslan, V. Nikitin, et al., "Time-coded aperture for x-ray imaging," *Optics Letters*, vol. 44, no. 11, p. 2803, 2019, DOI: 10.1364/ol.44.002803.

[71] S. Sun, Z. Zhang, L. Shuai, et al., "Development of a panorama coded-aperture gamma camera for radiation detection," *Radiation Measurements*, vol. 77, pp. 34–40, 2015, DOI: 10.1016/j.radmeas.2015.04.014.

[72] E. Caroli, J. B. Stephen, G. Di Cocco, et al., "Coded aperture imaging in X-and gamma-ray astronomy," *Space Science Reviews*, vol. 45, no. 3-4, pp. 349–403, 1987.

[73] H. Fujii, J. D. Idoine, S. Gioux, et al., "Optimization of Coded Aperture Radioscintigraphy for Sentinel Lymph Node Mapping," *Molecular Imaging and Biology*, vol. 14, no. 2, pp. 173–182, apr 2012, DOI: 10.1007/s11307-011-0494-2.

[74] M. J. Cieślak, K. A. Gamage, and R. Glover, "Coded-aperture imaging systems: Past, present and future development – A review," *Radiation Measurements*, vol. 92, pp. 59–71, sep 2016, DOI: 10.1016/j.radmeas.2016.08.002.

[75] E. E. Fenimore, "Coded aperture imaging: predicted performance of uniformly redundant arrays," *Applied Optics*, vol. 17, no. 22, p. 3562, 1978, DOI: 10.1364/ao.17.003562.

[76] S. R. Gottesman and E. E. Fenimore, "New family of binary arrays for coded aperture imaging," *Applied Optics*, vol. 28, no. 20, p. 4344, 1989, DOI: 10.1364/ao.28.004344.

[77] K. Amgarou, V. Paradiso, A. Patoz, et al., "A comprehensive experimental characterization of the iPIX gamma imager," *Journal of Instrumentation*, vol. 11, no. 8, 2016, DOI: 10.1088/1748-0221/11/08/P08012.

[78] M. Gmar, M. Agelou, F. Carrel, et al., "GAMPIX: A new generation of gamma camera," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 652, no. 1, pp. 638–640, 2011, DOI: 10.1016/j.nima.2010.09.003.

[79] S. Shifeng, Z. Zhiming, S. Lei, et al., "Far field 3D localization of radioactive hot spots using a coded aperture camera," *Applied Radiation and Isotopes*, vol. 107, pp. 177–182, jan 2016, DOI: 10.1016/j.apradiso.2015.10.021.

[80] O. P. Ivanov, I. A. Semin, V. N. Potapov, et al., "Extra-light gamma-ray imager for safeguards and homeland security," *2015 4th International Conference on Advancements in Nuclear Instrumentation Measurement Methods and their Applications, ANIMMA 2015*, no. December, 2015, DOI: 10.1109/ANIMMA.2015.7465584.

[81] R. Accorsi, F. Gasparini, and R. C. Lanza, "Optimal coded aperture patterns for improved SNR in nuclear medicine imaging," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 474, no. 3, pp. 273–284, 2001, DOI: 10.1016/S0168-9002(01)01326-2.

[82] I. Kaissas, C. Papadimitropoulos, A. Clouvas, et al., "Signal to Noise Ratio optimization for extended sources with a new kind of MURA masks," *Journal of Instrumentation*, vol. 15, no. 1, 2020, DOI: 10.1088/1748-0221/15/01/C01012.

[83] O. I. Vassilieva and R. C. Chaney, "Method for reducing background artifacts from images in single-photon emission computed tomography with a uniformly redundant array coded aperture," *Applied Optics*, vol. 41, no. 7, p. 1454, 2002, DOI: 10.1364/ao.41.001454.

[84] K. Jabbari, "Review of fast Monte Carlo codes for dose calculation in radiation therapy treatment planning," *Journal of Medical Signals and Sensors*, vol. 1, no. 1, pp. 73–86, 2011, DOI: 10.4103/2228-7477.83522.

[85] M. A. Alnafea, K. Wells, M. Guy, et al., "Near field corrections for coded aperture imaging in scintimammography," *IEEE Nuclear Science Symposium Conference Record*, vol. 5, pp. 2948–2953, 2006, DOI: 10.1109/NSSMIC.2006.356494.

[86] X. Li, Z. Zhang, D. Li, et al., "Comparison of the modified uniformly redundant array with the Singer array for near-field coded aperture imaging of multiple sources," *Nuclear Inst. and Methods in Physics Research, A*, vol. 1051, no. September 2022, p. 168230, 2023, DOI: 10.1016/j.nima.2023.168230.

[87] S. Agostinelli, J. Allison, K. Amako, et al., "GEANT4 - A simulation toolkit," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250–303, 2003, DOI: 10.1016/S0168-9002(03)01368-8.

[88] J. Perl, J. Shin, J. Schümann, et al., "TOPAS: An innovative proton Monte Carlo platform for research and clinical applications," *Medical Physics*, vol. 39, no. 11, pp. 6818–6837, 2012, DOI: 10.1118/1.4758060.

[89] M. Alnafea, K. Wells, N. M. Spyrou, et al., "Preliminary results from a Monte Carlo study of breast tumour imaging with low-energy high-resolution collimator and a modified uniformly-redundant array-coded aperture," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 563, no. 1, pp. 146–149, 2006, DOI: 10.1016/j.nima.2006.01.124.

[90] M. A. Alnafea, K. Shamma, and O. Kadri, "Experimental and Geant4 Simulation Study of MURA Mask for Scintimammography," *Applied Sciences (Switzerland)*, vol. 12, no. 10, 2022, DOI: 10.3390/app12104890.

[91] R. Accorsi, S. D. Metzler, J. R. Novak, et al., "Analytical derivation and experimental verification of a sensitivity formula for slit-slat SPECT collimation," *IEEE Nuclear Science Symposium Conference Record*, vol. 5, pp. 2976–2980, 2006, DOI: 10.1109/NSS-MIC.2006.356500.

[92] D. Hellfeld, P. Barton, D. Gunter, et al., "A Spherical Active Coded Aperture for 4 pi Gamma-Ray Imaging," *IEEE Transactions on Nuclear Science*, vol. 64, no. 11, pp. 2837–2842, 2017, DOI: 10.1109/TNS.2017.2755982.

[93] C. Domingo-Pardo, "A new technique for 3D gamma-ray imaging: Conceptual study of a 3D camera," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 675, pp. 123–132, 2012, DOI: 10.1016/j.nima.2012.02.008.

[94] S. Park, J. Boo, M. Hammig, et al., "Impact of aperture-thickness on the real-time imaging characteristics of coded-aperture gamma cameras," *Nuclear Engineering and Technology*, vol. 53, no. 4, pp. 1266–1276, 2021, DOI: 10.1016/j.net.2020.09.012.

[95] T. M. Cannon and E. E. Fenimore, "Tomographical imaging using uniformly redundant arrays," *Applied Optics*, vol. 18, no. 7, p. 1052, apr 1979, DOI: 10.1364/AO.18.001052.

[96] T. Meißner, S. Pietrantonio, W. Nahm, et al., "Towards a fast and accurate simulation framework for 3D spherical source localization in the near field of a coded aperture gamma camera," in *Translational Biophotonics: Diagnostics and Therapeutics III*, L. D. Lilge and Z. Huang, Eds., no. October. SPIE, oct 2023, p. 20, DOI: 10.1117/12.2670883.

[97] T. Buzug, "Algebraic and Statistical Reconstruction Methods," in *Computed Tomography*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 201–240, DOI: 10.1007/978-3-540-39408-2_6.

[98] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, pp. 234–241, 2015, DOI: 10.1007/978-3-319-24574-4_28.

[99] D. M. Pelt and J. A. Sethian, "A mixed-scale dense convolutional neural network for image analysis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 2, pp. 254–259, 2017, DOI: 10.1073/pnas.1715832114.

[100] C. Belthangady and L. A. Royer, "Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction," *Nature Methods*, vol. 16, no. 12, pp. 1215–1225, dec 2019, DOI: 10.1038/s41592-019-0458-z.

[101] J. He, Y. Wang, and J. Ma, "Radon Inversion via Deep Learning," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 2076–2087, jun 2020, DOI: 10.1109/TMI.2020.2964266.

[102] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., "Generative Adversarial Networks," *Advances in Neural Information Processing Systems*, vol. 3, no. January, pp. 2672–2680, jun 2014.

[103] J. R. Koza, F. H. Bennett, D. Andre, et al., "Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming," in *Artificial Intelligence in Design '96*. Dordrecht: Springer Netherlands, 1996, pp. 151–170, DOI: 10.1007/978-94-009-0279-4_9.

[104] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, Massachusetts ; London, England: The MIT Press, 2016.

[105] J. Patterson and A. Gibson, *Deep learning: A practitioner's approach*, vol. 29, no. 7553. "O'Reilly Media, Inc.", 2016.

[106] T. Buzug, *Computed Tomography*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, DOI: 10.1007/978-3-540-39408-2.

[107] H. Zhao, O. Gallo, I. Frosio, et al., "Loss Functions for Neural Networks for Image Processing," *arXiv*, pp. 1–11, nov 2015, DOI: 10.48550/arXiv.1511.08861.

[108] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–15, 2015.

[109] W. Pedrycz and S.-M. Chen, *Deep Learning: Algorithms and Applications*, Studies in Computational Intelligence, vol. 865, W. Pedrycz and S.-M. Chen, Eds. Cham: Springer International Publishing, 2020, DOI: 10.1007/978-3-030-31760-7.

[110] S. Saha, "A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way," https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53, 2018, Accessed: 21.03.2024 at 17:14.

[111] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," *arxiv.org*, vol. 12, pp. 818 –883, nov 2013, DOI: 10.48550/arXiv.1311.2901.

[112] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *Proceedings of the 32nd International Conference on Machine Learning*, Proceedings of Machine Learning Research, F. Bach and D. Blei, Eds., vol. 37. Lille, France: PMLR, 2015, pp. 448–456.

[113] S. Santurkar, D. Tsipras, A. Ilyas, et al., "How Does Batch Normalization Help Binary Training?" *Advances in Neural Information Processing Systems*, vol. 31, no. NeurIPS, sep 2018, DOI: 10.48550/arXiv.1909.09139.

[114] M. Du, K. Liang, Y. Liu, et al., "Investigation of domain gap problem in several deep-learning-based CT metal artefact reduction methods," *arXiv*, nov 2021, DOI: 10.48550/arXiv.2111.12983.

[115] X. Tao, H. Gao, X. Shen, et al., "Scale-Recurrent Network for Deep Image Deblurring," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8174–8182, DOI: 10.1109/CVPR.2018.00853.

[116] T. Meißner, V. Rozhkov, J. Hesser, et al., "Quantitative comparison of planar coded aperture imaging reconstruction methods," *Journal of Instrumentation*, vol. 18, no. 01, p. P01006, jan 2023, DOI: 10.1088/1748-0221/18/01/P01006.

[117] J. Ahrens, B. Geveci, and C. Law, "Paraview: An end-user tool for large data visualization," *The visualization handbook*, vol. 717, no. 8, 2005.

[118] T. Meißner, L. A. Cerbone, P. Russo, et al., "Assessment of the axial resolution of a compact gamma camera with coded aperture collimator," *EJNMMI Physics*, vol. 11, no. 1, p. 30, mar 2024, DOI: 10.1186/s40658-024-00631-5.

[119] E. Bertolucci, M. Maiorino, G. Mettivier, et al., "Preliminary test of an imaging probe for nuclear medicine using hybrid pixel detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 487, no. 1-2, pp. 193–201, jul 2002, DOI: 10.1016/S0168-9002(02)00965-8.

[120] R. Zhang, X. Tang, P. Gong, et al., "Low-noise reconstruction method for coded-aperture gamma camera based on multi-layer perceptron," *Nuclear Engineering and Technology*, vol. 52, no. 10, pp. 2250–2261, oct 2020, DOI: 10.1016/j.net.2020.03.024.

[121] O. Russakovsky, J. Deng, H. Su, et al., "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015, DOI: 10.1007/s11263-015-0816-y.

[122] G. Bradski, "The openCV library." *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, vol. 25, no. 11, pp. 120–123, 2000.

[123] H. Nam, H. Lee, J. Park, et al., "Reducing domain gap by reducing style bias," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8690–8699, DOI: 10.1109/cvpr46437.2021.00858.

[124] D. Schellingerhout, R. Accorsi, U. Mahmood, et al., "Coded Aperture Nuclear Scintigraphy: A Novel Small Animal Imaging Technique," *Molecular Imaging*, vol. 1, no. 4, p. 153535002002213, oct 2002, DOI: 10.1162/15353500200221362.

[125] F. Garibaldi, R. Accorsi, M. N. Cinti, et al., "Small animal imaging by single photon emission using pinhole and coded aperture collimation," *IEEE Transactions on Nuclear Science*, vol. 52, no. 3 I, pp. 573–579, 2005, DOI: 10.1109/TNS.2005.851428.

[126] R. Accorsi, A. S. Curion, P. Fralliccardi, et al., "Preliminary evaluation of the tomographic performance of the mediSPECT small animal imaging system," *Nuclear Instru-*

ments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 571, no. 1-2 SPEC. ISS., pp. 415–418, 2007, DOI: 10.1016/j.nima.2006.10.123.

[127] R. Accorsi, L. Celentano, P. Laccetti, et al., "High-resolution 125I small animal imaging with a coded aperture and a hybrid pixel detector," *IEEE Transactions on Nuclear Science*, vol. 55, no. 1, pp. 481–490, 2008, DOI: 10.1109/TNS.2007.909846.

[128] M. Jeong and M. D. Hammig, "Comparison of gamma ray localization using system matrixes obtained by either MCNP simulations or ray-driven calculations for a coded-aperture imaging system," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 954, no. July 2018, p. 161353, 2020, DOI: 10.1016/j.nima.2018.10.031.

[129] Z. Mu, L. W. Dobrucki, and Y. H. Liu, "SPECT Imaging of 2-D and 3-D Distributed Sources with Near-Field Coded Aperture Collimation: Computer Simulation and Real Data Validation," *Journal of Medical and Biological Engineering*, vol. 36, no. 1, pp. 32–43, feb 2016, DOI: 10.1007/s40846-016-0111-6.

[130] M. Abadi, A. Agarwal, P. Barham, et al., "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," *arXiv preprint*, mar 2016, DOI: 10.48550/arXiv.1603.04467.

[131] J. Beyerer, F. P. León, and C. Frese, *Machine vision: Automated visual inspection: Theory, practice and applications*. Springer, 2015, DOI: 10.1007/978-3-662-47794-6.

[132] K. H. Jin, M. T. McCann, E. Froustey, et al., "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017, DOI: 10.1109/TIP.2017.2713099.

[133] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and Checkerboard Artifacts," *Distill*, vol. 1, no. 10, pp. 1–4, 2017, DOI: 10.23915/distill.00003.

[134] H. Lim, I. Y. Chun, Y. K. Dewaraja, et al., "Improved Low-Count Quantitative PET Reconstruction With an Iterative Neural Network," *IEEE Transactions on Medical Imaging*, vol. 39, no. 11, pp. 3512–3522, nov 2020, DOI: 10.1109/TMI.2020.2998480.

[135] B. F. Hutton, M. Occhipinti, A. Kuehne, et al., "Development of clinical simultaneous SPECT/MRI," *The British Journal of Radiology*, vol. 91, no. 1081, p. 20160690, jan 2018, DOI: 10.1259/bjr.20160690.

[136] O. P. Ivanov, V. E. Stepanov, A. N. Sudarkin, et al., "Different methods of image reconstruction for portable X-ray and gamma-ray imager with coded aperture," *IEEE Nuclear Science Symposium & Medical Imaging Conference*, vol. 2, pp. 1743–1746, 1997, DOI: 10.1109/nssmic.1997.670653.

[137] Y. Xue, S. Cheng, Y. Li, et al., "Reliable deep-learning-based phase imaging with uncertainty quantification," *Optica*, vol. 6, no. 5, p. 618, 2019, DOI: 10.1364/optica.6.000618.

[138] V. Antun, F. Renna, C. Poon, et al., "On instabilities of deep learning in image reconstruction and the potential costs of AI," *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30 088–30 095, dec 2020, DOI: 10.1073/pnas.1907377117.

[139] D. J. Lingenfelter, J. A. Fessler, and Z. He, "Sparsity regularization for image reconstruction with Poisson data," *Computational Imaging VII*, vol. 7246, p. 72460F, 2009, DOI: 10.1117/12.816961.

[140] A. Gaitanis, G. Kontaxakis, G. Spyrou, et al., "PET image reconstruction: A stopping rule for the MLEM algorithm based on properties of the updating coefficients," *Computerized Medical Imaging and Graphics*, vol. 34, no. 2, pp. 131–141, 2010, DOI: 10.1016/j.compmedimag.2009.07.006.

[141] J. Su, D. V. Vargas, and K. Sakurai, "One Pixel Attack for Fooling Deep Neural Networks," *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 5, pp. 828–841, 2019, DOI: 10.1109/TEVC.2019.2890858.

[142] E. Ozkan and A. Eroglu, "The Utility of Intraoperative Handheld Gamma Camera for Detection of Sentinel Lymph Nodes in Melanoma," *Nuclear Medicine and Molecular Imaging*, vol. 49, no. 4, pp. 318–320, dec 2015, DOI: 10.1007/s13139-015-0341-5.

[143] T. Meißner, W. Nahm, J. Hesser, et al., "Simulation Study on Super-Resolution for Coded Aperture Gamma Imaging," *arXiv*, jun 2023, DOI: 10.48550/arXiv.2306.08483.

[144] R. F. Marcia and R. M. . Willett, "Compressive Coded Aperture Superresolution Image Reconstruction," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 833–836, 2008, DOI: 10.1109/ICASSP.2008.4517739.

[145] M. Ljungberg and P. H. Pretorius, "SPECT/CT: an update on technological developments and clinical applications," *The British Journal of Radiology*, vol. 91, no. 1081, p. 20160402, jan 2018, DOI: 10.1259/bjr.20160402.

[146] K. Okuda, K. Nakajima, H. Yoneyama, et al., "Impact of iterative reconstruction with resolution recovery in myocardial perfusion SPECT: phantom and clinical studies," *Scientific Reports*, vol. 9, no. 1, pp. 1–9, 2019, DOI: 10.1038/s41598-019-56097-4.

[147] L. A. Cerbone, L. Cimmino, A. Sarno, et al., "Monte Carlo and experimental evaluation of a Timepix4 compact gamma camera for coded aperture nuclear medicine imaging with depth resolution," *Physica Medica*, vol. 113, no. May, p. 102663, sep 2023, DOI: 10.1016/j.ejmp.2023.102663.

[148] T. Buzug, "Algebraic and Statistical Reconstruction Methods," in *Computed Tomography*. Berlin, Heidelberg: Springer, 2008, pp. 201–240, DOI: 10.1007/978-3-540-39408-2_6.

[149] T. Meißner, L. A. Cerbone, P. Russo, et al., "3D-localization of single point-like gamma sources with a coded aperture camera," *Physics in Medicine & Biology*, vol. 69, no. 16, p. 165004, aug 2024, DOI: 10.1088/1361-6560/ad6370.

[150] I. Assam, S. P. Dierck, Y. Zhao, et al., "Evaluation of sentinel lymph node localization in malignant melanoma by preoperative semiconductor gamma camera and planar lymphoscintigraphy," *Journal of Applied Clinical Medical Physics*, vol. 24, no. 8, pp. 1–13, aug 2023, DOI: 10.1002/acm2.14077.

[151] Y. Kalambet, Y. Kozmin, K. Mikhailova, et al., "Reconstruction of chromatographic peaks using the exponentially modified Gaussian function," *Journal of Chemometrics*, vol. 25, no. 7, pp. 352–356, jul 2011, DOI: 10.1002/cem.1343.

[152] D. Hellfeld, "Free-moving Omnidirectional 3D Gamma-ray Imaging and Localization," PhD thesis, UC Berkeley, 2020.

[153] C. Papadimitropoulos, I. Kaissas, K. Karafasoulis, et al., "A setup for gamma ray sources localization using coded apertures and CdTe detectors," *2014 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2014*, pp. 1–6, 2016, DOI: 10.1109/NSSMIC.2014.7431273.

[154] R. Willingale, M. Sims, and M. Turner, "Advanced deconvolution techniques for coded aperture imaging," *Nuclear Instruments and Methods in Physics Research*, vol. 221, no. 1, pp. 60–66, mar 1984, DOI: 10.1016/0167-5087(84)90180-7.

[155] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *ICLR 2021 - 9th International Conference on Learning Representations*, oct 2020, DOI: 10.48550/arXiv.2010.11929.

[156] Y. Yang, L. Jiao, X. Liu, et al., "Transformers Meet Visual Learning Understanding: A Comprehensive Review," *arXiv*, pp. 1–20, mar 2022, DOI: 10.48550/arXiv.2203.12944.

[157] G. Daniel and O. Limousin, "Extended sources reconstructions by means of coded mask aperture systems and deep learning algorithm," *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 1012, no. January, p. 165600, 2021, DOI: 10.1016/j.nima.2021.165600.

[158] I. Anzai, T. Inoue, T. Ito, et al., "Analysis of Tomographical Information of Tracer Distribution by Using a CdZnTe Semiconductor-Based Gamma Counter Equipped with Using Coded Aperture," *RADIOISOTOPES*, vol. 51, no. 11, pp. 505–508, 2002, DOI: 10.3769/radioisotopes.51.505.

List of Publications and Supervised Theses

Journal Articles

- **Meißner, Tobias**, Rozhkov, Vladislav, Hesser, Jürgen, Nahm, Werner *Quantitative comparison of planar coded aperture imaging reconstruction methods*, Journal of Instrumentation, 2023, 18(01), P01006, https://doi.org/10.1088/1748-0221/18/01/P01006.
- **Meißner, Tobias**, **Cerbone, Laura Antonia**, Russo, Paolo, Nahm, Werner, Hesser, Jürgen *Assessment of the axial resolution of a compact gamma camera with coded aperture collimator*, European Journal for Nuclear Medicine and Molecular Imaging, 2024, 11(1), 30, https://doi.org/10.1186/s40658-024-00631-5.
- **Meißner, Tobias**, Cerbone, Laura Antonia, Russo, Paolo, Nahm, Werner, Hesser, Jürgen *3D-localization of single point-like gamma sources with a coded aperture camera*, Physics in Medicine & Biology, 2024, 69(16), 165004, https://doi.org/10.1088/1361-6560/ad6370.

Refereed Conference Articles

- **Meißner, Tobias**, Nahm, Werner, Hesser, Jürgen, Löw, Nicolas *Simulation study on super-resolution for coded aperture gamma imaging*, submitted, reviewed but not yet published in 4EU+ International Workshop

on Recent Advancements in Artificial Intelligence, Milano, Italy, 2022, http://arxiv.org/abs/2306.08483(preprint).

- **Meißner, Tobias**, Pietrantonio, Saverio, Nahm, Werner, Hesser, Jürgen, *Towards a fast and accurate simulation framework for 3D spherical source localization in the near field of a coded aperture gamma camera*, SPIE - Translational Biophotonics: Diagnostics and Therapeutics III, 2023, Issue October, p. 20, Munich, Germany, https://doi.org/10.1117/12.2670883.

Conference Presentations

- **Meißner, Tobias**, Nahm, Werner, Hesser, Jürgen, Löw, Nicolas *Planar Coded Aperture Reconstruction for Gamma-Imaging via Machine Learning and the TOPAS Simulation Toolkit*, IV. Geant4 International User Conference, Naples, Italy, 2022.
- **Meißner, Tobias**, Pietrantonio, Saverio, Nahm, Werner, Hesser, Jürgen, *Towards a fast and accurate simulation framework for 3D spherical source localization in the near field of a coded aperture gamma camera*, SPIE - Translational Biophotonics: Diagnostics and Therapeutics III, 2023, Issue October, p. 20, Munich, Germany, https://doi.org/10.1117/12.2670883.
- **Meißner, Tobias**, Hesser, Jürgen, Nahm, Werner *Potential and challenges of coded aperture collimation for intraoperative gamma cameras*, 58$^{\text{th}}$ Annual Conference of the German Society for Biomedical Engineering, Stuttgart, 2024.

Reports and Theses

- **Stegmeier, Elisa** *Evaluation of different CNN architectures for planar coded aperture image reconstruction*, Mandatory research internship, Mannheim Institute for Intelligent Systems in Medicine, Heidelberg University, 2022.

Supervised Student Theses

- **Seubert, Pauline Bonaventúra** *Interpretable machine learning: application of integrated gradients on an image-to-image network*, Bachelor Thesis, Mannheim Institute for Intelligent Systems in Medicine, Heidelberg University, 2021.
- **Zehender, Dominik** *Compression of a convolutional encoder-decoder network for coded aperture imaging*, Master Thesis, Mannheim Institute for Intelligent Systems in Medicine, Heidelberg University, 2022.
- **Pietrantonio, Saverio** *Convolutional simulation and 3D MLEM reconstruction for coded aperture imaging*, Master Thesis, Mannheim Institute for Intelligent Systems in Medicine, Heidelberg University, 2022.

Image License

DISSERTATION COPYRIGHT PERMISSION FORM

Title(s) of the Image(s): Terese Winslow LLC owns the copyright to the following image(s):

Title(s) of illustration(s): *"Biopsy, Sentinel Lymph Node, Breast (3-Panel)", "Female Breast Anatomy 2011", "Anatomy of the Female Breasts"*

Description of the Work: Terese Winslow LLC hereby grants permission to reproduce the above image(s) for use in the work specified:

Dissertation title: Advances in Coded Aperture Imaging for Intraoperative Gamma Cameras: Quantitative Comparison of Planar Reconstruction Methods and Transition to 3D-Localization of Point Sources
University: Karlsruhe Institute for Technology (KIT), Karlsruhe, Germany

Digital object identifier (DOI), if available: -
DOI is a unique alphanumeric string assigned by a registration agency (the International DOI Foundation) to identify content and provide a persistent link to its location on the Internet. The publisher assigns a DOI when your article is published and made available electronically.

License Granted: Terese Winslow LLC hereby grants limited, non-exclusive worldwide print and electronic rights only for use in the Work specified. Terese Winslow LLC grants such rights "AS IS" without representation or warranty of any kind and shall have no liability in connection with such license.

Restrictions: Reproduction for use in any other work, derivative works, or by any third party by manual or electronic methods is prohibited. Ownership of original artwork, copyright, and all rights not specifically transferred herein remain the exclusive property of Terese Winslow LLC. Additional license(s) are required for ancillary usage(s).

Credit must be placed adjacent to the image(s) as follows:

~~For the National Cancer Institute~~ © (*copyright year*) Terese Winslow LLC, U.S. Govt. has certain rights

Permission granted to:
Name: *Tobias Meissner*
Mailing address: [redacted]

Email address: *tobias.meissner@medma.uni-heidelberg.de*
Phone number: *0049*[redacted]

Signature_____ ***Date*** *March 15, 2024*
Name

Signature_____ **Date** March 15, 2024
Terese Winslow, CMI, Member

Terese Winslow LLC, Medical Illustration
714 South Fairfax Street, Alexandria, Virginia 22314
(703) 836-9121
terese@teresewinslow.com
www.teresewinslow.com

Karlsruhe Transactions on Biomedical Engineering (ISSN 1864-5933)

Karlsruhe Institute of Technology / Institute of Biomedical Engineering (Ed.)

Band 2 Matthias Reumann
Computer assisted optimisation on non-pharmacological treatment of congestive heart failure and supraventricular arrhythmia.
ISBN 978-3-86644-122-4

Band 3 Antoun Khawaja
Automatic ECG analysis using principal component analysis and wavelet transformation.
ISBN 978-3-86644-132-3

Band 4 Dmytro Farina
Forward and inverse problems of electrocardiography: clinical investigations.
ISBN 978-3-86644-219-1

Band 5 Jörn Thiele
Optische und mechanische Messungen von elektrophysiologischen Vorgängen im Myokardgewebe.
ISBN 978-3-86644-240-5

Band 6 Raz Miri
Computer assisted optimization of cardiac resynchronization therapy.
ISBN 978-3-86644-360-0

Band 7 Frank Kreuder
2D-3D-Registrierung mit Parameterentkopplung für die Patientenlagerung in der Strahlentherapie.
ISBN 978-3-86644-376-1

Band 8 Daniel Unholtz
Optische Oberflächensignalmessung mit Mikrolinsen-Detektoren für die Kleintierbildgebung.
ISBN 978-3-86644-423-2

Band 9 Yuan Jiang
Solving the inverse problem of electrocardiography in a realistic environment.
ISBN 978-3-86644-486-7

Band 10 Sebastian Seitz
Magnetic Resonance Imaging on Patients with Implanted Cardiac Pacemakers.
ISBN 978-3-86644-610-6

Band 11 Tobias Voigt
Quantitative MR Imaging of the Electric Properties and Local SAR based on Improved RF Transmit Field Mapping.
ISBN 978-3-86644-598-7

Band 12 Frank Michael Weber
Personalizing Simulations of the Human Atria: Intracardiac Measurements, Tissue Conductivities, and Cellular Electrophysiology.
ISBN 978-3-86644-646-5

Band 13 David Urs Josef Keller
Multiscale Modeling of the Ventricles: from Cellular Electrophysio-logy to Body Surface Electrocardiograms.
ISBN 978-3-86644-714-1

Band 14 Oussama Jarrousse
Modified Mass-Spring System for Physically Based Deformation Modeling.
ISBN 978-3-86644-742-4

Band 15 Julia Bohnert
Effects of Time-Varying Magnetic Fields in the Frequency Range1 kHz to 100 kHz upon the Human Body: Numerical Studies and Stimulation Experiment.
ISBN 978-3-86644-782-0

Band 16 Hanno Homann
SAR Prediction and SAR Management for Parallel Transmit MRI.
ISBN 978-3-86644-800-1

Band 17 Christopher Schilling
Analysis of Atrial Electrograms.
ISBN 978-3-86644-894-0

Band 18 Tobias Baas
ECG Based Analysis of the Ventricular Repolarisation in the Human Heart.
ISBN 978-3-86644-882-7

Band 19 Martin Wolfgang Krüger
Personalized Multi-Scale Modeling of the Atria: Heterogeneities, Fiber Architecture, Hemodialysis and Ablation Therapy.
ISBN 978-3-86644-948-0

Band 20 Mathias Wilhelms
Multiscale Modeling of Cardiac Electrophysiology: Adaptation to Atrial and Ventricular Rhythm Disorders and Pharmacological Treatment.
ISBN 978-3-7315-0045-2

Band 21 Matthias Keller
Formation of Intracardiac Electrograms under Physiological and Pathological Conditions.
ISBN 978-3-7315-0228-9

Band 22 Walther H. W. Schulze
ECG Imaging of Ventricular Activity in Clinical Applications.
ISBN 978-3-7315-0374-3

Band 23 Axel Loewe
Modeling Human Atrial Patho-Electrophysiology from Ion Channels to ECG. Substrates, Pharmacology, Vulnerability, and P-Waves.
ISBN 978-3-7315-0528-0

Band 24 Stefan Pollnow
Characterizing Cardiac Electrophysiology during Radiofrequency Ablation: An Integrative *Ex vivo*, *In silico*, and *In vivo* Approach.
ISBN 978-3-7315-0886-1

Band 25 Jorge Patricio Sánchez Arciniegas
A Multiscale In Silico Study to Characterize the Atrial Electrical Activity of Patients With Atrial Fibrillation: A Translational Study to Guide Ablation Therapy.
ISBN 978-3-7315-1170-0

Band 26 Nicht erschienen

Band 27 Claudia Nagel
Multiscale Cohort Modeling of Atrial Electrophysiology: Risk Stratification for Atrial Fibrillation through Machine Learning on Electrocardiograms.
ISBN 978-3-7315-1281-3

Band 28 María del Carmen Martínez Antón
Local Impedance Characterization for Scar and Fibrosis Detection: Towards a New Substrate Assessment for Atrial Mapping.
ISBN 978-3-7315-1368-1

Band 29 Tobias Meißner
Development and Evaluation of Coded Aperture Reconstruction Methods for Intraoperative Gamma Cameras.
ISBN 978-3-7315-1394-0

INSTITUTE OF BIOMEDICAL ENGINEERING
KARLSRUHE INSTITUTE OF TECHNOLOGY (KIT)

Female breast cancer has surpassed lung cancer as the most commonly diagnosed cancer. The biopsy of lymph nodes that receive drainage from the primary tumor are guided by Intraoperative Gamma Cameras (IGCs). Coded Aperture Imaging (CAI) has been proposed as an alternative collimation technique to produce an image, because it offers a better trade-off between sensitivity and spatial resolution. However, it requires image reconstruction. Therefore, a Convolutional Encoder-Decoder Network (CED) was developed and trained on synthetic source images. It was quantitatively compared based on the reconstruction quality and the runtime with state-of-the-art methods, such as MURA Decoding and the iterative Maximum Likelihood Expectation Maximization (MLEM) algorithm. Afterwards the axial resolution and 3D localization accuracy of the most commonly used methods were analyzed. In conclusion, the CED outperforms all methods regarding the quality. From the analytical methods the MLEM algorithm yields a higher quality and a better axial resolution, but is not suitable for IGCs in its current form, due to its computational complexity. MURA Decoding provides robust reconstructions. Its fast computation enables 3D reconstruction which allows the localization of point-like sources with an accuracy comparable to that of stereoscopic cameras.

ISSN 1864-5933
ISBN 978-3-7315-1394-0

Gedruckt auf FSC-zertifiziertem Papier

ISBN 978-3-7315-1394-0
9 783731 513940